Advances in Fuzzy Systems

Hybrid Biomedical Intelligent Systems

Guest Editors: Maysam Abbod, Jiann-Shing Shieh, and Hak-Keung Lam

Hybrid Biomedical Intelligent Systems

Advances in Fuzzy Systems

Hybrid Biomedical Intelligent Systems

Guest Editors: Maysam Abbod, Jiann-Shing Shieh, and Hak-Keung Lam

Editorial Board

Contents

Hybrid Biomedical Intelligent Systems, Maysam Abbod, Jiann-Shing Shieh, and Hak-Keung Lam
Volume 2012, Article ID 320428, 1 page

Optimization the Initial Weights of Artificial Neural Networks via Genetic Algorithm Applied to Hip Bone Fracture Prediction, Yu-Tzu Chang, Jinn Lin, Jiann-Shing Shieh, and Maysam F. Abbod
Volume 2012, Article ID 951247, 9 pages

Artificial Neural Network-Statistical Approach for PET Volume Analysis and Classification, Mhd Saeed Sharif, Maysam Abbod, Abbes Amira, and Habib Zaidi
Volume 2012, Article ID 327861, 10 pages

Classifying High-Dimensional Patterns Using a Fuzzy Logic Discriminant Network, Nick J. Pizzi and Witold Pedrycz
Volume 2012, Article ID 920920, 7 pages

A Novel Technique for Identifying Patients with ICU Needs Using Hemodynamic Features, A. Jalali, P. Ghorbanian, A. Ghaffari, and C. Nataraj
Volume 2012, Article ID 696194, 9 pages

An Intelligent Dynamic MRI System for Automatic Nasal Tumor Detection, Wen-Chen Huang and Chun-Liang Liu
Volume 2012, Article ID 272570, 7 pages

Medical Pattern Recognition: Applying an Improved Intuitionistic Fuzzy Cross-Entropy Approach, Kuo-Chen Hung
Volume 2012, Article ID 863549, 6 pages

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2012, Article ID 320428, 1 page
doi:10.1155/2012/320428

Editorial
Hybrid Biomedical Intelligent Systems

Maysam Abbod,[1] Jiann-Shing Shieh,[2] and Hak-Keung Lam[3]

[1] *School of Engineering and Design, Brunel University, Uxbridge UB8 3PH, UK*
[2] *Department of Mechanical Engineering, Yuan Ze University, Chung-Li 32003, Taiwan*
[3] *Division of Engineering, King's College London, London WC2R 2LS, UK*

Correspondence should be addressed to Maysam Abbod, maysam.abbod@brunel.ac.uk

Received 20 March 2012; Accepted 20 March 2012

Computerised systems are one of the fastest growing areas within medicine at present, both in the clinical setting in hospitals and in research and development (R&D). The importance of such systems for patient management is growing steadily and plays an increasingly important role in treatment and diagnosis. Intelligent systems are becoming the core for such system where they provide more rapid and precise patients' management, aid diagnosis, and improved outcome. Furthermore, intelligent systems benefit the health care system through improved accuracy and more cost-effective solutions.

Initially, singular approaches were used, but increasingly this has moved towards hybridisation, in particular that of artificial intelligence techniques. The current emphasis is on hydride models (quantitative/qualitative) and hybrid intelligent algorithms (fuzzy logic, neural networks, and neurofuzzy systems) applied into biomedical applications. The purpose of this special issue is to promote research and developments of the best work in the field of hybrid intelligent systems for biomedical applications.

One paper in this special issue addresses an improved intuitionistic fuzzy cross-entropy for medical pattern recognition. The improved nonprobabilistic entropy approach supports doctors examining the work of the preliminary diagnosing. Another paper is concerned with classifying high-dimensional patterns using fuzzy logic discriminate network. The technique is based on an adaptive network of fuzzy logic connectives to combine class boundaries generated by sets of discriminate functions. Another different paper investigates the use of an intelligent dynamic MRI (DMRI) system for automatic nasal tumour detection. The dynamic MRI is one of the major tools for diagnosing nasal tumours in recent years. The purposed system detects and enhances the tumour region in DMRI automatically by using Adaboost, SVM, and Bayes-Gaussian classifier.

One other paper presents a novel technique for identifying patients with ICU needs using hemodynamic features. The system can classify the types of patients and their ICU need based on hemodynamic features such as heart rate and blood pressure. Further analyses are done based on the initial measurements such as circle criterion, estimation error criterion, Poincare plot deviation, and autonomic response delay criterion. All the features are utilised to classify the patient using a fuzzy system. There is a paper that investigates the effect of the initial random weight of the neural network on the accuracy of the trained model. Hip fracture data were used to generate a model to predict the patient status using a neural network model. The initial weights of the neural network have been optimised using genetic algorithm. Results show that the initialisation process has a big effect on the accuracy of the trained model. Another paper addresses the use of neural networks for PET volume analysis and segmentation; different images were used to train neural network for classifying tumour data on PET images. The developed classifier shows that the system has a high accuracy such that a single tumour pixel can be detected and classified accurately.

Maysam Abbod
Jiann-Shing Shieh
Hak-Keung Lam

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2012, Article ID 951247, 9 pages
doi:10.1155/2012/951247

Research Article

Optimization the Initial Weights of Artificial Neural Networks via Genetic Algorithm Applied to Hip Bone Fracture Prediction

Yu-Tzu Chang,[1] Jinn Lin,[2] Jiann-Shing Shieh,[1,3] and Maysam F. Abbod[4]

[1] *Department of Mechanical Engineering, Yuan Ze University, 32003 Chungli, Taiwan*
[2] *Department of Orthopaedic Surgery, National Taiwan University Hospital, Taipei, Taiwan*
[3] *Center for Dynamical Biomarkers and Translational Medicine, National Central University, Taoyuan City, Taiwan*
[4] *School of Engineering and Design, Brunel University, London, UK*

Correspondence should be addressed to Maysam F. Abbod, maysam.abbod@brunel.ac.uk

Received 25 December 2011; Accepted 23 January 2012

Academic Editor: Hak-Keung Lam

This paper aims to find the optimal set of initial weights to enhance the accuracy of artificial neural networks (ANNs) by using genetic algorithms (GA). The sample in this study included 228 patients with first low-trauma hip fracture and 215 patients without hip fracture, both of them were interviewed with 78 questions. We used logistic regression to select 5 important factors (i.e., bone mineral density, experience of fracture, average hand grip strength, intake of coffee, and peak expiratory flow rate) for building artificial neural networks to predict the probabilities of hip fractures. Three-layer (one hidden layer) ANNs models with back-propagation training algorithms were adopted. The purpose in this paper is to find the optimal initial weights of neural networks via genetic algorithm to improve the predictability. Area under the ROC curve (AUC) was used to assess the performance of neural networks. The study results showed the genetic algorithm obtained an AUC of 0.858 ± 0.00493 on modeling data and 0.802 ± 0.03318 on testing data. They were slightly better than the results of our previous study (0.868 ± 0.00387 and 0.796 ± 0.02559, resp.). Thus, the preliminary study for only using simple GA has been proved to be effective for improving the accuracy of artificial neural networks.

1. Introduction

With increment of the life expectancy among the world, osteoporosis becomes more and more prevalent and may lead to disastrous pathological fractures. For the year 2000, there were an estimated 9 million new osteoporotic fractures, 1.6 million at the hip, 1.7 million at the forearm, and 1.4 million at the vertebrae. Europe and the Americas accounted for 51% of these osteoporotic fractures, while most of the remainder occurred in the Western Pacific region and Southeast Asia [1]. Hip fractures cause the most morbidity with a reported mortality rate up to 20–24% in the first year after a hip fracture [2, 3], and greater risk of dying may persist for at least 5 years afterwards [4]. Hip fractures are invariably associated with severe chronic pain, reduced mobility, disability, and an increasing degree of dependence [5]; even if the patients survive after the incidence, some of them still suffer its subsequent complications [6]. Furthermore, the patients have to shoulder the huge health and economic burdens that caused a high health care expenditure.

In order to reduce the occurrence of this preventable injury and its subsequent complication, we expected to find out the risk factors that are important for fracture prevention and health promotion and then build a predictor for the probability of hip bone fracture. Recently, support vector machines (SVMs) have been applied in machine learning techniques and are state-of-the-art machine learning techniques for risk minimization [7]. Since their invention, research on SVMs has exploded both in theory and applications. In practice, SVMs have been successfully applied to many real-world domains [8, 9]. However, in dealing with

highly nonlinear and complex system like hip fracture, artificial neural network (ANN) is still better than SVMs because so many hidden layers, nodes, and parameters (e.g., learning constant, learning algorithms, initial weights, etc.) can be adjusted in ANN. Also, in previous study [10], although many potential risk factors for hip fracture have been identified, these risk factors may vary geographically and combined effects of different risk factors have not been well understood. Then, they established the artificial neural network (ANN) to predict the risk of hip bone fracture according to the advantages of nonlinearity, fault tolerance, universality, and real-time operation.

ANNs are computer programs that simulate some of the higher level functions of the human brain. There are neurons and synapses in the brain, with various synaptic connection strengths—called "weights"—between the connected neuron pairs. The so-called input and output neurons for each problem correspond to the inputs and to outputs from a traditional computer program. The other, called "hidden" neurons, along with the synapses and weights, comes between the input and output neurons corresponding to the instructions in a traditional program. Use of ANNs as clinical prediction models has been explored in many areas of medicine, including nephrology, and microbiology, radiology, neurology.

Backpropagation is a topology of artificial neural network; it adjusts the network's weights and biases by calculating the gradient of the error. Usually, backpropagation neural networks are applied with random initial weight setting because of symmetry breaking [11]. However, training the neural networks with random initial weights may cause two main drawbacks: trapping into local minima and converging slowly [11, 12]. In view of these limitations of back-propagation neural networks, global search techniques (e.g., genetic algorithm and particle swarm optimization) have been presented to overcome these shortcomings [13, 14]. So far a number of works compare the evaluation between back-propagation neural network and genetic algorithm for training neural networks [15, 16], both of them are techniques for optimization and learning.

Genetic algorithms (GAs) developed to mimic some of the processes observed in natural evolution are a class of global search algorithms techniques. They have been shown in practice to be very effective at function optimization, and searching large or complex (multimodal, discontinuous, etc.) spaces to find nearly global optimum efficiently [11]. Therefore, this study tried to find the optimal initial weights of artificial neural network via genetic algorithm so that the predictor could enhance the ability of predicting the risk of hip bone fracture.

This paper is arranged as follows. Section 2 gives an overview of artificial neural networks and genetic algorithm. The materials and the artificial neural network prediction model are used in Section 3. In Section 4, the proposed genetic algorithm model is explained. Results and discussion are in Section 5. Conclusions and future work are in Section 6.

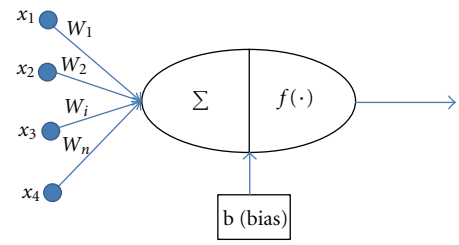


Figure 1: The diagram of one neuron.

2. An Overview of Neural Networks and Genetic Algorithm

2.1. Artificial Neural Networks. Artificial neural networks are a system that emulates the process of biological neural networks. Artificial neural networks generally consist of five components.

(1) The directed graph of the ANN topology.

(2) A state variable associated with each neuron.

(3) A real-valued weight associated with each link.

(4) A real-valued bias associated with each neuron.

(5) The output of each neuron $f(\sum w_i x_i - b)$, which is the input for next layer, where f is the transfer function, w_i are the weights connected with each neuron at the last layer, x_i are the input values of the neurons, and b is the bias of the neuron (Figure 1).

Artificial neural networks have become very popular for a few reasons. Firstly, they have the capability of learning what adjusts the weights and biases between the nodes. If the prediction is correct, the weights of the connections will be increased and vice versa [17]. Secondly, artificial neural networks are a parallel system that can deal with missing data that the linear program cannot deal with. Thirdly, with multiple layers, artificial neural networks can process nonlinearity even though the relationships between multifactor variables have not been exactly understood.

Feedforward network is one of the artificial neural network topologies. It usually consists of multiple layers, and the information will just be communicated to the next layer (i.e., output nodes have no arcs away from them). By different tactics for modifying the weights in training networks, some types of feedforward are presented such as back-propagation neural network. Back-propagation neural network is one that calculates the gradient of the error and then propagates error backward through the network to modify the weights and biases.

2.2. Genetic Algorithm. Genetic algorithms are developed to mimic some of the processes inspired by natural evolution. There are five components that we should define first [18, 19]:

(1) a way of coding solution to the problem on chromosomes;

(2) a fitness function which returns a value for each chromosome given to it;

(3) a way of initializing the population of chromosomes;

(4) operators that may be applied to parents when they reproduce to alter their genetic composition standard operators are mutation and crossover;

(5) parameter settings for the algorithm, the operators, and so forth.

With these definitions, genetic algorithm operates in the following steps.

(1) Encode the problem in a string and generate the initial population using initialization procedure.

(2) Reckon the fitness value for each chromosome. It will directly react on the distance to the optimum.

(3) Reproduce until a stopping criterion is met; reproduction consists of iterations of the following steps.

   (a) Choose a number of parents to reproduce; selection is stochastic, but the individuals with the highest evaluations are favored in the selection.

   (b) Apply the genetic algorithms (e.g., crossover, mutation) to the parents.

   (c) Accumulate the children and evaluate the fitness value. Insert children into the population to replace worse individuals of the current population.

In genetic algorithm, fitness function will evaluate the adaptation of each individual; it is a key point to decide if the outcome is good or not. Selection operator will choose adaptive parents depending on their fitness values. By this step, the population tends towards better individuals. Crossover operator and mutation operator make the chromosomes reach a wider search space [20].

3. Materials and Prediction Model

3.1. Study Sample. The sample data were gathered in the previous case-control matched study for the analysis of risk factors of hip fracture for older adults aged 60 and older [21] and predicting the risk of hip bone fracture for elders in Taiwan by ensemble backpropagation neural networks [10]. The sample included 228 cases who were the patients admitted to the National Taiwan University Hospital with first low-trauma hip fracture and 215 patients in the same hospital, but without hip fracture.

Both cases and controls were interviewed by trained interviewers with the same standardized questionnaire, which included questions about basic and social demography, history of diseases and conditions, self-rated overall health, health habits, intake of food and nutritional supplements, falls and fracture experiences, living environment and potential home hazards, physical functioning and use of assistive devices, and cognitive, and other functions. Athropometric measures and physical assessments less influenced by lower extremity function were performed after the questionnaire interview, including body height, weight, handgrip strength, peak expiratory flow, and coordination test. Bone mineral density (BMD) was examined at the nonfractured side of proximal femur for cases and the same side for matched controls by dual-energy X-ray absorptiometry (DXA) using the same machine of DEXA (Model: QDR 4500A; Hologic, Waltham, MA, US), and read by the same radiologist in 153 cases and 197 hospital controls. Leisure-time physical activity in the health habits was measured using total energy expended on all leisure-time activities in a week. The physical functions were measured by questions on the level of difficulty in performing 5 ADL, 6 IADL, and 8 mobility tasks. Cognitive function was measured with the Mini-Mental State Examination (MMSE). Height and weight were measured using electronic scales. BMI was calculated as weight in kg/height in m^2. Grip strength was measured with a hand-held hydraulic dynamometer (Model: NC70142; North Coast Medical, Morgan Hill, CA, US). The participants used the dominant hand, and three maximal values were averaged. Peak expiratory flow was assessed by using a peak flow meter (Model: Standard Mini Wright; Clement Clarke International, Harlow, Essex, UK). The participants took a deep breath and blew as fast and vigorously as possible. The maximum of three trials was chosen as the peak flow. The finger-nose-finger test was conducted by asking the participants to use their finger to alternately touch their own nose and the interviewers' finger as quickly as possible for assessing the coordination. A total of 78 variables were measured.

Because the number of variables was too large to be collected rapidly in clinic, logistic regression was applied to filter out irrelevant factors with two steps: univariable analysis and multivariable analysis. After these analyses, five significant factors (i.e., bone mineral density, experience of fracture, average hand grip strength, intake of coffee, and peak expiratory flow rate) remained to be the variables of neural networks. Typically, the data for artificial neural networks were divided into two parts: modeling set and testing set. Then the modeling set was further divided into training group and validation group.

However, artificial neural networks were unstable predictors that, with small changes in training data, may result in very different models. To reduce the influences from unstable predictors, k-fold cross-validation method was applied here [19]. The study divided the database into five equal parts. One part was for testing (i.e., testing data), and the other four parts were combined to be modeling. This cross-validation procedure was repeated five times so that we got five data sets with different testing data.

3.2. Architecture of Prediction Model. Back-propagation neural network is the most popular training algorithm with gradient techniques [22]. In previous study [7], back-propagation neural network comprised an input layer (with 5 input variables), a hidden layer (with 10 nodes), and an output layer (with 1 nodes). The ensemble artificial neural networks method was utilized to improve the generalization of the back-propagation neural network [23]: the previous

Table 1: Standard parameter set for training.

Parameter	Value
Transfer function of the hidden neurons	Tan sigmoid
Transfer function of the output neurons	Log sigmoid
Training function	Trainscg
Maximal fail	1
Encoding	Real (decimal)
Chromosome length	71
Population size	30
Weight initialization routine	Rand
Initial range	$-1 \sim 1$
Fitness function	Mean square error
Selection operation	Roulette whe5el
Crossover	BLX − 0.5
Mutation	Non-uniform
Elitist	2
Stopping criterion	100 iterations

study structured 20 neural networks with different training data and validation data then trained each of them for 15 times with random initial weights and then ensembled all the neural networks with the best performance (Figure 2). In this study, genetic algorithm tried to find the optimization initial points instead of 15 random initial weights for back-propagation neural networks training (Figure 3).

4. The Proposed Genetic Algorithm Model in This Case

During the study, there were many methods, operators, or ideals that tried to reach the optimum. The processes would be presented below and the parameter set was listed in Table 1.

4.1. Modeling Strategy. The modeling strategy was the skeleton about how to optimize the initial weights of the artificial neural network model and the study tried two types. At the beginning of the study (i.e., type 1), the genetic algorithm would evolve the population each iteration with different training data (validation data is also) and then choose 15 the best chromosomes into each artificial neural networks instead of 15 random initial weights (Figure 3(a)). However, neural network is unstable with different results because of a small change in training data. Therefore, the second strategy (i.e., type 2) was presented: the training data were defined for each artificial neural network, and then the genetic algorithm would find the optimum initial weights of ANNs, respectively (Figure 3(b)).

4.2. Initial Population. In genetic algorithms, the binary-code and the real code are the primary schemes to describe a chromosome. But, because the binary-coded scheme is neither necessary nor beneficial [22, 24], and according to the advantages of intuitiveness, resolution, and facility (i.e., need not to decode) for real code, the study used the real-coded method for describing the chromosomes. There were 30 chromosomes generated in each generation, and each of chromosomes consisted of 60 weights and 11 biases represented by one digit. The range for the initial population will affect the search efficiency, so we tried three levels. At first, the range between −2 and 2 was used, because all of the weights fell in this range after training by back-propagation neural network. But later, the study tried to set the value between −1 and 1 to compare with back-propagation neural network. And last, because the crossover operator could search over the initial range, we tried to narrow the range again between −0.5 and 0.5.

4.3. Evaluation. Each member of the current population was evaluated by fitness function based on the mean square error value to assign the probability of being selected in matting pool. The fitness function here was the back-propagation neural network. The study inserts the solutions into the networks and then calculated the error after the training. The mean square error value represents how the solution is fit for the problem, but it does not mean the solution is suitable for being redrawn in the next generation. The reason is that it is not difficult to find the optimal chromosomes with minimal errors for training data, but it is difficult for validation data. In other words, because the lowest mean square error in training data was the goal of the networks, but the mean square error in validation data used for preventing overfitting of the neural networks should also be considered.

To avoid the above state, the study sets the limitation: no matter how lower the mean square error was, if the error resulted from validation data is higher than the threshold, the chromosome would not be chosen. The threshold was defined as with the optimal solution in last generation, the network would not stop training until the validation error went up, and the error was the threshold of the next generation [25].

4.4. Reproduction. A mating pool of 30 chromosomes was created by Roulette wheel selection operator according to the probability of each chromosome in the current population. The steps of the procedure were as follows: firstly, select the random number between 0 and 1. Secondly, chose the chromosome whose cumulative probability is a little more than the random number into the mating pool. Finally, repeat above steps until 30 new chromosomes are created in the population.

4.5. Crossover. The process was described as below: firstly, randomly select two chromosomes from the matting pool. Secondly, choose four random positions and exchange genes between the first two positions and two last positions. Thirdly, randomly choose the numbers of the interval $[c_{\min} - I \cdot 0.5,\ c_{\max} + I \cdot 0.5]$ for four positions separately, where $c_{\min}$ is the minimal value between two parents, I is the range between $c_{\min}$ and $c_{\max}$ [26]. The last two springs were

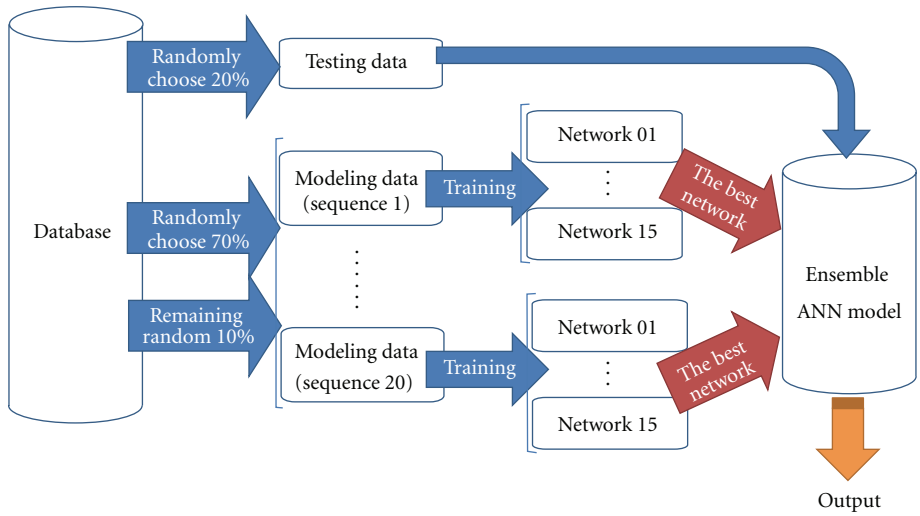
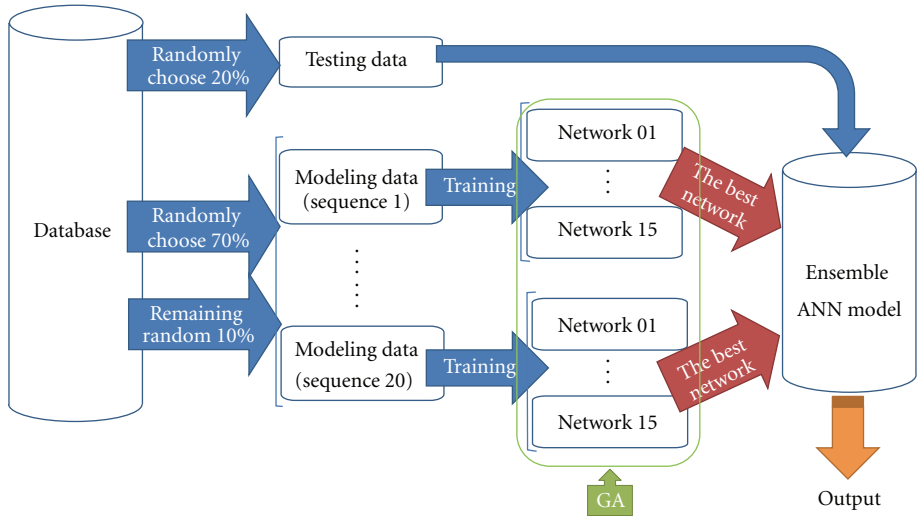
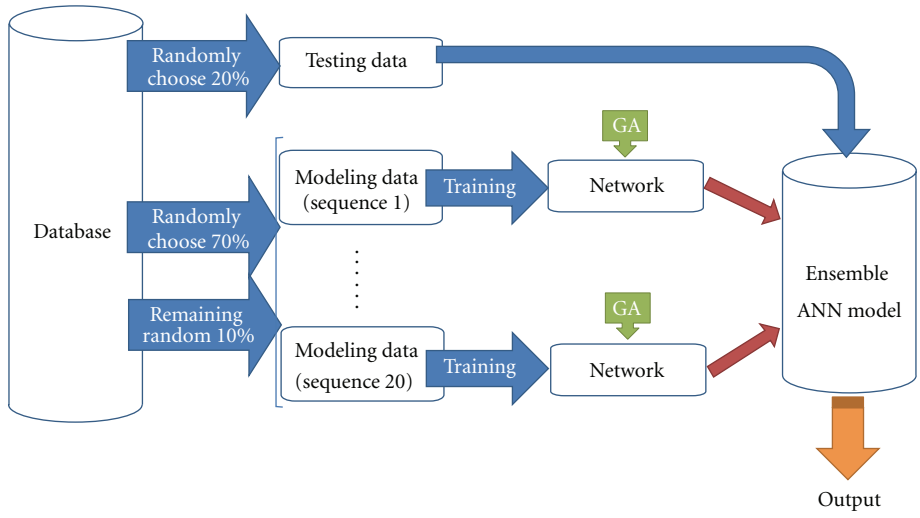


Figure 2: The skeleton with initial weights.



(a)



(b)

Figure 3: The skeleton with GA-based initial weights: (a) type 1 and (b) type 2.

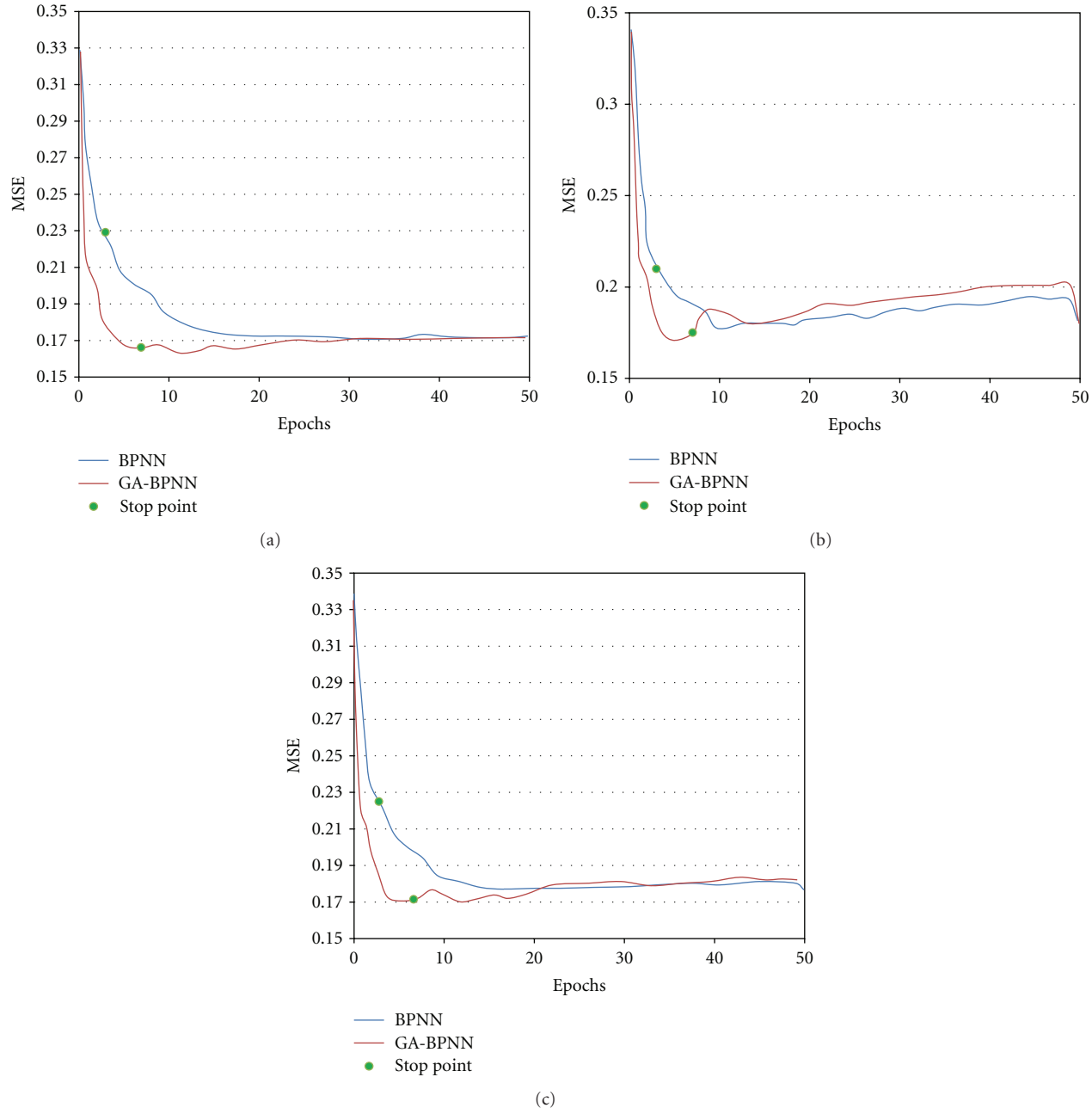


Figure 4: Comparison of the MSE between BPNN and GA-based BPNN for different data: (a) training data, (b) testing data, (c) database (including training and testing data).

generated into matting pool. The step was repeated until four-fifth of population was altered.

The features of the operator were that using four crossover points could match uniformly, in other words, the beginning of the string would not always separate from the end of the string. Secondly, chromosomes might include genes that never appear; it was because our paper used the blend crossover method (α value = 0.5) [27].

4.6. Mutation. The mutation operator used in this study was nonuniform mutation [28]. Compared with random mutation, the nonuniform mutation could change the interval for mutating depending on iterations. The genes would be mutated by (1)

$$c_i' = \begin{cases} c_i + \Delta(t, b_i - c_i), & \text{if } \tau = 0 \\ c_i - \Delta(t, c_i - a_i), & \text{if } \tau = 1, \end{cases} \quad (1)$$

where t is current generation, a_i and b_i are the initial ranges of lowest and highest limits, τ is a random number which may have a value of zero or one, and calculate Δ using (2)

$$\Delta(t, y) = y\left(1 - r^{(1-(t/g_{\max}))^b}\right), \quad (2)$$

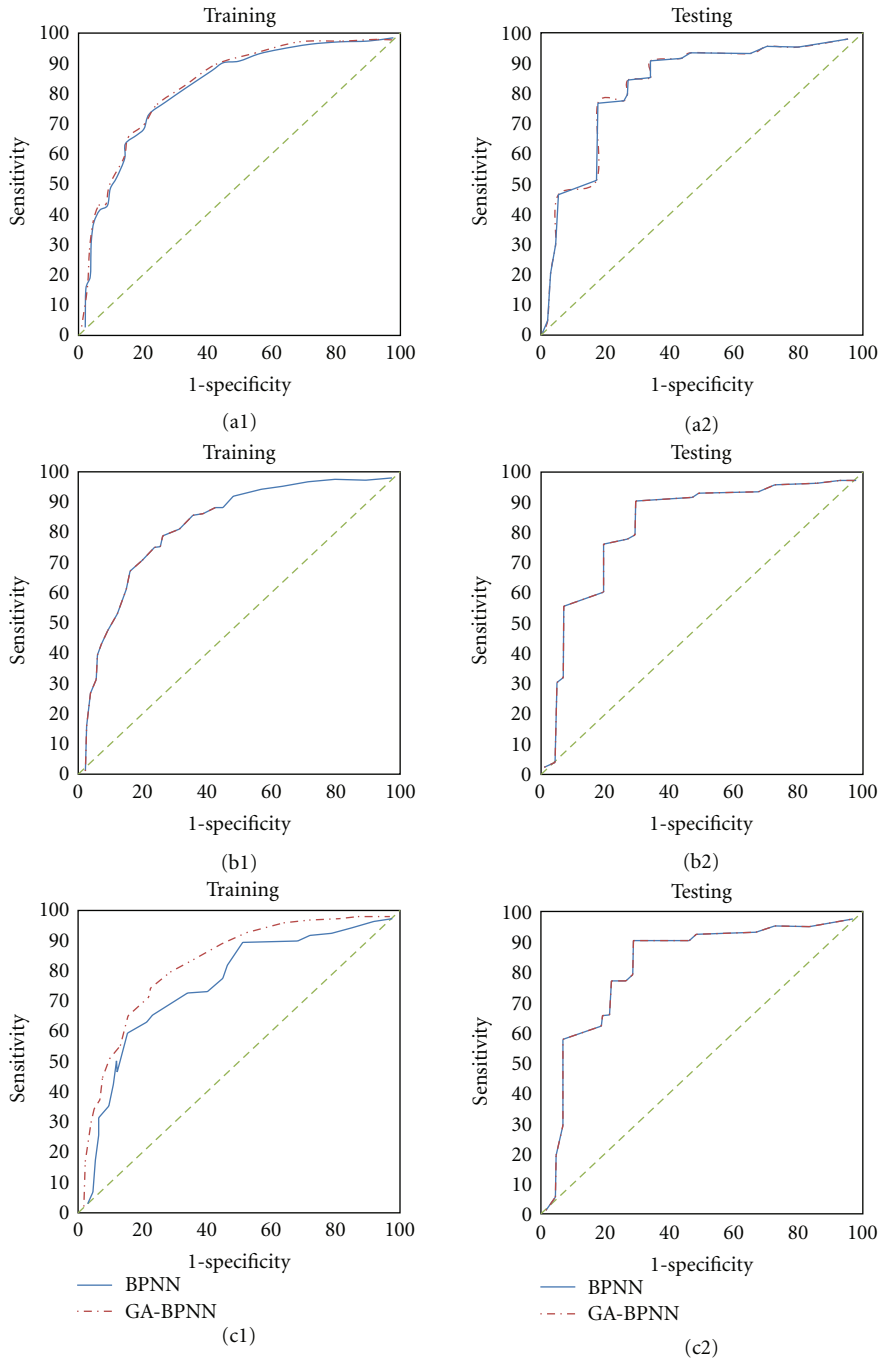


Figure 5: The ROC curves of the original BPNN and the GA-based BPNN for different ranges, (a1) and (a2) −0.5 to 0.5 range for the training and testing data, (b1) and (b2) −1 to 1 range for the training and testing data, (c1) and (c2) −5 to 5 range for the training and testing data.

where r is a random number from the interval [0,1], $g_{\max}$ is the maximum number of generations, and b is a parameter which determines the level of dependency on the number of iterations (it is equal to five here) [29].

The feature of the operator was as follows: the operator would make a uniform search in the initial space when t was small and become narrower in later generations.

4.7. Stopping Criterion. The algorithm would terminate after 100 generations, because it had almost converged.

5. Results and Discussion

In this paper, the normal backpropagation algorithm has been used in ANN. Regarding the learning rate chosen 0.01,

the chosen 10 nodes for hidden layer and using 20% dataset for testing, have been reported in our previous study [10]. Moreover, in order to avoid the overfitting, the neural network would stop training when the validation error started to go up (see Figure 4, the point signed with green point). In these figures, we could make sure that the neural networks converged rapidly with pretraining by genetic algorithm.

The paper calculated the area under ROC curve (AUC) with different initial range (Figure 5). The differences between Figure 5 (a1) and (a2) ($AUC_{modeling}$ = 0.858 and $AUC_{testing}$ = 0.802), and Figure 5 (b1) and (b2) ($AUC_{modeling}$ = 0.849 and $AUC_{testing}$ = 0.831) seemed to be similar, but pretty different in Figure 5 (c1) and (c2) ($AUC_{modeling}$ = 0.778 and $AUC_{testing}$ = 0.849). It means that if the initial range of GA parameter is smaller it is able to get better testing result.

To find the optimal initial weights however was a difficult task. Firstly, it had misgivings about not only the over-fitting from backpropagation in neural networks but also the tendency toward minimal mean square error in training data only without consideration of validation data in genetic algorithms. Secondly, the searching space of genetic algorithm might be limited by the initial range of the initial weights. Last, the advantages of using genetic algorithms compared with our previous study [10] should be based on the performance of the neural networks on the testing datasets, instead of the minimal square error only in the modeling datasets.

Another consideration of minimal improvement of genetic algorithms in this study was the ratio between the number of chromosomes in a generation (i.e., population size) and the length of a chromosome was small. It might be the reason why the genetic algorithm cannot search extensively to reach the optimum.

6. Conclusions and Future Work

The study results showed that the genetic algorithm obtained a good result of AUC of 0.858 ± 0.00493 on modeling data and 0.802 ± 0.03318 on testing data for small range of initial parameter. They were slightly better than the results of our previous study (0.868 ± 0.00387 and 0.796 ± 0.02559, resp.). Thus, the preliminary study for only using simple GA has been proved to be effective for improving the accuracy of artificial neural networks. However, the genetic algorithm should be further modified to improve the performance because of the data of our hip fracture cases were highly nonlinear and complex. Our future work is to try different ways of coding schemes to increase the efficiency of genetic operations, change the ratio between chromosome length and population size to extend the searching space, and investigate the effects of different initial ranges of initial weights on the network performance.

The works that should be done to improve the algorithms in the future are as follows: firstly, a better fitness function is designed to prevent the overfitting in genetic algorithm. Secondly, the criterion of stopping the algorithm can try some other methods, for example, to stop when the best chromosome does not change for a certain number of generations or the chromosomes with similar minimal mean square errors reach a certain number. Finally, other classification methods, such as neurofuzzy algorithms [30], support vector machines [9], and particle swarm optimizations [31] would be a good candidate for improving this prediction accuracy.

Acknowledgments

This research was supported by the National Science Council (NSC) of Taiwan (Grant no. NSC98-2915-I-155-005), the Department of Education grant of Excellent Teaching Program of Yuan Ze University (Grant no. 217517) and the Center for Dynamical Biomarkers and Translational Medicine supported by National Science Council (Grant no. NSC 100-2911-I-008-001).

References

[1] O. Johnell and J. A. Kanis, "An estimate of the worldwide prevalence and disability associated with osteoporotic fractures," *Osteoporosis International*, vol. 17, no. 12, pp. 1726–1733, 2006.

[2] C. Cooper, E. J. Atkinson, S. J. Jacobsen, W. M. O'Fallon, and L. J. Melton, "Population-based study of survival after osteoporotic fractures," *American Journal of Epidemiology*, vol. 137, no. 9, pp. 1001–1005, 1993.

[3] C. L. Leibson, A. N. A. Tosteson, S. E. Gabriel, J. E. Ransom, and L. J. Melton, "Mortality, disability, and nursing home use for persons with and without hip fracture: a population-based study," *Journal of the American Geriatrics Society*, vol. 50, no. 10, pp. 1644–1650, 2002.

[4] J. Magaziner, E. Lydick, W. Hawkes et al., "Excess mortality attributable to hip fracture in White women aged 70 years and older," *American Journal of Public Health*, vol. 87, no. 10, pp. 1630–1636, 1997.

[5] G. S. Keene, M. J. Parker, and G. A. Pryor, "Mortality and morbidity after hip fractures," *British Medical Journal*, vol. 307, no. 6914, pp. 1248–1260, 1993.

[6] A. M. Jette, B. A. Harris, P. D. Cleary, and E. W. Campion, "Functional recovery after hip fracture," *Archives of Physical Medicine and Rehabilitation*, vol. 68, no. 10, pp. 735–740, 1987.

[7] J. Shawe-Taylor, P. L. Bartlett, R. C. Williamson, and M. Anthony, "Structural risk minimization over data-dependent hierarchies," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1926–1940, 1998.

[8] C. Campbell, "Kernel methods: a survey of current techniques," *Neurocomputing*, vol. 48, pp. 63–84, 2002.

[9] J. Shawe-Taylor and S. Sun, "A review of optimization methodologies in support vector machines," *Neurocomputing*, vol. 74, no. 17, pp. 3609–3618, 2011.

[10] Y. C. Chou, *Predicting the risk of hip bone fracture for elders in Taiwan by ensemble back-propagation neural networks*, BS thesis, Department of Mechanical Engineering, University of Yuan Ze University, 2009.

[11] G. Li, H. Alnuweiri, Y. Wu, and H. Li, "Acceleration of back propagation through initial weight pre-training with delta rule," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 580–585, April 1993.

[12] Y. Lee, S. H. Oh, and M. W. Kim, "The effect of initial weights on premature saturation in back-propagation learning," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '91)*, pp. 765–770, July 1991.

[13] A. J. Al-Shareef and M. F. Abbod, "Neural networks initial weights optimisation," in *Proceedings of the 12th International Conference on Modelling and Simulation (UKSim '10)*, pp. 57–61, March 2010.

[14] D. Venkatesan, K. Kannan, and R. Saravanan, "A genetic algorithm-based artificial neural network model for the optimization of machining processes," *Neural Computing and Applications*, vol. 18, no. 2, pp. 135–140, 2009.

[15] R. S. Sexton and J. N. D. Gupta, "Comparative evaluation of genetic algorithm and backpropagation for training neural networks," *Information Sciences*, vol. 129, no. 1–4, pp. 45–59, 2000.

[16] D. J. Montana and L. Davis, "Training feedforward neural networks using genetic algorithms," in *Proceedings of the Joint Conference on Artificial Intelligence*, vol. 1, pp. 762–767, 1989.

[17] B. G. Eriksson and V. Sundh, "Prediction of seven-year survival by artificial neural network and logistic regression: a comparison of results from medical and social data among 70-year-olds in Göteborg, Sweden," Eriksson & Sundh, 2010.

[18] D. J. Montana, *Neural Network Weight Selection Using Genetic Algorithms*, Bolt Beranek and Newman, 1994.

[19] L. Davis, Ed., *Genetic Algorithms and Simulated Annealing*, Pitman, London, UK, 1991.

[20] D. E. Goldberg, *Genetic Algorithms in Machine Learning*, Addison-Wesley, 1988.

[21] T. Y. Lan, S. M. Hou, C. Y. Chen et al., "Risk factors for hip fracture in older adults: a case-control study in Taiwan," *Osteoporosis International*, vol. 21, no. 5, pp. 773–784, 2010.

[22] L. M. Salchenberger, E. M. Cinar, and N. A. Lash, "Neural networks: a new tool for predicting thrift failures," *Decision Sciences*, vol. 23, pp. 899–916, 1992.

[23] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.

[24] P. Cunningham, J. Carney, and S. Jacob, "Stability problems with artificial neural networks and the ensemble solution," *Artificial Intelligence in Medicine*, vol. 20, no. 3, pp. 217–225, 2000.

[25] W. S. Sarle, "Stopping training and other remedies for overfitting," in *Proceeding of the 27th Symposium on the Inter Face*, 1995.

[26] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling real-coded genetic algorithms: operators and tools for behavioural analysis," *Artificial Intelligence Review*, vol. 12, no. 4, pp. 265–319, 1998.

[27] L. J. Eshelman and J. D. Schaffer, "Real-coded genetic algorithms and interval-schemata," *Foundation of Genetic Algorithms*, pp. 187–202, 1993.

[28] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1992.

[29] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, 1991.

[30] J. W. F. Catto, M. F. Abbod, D. A. Linkens, and F. C. Hamdy, "Neuro-fuzzy modeling: an accurate and interpretable method for predicting bladder cancer progression," *Journal of Urology*, vol. 175, no. 2, pp. 474–479, 2006.

[31] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, pp. 33–37, 2007.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2012, Article ID 327861, 10 pages
doi:10.1155/2012/327861

Research Article

Artificial Neural Network-Statistical Approach for PET Volume Analysis and Classification

Mhd Saeed Sharif,[1] Maysam Abbod,[1] Abbes Amira,[2] and Habib Zaidi[3, 4, 5]

[1] *Department of Electronic and Computer Engineering, School of Engineering and Design, Brunel University, Uxbridge UB83PH, UK*
[2] *Nanotechnology and Integrated BioEngineering Centre, University of Ulster, Newtownabbey BT37OQB, UK*
[3] *Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland*
[4] *Geneva Neuroscience Center, Geneva University, CH-1205 Geneva, Switzerland*
[5] *Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands*

Correspondence should be addressed to Mhd Saeed Sharif, mhd.sharif@brunel.ac.uk

Received 14 September 2011; Accepted 12 January 2012

Academic Editor: Jiann-Shing Shieh

The increasing number of imaging studies and the prevailing application of positron emission tomography (PET) in clinical oncology have led to a real need for efficient PET volume handling and the development of new volume analysis approaches to aid the clinicians in the clinical diagnosis, planning of treatment, and assessment of response to therapy. A novel automated system for oncological PET volume analysis is proposed in this work. The proposed intelligent system deploys two types of artificial neural networks (ANNs) for classifying PET volumes. The first methodology is a competitive neural network (CNN), whereas the second one is based on learning vector quantisation neural network (LVQNN). Furthermore, Bayesian information criterion (BIC) is used in this system to assess the optimal number of classes for each PET data set and assist the ANN blocks to achieve accurate analysis by providing the best number of classes. The system evaluation was carried out using experimental phantom studies (NEMA IEC image quality body phantom), simulated PET studies using the Zubal phantom, and clinical studies representative of nonsmall cell lung cancer and pharyngolaryngeal squamous cell carcinoma. The proposed analysis methodology of clinical oncological PET data has shown promising results and can successfully classify and quantify malignant lesions.

1. Introduction

Positron emission tomography (PET) volume analysis is vital for various clinical applications including artefact reduction and removal, tumour quantification in staging, a process which analyses the development of tumours over time, and to aid in radiotherapy treatment planning [1, 2]. PET has been progressively incorporated into the management of patients. Results of clinical studies using fluorodeoxyglucose (FDG)-PET have demonstrated its added value in the diagnosis, staging, and evaluation of response to therapy [3–5]. The utilisation of advanced high performance analysis approaches will be useful in aiding clinicians in diagnosis and radiotherapy planning. Although the task of medical volume analysis appears simple, the reality is that an indepth knowledge of the anatomy and physiology is required to perform such task on clinical medical images. Essentially, the expert observes a particular slice, determines borders between regions, and classifies each region. This is commonly completed slice by slice for a 3D volume and requires a reslicing of data into the transaxial, sagittal, and coronal planes. In addition to this, identifying smaller slice features and contrast modifications are often required. Although, for a typical 3D data set, the entire expert manual analysis can take several hours to complete, this approach is perhaps the most reliable and accurate method of medical volume analysis. This is due to the immense complexity of the human visual system, a system well suited to this task [6–9].

The main challenges associated with PET are the statistical noise and the low resolution which results in a significant partial volume effect. This effect should be reduced to the minimum level so that the required information can be precisely extracted from the analysed volume. Analysing and extracting the proper information from PET volumes can be performed by utilising analysis and classification approaches which provide rich information compared to what can be extracted from visual interpretation of the PET volumes alone. The need for accurate and fast analysis approaches of imaging data motivated the exploitation of artificial intelligence (AI) technologies. Artificial neural network (ANN) is one of the powerful AI techniques that has the capability to learn from a set of data and construct weight matrices to represent the learning patterns. The ANN is a mathematical model which emulates the activity of biological neural networks in the human brain.

ANNs had great success in many applications including pattern classification, decision making, forecasting, and adaptive control [10]. Competitive neural networks with wavelet invariant moments have been used in [11] to detect the arbitrary pose of the face and verify the candidate face regions. In this work, the user selects some wavelet invariant moments and feed them into the neural network. There is just one active output out of three outputs of the network which correspond to frontal face, nonface, and profile face. Therefore there is no full classification for the whole image since it is just restricted to some selected features. Supervised competitive learning algorithm has been used to train competitive learning neural network with the extracted features set for handwritten Chinese character recognition [12]. A number of research studies has been carried out in the medical field utilising ANN for image segmentation and classification using various medical imaging modalities. Multilayer perceptron (MLP) neural network have been used in [13] to identify breast nodule malignancy using sonographic images. A multiple classifier system using five types of ANNs and five sets of texture features extraction for the characterisation of hepatic tissue from CT images was presented in [14]. Kohonen self-organising neural network for segmentation and a multilayer backpropagation neural network for classifying multispectral MR images have been used in [15]. Kohonen ANN was also used for image segmentation in [16]. Computer-aided diagnostic (CAD) scheme to detect nodules in lung using multiresolution massive training artificial neural network (MTANN) is presented in [17].

Many other approaches were used for medical image segmentation. A fuzzy locally adaptive Bayesian (FLAB) segmentation for automatic lesion volume delineation has been proposed in [18]. The FLAB approach was compared with a threshold approach as well as fuzzy hidden Markov chains (FHMCs) and the fuzzy C-Means (FCM) algorithms. In this comparison, phantom data sets were used to validate the performance of the proposed algorithm. A new fuzzy segmentation technique adapted to typical PET data was also recently proposed [19]. First, PET images smoothed using a nonlinear anisotropic diffusion filter are added as a second input to the FCM algorithm to incorporate spatial information. Thereafter, a methodology was developed to integrate the a trous wavelet transform in the standard FCM algorithm to allow handling of heterogeneous lesions' uptake. An unsupervised MRI segmentation method based on self-organising feature map has been reported in [20]. An extra spatial information about a pixel region was obtained using a Markov random field (MRF) model. The utilisation of MRF term improves the segmentation results without extra data samples in the training set. The cooperation of MRF into SOFM has shown its great potentials as MRF term models the smoothness of the segmented regions. It verifies that the neighboring pixels should have similar segmentation assignment unless they are on the boundary of two distinct regions. However, it is not clear how the proposed approach is able to differentiate between two regions.

This paper aims to develop a robust PET volume analysis system using ANN combined with Bayesian information criterion (BIC). The initial investigation of this system was published in [21]. Two methodologies have been investigated using the competitive neural network (CNN) and the learning vector quantisation neural network (LVQNN) for classifying PET volumes. BIC has been used in this system to select the optimal number of classes for each PET data set and feed it into the ANN blocks. The CNN and LVQNN outputs have been evaluated using two PET phantom data sets, a clinical PET volume of nonsmall cell lung cancer patient, and PET volumes from seven patients with laryngeal tumours.

This paper is organised as follows. Section 2 presents theoretical background for the main system components including BIC, CNN, and LVQNN. The proposed medical volume analysis system is described in Section 3. Results and analysis are illustrated in Section 4, and finally conclusions are presented in Section 5.

2. Theoretical Background

2.1. Bayesian Information Criterion. Bayesian information criterion (BIC) is employed to approximate the Bayes factor which is consequently used to compare a series of rival theories. BIC is one hypothesis testing approach which uses Bayesian inference. BIC has gained notoriety as a significant approach for model selection and has been used in contexts varying from image processing and analysis [22], to biological and sociological research [23, 24]. Although the BIC does not allow for spatial correlations between voxels to be considered, it does provide a useful strategy for the comparison of contesting models. The model which applied to the data set is denoted by M_K and is defined from the Gaussian distribution utilised and its associated parameters θ_K, where K is a prior number of classes. To compare two competing hypotheses M_K and $M_{K'}$, the posterior probability of each model with voxel labelling Y is computed as follows [23]:

$$p(M_K \mid Y) = \frac{p(Y \mid M_K)p(M_K)}{\sum_{l=1}^{K} p(Y \mid M_l)p(M_l)}, \tag{1}$$

$p(M_K)$ is the prior probability of model M_K, and in this case the number of classes considered is taken to be equally likely a priori, therefore,

$$p(M_K) = \frac{1}{K_{\max}} \quad (K = 1, 2, \ldots, K_{\max}). \tag{2}$$

The ratio of posteriors, $p(Y \mid M_K)/p(YM_L)$, is commonly referred to as the Bayes factor for model M_K versus model M_L. This factor equates to the posterior odds of one hypothesis when the prior probabilities of the two hypotheses are likely to be equal. This also provides a measure of evidential weight provided for the data or against the null hypothesis. $p(Y \mid M_K)$ represents the integrated likelihood of model M_K rather than the maximised likelihood and is given by:

$$p(Y \mid M_K) = \int p(Y \mid \theta_K, M_K) p(\theta_K) d\theta_K, \tag{3}$$

$p(Y \mid \theta_K, M_K)$ is the usual likelihood, and $p(\theta_K)$ is the prior assumed to be equally probable for all M_K. Evaluating this integral is combinatorially difficult; however, a good approximation to the integrated likelihood is given by the BIC:

$$\mathrm{BIC} = 2 \log p\left(Y \mid \hat{\theta}_K, M_K\right) - D_K \log N, \tag{4}$$

where $\hat{\theta}_K$ is the maximum likelihood estimator of θ_K obtained from Gaussian mixture fitting:

$$p\left(Y \mid \hat{\theta}_K, M_K\right) = \prod_{i=1}^{N} \sum_{j=1}^{K} \hat{P}_j \Phi_j\left(Y_i \mid \hat{\theta}_j\right), \tag{5}$$

where N is the dimensionality of the data vectors, and D_K is the cardinality of the parameter set employed. The Bayes factor shown in (6) can be approximated by computing the difference of BIC terms, which are the results of model fitting for different numbers of classes, K and L [23]:

$$2 \log \frac{p(Y \mid M_K)}{p(Y \mid M_L)} \approx \mathrm{BIC}(K) - \mathrm{BIC}(L). \tag{6}$$

Although the absolute value of the BIC is not individually informative due to comparison with the null hypothesis, the disparity between BIC values for competing models provides evidence specifying the use of one model against another.

Expectation maximisation (EM) algorithm is used to find the maximum likelihood estimation for each class in the processed PET volume. The maximum likelihood estimation of X (MLE) based on the incomplete observed data Y can be defined as follows:

$$\mathrm{MLE} = \arg\max\{\log p(Y \mid X)\}, \tag{7}$$

where $\log p(Y \mid X)$ is the log likelihood of Y given X. The maximum likelihood estimation is calculated first based on Gaussian distribution, which is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to classify around a single mean value:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \tag{8}$$

where μ is the mean for each class, and σ is the standard deviation.

The maximum likelihood estimation for each segment is finally obtained utilising this probability beside the histogram of each level in the processed slice. The mean and standard deviation, for each class are also calculated based on histogram calculation, and according to these statistical details mthe signal-to-noise ratio (SNR) for each class is obtained as well to evaluate the level of the signal in each segment [25]. SNR can be calculated according to the following equation:

$$\mathrm{SNR} = \frac{\mu}{\sigma}. \tag{9}$$

2.2. Competitive Neural Network. Competitive neural networks can learn to detect regularities and correlations in their input and adapt their future responses to that input accordingly. The neurons of competitive networks learn to recognise groups of similar input vectors. Self-organising maps learn to recognise groups of similar input vectors in such a way that neurons physically near each other in the neuron layer respond to similar input vectors.

CNN consists of a single layer, the N neurons in this competitive layer distribute themselves to recognise frequently presented input vector. The weights are applied to the input vector using negative Euclidean distance approach. The layer's net input is calculated by combining its weighted inputs and biases. CNN is trained using two approaches: the first one is sequential order incremental training which trains the network with weight and bias learning rules with sequential updates. The other approach is random order incremental training which trains a network with weight and bias learning rules with incremental updates after each presentation of an input. Where in this type of neural network inputs are presented in random order [26].

2.2.1. Competitive Learning. The learning rule used for CNN is based on Kohonen rule [27, 28], the neuron, whose weight vector was closest to the input vector I, is updated to be even closer. The result is that the winning neuron is more likely to win the competition the next time a similar vector is presented, and less likely to win when a very different input vector is presented. As more and more inputs are presented, each neuron in the layer closest to a group of input vectors soon adjusts its weight vector toward those input vectors. Eventually, if there are enough neurons, every class of similar input vectors will have a neuron that outputs 1 when a vector in the class is presented, while outputting a 0 at all other times. Thus, the competitive network learns to categorise the input vectors it sees each time. The weight W_i of a neuron i at iteration q is adjusted as follows:

$$W_i(q) = W_i(q-1) + \alpha(I(q) - W_i(q-1)), \tag{10}$$

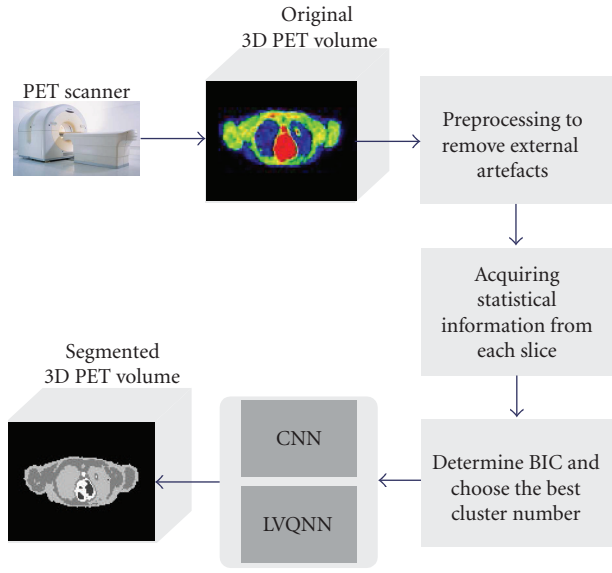


Figure 1: Proposed system for oncological PET volume analysis.

where the learning rate α is set in the proposed PET application to 0.6. This value has been chosen by experiment, as the learning rate value near zero results in slow learning, however, values near one result in faster learning, but the weight vector stays unstable. To solve the problem of the number of classes in the CNN, BIC has been utilised to determine the optimal number of classes for each type of the processed data set.

2.3. Learning Vector Quantisation Neural Network. Learning vector quantisation neural network is a hybrid network, it uses unsupervised and supervised learning to form the classification. LVQNN has two layers: the first layer calculates weighted inputs using negative Euclidean distance approach. The second layer has neurons with pure-line activation function and calculates weighted input using dot product weight approach. There are no biases used in LVQNN. LVQ learning in the competitive layer is based on a set of input/target pairs. Each target vector has a single 1, and the rest of its elements are 0. The 1 tells the right classification of the associated input.

LVQNN is more efficient than CNN in case of large number of inputs. The optimisation procedure implicitly in this network yields the class means by estimating optimised assignment for each class [29].

2.3.1. LVQ Learning Rule. The learning rule in LVQNN combines competitive learning with supervised learning approach [30, 31], which requires a set of examples of the suitable network behavior as follows:

$$\{I_1, t_1\}, \{I_2, t_2\}, \ldots, \{I_Q, t_Q\}. \tag{11}$$

Each of these target vectors should contain all zeros except for a single 1. This 1 indicates the class to which the assigned input vector belongs. At each iteration, an input vector I is presented to the LVQNN, and the distance from I to each prototype vector is calculated. After competition, one neuron wins, and the ith element of the competition outputs vector is set to 1, then this outputs are multiplied by the weight matrix to get the final output, which has one nonzero element indicating the vector class. Kohonen rule is used as a learning rule; so if the input vector I is classified correctly, then the weight of the winning neuron needs to be moved toward I as follows:

$$W_i(q) = W_i(q-1) + \alpha(I(q) - W_i(q-1)), \tag{12}$$

however, if I is classified incorrectly then the weight of the winning neuron needs to be moved away from I as follows:

$$W_i(q) = W_i(q-1) - \alpha(I(q) - W_i(q-1)). \tag{13}$$

3. The Proposed System

The proposed medical volume analysis system is illustrated in Figure 1. The 3D PET volume acquired from PET scanner goes through the preprocessing block, where thresholding, histogram equalisation, and median filter are utilised to remove external artefacts and enhance the quality of each slice features. The EM algorithm is then used to find the maximum likelihood estimation for each class in the enhanced volume. The normal Gaussian density is then calculated according to (8) for the values 0–255 of each class. The mean standard deviation and the class probability (CP) for each generated class are also calculated, according to these statistical details, the SNR for each class is also obtained. The classes probabilities must add up to one. The BIC values are calculated and plotted against different values of K to determine the optimal number of classes for each processed slice. This number is fed to CNN and LVQNN, where the processed volume can be then classified. The outputs of these approaches are compared in the next step, and the best classified outputs are selected and displayed.

Table 1: Statistical information about the best class number for experimental phantom data set.

CN	μ	σ	SNR	CP
1	1.000986	0.250000	4.003944	0.469466
2	3.528585	1.602465	2.201973	0.240734
3	13.024477	7.048304	1.847888	0.247406
4	57.982584	45.253936	1.281271	0.042394

4. Results and Analysis

4.1. Phantom Studies

4.1.1. Experimental Phantom Studies. The first data set used in this study is obtained using the NEMA IEC image quality body phantom which consists of an elliptical water filled cavity with six spherical inserts suspended by plastic rods of volumes 0.5, 1.2, 2.6, 5.6, 11.5, and 26.5 mL. The inner diameters of these spheres are 10, 13, 17, 22, 28, and 37 mm. The PET image volume consists of 168 × 168 × 66 voxels, each voxel has dimensions of 4.07 mm × 4.07 mm × 5 mm corresponding to voxel volume of 0.0828 mL. This phantom was extensively used in the literature for the assessment of image quality and validation of quantitative procedures [32–35]. Other variants of the multisphere phantoms have also been suggested [36]. The PET scanner used for acquiring the data is the Biograph 16 PET/CT scanner (Siemens Healthcare, Erlangen, Germany) operating in 3D mode [37]. Following Fourier rebinning and model-based scatter correction, PET images were reconstructed using two-dimensional iterative normalised attenuation-weighted ordered subsets expectation maximisation (NAW-OSEM). CT-based attenuation correction was used to reconstruct the PET emission data. The default parameters used were ordered OSEM iterative reconstruction with four iterations and eight subsets followed by a postprocessing Gaussian filter (kernel full-with half-maximal height, 5 mm).

To choose the optimal number of classes for each slice in the processed PET phantom volume, different values of K have been used. For the proposed application, BIC values are calculated incrementally increasing from $K = 2$ to $K = 8$. K is not further increased, as in this medical application, any additional separation is unnecessary. BIC values tend to increase indefinitely as the number of components increases. An increase in BIC value indicates improved model fit; however, these values typically stabilise on an approximate curve plateau, the beginning of which is usually taken to indicate the optimal K value. The plot of BIC values with K for experimental phantom data set shows that the best K value can be obtained at 4, as illustrated in Figure 2.

The mean standard deviation, SNR, and class probability (CP) for each slice have been calculated to analyse all the recommended classes in each slice. Table 1 presents these values for the optimal class number (CN) for experimental phantom data set.

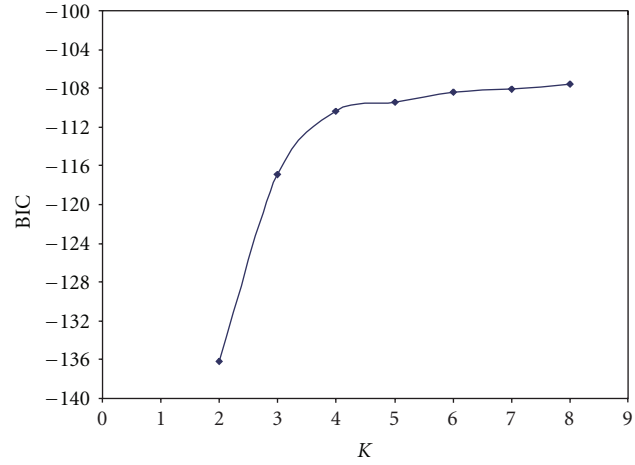


Figure 2: Plot of BIC values for experimental phantom data set, scaled by a factor of 1000.

Table 2: Tumours characteristics for the second data set with 2 types of voxels.

Tumours	Isotropic voxels		Nonisotropic voxels	
	Position	Size	Position	Size
1	Slice 68	2 voxels	Slice 142	2 voxels
2	Slice 57	3 voxels	Slice 119	3 voxels
3	Slice 74	2 voxels	Slice 155	2 voxels

4.1.2. Simulated Phantom Studies. The second data set consists of Monte Carlo simulations of the Zubal anthropomorphic model where two volumes were generated. The first volume contains a matrix with isotropic voxels, the size of this volume is 128 × 128 × 180 voxels. The second volume contains the same matrix of the first one but without isotropic voxels, it has a size of 128 × 128 × 375 voxels. The voxel size in both volumes is 5.0625 mm × 5.0625 mm × 2.4250 mm. The second data volume has 3 tumours in the lungs whose characteristics are given in Table 2 [38].

For isotropic voxels in simulated phantom data set, the optimal class number obtained from BIC plot is 5 classes, as shown in Figure 3. The plot of BIC values flattens to an approximate plateau at $K = 5$, and for this reason, this statistical model selection test determines 5 to be the most appropriate number of labels for classifying this data set. The optimal number of classes is the same for tumours 1, 2, and 3. For volume with nonisotropic voxels in the simulated phantom data set, the optimal number of classes obtained from BIC plot is also 5 classes for all the three tumours.

Table 3 shows the statistical information about the best number of classes for simulated phantom data set, tumour 1, with isotropic voxels. While the statistical details about tumour 2 are presented in Table 4.

Analysing the statistical details about tumour 3 shows that there is a small difference between the SNR values calculated for classes 2, 3, 4, and 5, as presented in Table 5. Similar analysis has been performed for the simulated

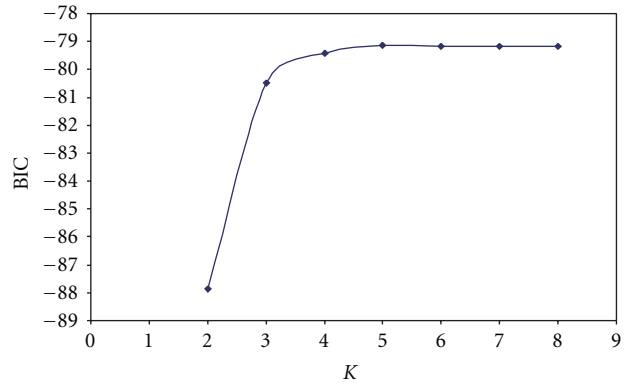


Figure 3: Plot of BIC values for isotropic phantom data set, scaled by a factor of 1000.

Table 3: Statistical information about the best class number for the simulated phantom data set, tumour 1.

CN	μ	σ	SNR	CP
1	1.003849	0.250000	4.015396	0.409114
2	16.288944	6.399162	2.545480	0.247941
3	26.890255	9.694972	2.773628	0.197655
4	48.774962	18.579216	2.625243	0.120429
5	137.468019	57.911212	2.373772	0.024862

Table 4: Statistical information about the best class number for the simulated phantom data set, tumour 2.

CN	μ	σ	SNR	CP
1	1.003102	0.250000	4.012408	0.399154
2	14.181862	6.814574	2.081107	0.207690
3	26.481251	9.468495	2.796775	0.227157
4	51.409594	19.963755	2.575146	0.142332
5	133.638905	56.426063	2.368389	0.023667

Table 5: Statistical information about the best class number for the simulated phantom data set, tumour 3.

CN	μ	σ	SNR	CP
1	1.002889	0.250000	4.011556	0.407105
2	14.616784	6.508255	2.245883	0.207151
3	25.914234	8.984076	2.884462	0.223710
4	47.367310	18.536074	2.555412	0.136315
5	141.792551	60.934125	2.326980	0.025720

phantom data set with nonisotropic voxels. Table 6 compares the SNR and class probability for tumours 1, 2, and 3 with nonisotropic voxels.

The optimum chosen class number is fed to both CNN and LVQNN, where both have clearly classified all spheres in experimental phantom data set, as illustrated in Figure 4. Clear classes have been also obtained for the simulated phantom data set with isotropic and nonisotropic voxels. For example, Figure 5 presents the original and classified slice for the simulated phantom data set with isotropic voxels and tumour 1.

Better performance has been achieved using LVQNN rather than CNN; however, the required time for classifying each slice in the processed volume using LVQNN is higher than the time required for processing each slice using CNN.

Two performance metrics have been employed to evaluate the performance of the proposed ANNs. A confusion matrix is a visualisation tool typically used in supervised and unsupervised learning approaches. Each row of the matrix represents the instances in a predicted class, while each column represents the instances in an actual class. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes: (the tumour and the remaining tissues in case of clinical PET data sets). The other performance metric approach is receiver-operating characteristic (ROC). This approach can be represented by plotting the fraction of true positives rate (TPR) versus the fraction of false positives rate (FPR), where the perfect point in the ROC curve is the point (0,1) [39]. Both ANNs have performed well in classifying both experimental and simulated data sets.

4.2. Clinical PET Volume

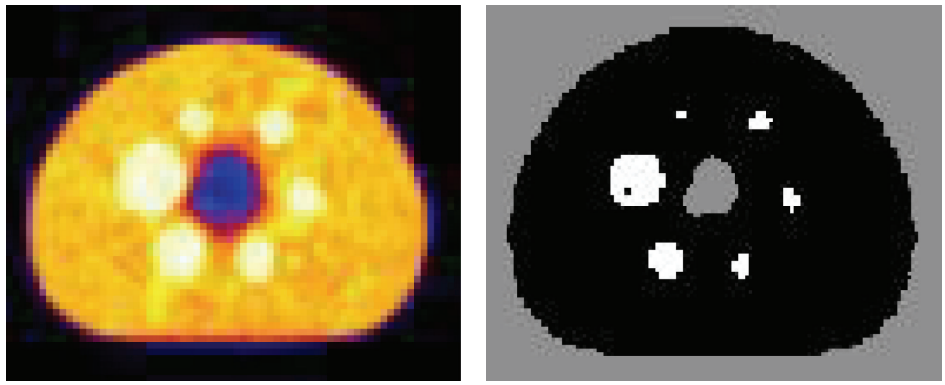
4.2.1. Clinical Data Set 1. Clinical PET volume of patient with histologically proven NSCLC (clinical Stage Ib-IIIb) who has undertaken a diagnostic whole-body PET/CT scan was used for assessment of the proposed classification technique. Patient fasted not less than 6 hours before PET/CT scanning. The standard protocol involved intravenous injection of ^{18}F-FDG followed by a physiologic saline (10 mL). The injected FDG activity was adjusted according to patient's weight using the following formula:

$$A(\text{MBq}) = \text{weight} * 4 + 20, \tag{14}$$

where MBq (megabecquerel) is the international system of units for radioactivity. As One Bq is defined as the activity of a quantity of radioactive material in which one nucleus decays per second. After 45 min uptake time, free-breathing PET and CT images were acquired. The data were reconstructed using the same procedure described for the phantom studies. The maximal tumour diameters measured from the macroscopic examination of the surgical specimen served as ground truth for comparison with the maximum diameter estimated by the proposed classification technique.

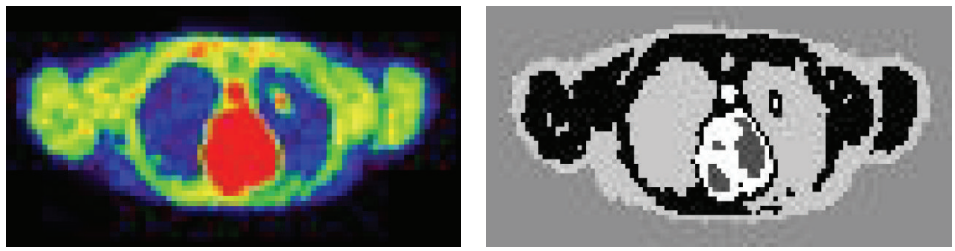
The optimal number of classes obtained from BIC plot for clinical PET volume of nonsmall cell lung cancer patient is 5 classes, as illustrated in Figure 6. While Table 7 shows the statistical information about these classes. A subjective evaluation based on the clinician knowledge has been carried out for the output of the proposed approaches. The classified slice using the proposed approaches has clear detection of the region of interest (ROI) as illustrated in Figure 7.

4.2.2. Clinical Data Set 2. The second clinical data set used in this study is PET volumes from seven patients with T3-T4 laryngeal squamous cell carcinoma. Prior to treatment, each patient underwent an FDG-PET study. Patients



(a) (b)

Figure 4: Experimental phantom data set: (a) original PET image and (b) classified image.



(a) (b)

Figure 5: Simulated phantom data set (tumour 1): (a) original PET and (b) classified image.

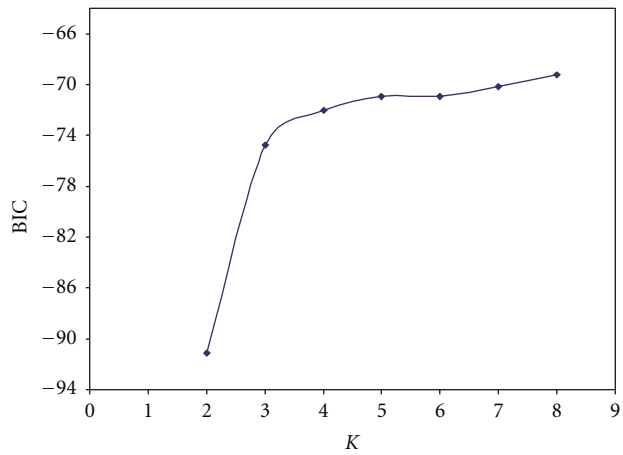


Figure 6: Plot of BIC values for clinical PET volumes of nonsmall cell lung cancer patient, scaled by a factor of 1000.

were immobilised with a customised thermoplastic mask (Sinmed, Reeuwijk, The Netherlands) fixed to a flat table-top to prevent complex neck movements. First, a 10 min transmission scan was obtained on the Siemens Exact HR camera (CTI, Knoxville, USA). Immediately after intravenous injection of 185–370 MBq (5–10 mCi) of FDG, a 1h dynamic 3D emission scan was performed. It consisted of eight frames with variable duration ranging from 90 to 600 s.

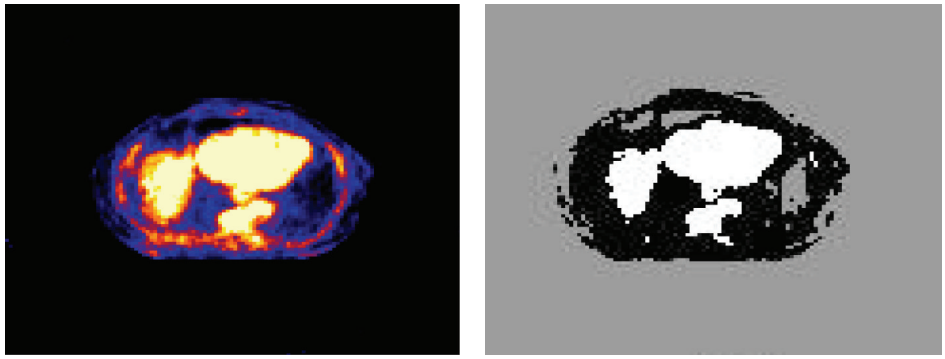
Table 6: SNR and CP for tumours 1, 2 and 3 in the simulated phantom data set with nonisotropic voxels.

CN	Tumour 1		Tumour 2		Tumour 3	
	SNR	CP	SNR	CP	SNR	CP
1	4.016184	0.407899	4.013160	0.402693	4.012848	0.396036
2	2.504101	0.239490	2.110565	0.190559	1.962133	0.176196
3	2.814225	0.205216	2.846712	0.222405	2.965170	0.237096
4	2.671825	0.123275	2.557056	0.158497	2.515199	0.161232
5	2.488246	0.024121	2.286361	0.025846	2.224181	0.029441

Table 7: Statistical information about the best class number for clinical PET volume of nonsmall cell lung cancer patient.

CN	μ	σ	SNR	CP
1	1.000840	0.250000	4.003360	0.509741
2	2.911473	1.142062	2.549312	0.163043
3	9.035780	4.213254	2.144608	0.146897
4	32.665878	16.455158	1.985145	0.078470
5	167.509276	52.633532	3.182558	0.101848

All images were corrected for dead time, random, scatter attenuation and decay and then reconstructed using a 3D OSEM algorithm, as used in the clinics for patients with



(a) (b)

Figure 7: Clinical PET data: (a) original PET image and (b) classified image.

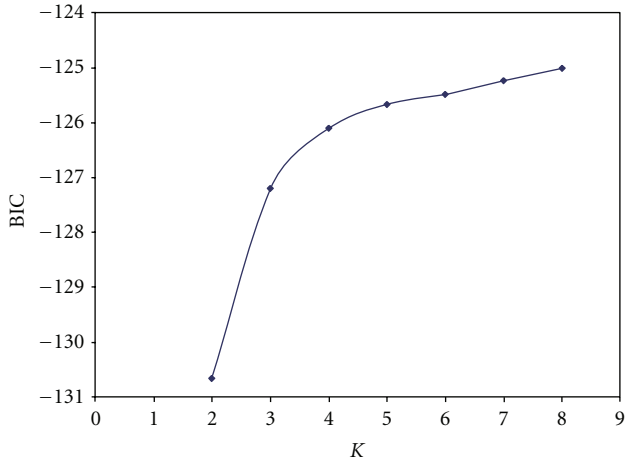


Figure 8: Plot of BIC values for clinical PET data set 2, scaled by a factor of 1000.

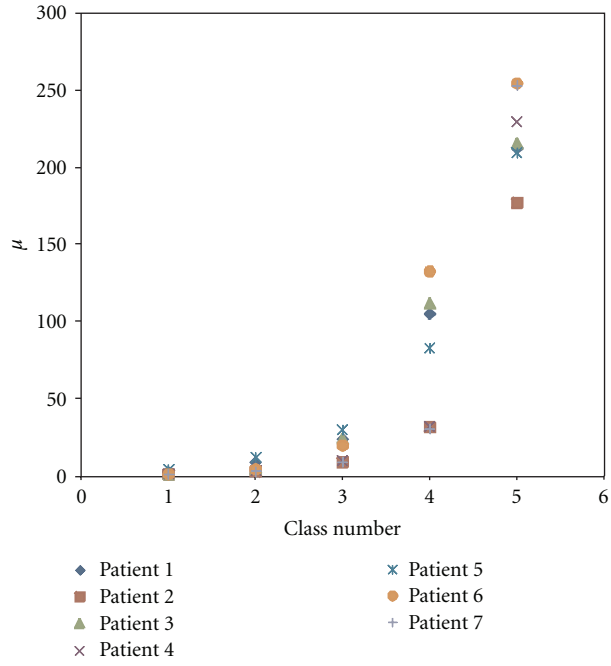


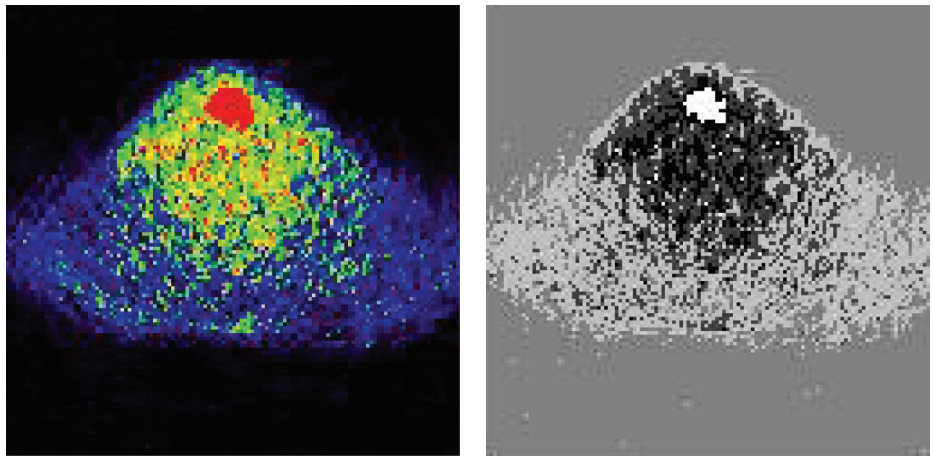
Figure 9: Clinical PET data set 2: the μ values for classes one to five.

head and neck tumours [40–42]. The size of this data set is $128 \times 128 \times 47$ voxels for each patient.

In the case of clinical data set, the plot of BIC values flattens to an approximate plateau at $K = 5$, and for this reason, this statistical model selection test determines 5 to be the most appropriate number of labels for the classification of this data set. For the data set in question, the K determined by the BIC plot corresponds precisely to the number of classification levels recommended by clinicians specifically for tumour quantification. The BIC provides a useful objective methodology for classification level selection. In particular, the BIC works efficiently for tumour quantification in oncological PET data and can be computed very rapidly. In the case of a volume with dimensionality $128 \times 128 \times 47$ voxels, histogram computation takes approximately 1.9 seconds, and thereafter, one BIC value associated with a specific value of K can be calculated every 1.2 seconds. The total BIC model selection procedure takes approximately 8.4 seconds (for $K = 2, \ldots, 8$). The model timings are obtained using a single processor, 2.66 GHz, with 3 GB of RAM.

Using the proposed approach the optimum CN for each patient has been chosen. According to Figure 8 which illustrates the BIC values for different K values (from 2–8), the best CN for this data set is 5 classes. This CN is fed to both CNN and LVQNN to do the classification for each slice in the processed clinical data set.

All the statistical information about each class in this data set has been calculated, Figure 9 illustrates the μ values for each class of clinical data set 2 from patient number one to patient number seven. The calculated SNR was important to refer to the class which contains the ROI. A subjective evaluation based on the clinician knowledge has been carried out for the output of the proposed approach. The classified volumes using the proposed approaches have a clear detection of ROI. Laryngeal tumours from seven patients were clearly classified, Figure 10 illustrates the original and classified slice for patient 1 data set.



(a) (b)

Figure 10: Clinical PET data set 2 (patient 1): (a) original PET (128 × 128) and (b) classified slice (128 × 128).

5. Conclusion

An artificial intelligent statistical approach based on CNN and LVQNN was proposed for 3D oncological PET volume analysis. Experimental, simulated, and clinical PET studies of nonsmall cell lung cancer and pharyngolaryngeal squamous cell carcinoma were used to evaluate the performance of the proposed system. BIC and EM approaches were deployed to obtain the optimal number of classes, which was used by CNN and LVQNN to classify each slice in the processed PET volume. The mean, standard deviation, signal-to-noise ratio, and class probability were also calculated for each class. A detailed objective assessment together with subjective evaluation based on clinical knowledge was performed to characterise the performance of the proposed approach. Promising results were obtained, and the system appears to successfully classify and quantify lesions from clinical oncological PET studies.

Acknowledgments

This paper was supported by the Swiss National Science Foundation under Grant SNSF 31003A-125246, Geneva Cancer League, and the Indo Swiss Joint Research Programme ISJRP 138866.

References

[1] H. Zaidi and I. El Naqa, "PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 11, pp. 2165–2187, 2010.

[2] H. Zaidi, M. Diaz-Gomez, A. Boudraa, and D. O. Slosman, "Fuzzy clustering-based segmented attenuation correction in whole-body PET imaging," *Physics in Medicine and Biology*, vol. 47, no. 7, pp. 1143–1160, 2002.

[3] I. J. Kalet and M. M. Austin-Seymour, "Use of medical imagery in the delivery and planning of radiation therapy," *Journal of the American Medical Informatics Association*, vol. 4, no. 5, pp. 327–339, 1997.

[4] D. W. G. Montgomery and A. Amira, "Automated multiscale segmentation of oncological cerebral MR image volumes," in *Proceedings of the IEEE International Conference on Computer Systems and Information Technology*, 2005.

[5] D. Delbeke and W. H. Martin, "Positron emission tomography imaging in oncology," *Radiologic Clinics of North America*, vol. 39, no. 5, pp. 883–917, 2001.

[6] D. A. Mankoff, M. Muzi, and H. Zaidi, "Quantitative analysis in nuclear oncologic imaging," in *Quantitative Analysis of Nuclear Medicine Images*, H. Zaidi, Ed., pp. 494–536, Springer, New York, NY, USA, 2006.

[7] M. Aristophanous, B. C. Penney, and C. A. Pelizzari, "The development and testing of a digital PET phantom for the evaluation of tumor volume segmentation techniques," *Medical Physics*, vol. 35, no. 7, pp. 3331–3342, 2008.

[8] H. Vees, S. Senthamizhchelvan, R. Miralbell, D. C. Weber, O. Ratib, and H. Zaidi, "Assessment of various strategies for 18F-FET PET-guided delineation of target volumes in high-grade glioma patients," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 36, no. 2, pp. 182–193, 2009.

[9] S. Basu, "Selecting the optimal image segmentation strategy in the era of multitracer multimodality imaging: a critical step for image-guided radiation therapy," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 36, no. 2, pp. 180–181, 2009.

[10] G. F. Luger, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Pearson Education, 2009.

[11] Y. Zhengchun and L. Hongji, "Face detection based on SCNN and wavelet invariant moment in color image," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR '07)*, pp. 783–787, November 2007.

[12] L. Sun and S. Wu, "Handwritten Chinese character recognition based on supervised competitive learning neural network and block-based relative fuzzy feature extraction," in *Applications of Neural Networks and Machine Learning in Image Processing IX*, Proceedings of SPIE, pp. 65–70, January 2005.

[13] S. Joo, W. K. Moon, and H. C. Kim, "Computer-aidied diagnosis of solid breast nodules on ultrasound with digital image processing and artificial neural network," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '04)*, pp. 1397–1400, September 2004.

[14] S. G. Mougiakakou, I. Valavanis, K. S. Nikita, A. Nikita, and D. Kelekis, "Characterization of CT liver lesions based on texture features and a multiple neural network classification scheme," in *Proceddings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1287–1290, September 2003.

[15] W. E. Reddick, J. O. Glass, E. N. Cook, T. David Elkin, and R. J. Deaton, "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 911–918, 1997.

[16] C. C. Reyes-Aldasoro and A. L. Aldeco, "Image segmentation and compression using neural networks," in *Advances in Artificial Perception and Robotics (CIMAT '00)*, October 2000.

[17] K. Suzuki, H. Abe, H. MacMahon, and K. Doi, "Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN)," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 406–416, 2006.

[18] M. Halt, C. C. Le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET," *IEEE Transactions on Medical Imaging*, vol. 28, no. 6, pp. 881–893, 2009.

[19] S. Belhassen and H. Zaidi, "A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET," *Medical Physics*, vol. 37, no. 3, pp. 1309–1324, 2010.

[20] Y. Li and Z. Chi, "MR brain image segmentation based on self-organizing map network," *International Journal of Information Technology*, vol. 11, no. 8, 2005.

[21] M. S. Sharif, A. Amira, and H. Zaidi, "3D oncological PET volume analysis using CNN and LVQNN," in *Proceedings of IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems (ISCAS '10)*, pp. 1783–1786, June 2010.

[22] C. Collet and F. Murtagh, "Multiband segmentation based on a hierarchical Markov model," *Pattern Recognition*, vol. 37, no. 12, pp. 2337–2347, 2004.

[23] D. L. Weakliem, "A critique of the Bayesian information criterion for model selection," *Sociological Methods and Research*, vol. 27, no. 3, pp. 359–397, 1999.

[24] C. T. Volinsky and A. E. Raftery, "Bayesian information criterion for censored survival models," *Biometrics*, vol. 56, no. 1, pp. 256–262, 2000.

[25] J. T. Bushberg et al., *The Essential Physics of Medical Imaging*, Lippincott Williams and Wilkins, Philadelphia, Pa, USA, 2006.

[26] G. Dreyfus, *Neural Networks Methodology and Applications*, Springer, Berlin, Germany, 2005.

[27] B. Curry and P. H. Morgan, "Evaluating Kohonen's learning rule: an approach through genetic algorithms," *European Journal of Operational Research*, vol. 154, no. 1, pp. 191–205, 2004.

[28] S. Clippingdale and R. Wilson, "Self-similar neural networks based on a Kohonen learning rule," *Neural Networks*, vol. 9, no. 5, pp. 747–763, 1996.

[29] M. A. Arbib, *The Handbook of Brain Theory and Neural Networks*, Massachusetts Institute of Technology, 2003.

[30] M. Swiercz, J. Kochanowicz, J. Weigele et al., "Learning vector quantization neural networks improve accuracy of transcranial color-coded duplex sonography in detection of middle cerebral artery spasm—preliminary report," *Neuroinformatics*, vol. 6, no. 4, pp. 279–290, 2008.

[31] P. Somervuo and T. Kohonen, "Self-organizing maps and learning vector quantization forfeature sequences," *Neural Processing Letters*, vol. 10, no. 2, pp. 151–159, 1999.

[32] C. Jonsson, R. Odh, P. O. Schnell, and S. A. Larsson, "A comparison of the imaging properties of a 3- and 4-ring biograph PET scanner using a novel extended NEMA phantom," in *Proceedings of IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC '07)*, pp. 2865–2867, November 2007.

[33] M. D. R. Thomas, D. L. Bailey, and L. Livieratos, "A dual modality approach to quantitative quality control in emission tomography," *Physics in Medicine and Biology*, vol. 50, no. 15, pp. N187–N194, 2005.

[34] H. Herzog, L. Tellmann, C. Hocke, U. Pietrzyk, M. E. Casey, and T. Kawert, "NEMA NU2-2001 guided performance evaluation of four siemens ECAT PET scanners," *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2662–2669, 2004.

[35] H. Bergmann, G. Dobrozemsky, G. Minear, R. Nicoletti, and M. Samal, "An inter-laboratory comparison study of image quality of PET scanners using the NEMA NU 2-2001 procedure for assessment of image quality," *Physics in Medicine and Biology*, vol. 50, no. 10, pp. 2193–2207, 2005.

[36] J. M. Wilson and T. G. Turkington, "Multisphere phantom and analysis algorithm for PET image quality assessment," *Physics in Medicine and Biology*, vol. 53, no. 12, pp. 3267–3278, 2008.

[37] H. Zaidi, F. Schoenahl, and O. Ratib, "Geneva PET/CT facility: design considerations and performance characteristics of two commercial (biograph 16/64) scanners," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, supplement 2, p. S166, 2007.

[38] M. S. Sharif, M. Abbod, A. Amira, and H. Zaidi, "Artificial neural network-based system for PET volume segmentation," *International Journal of Biomedical Imaging*, vol. 2010, Article ID 105610, 11 pages, 2010.

[39] M. Vuk and T. Curk, "ROC curve, lift chart and calibration plot," *Metodoloski zvezki*, vol. 3, no. 1, pp. 89–108, 2006.

[40] J. -F. Daisne, M. Sibomana, A. Bol, G. Cosnard, M. Lonneux, and V. Grégoire, "Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency," *Radiotherapy and Oncology*, vol. 69, no. 3, pp. 237–245, 2003.

[41] J. F. Daisne, T. Duprez, B. Weynand et al., "Tumor volume in pharyngolaryngeal squamous cell carcinoma: Comparison at CT, MR imaging, and FDG PET and validation with surgical specimen," *Radiology*, vol. 233, no. 1, pp. 93–100, 2004.

[42] X. Geets, J. A. Lee, A. Bol, M. Lonneux, and V. Grégoire, "A gradient-based method for segmenting FDG-PET images: methodology and validation," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 9, pp. 1427–1438, 2007.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2012, Article ID 920920, 7 pages
doi:10.1155/2012/920920

Research Article

Classifying High-Dimensional Patterns Using a Fuzzy Logic Discriminant Network

Nick J. Pizzi[1] and Witold Pedrycz[2]

[1] *Department of Computer Science, University of Manitoba, Winnipeg MB, Canada R3T 2N2*
[2] *Department of Electrical and Computer Engineering, University of Alberta, Edmonton AB, Canada T6R 2G7*

Correspondence should be addressed to Nick J. Pizzi, pizzi@cs.umanitoba.ca

Received 28 July 2011; Accepted 8 December 2011

Academic Editor: Maysam Abbod

Although many classification techniques exist to analyze patterns possessing straightforward characteristics, they tend to fail when the ratio of features to patterns is very large. This "curse of dimensionality" is especially prevalent in many complex, voluminous biomedical datasets acquired using the latest spectroscopic modalities. To address this pattern classification issue, we present a technique using an adaptive network of fuzzy logic connectives to combine class boundaries generated by sets of discriminant functions. We empirically evaluate the effectiveness of this classification technique by comparing it against two conventional benchmark approaches, both of which use feature averaging as a preprocessing phase.

1. Introduction

Biomedical spectroscopic modalities produce information-rich but complex, voluminous data [1]. For instance, magnetic resonance spectroscopy, which exploits the interaction between an external homogenous magnetic field and a nucleus that possesses spin, is a reliable and versatile spectroscopic modality [2, 3]. Coupled with robust multivariate discrimination methods, it is especially useful in the interpretation and classification of high-dimensional biomedical spectra (patterns) of tissues and biofluids [4]. However, the ratio of the number of features to the number of patterns for these data is typically very large; the feature space dimensionality is $O(10^3\text{-}10^4)$ while the number of patterns is O(10–100). This "curse of dimensionality" [5, 6] is a serious challenge for the classification of complex biomedical spectra: the excess degrees of freedom tend to cause overfitting, which significantly affects the reliability of the chosen classifier by diminishing its capability to determine effective generalizations.

We present a pattern classification technique, an extension to a method described in [7], that attenuates the confounding effects of the curse of dimensionality using an adaptive network of fuzzy logic connectives to combine pattern class boundaries generated by sets of discriminant functions based on sets of feature regions possessing high discriminatory power. We empirically evaluate the effectiveness of this classification technique by comparing it against two conventional benchmark approaches, both of which use feature averaging as a preprocessing phase.

Section 2 presents a brief discussion on pattern classification including pattern mapping, validation, discriminant analysis, and dimensionality reduction approaches. Details of our technique are presented in Section 3. Datasets, experiment design, and results are discussed in Section 4 followed by some concluding remarks.

2. Biomedical Pattern Classification

2.1. Mappings and Validation. We begin by defining some formal notation to precisely describe the problem of pattern classification where N is the number of patterns (samples, vectors, individuals, or cases), n is the number of features (dimensions, attributes, or measurements), and c is the number of classes (groups). Let $\mathbf{X} = \{(\mathbf{x}_k, \omega_k),\ k = 1, 2, \ldots, N\}$ be a set of N labeled patterns where $\mathbf{x}_k \in \Re^n$ and $\omega_k \in \Omega$. Typically, $\Omega = \{1, 2, \ldots, c\}$; however, it is often advantageous [8] to use 1-of-c encoding for the class labels for iterative

classifiers such as artificial neural networks [2]; namely, $\Omega = \{\gamma_1, \gamma_2, \ldots, \gamma_c\}$, where, for $\mathbf{x}_i$, $\gamma_{\omega_i} = 1$ and $\gamma_{\omega_j} = 0$ $(\forall\, \omega_i \neq \omega_j)$. A classifier is a system that determines a mapping, $f : \mathbf{X} \rightarrow \Omega$. Using f, if a classifier predicts that the class label for $\mathbf{x}_i$ is ω_p, then a correct classification occurs when $\omega_p = \omega_i$. It is considered a misclassification (a classification error) if $\omega_p \neq \omega_i$.

Unfortunately, many investigations involving pattern classification are biased as they use the entire dataset to determine the mapping. This approach leads to overly optimistic pattern classification results and do not take into account the possibility of overfitting; that is, the mapping becomes a simple table lookup between the given patterns and class labels, thereby possessing no generalized predictive power for new (unseen) patterns. To compensate for this bias, it is essential to perform some type of validation [9, 10]. For instance, patterns in $\mathbf{X}$ may be randomly allocated to a design (training) subset, $\mathbf{X}^D$ containing N^D patterns, or a validation (test) subset, $\mathbf{X}^V$ containing N^V patterns ($N^D + N^V = N$). Now, a mapping is determined using only design patterns, $f' : \mathbf{X}^D \rightarrow \Omega$, but the classification performance is measured using f' with the validation patterns.

Classification performance is measured using the $c \times c$ "confusion matrix" of the desired class labels versus the predicted class labels. If the class prediction for $\mathbf{x}_i$ is ω_p, then the element, $[\omega_p, \omega_i]$, of the confusion matrix is incremented by one (perfect accuracy is reflected by zeroes on the off-diagonal and nonzeroes on the diagonal). The conventional performance measure is the ratio of correctly classified patterns to the total number of patterns, $P_O = (\sum_i r_{ii})/N^V (i = 1, 2, \ldots, c)$, where r_{ij} is the number of class i validation patterns predicted, by the mapping f, to belong to class j. While other measures exist, such as the average class-wise accuracy, receiver operating characteristics graphs (ROC curves) [11], or the kappa score (a chance corrected measure of agreement) [12], for the sake of clarity during the discussion of the experiment results, we will use P_O.

2.2. Discriminant Functions. Linear discriminant analysis (LDA) [13] is a conventional classification approach that determines linear boundaries between c classes while taking into account inter class and intra class variances. If the error distributions for the classes are the same (identical covariance matrices), LDA constructs the optimal linear boundary between the classes. In real-world situations, this optimality is seldom achieved since different classes typically give rise to different distributions.

LDA allocates a pattern, $\mathbf{x}$, to class i for which the probability distribution, $p_i(\mathbf{x})$, is greatest. That is, $\mathbf{x}$ is allocated to class i, if $q_i p_i(\mathbf{x}) \geq q_j p_j(\mathbf{x})$ $(\forall\, j = 1, 2, \ldots, c\ [j \neq i])$, where q_i is the class' prior (or proportional) probability. The discriminant function for class i is

$$D_i(\mathbf{x}) = \log q_i + \boldsymbol{\mu}_i^{\mathrm{T}} \mathbf{W}^{-1} \left(x - \frac{1}{2} \boldsymbol{\mu}_i \right), \tag{1}$$

where $\boldsymbol{\mu}_i$ is the mean for class i and $\mathbf{W}$ is the covariance matrix of the patterns in $\mathbf{X}$. The feature space hyperplane separating class i from class j is defined by $F_{ij}(\mathbf{x}) = D_i(\mathbf{x}) - D_j(\mathbf{x}) = 0$. Figure 1 illustrates the class boundaries defined by a set of linear discriminant functions for a two-dimensional dataset with three classes ($N = 33, n = 2, c = 3$). As mentioned in Section 2.2, when LDA is used for pattern classification, it is imperative to define the discriminant functions using the design patterns, $\mathbf{X}^D$, but to validate the performance using the validation patterns, $\mathbf{X}^V$.

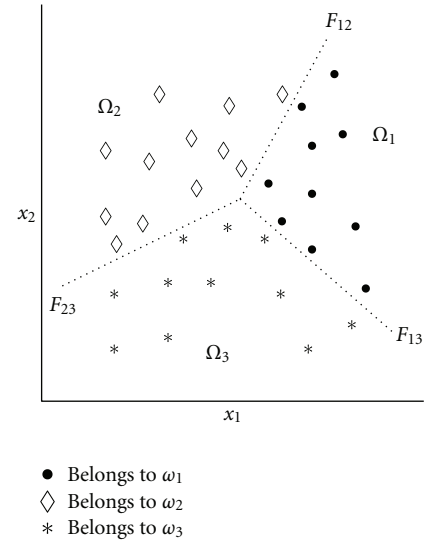


Figure 1: Class boundaries defined by linear discriminant functions for three classes of 2-dimensional patterns.

The support vector machine (SVM) [14, 15] is an important family of supervised learning algorithms that select models that maximize the error margin of a training subset. This approach has been successively used in a wide range of data classification problems [16]. Given a set of patterns that belong to one of two classes, an SVM finds the hyperplane leaving the largest possible fraction of patterns of the same class on the same side while maximizing the distance of either class from the hyperplane. The approach is usually formulated as a constrained optimization problem and solved using constrained quadratic programming. While the original approach [17] could only be used for linearly separable problems, it may be extended by employing a "kernel trick" [18] that exploits the fact that a nonlinear mapping of sufficiently high dimension can project the patterns to a new parameter space in which classes can be separated by a hyperplane. In general, it cannot be determined a priori which kernel will contribute to producing the best classification results for a given dataset, and one must rely on heuristic (trial and error) experimentation. Common kernel functions $K(\mathbf{x}, \mathbf{y})$, for patterns $\mathbf{x}$ and $\mathbf{y}$, are power, $(\mathbf{x} \cdot \mathbf{y})^d$; polynomial, $(a\mathbf{x} \cdot \mathbf{y} + b)^d$; sigmoid, $\tanh(a\mathbf{x} \cdot \mathbf{y} + b)$; Gaussian, $\exp(-0.5|\mathbf{x} - \mathbf{y}|^2/\sigma)$.

2.3. Feature Reduction. As with any pattern classifier, LDA becomes unreliable when there are a large number of features. Even when using stable methods such as singular value decomposition, the inversion of $\mathbf{W}$ in (1) becomes unstable, so it becomes imperative to preprocess the features.

A preprocessing strategy to use when n/N is very large (curse of dimensionality) is to reduce the dimensionality of the feature space of the patterns; that is, we find a mapping (transformation) $f' : \mathbf{X} \rightarrow \mathbf{Y} = \{(\mathbf{y}_k, \omega_k)\}$ where $\mathbf{y}_k \in \Re^m$ and $m \ll n$. Now, the classification mapping becomes $f : \mathbf{Y} \rightarrow \Omega$. A standard approach to feature space reduction is to take the averages of a fixed number of contiguous feature regions. Although this type of averaging may often work well in attenuating the effects of the curse of dimensionality, it also has a tendency to sometimes wash away information content. Other feature reduction approaches do not transform the original feature space but rather attempt to find those features that possess the greatest discriminatory power [19–22]. One example of this type of approach is stochastic feature selection.

2.4. Stochastic Feature Selection. Stochastic feature selection (SFS) [23] is a feature selection/reduction preprocessing strategy that may be used with any homogeneous or heterogeneous set of classifiers (e.g., LDA, artificial neural networks, support vector machines). Essentially, SFS iteratively presents, in a highly parallelized fashion, many feature regions (contiguous subsets of pattern features) to the set of classifiers retaining the best set of classifier/region pairs. While SFS has a rich set of parameters to control many different aspects of the classification process, here we present only those aspects that are relevant to this discussion and refer the reader to [23] for a thorough description of this strategy. For a pattern $\mathbf{x} = [x_1, x_2, \ldots, x_n]$, we define a region to be a contiguous subset of its features, $\mathbf{x}^{rs} = [x_r, x_{r+1}, \ldots, x_s]$ $(1 \leq r \leq s \leq n)$. The user specifies the minimum and maximum number of regions to be selected for each classification iteration as well as the minimum and maximum length for a feature region $(s - r + 1)$. SFS exploits the quadratic combination of (disjoint or overlapping) feature regions. The intent is that if the original feature space has nonlinear boundaries between classes, the new (quadratic) parameter space may have boundaries that are more linear. Given the feature region $\mathbf{x}^{rs} = [x_r, x_{r+1}, \ldots, x_s]$, SFS has three categories of quadratic combinations: using the original feature region, $\mathbf{x}^{rs}$; squaring the feature values for $\mathbf{x}^{rs}$ $[x_r^2, x_{r+1}^2, \ldots, x_s^2]$, or using all pair-wise feature cross products from two regions, $\mathbf{x}^{rs}$ and $\mathbf{x}^{tu} = [x_t, x_{t+1}, \ldots, x_u]$ $(1 \leq t \leq u \leq n)$, producing the result $[x_r x_t, x_r x_{t+1}, \ldots, x_r x_u, \ldots, x_s x_t, x_s x_{t+1}, \ldots, x_s x_u]$. The fitness function (classification performance measure) is P_O. In this study, the only classifier that is used is LDA. When SFS is finished, it returns the best set of classifier results (the cardinality of the set is user-specified) where each result contains (i) the value of P_O, (ii) the indices (to the original features) of the set of feature regions selected, and (iii) the discriminant functions for each class as determined by LDA using the selected feature regions.

2.5. Fuzzy Adaptive Logic Network. Our approach builds upon the fuzzy adaptive logic network (cf. [24] for a thorough description). This approach, which can be used for pattern classification, combines two different subsystems within its general architecture. A neurocomputing subsystem uses a set of perceptrons to construct class boundaries to delineate patterns from different classes. Via a set of respective weights and inputs, a perceptron is defined as $P(\mathbf{x}, \mathbf{w}) = f(\sum_i w_i x_i + w_0)$ where f is a transfer function (any sigmoid function but often the logistic function), which describes an n-dimensional hyperplane. This geometric information is then presented to the logic processing subsystem that comprises a layer of fuzzy conjunctions ("and" elements) and another layer of fuzzy disjunctions ("or" elements). The intent is to use these fuzzy logic connectives to combine the hyperplanes from the neurocomputing subsystem to form convex hull-like topologies. For instance, a convex region delineated by p perceptrons may be represented by the compound logic predicate, $Q = P_1(\mathbf{x}, \mathbf{w}_1), P_2(\mathbf{x}, \mathbf{w}_2), \ldots, P_p(\mathbf{x}, \mathbf{w}_p)$, which produces values close to one (meaning it becomes true) when all contributing predicates are *true* (i.e., the respective perceptrons produce high outputs). To capture the geometric notion of disjoint regions, one may take a union (in the set theoretic sense) of the individual regions described by the Q's: $V = Q_1$ or Q_2 or $\ldots$ or Q_q. To implement these fuzzy predicates, one uses t-norms to model the *and* logic connectives and s-norms to model the *or* logic connectives. A t-norm, $\wedge$, is a function $[0,1]^2 \rightarrow [0,1]$ that is commutative, symmetric, monotonic, and satisfies the boundary conditions $x \,\mathrm{t}\, 0 = 0$ and $x \,\mathrm{t}\, 1 = x$, while the boundary conditions for the s-norm, $\vee$, are $x \,\mathrm{s}\, 0 = x$ and $x \,\mathrm{s}\, 1 = 1$. The fuzzy *or* and *and* connectives may now be defined as

$$\begin{aligned} OR(\mathbf{x};\ \mathbf{w}) &= \wedge_i (w_i \vee x_i), \\ AND(\mathbf{x};\ \mathbf{w}) &= \vee_i (w_i \wedge x_i), \end{aligned} \tag{2}$$

where $\mathbf{x}$ is the input and $\mathbf{w}$ are the corresponding adjustable weights (connections) confined to the unit interval. In the case of $OR(\mathbf{x}; \mathbf{w})$, the greater the weight value the more relevant the respective input (if all weights are 1, it becomes a standard *or* gate). In the case of $AND(\mathbf{x}; \mathbf{w})$, the greater the weight value, the less relevant the respective input (if all weights are 0, it becomes a standard *and* gate). If we restrict ourselves to differentiable t- and s-norms, a gradient descent strategy can be used to train a fuzzy adaptive logic network (cf. [24] for details).

3. Fuzzy Logic Network with Linear Discriminants

Building upon the concepts described in Section 2, we now describe our pattern classification algorithm, FLND (fuzzy logic network with discriminants). There are four major steps to the FLND algorithm: (i) use SFS to find the best κ sets of feature regions using the patterns from the design subset, $\mathbf{X}^D$; (ii) for each set of feature regions, compute the linear discriminant function for each class and then compute the discriminant values for each design pattern; (iii) use a genetic algorithm to determine the optimal weights for the fuzzy logic network given the design pattern discriminant values found in (ii); (iv) use the patterns from the validation subset, $\mathbf{X}^V$, to assess the classification performance, P_O, using the selected feature regions and discriminant function values. Figure 2 illustrates the architecture of the FLND system.

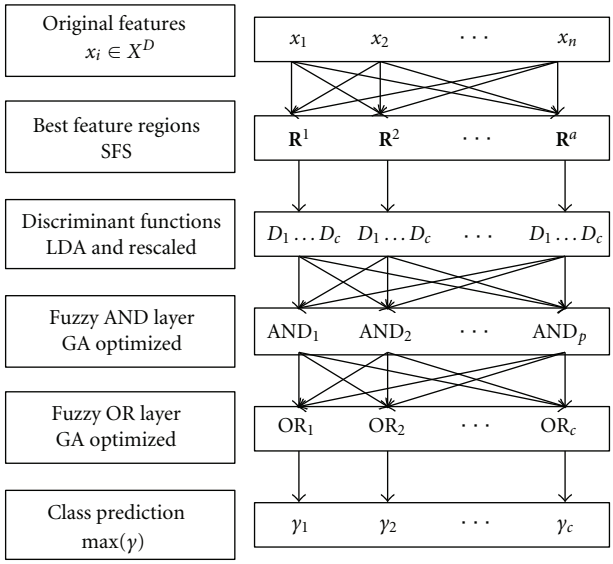


Figure 2: FLND architecture.

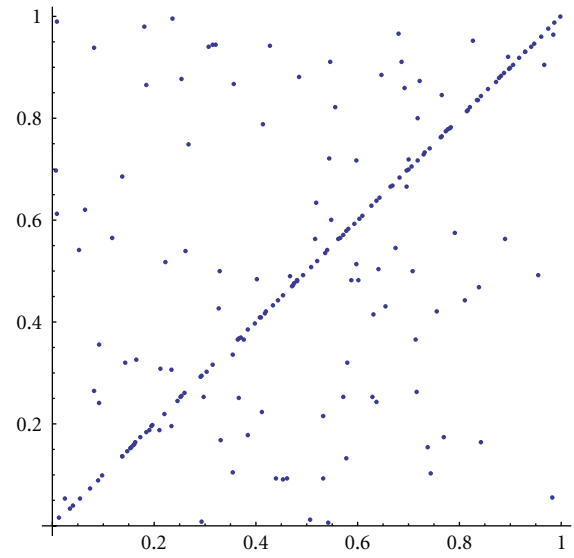


Figure 3: Plot of the first two features of the second synthetic dataset.

Let us now look at each algorithmic step in more detail. In the experiments described in Section 4, SFS uses LDA as the sole classifier and P_O is the performance measure. After a set number of iterations, SFS returns κ sets of feature regions, $\mathbf{R}^\alpha$ ($\alpha = 1, 2, \ldots, \kappa$), and the respective discriminant functions for each class, D_i^α ($i = 1, 2, \ldots, c$), using the feature regions (feature regions are sorted by P_O). The set of feature regions is of the form $\mathbf{R}^\alpha = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{\beta_\alpha}\}$ where β_α is the total number of regions for set α and $\mathbf{r}$ is a single contiguous feature region as described in Section 2.4. The discriminant functions are computed using $\mathbf{R}$ rather than all n features. The input space is now no longer the original n features but rather the respective values of $D(\mathbf{R})$ for each class and each feature region set, which is a significant reduction in the dimensionality of the input space ($c \times \alpha \ll n$).

The fuzzy logic network component of FLND uses the product ($x_1 \times x_2$) and probabilistic sum ($x_1 + x_2 - x_1 \times x_2$) for the t- and s-norms, respectively, with p (user selected) *AND* connectives and *cOR* connectives. There are two deficiencies with this component that does not exist with the fuzzy adaptive logic network described in Section 2.5. First, while perceptron output maps onto the unit interval (due to the sigmoidal nature of its transfer function), which is necessary for input into a fuzzy logic *AND* connective, values from linear discriminant functions map onto $\Re$. This can be easily dealt with by rescaling the linear discriminant values prior to presentation to the fuzzy logic network ($(x - \min) \div (\max - \min)$, where min and max are the respective minimum and maximum for all discriminant function values).

The second, more serious, issue is that a gradient descent strategy cannot be used to minimize the network error (i.e., optimize the weights) since the weight adjustments are now based on discrete sets of discriminant functions rather than differentiable perceptron output. We deal with this issue by using a straightforward implementation of a genetic algorithm (GA) [9, 25, 26] to perform the structural optimization of the network. While much slower than a gradient descent approach, it still provides more than adequate computational performance. We implemented a conventional genetic algorithm as described in [27], but other more sophisticated GA variants could certainly be explored. The crossover rate was set to 0.10, and the mutation rate was set to 0.007 for all experiments listed in Section 4.

Finally, all performance results, using P_O, are based on the class predictions of FLND using the patterns from the validation subset. Further, the results are also benchmarked against conventional applications of LDA and SVM.

4. Experiments and Discussion

4.1. Synthetic Datasets. We begin our experiments with the two-dimensional exclusive-or problem ($n = 2, c = 2, N = 4$). Intuitively, one expects that LDA would perform poorly in this case as no hyperplane can act as a class boundary to perfectly separate the two classes of patterns ($\{\{0, 0\}, \{1, 1\}\}$ versus $\{\{0, 1\}, \{1, 0\}\}$). Using LDA, this is actually the case with $P_O = 0.5$ (one pattern misclassification for each class). As this is a strictly pedagogical experiment, we skip the validation exercise and do not bother with SFS and move directly to the fuzzy logic network. Setting the initial GA population to 200, the number of iterations to 100, and the number of *AND* connectives to 2, we now get perfect accuracy, $P_O = 1.0$. The weights for the two *AND* connectives are $\{0.09, 0.04\}$ and $\{0.0, 1.0\}$. The weights for the subsequent two *OR* connectives are $\{0.37, 0.04\}$ and $\{0.08, 0.25\}$.

This synthetic dataset is a variant of the exclusive-or dataset described above ($n = 10$, $c = 2$, $N = 500$, $N^D = N^V = 250$, $\mathbf{x} \in [0, 1]^n$). A pattern belongs to the first class, if all of its features are identical; otherwise, it belongs to the second class. Figure 3 is a plot of the first two features of this dataset. The initial GA population is 800, the number

Table 1: Confusion matrices for LDA and FLND using $\mathbf{X}^V$.

Desired versus predicted	LDA (P_O=0.5)		FLND (P_O = 0.8)	
	Class 1	Class 2	Class 1	Class 2
Class 1	65	60	88	37
Class 2	63	62	14	111

Table 2: FLND confusion matrices for patterns $\mathbf{X}^D$ (N^D = 80) and $\mathbf{X}^V$ (N^V = 70).

Desired versus predicted	Design patterns (P_O = 0.84)		Validation patterns (P_O = 0.83)	
	Normal	Abnormal	Normal	Abnormal
Normal	33	7	40	9
Abnormal	6	34	3	18

of iterations is 100, and the number of *AND* connectives is 10 (as with the previous experiment, we do not use SFS). In this case, LDA again performed extremely poorly, $P_O = 0.51$, while FLND produced a significantly superior classification accuracy, $P_O = 0.80$. Table 1 lists the confusion matrices for LDA and FLND using this dataset. For completeness, we also list the weights for the *AND* connectives, $\{0.06, 0.34\}$, $\{0.04, 0.15\}$, $\{0.11, 0.07\}$, $\{0.05, 0.08\}$, $\{0.04, 0.08\}$, $\{0.09, 0.09\}$, $\{0.04, 0.08\}$, $\{0.05, 0.20\}$, $\{0.04, 0.05\}$, $\{0.12, 0.20\}$, and the *OR* connectives, $\{0.07, 0.04, 0.08, 0.05, 0.50, 0.04, 0.04, 0.05, 0.03, 0.25\}$, $\{0.04, 0.05, 0.15, 0.20, 0.04, 0.04, 0.05, 0.40, 0.25, 0.04\}$.

4.2. Magnetic Resonance Spectra. Magnetic resonance spectra (patterns) of a biofluid ($n = 4255$) were acquired and used to measure the effectiveness of FLND for the classification of a complex, voluminous, "real world" biomedical dataset. In this case, $N = 150$ with 89 spectra belonging to class 1 ("normal") and 61 spectra belonging to class 2 ("abnormal"). These spectra were randomly allocated to the design subset ($N^D = 80$ with 40 normal spectra and 40 abnormal spectra) or the validation ($N^V = 70$ with the remaining 49 normal spectra and 21 abnormal spectra) subset.

For this dataset, the following SFS parameters were used with FLND: the range for the number of feature regions, 2–5; the range for the number of features within a region, 2–20; $\kappa = 10$; 10^4 iterations. The fuzzy logic network parameters were $p = 7$; crossover rate, 0.10; mutation rate, 0.008; size of GA population, 1200; 50 GA iterations.

Table 2 lists the confusion matrices for FLND with the design patterns and validation patterns. For the design patterns, $P_O = 0.84$, while $P_O = 0.83$ for the validation patterns. Moreover, 82% of the normal (class 1) validation patterns were correctly classified and 86% of the abnormal (class 2) validation patterns were correctly classified. The latter result is especially advantageous as, for many confirmatory biomedical data analysis problems, it is important to have a low false positive rate (i.e., predictions for abnormal conditions should be as accurate as possible).

Table 3 lists the $\kappa = 10$ best sets of discriminatory feature regions, $\mathbf{R}$, found by FLND. For each entry, we list the specific regions selected, how those regions were combined, and the total number of individual features used. Interestingly, over half of the selected discriminatory regions fell in the approximate range 3050–3850, which likely indicates that the biological metabolites represented by this spectral region are particularly germane in distinguishing between normal and abnormal states for the underlying biofluid being investigated. Also important to note is that most of the entries used quadratic combinations of the corresponding feature regions, with the top three results using the pair-wise cross products of the respective regions. Finally, the dimensionality of the feature space is only 4% that of the original space (180 quadratically combined features versus 4255 original spectra features).

4.3. Benchmark Comparisons. We now compare the FLND results from Section 4.2 with two classifier benchmarks, SVM and LDA. First, we use SVM and LDA to construct mappings using all $n = 4255$ features. Subsequently, for each classifier, feature averaging is used as a preprocessing technique, which is a typical strategy for voluminous biomedical spectra, in order to reduce the complexity of the classification problem [28–31]. By reducing the dimensionality of the feature space, we hope to address the curse of dimensionality. Furthermore, averaging has a tendency to attenuate noise signatures. In our specific case, the original features are contiguously averaged using varying window sizes (with no overlap) to produce six sets of averaged features of size 851, 185, 115, 37, 23, and 5, respectively. We use proportional class probabilities for LDA and all SVM kernels listed in Section 2.2. For clarity, in the case of SVM, we report only the best results for each averaged feature set. Table 4 lists the validation subset classification results (confusion matrices and P_O) using the benchmarks with feature averaging. In no case did the benchmarks outperform FLND. Using all of the original features, both benchmarks performed poorly: $P_O = 0.64$ for SVM and $P_O = 0.60$ for LDA. For each benchmark, the best results occurred with 185 averaged features: $P_O = 0.77$ for SVM and $P_O = 0.74$ for LDA. We also note that classification results begin to degrade as the window size increases (i.e., the number of averaged features decreases). This is not uncommon as feature averaging can cause a washing away of information content present in biomedical spectra.

Table 3: Discriminatory feature regions selected by FLND.

κ	Feature regions, $\mathbf{R}^1$–$\mathbf{R}^{10}$	No. of Features	Combination
1	$[x_{1948} \cdots x_{1962}][x_{3642} \cdots x_{3653}]$	180	Cross product
2	$[x_{1207} \cdots x_{1223}][x_{3058} \cdots x_{3073}]$	272	Cross product
3	$[x_{987} \cdots x_{1005}][x_{3544} \cdots x_{3559}]$	304	Cross product
4	$[x_{3817} \cdots x_{3835}]$	19	Square
5	$[x_{3198} \cdots x_{3216}]$	19	Square
6	$[x_{2175} \cdots x_{2193}][x_{3233} \cdots x_{3252}][x_{3849} \cdots x_{3868}]$	59	Original
7	$[x_{3408} \cdots x_{3424}][x_{3441} \cdots x_{3459}]$	323	Cross product
8	$[x_{2349} \cdots x_{2364}][x_{2993} \cdots x_{3004}][x_{3836} \cdots x_{3854}]$	47	Original
9	$[x_{3635} \cdots x_{3649}][x_{3912} \cdots x_{3928}]$	255	Cross product
10	$[x_{3107} \cdots x_{3125}][x_{3782} \cdots x_{3799}][x_{3849} \cdots x_{3868}]$	57	Original

Table 4: Validation patterns ($\mathbf{X}^V$) confusion matrices for benchmark classifiers using averaged features.

No. Features	Desired versus Predicted	SVM			LDA		
		Normal	Abnormal	P_O	Normal	Abnormal	P_O
4255	Normal	33	16	0.64	30	19	0.60
	Abnormal	9	12		9	12	
851	Normal	35	14	0.70	35	14	0.71
	Abnormal	7	14		7	14	
185	Normal	38	11	0.77	37	12	0.74
	Abnormal	5	16		6	15	
115	Normal	37	12	0.76	37	12	0.73
	Abnormal	5	16		7	14	
37	Normal	35	14	0.70	36	13	0.69
	Abnormal	7	14		9	12	
23	Normal	35	14	0.70	35	14	0.71
	Abnormal	7	14		7	14	
5	Normal	30	19	0.60	30	19	0.60
	Abnormal	9	12		9	12	

5. Conclusion

We have empirically demonstrated the effectiveness of a classification technique that uses an adaptive network of fuzzy logic connectives to combine class boundaries generated by sets of discriminant functions based on collections of feature regions possessing high discriminatory power. Using a complex, voluminous "real world" biomedical dataset, FLND outperformed all classifier benchmarks in the classification of patterns from a validation subset. It achieved an 8% improvement in classification accuracy compared against the best benchmark result (0.83 versus 0.77 for SVM using feature averaging with a window size of 23). This increase in classification accuracy is achieved by taking the class boundaries described by the discriminant functions and using layers of fuzzy logic connectives to combine these boundaries into convex, nonlinear boundaries. This new method also significantly reduces the dimensionality of the input space as the original set of spectral features is replaced by a much smaller set of class discriminant values. This is a particularly useful characteristic when dealing with the curse of dimensionality (large feature to sample ratio), which is a prevalent property of many complex biomedical datasets acquired using current spectroscopic modalities.

While this classification technique has demonstrated the utility of merging fuzzy logic connectives with multivariate statistical discrimination, the investigation has also led to the identification of future areas of research to potentially improve its overall effectiveness and computational performance. First, rather than setting the number of fuzzy *and* connectives by the user *a priori*, it would be worthwhile to investigate a cascade approach to determining an optimal number of *and* connections that would be completely data-driven. Second, alternate structural optimizations to the fuzzy logic network need to be examined beginning with more sophisticated evolutionary computational approaches or exploiting recent advances in stochastic optimization techniques. Finally, a more intelligent rescaling strategy for the discriminant function values needs to be investigated. For instance, this may include a fuzzified (weighted) distance measure based on the proximity (belongingness) of a sample to all class boundaries.

Acknowledgments

Conrad Wiebe and Aleksander Demko are gratefully acknowledged for the implementation of the stochastic feature selection algorithm. The aythors also thank the Natural Sciences and Engineering Research Council (NSERC) for its support of this investigation.

References

[1] D. L. Pavia, G. M. Lampman, G. S. Kriz, and J. A. Vyvyan, *Introduction to Spectroscopy*, Harcourt Brace College, Fort Worth, Tex, USA, 2008.

[2] M. Anthony and P. L. Bartlett, *Neural Network Learning: Theoretical Foundations*, Cambridge University Press, Cambridge, UK, 2009.

[3] W. Pedrycz, D. J. Lee, and N. J. Pizzi, "Representation and classification of high-dimensional biomedical spectral data," *Pattern Analysis & Applications*, vol. 13, no. 4, pp. 423–436, 2010.

[4] N. J. Pizzi and W. Pedrycz, "Aggregating multiple classification results using fuzzy integration and stochastic feature selection," *International Journal of Approximate Reasoning*, vol. 51, no. 8, pp. 883–894, 2010.

[5] F. Y. Kuo and I. H. Sloan, "Lifting the curse of dimensionality," *Notices of the American Mathematical Society*, vol. 52, no. 11, pp. 1320–1328, 2005.

[6] I. V. Oseledets and E. E. Tyrtyshnikov, "Breaking the curse of dimensionality, or how to use SVD in many dimensions," *SIAM Journal on Scientific Computing*, vol. 31, no. 5, pp. 3744–3759, 2009.

[7] N. J. Pizzi and W. Pedrycz, "A fuzzy logic network for pattern classification," in *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society*, pp. 53–58, Cincinnati, Ohio, USA, June 2009,.

[8] N. Pizzi, L. P. Choo, J. Mansfield et al., "Neural network classification of infrared spectra of control and Alzheimer's diseased tissue," *Artificial Intelligence in Medicine*, vol. 7, no. 1, pp. 67–79, 1995.

[9] U. M. Braga-Neto and E. R. Dougherty, "Is cross-validation valid for small-sample microarray classification?" *Bioinformatics*, vol. 20, no. 3, pp. 374–380, 2004.

[10] B. Efron and G. Gong, "A leisurely look at the bootstrap, the jackknife, and cross-validation," *The American Scientist*, vol. 37, no. 1, pp. 36–48, 1983.

[11] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, no. 4857, pp. 1285–1293, 1988.

[12] B. S. Everitt, "Moments of the statistics kappa and weighted kappa," *The British Journal of Mathematical and Statistical Psychology*, vol. 21, pp. 97–103, 1968.

[13] G. A. F. Seber, *Multivariate Observations*, John Wiley & Sons, Hoboken, NJ, USA, 2004.

[14] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, UK, 2002.

[15] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, NY, USA, 1998.

[16] L. Wang, *Support Vector Machines: Theory and Applications*, Springer, Berlin, Germany, 2005.

[17] V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 3rd edition, 2007.

[19] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.

[20] N. K. Kasabov and Q. Song, "DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 144–154, 2002.

[21] Q. Liu, A. H. Sung, Z. Chen, and J. Xu, "Feature mining and pattern classification for steganalysis of LSB matching steganography in grayscale images," *Pattern Recognition*, vol. 41, no. 1, pp. 56–66, 2008.

[22] E. K. Tang, P. N. Suganthan, and X. Yao, "Gene selection algorithms for microarray data based on least squares support vector machine," *BMC Bioinformatics*, vol. 7, article 95, 2006.

[23] N. J. Pizzi, "Classification of biomedical spectra using stochastic feature selection," *Neural Network World*, vol. 15, no. 3, pp. 257–268, 2005.

[24] W. Pedrycz, A. Breuer, and N. J. Pizzi, "Fuzzy adaptive logic networks as hybrid models of quantitative software engineering," *Intelligent Automation and Soft Computing*, vol. 12, no. 2, pp. 189–209, 2006.

[25] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.

[26] R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, John Wiley & Sons, Hoboken, NJ, USA, 2004.

[27] C. Jacob, *Illustrating Evolutionary Computation with Mathematica*, Academic Press, San Diego, Calif, USA, 2001.

[28] R. M. Rangayyan, *Biomedical Signal Analysis: A Case-Study Approach*, Wiley-IEEE Press, New York, NY, USA, 2001.

[29] R. L. Somorjai, M. E. Alexander, R. Baumgartner et al., "A data-driven, flexible machine learning strategy for the classification of biomedical data," in *Artificial Intelligence Methods and Tools for Systems Biology*, W. Dubitzky and F. Azuaje, Eds., pp. 67–85, Springer, Dordrecht, The Netherlands, 2004.

[30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd, Springer, New York, NY, USA, 2009.

[31] B. C. Wheeler and W. J. Heetderks, "A comparison of techniques for classification of multiple neural signals," *IEEE Transactions on Biomedical Engineering*, vol. 29, no. 12, pp. 752–759, 1982.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2012, Article ID 696194, 9 pages
doi:10.1155/2012/696194

Research Article

A Novel Technique for Identifying Patients with ICU Needs Using Hemodynamic Features

A. Jalali,[1] P. Ghorbanian,[1] A. Ghaffari,[2] and C. Nataraj[1]

[1] *Department of Mechanical Engineering, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085, USA*
[2] *Department of Mechanical Engineering, K.N.Toosi University of Technology, No. 19, Pardis street, Mollasadra Avenue, Vanak Square, Tehran 19991, Iran*

Correspondence should be addressed to A. Jalali, ali.jalali@villanova.edu

Received 1 August 2011; Accepted 11 November 2011

Academic Editor: Hak-Keung Lam

Identification of patients requiring intensive care is a critical issue in clinical treatment. The objective of this study is to develop a novel methodology using hemodynamic features for distinguishing such patients requiring intensive care from a group of healthy subjects. In this study, based on the hemodynamic features, subjects are divided into three groups: healthy, risky and patient. For each of the healthy and patient subjects, the evaluated features are based on the analysis of existing differences between hemodynamic variables: Blood Pressure and Heart Rate. Further, four criteria from the hemodynamic variables are introduced: circle criterion, estimation error criterion, Poincare plot deviation, and autonomic response delay criterion. For each of these criteria, three fuzzy membership functions are defined to distinguish patients from healthy subjects. Furthermore, based on the evaluated criteria, a scoring method is developed. In this scoring method membership degree of each subject is evaluated for the three classifying groups. Then, for each subject, the cumulative sum of membership degree of all four criteria is calculated. Finally, a given subject is classified with the group which has the largest cumulative sum. In summary, the scoring method results in 86% sensitivity, 94.8% positive predictive accuracy and 82.2% total accuracy.

1. Introduction

Hemodynamic instability is most commonly associated with abnormal or unstable blood pressure (BP), especially hypotension, or more broadly associated with inadequate global or regional perfusion. Inadequate perfusion may compromise important organs, such as the heart and the brain, due to limits on coronary and cerebral autoregulation, and may cause life-threatening illnesses or even death. Therefore, it is crucial to identify patients who are likely to become hemodynamically unstable for early detection and treatment of these life-threatening conditions [1]. The modern intensive care units (ICUs) typically employ continuous hemodynamic monitoring (e.g., heart rate (HR) and invasive arterial BP measurements) to track the state of patients. However, clinicians in a busy ICU are overwhelmed with the task of assimilating and interpreting the tremendous volumes of data into working hypotheses. Consequently, it is important to have automated algorithms that can accurately process and classify the large amount of monitoring data and identify patients who are on the verge of becoming unstable [1].

Modern ICUs are equipped with a large array of alarmed monitors and devices which are used in an attempt to detect clinical changes at the earliest possible moment, so as to prevent any further deterioration in a patient's condition. The effectiveness of these systems depends on the sensitivity and specificity of the alarms as well as on the responses of the ICU staff to the alarms. However, when large numbers of alarms are either technically false, or true, but clinically irrelevant, response efficiency can be decreased, reducing the quality of patient care and increased patient (and family) anxiety [2].

Medical and technical progress has extended the therapeutic possibilities of ICUs tremendously. A multitude of devices are available for monitoring and treatment in an individual assembly according to the requirements of the situation [3]. Due to limited physiological monitoring and a

patient's individual pathophysiology, intensive care medicine has to cope with a high amount of uncertainty. Unusual circumstances caused by patients, clinicians, and technology occur frequently and must be controlled and managed adequately to prevent a bad outcome and to achieve system reliability [3].

Cao et al. [1] have used ICU minute-by-minute heart rate (HR) and invasive arterial blood pressure (BP) monitoring trend data collected from the MIMIC II database to predict hemodynamic instability at least two hours before a major clinical intervention. They derived additional physiological parameters of shock index, rate pressure product, heart rate variability, and two measures of trending based on HR and BP, and they applied multivariable logistic regression modeling to carry out classification and implemented validation via bootstrapping, resulting in 75% sensitivity and 80% specificity. Eshelman et al. [4] have developed an algorithm for identifying ICU patients who are likely to become hemodynamically unstable. Their algorithm consists of a set of rules that trigger alerts and uses data from multiple sources; it is often able to identify unstable patients earlier and with more accuracy than alerts based on a single threshold. The rules were generated using the machine learning techniques of support vector machines and neural network and were tested on retrospective data in the MIMIC II ICU database, yielding a specificity of approximately 90% and a sensitivity of 60%.

Several investigations have been reported in the literature in the area of cardiovascular fault diagnosis using hemodynamic features. Javorka et al. [18] compared heart rate and blood pressure variability among young patients with type I diabetes mellitus (DM) and control subjects by using Poincare plots, which are the standard tools of nonlinear dynamic analysis. They found significant reduction of all HRV Poincare plot measure in patients with type I diabetes mellitus, indicating heart rate dysregulation. The study carried out by Pagani et al. [5] concerned patients suffering from hypertension. They showed that baroreflex gain decreases with the presence of hypertension. Blasi et al. [6] studied the effects of arousal from sleep on cardiovascular variability. They performed time-varying spectral analyses of heart rate variability (HRV) and blood pressure variability (BPV) records during acoustically induced arousals from sleep. They found that arousal-induced changes in parasympathetic activity are strongly coupled to respiratory patterns, and that the sympathoexcitatory cardiovascular effects of arousal are relatively long lasting and may accumulate if repetitive arousals occur in close succession.

Advances in knowledge-based systems have also enhanced the functionality of intelligent alarm systems and ICU needed patient detection. Using the knowledge of a domain expert to formulate rules or an expertly classified data set to train an adaptive algorithm has proven useful for intelligent processing of clinical alarms [2]. Expert systems such as neural network [7], knowledge-based decision trees [8, 9], and neurofuzzy systems [10] that encode the decisions of an expert clinician all show significant statistical improvement in the classification of alarms and ICU needed patients. Singh and Guttag [11] proposed a classification algorithm based on a decision tree method for cardiovascular risk stratification. They have shown that the decision tree method can improve the performance of the classification algorithm. They have reported that the decision tree models outperform the radial basis function (RBF) kernel-based support vector machine (SVM) classifiers. Timms et al. [12] have used a Mock circulation loop for hemodynamic modeling of the cardiovascular system in order to test cardiovascular devices, which are used in the ICU and can provide a better indication of patient's condition for the nursing staff. Also, Laramee et al. [2] have described an integrated systems methodology to extract clinically relevant information from physiological data. Such a method would aid significantly in the reduction of false alarms and provide the nursing staff with a more reliable indicator of patient condition.

Regarding suitable classifier structures, several studies have been reported in the literature. Ghorbanian et al. [13] proposed an algorithm based on a neural network classifier for heart arrhythmias detection. Their results show that the multilayer perceptron neural network (MLPNN) structure is a strong and precise classifier. They used several preprocessing techniques in their algorithm to improve the performance of the NN classifier. Rajendra Acharya et al. [14] proposed an algorithm based on a neural network classifier and fuzzy cluster for classification of heart arrhythmias. They compared these two classifiers and they reported that the fuzzy cluster is a better classifier in comparison with the neural one. Also, Özbay et al. [15] proposed a comparative study of the classification accuracy cardiovascular diseases using a well-known neural network architecture, a multilayered perceptron (MLP) structure, and a new fuzzy clustering NN architecture (FCNN) for early diagnosis. Based on their test results, they suggested that a new proposed FCNN architecture can generalize better than ordinary MLP architecture and also learn better and faster.

The method for classification of subjects into two categories of normal and abnormal subjects, as described in this paper, is based on the hypothesis that there should be differences between the hemodynamic data collected from normal subjects and abnormal patients. This hypothesis is constructed on the same foundation as all developed scoring methods for ICU patients. The idea behind all patient scoring methods in ICU is that critically ill patients in ICU are typically characterized by disturbance of the body's homeostasis. These disturbances can be estimated by measuring to what extent one or many physiologic variables differ from the normal range [16].

2. Methodology

While this paper shares some fundamental ideas with traditional scoring methods, the proposed method differs from them in two key areas. The first difference comes from fact that the patient scoring methods are based on the wide variety of data ranging from cardiovascular and respiratory systems to neurologic and renal systems variables. However, in our method we use a small subset of hemodynamic data, namely, HR and SBP. The principal objection to this could be that such a small amount of data could be insufficient

for identifying the patient state; the answer to this objection leads us to the second major difference of the proposed method with the scoring methods. Scoring methods just look at the data as they are being collected in the ICU; they ignore information hidden in the different time scales. In our proposed method on the other hand, this hidden information is extracted which gives us better insight into the patient's physiological condition.

The data used in this study is collected from the Physionet database. Data are collected from two databases: MIT-BIH Polysmonographic and MIMIC II databases within Physionet archive. Twenty-five subjects from these databases were collected for training. For each subject, ECG signal and blood pressure waveform, in a five-hour range of the total data were collected. For the first part of the study, the HR and SBP series for each subject are derived from ECG and arterial pressure waveforms, respectively.

The algorithm of the method developed in this study is shown in Figure 1. The proposed algorithm consists of two stages: training and testing stages. In the training stage, 25 subjects' data are used for the training purpose of the algorithm. In this stage and after collecting the data, four features which highlight the differences between normal subjects and patients, are extracted from data. We then define four criteria based on the extracted features. These four criteria which form the basis of our classification algorithm are: circle criterion, estimation error criterion, Poincare care plot deviation, and autonomic response delay criterion. In the next step and for the task of classification, we define three groups, namely, healthy, high risk, and patient. Then we design three fuzzy membership functions for each criterion to find the subject degree of membership to each group. Finally, a scoring method is developed based on the degree of membership of each case, and subjects are classified based on this scoring method.

In the following sections, we provide a step-by-step description of our method, beginning with the definition of the proposed criteria.

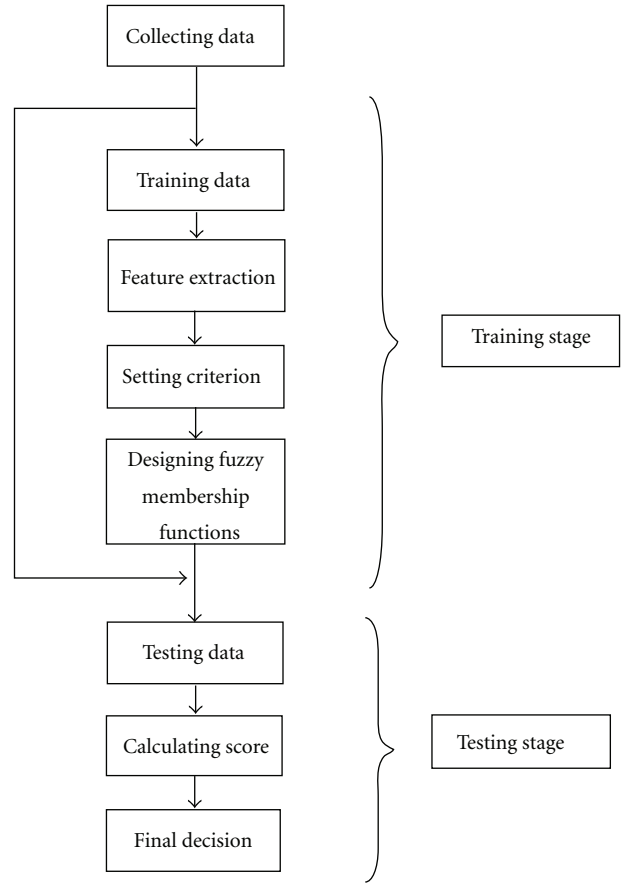


Figure 1: Schematic of the proposed algorithm. The proposed algorithm consists of two stages: training and testing. In the training stage 25 subjects' data are used to extract features to classify patients from healthy subjects. In the test stage, subjects will be divided into three predefined groups of healthy, high risk, and patient, based on their assigned score.

2.1. Circle Criterion. To evaluate the differences between healthy and patients, the SBP against HR diagram for each subject is plotted. Figure 2 shows these plots for healthy and patient cases. Clearly, the plots show a significant difference between normal subjects and abnormal patients: the data for normal subjects are concentrated, while those of the patients are scattered.

The mean value of SBP and HR for each normal subject and abnormal patient is then calculated and plotted in one diagram. Figure 3 shows the mean values for all the subjects in one diagram. The principal difference between the two groups is quite clear. This diagram reveals the fact that there are differences between the HR and SBP data in normal subjects and abnormal ones. The plot shows that the data for the normal subjects is clustered and limited in a specific area, while those of the patients are spread out through the whole plot. The first criterion is named the "circle criterion". The center of the circle is located at point "O" where its coordinates are the mean values of HR and SBP of normal patients and, in this case, is (83, 120). The radius of this circle is calculated based on Euclidian distance between the center and the outer limit of the circle.

A given subject would be considered to be a patient if its corresponding mean (HR and SBP) point is out of the healthy subject's circle (the limited area).

2.2. Estimation Error Criterion. As the second feature, a system identification method is used for the prediction of the next HR based on the current and previous HR and SBP data. A nonlinear ARX or NARX model is employed to estimate HR series [17]. NARX models in general are represented by the following equation:

$$y(t) = F\big(y(t-1), y(t-2), \ldots, y(t-n_a), u(t-n_k), \ldots, u(t-n_k-n_b+1)\big), \tag{1}$$

where, $y(t)$ and $u(t)$ are the output and input of the system, respectively. In (1) the matrix $[n_a \quad n_b \quad n_k]$ is the same as the order of the model. Model order is selected by use of the A-Information Criterion (AIC) method. This is the traditional

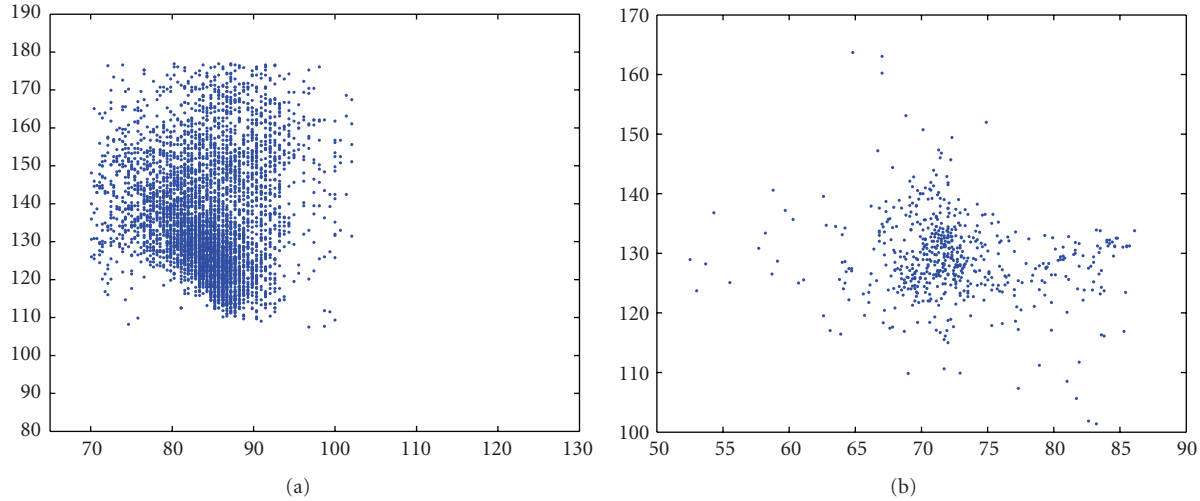


Figure 2: SBP against HR for a healthy (a) and an abnormal (b) case.

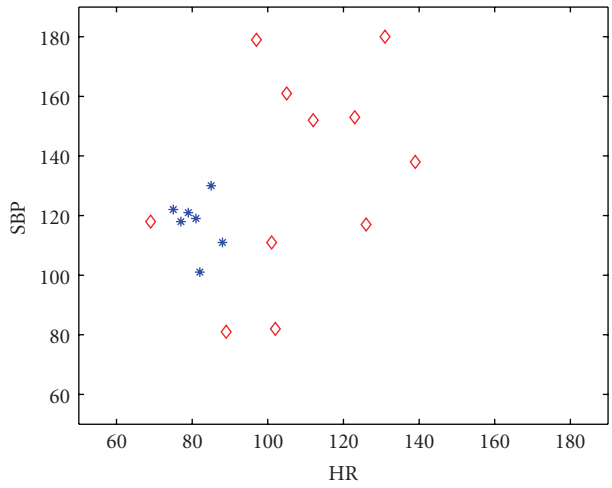


Figure 3: Mean values of SBP versus HR for all subjects.

Table 1: Error estimation for identification of HR baroreflex.

Group	Mean	Max	Min
Normal	0.193	0.238	0.119
Abnormal	0.367	0.473	0.263

Table 2: Deviation from line $y = x$ in Poincare plot.

Group	Mean	Max
Healthy	8%	13%
Patient	19%	24%

method for model order selection in cardiovascular system identification research. Model order for data in this research has been calculated to be [9 6 3].

In this criterion, an artificial neuro fuzzy inference system (ANFIS) structure is employed for the identification purpose. The model has 15 inputs and one output. Membership functions for inputs are designed based on physiological facts. Since the nervous system consists of sympathetic and parasympathetic nerves, for each input, two generalized bell-shaped membership functions are assigned to designate the sympathetic and parasympathetic functions.

The system identification results are described in Table 1. The results in this table show that differences exist in the normalized root mean square error (NRMSE) with respect to the estimation of the HR for the two groups under study. In particular, the results indicate that NRMSE is smaller for normal subjects than for patients. These differences are due to the fact that the model is designed for normal subjects, thus, the output of the model for patients has higher errors than that for normal subjects.

Based on these results and noting that the maximum error for healthy subject is 0.238, while the minimum error for patient is 0.263, we define a second criterion called "estimation error criterion." According to this criterion, the subject would be flagged as abnormal if the calculated error in HR estimation raise is more than 0.25.

2.3. Poincare Plot Deviation. A Poincare plot, named after Henri Poincare, is used to quantify self-similarity in processes which are usually characterized by periodic functions. This plot is commonly used in heart rate variability (HRV) analysis. The Poincare plot is a graph in which each heart rate episode is plotted as a function of previous HR, and then the line $y = x$ is fitted to the data. In [18], this method is also applied to classify patients with type I DM from healthy subjects. Drawing the Poincare plot for healthy and abnormal subjects, it is found that the deviation from the mentioned line is less in healthy subjects than in abnormal subjects. These plots are shown in Figure 4.

The deviation from the line $y = x$ in the Poincare plot for the two groups under study is shown in Table 2. Therefore, we define the third criterion using this deviation to characterize abnormality. Based on this criterion, subjects

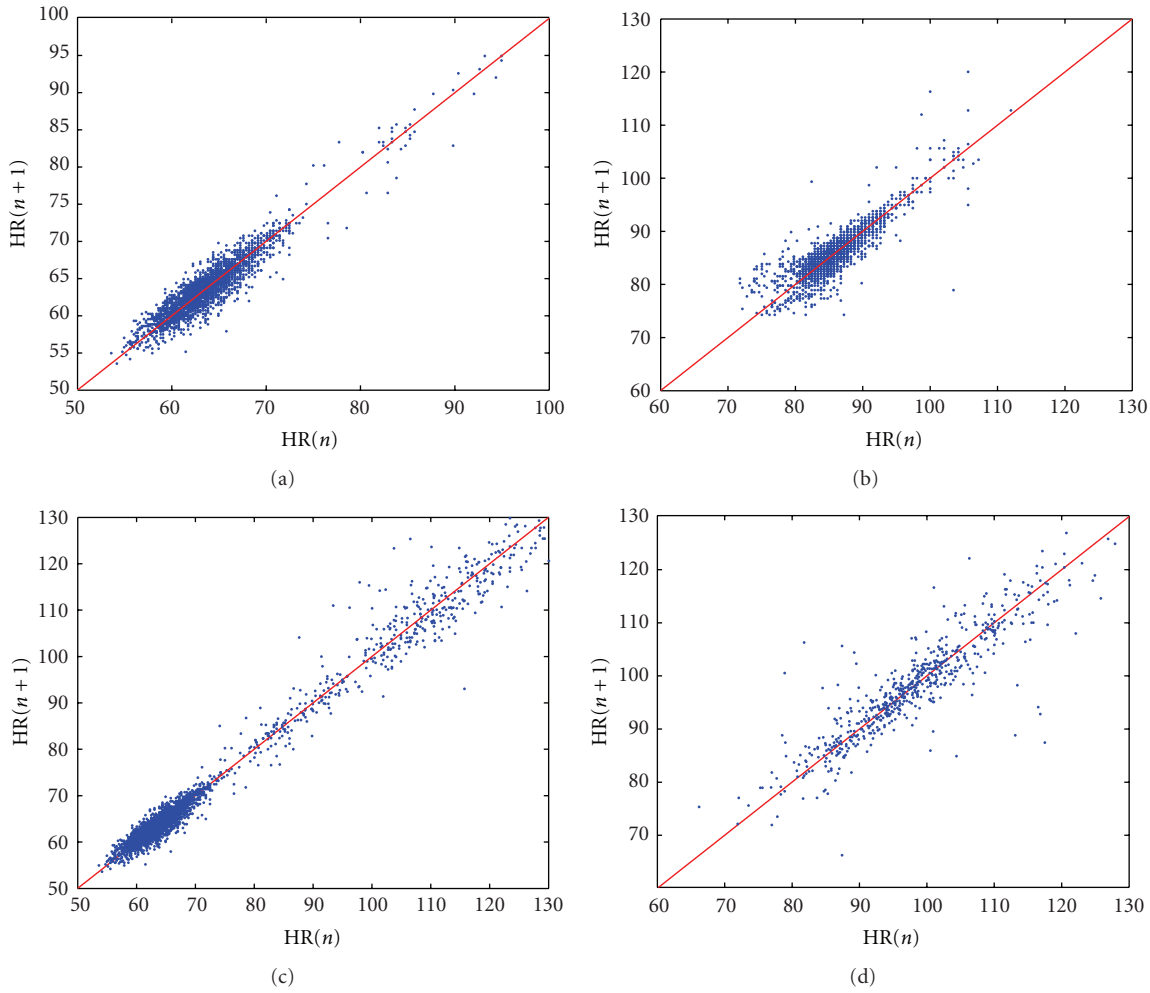


Figure 4: Poincare plots of HR for two healthy (up) and two abnormal (down) cases. The Poincare plot is plot of HR($n + 1$) versus HR(n). Line $y = x$ is illustrated in all pictures.

would be called abnormal if deviation from line $y = x$ is more than 15%.

2.4. Autonomic Response Delay Criterion. The normally occurring delay in the autonomic response to a stimulus has its origins in the parasympathetic nervous system. Calculating the delay for healthy subjects and patients, we can infer that response delays in abnormal subjects are remarkably higher than in healthy subjects. The results of calculating the delay in the autonomic response are shown in Figure 5. Fifteen abnormal patients and ten healthy subjects were involved in the training group.

The results of the delay calculations in the autonomic response are also presented in Table 3. Based on the above results, we define the fourth criterion where the subject is characterized as abnormal if the calculated delay in the autonomic response increases to more than 0.021 second.

After deriving the four criteria discussed above, an algorithm is designed to classify healthy subjects from patients. In the following section, we describe the proposed algorithm.

Table 3: Delay in autonomic response for two groups.

Group	Mean delay (sec)	Max delay (sec)
Healthy	0.015	0.02
Patient	0.038	0.06

3. Scoring Method and Classification Algorithm

Based on the evaluated criteria from training data, an algorithm is developed to automatically distinguish patients from healthy subjects. The algorithm is based on a fuzzy decision-making method. First, for each criterion, three Gaussian bell membership functions are designed as an indicator of three major groups: healthy, high risk, and patient. Since this algorithm is designed for clinical use and since there exists a high degree of uncertainty in clinical applications, we added the high-risk groups to our predefined healthy and patient groups, to account the cases that do not completely belong to the healthy or patient groups. For the training part, we first made a general guess for the shape of the membership functions. The membership functions during the training

Table 4: Results of algorithm testing on the first group. The CS is abbreviation for cumulative sum.

Case Number	Status	Healthy CS	High-risk CS	Patient CS	Category
1	Patient	0.4	0.7	2.8	Patient
2	Patient	0.2	0.6	3.1	Patient
3	Healthy	3.2	0.9	0.1	Healthy
4	Healthy	2.5	1.4	0.2	Healthy
5	Patient	0.1	1.1	2.6	Patient
6	Patient	0.3	0.4	3.4	Patient
7	Patient	0.3	2.2	1.6	High risk
8	Patient	0.5	0.8	2.9	Patient
9	Healthy	3.6	0.4	0.1	Healthy
10	Patient	0.2	2.2	1.7	High risk
11	Patient	0.2	0.6	3.0	Patient
12	Healthy	2.7	0.9	0.3	Healthy
13	Patient	0.3	0.6	3.2	Patient
14	Patient	0.1	0.3	3.5	Patient
15	Patient	2.9	0.7	0.3	Healthy

Table 5: Results of algorithm testing on the second group.

Case Number	Status	Healthy CS	High-risk CS	Patient CS	Category
1	Healthy	2.7	0.6	0.6	Healthy
2	Patient	0.3	0.8	2.9	Patient
3	Patient	0.2	0.8	3.0	Patient
4	Patient	0.3	0.4	3.3	Patient
5	Healthy	2.8	0.9	0.4	Healthy
6	Patient	0.3	0.4	3.4	Patient
7	Patient	0.4	0.8	2.8	Patient
8	Healthy	1.5	1.9	0.6	High risk
9	Patient	0.5	0.8	2.8	Patient
10	Patient	0.4	0.6	3.1	Patient
11	Patient	0.1	0.9	3.1	Patient
12	Patient	0.2	2.5	1.2	High risk
13	Patient	0.1	0.9	3.1	Patient
14	Patient	0.4	0.5	3.1	Patient
15	Healthy	3.1	0.8	0.2	Healthy

Table 6: Results of algorithm testing on the third group.

Case Number	Status	Healthy CS	High-risk CS	Patient CS	Category
1	Patient	0.1	0.7	3.2	Patient
2	Patient	0.3	0.8	2.9	Patient
3	Healthy	1.3	2.4	0.3	High risk
4	Patient	0.2	0.7	3.1	Patient
5	Patient	0.4	1.9	1.6	High risk
6	Patient	0.4	0.8	2.9	Patient
7	Healthy	2.7	0.9	0.3	Healthy
8	Patient	0.1	0.4	3.5	Patient
9	Patient	0.3	2.2	1.6	High risk
10	Patient	0.4	0.6	3.1	Patient
11	Healthy	3.0	0.7	0.3	Healthy
12	Patient	0.4	0.8	2.7	Patient
13	Patient	0.7	1.0	2.8	Patient
14	Healthy	3.2	0.6	0.2	Healthy
15	Patient	0.5	0.8	2.8	Patient

Table 7: Results of testing the algorithm on Physionet database.

Group	I	II	III	All
All	15	15	15	45
TP	12	13	12	37
FN	3	1	2	6
FP	0	1	1	2
Se (%)	80	92.8	85.7	86
PPA (%)	100	92.8	92.3	94.8
TA (%)	80	86	80	82.2

round then adapt their shape parameters to the incoming data for best classification performance. Now, the classifier is designed and ready for the testing stage. Figure 6 represents the adapted membership functions for each criterion based on the training data.

To test the developed algorithm, in the first step for each subject, all the mentioned features that form the basis of four criteria are extracted and used as an input for the four abnormality criteria. Then, for each criterion, the subject's degree of membership to all groups is evaluated. In this step, for each subject, we have 12 degrees of membership to the designed three groups, meaning four degrees of membership for each group. After evaluating the degree of memberships, the cumulative sum of the four degree of memberships of each group is calculated. In this stage, we have three numbers indicating the subject's degree of membership to each group. We call these numbers the subject's "score" for each group. A given subject will belong to the group whose score is the largest.

4. Results

From a total of seventy subject data which were collected from MIMIC II database, the algorithm was first trained with twenty-five subjects including ten healthy and fifteen patients. The training data was selected randomly to avoid bias towards a specific disease. Then, three groups of subjects were tested with each group consists of four healthy individuals and eleven patients.

The results of the test are presented in Tables 4, 5, and 6. The proposed method was applied to 45 cases from Physionet database, containing 12 healthy subjects and 33 patients. From all cases, 37 cases were accurately detected, while there was one false detection. Furthermore, in five cases a patient subject was classified as high risk and, in two cases, a healthy subject was classified as high risk.

To assess the accuracy of the classifier, sensitivity (Se), positive predictive accuracy (PPA), and total accuracy (TA) have been calculated. These are defined as follows.

$$\begin{aligned} \text{Se} &= \frac{\text{TP}}{(\text{TP}+\text{FN})}, \\ \text{PPA} &= \frac{\text{TP}}{(\text{TP}+\text{FP})}, \\ \text{TA} &= \frac{\text{TP}}{(\text{TP}+\text{FN}+\text{FP})}. \end{aligned} \tag{2}$$

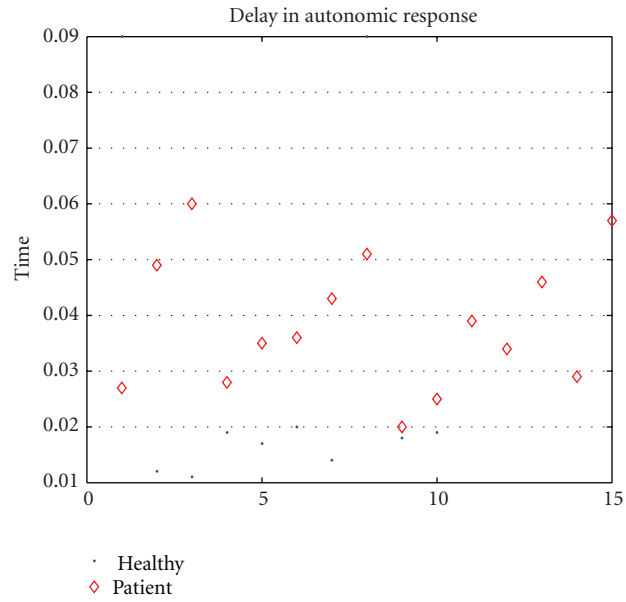


Figure 5: Delay in autonomic response.

Here, TP is the number of true positive detections, FN stands for the number of false negative detections, and FP stands for the number of false positive misdetections. Table 7 shows the overall result of the classification for all 45 cases of the 3 groups. The FP is the healthy subject who is misclassified as a high-risk subject, and FN is the patient who is misclassified as a high-risk subject. According to this table, the scoring method of the proposed algorithm results in 86% sensitivity, 94.8% positive predictive accuracy, and 82.2% total accuracy.

A comprehensive comparison between the results of different studies in the field of identified ICU needed patients by the use of hemodynamic features is very difficult since the database, signals under study, the algorithm structure, and the data processing methods are not the same in the various studies. However, in order to present an estimate of the performance of our algorithm and our classifier we show the results of this study versus the reported results of two other well-known studies in the area of ICU needed patients identifying in Table 8. As seen from this table, the algorithm in the present study shows reasonably accurate results and compares favorably with other studies. The goal of this study, which was identifying patients with ICU needs by use of the hemodynamic features, has clearly been achieved.

5. Discussion and Conclusions

In this paper a scoring method based on fuzzy logic and feature extraction is proposed to distinguish patients from healthy subjects. The method is based on the same principle that the ICU scoring methods follow: that of finding differences between hemodynamic data of healthy subjects and patients. Four different criteria are proposed to detect and identify patients from a group of subjects. For each criterion a fuzzy classifier is designed such that

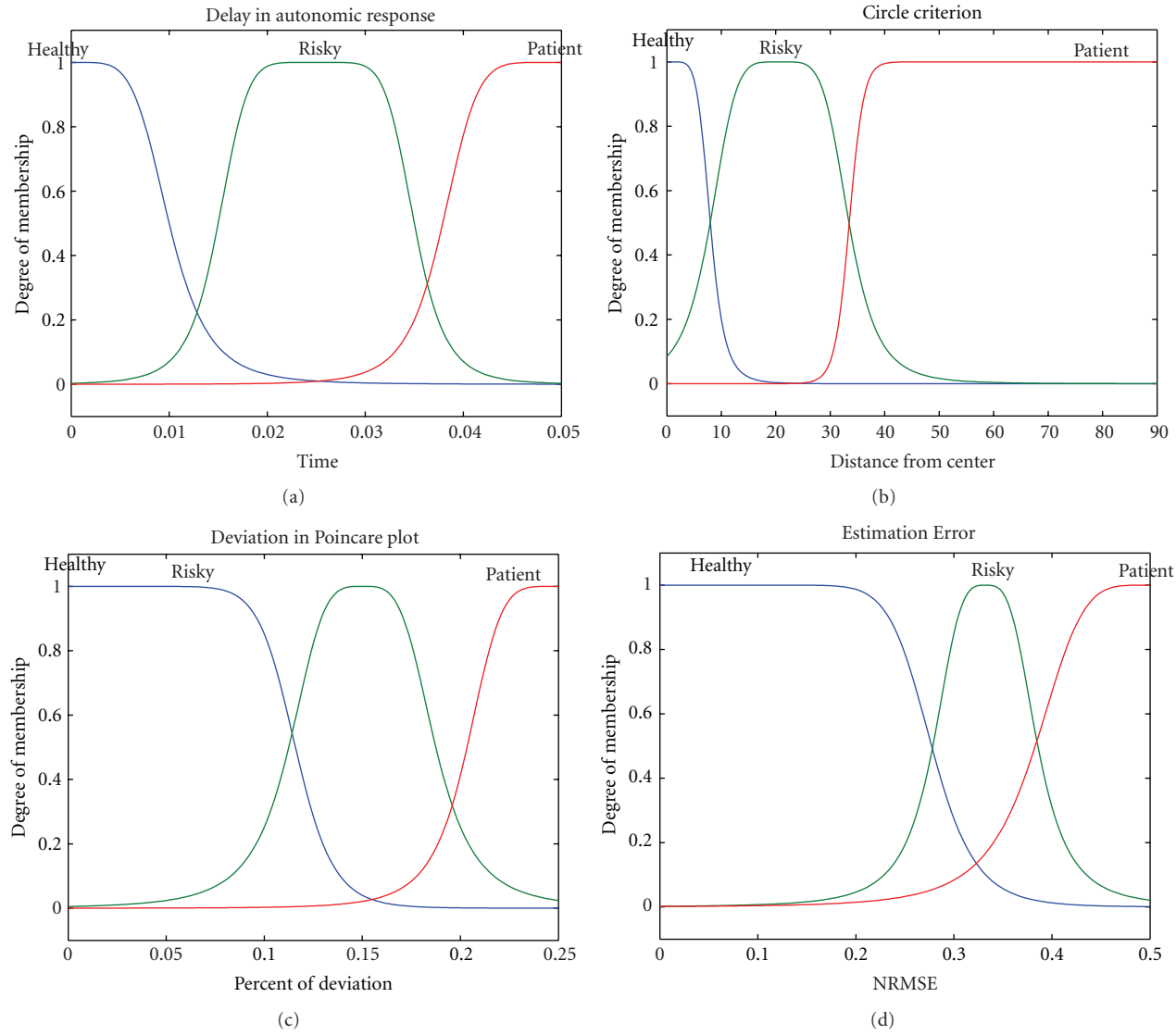


Figure 6: The designed membership functions for each criterion.

Table 8: Comparison of several classifier performances on MIMIC II ICU database (blank boxes have not been reported).

Study	Se (%)
Cao et al. [1]	75
Eshelman et al. [4]	60
This study	86

the individuals are classified into the healthy, high risk, and patient fuzzy groups. In other words, a given person may have a membership grade in all three classes. A score is assigned to the subject for a group as being equal to the sum of degree of memberships. The algorithm calculates a combined criterion based on the results of the four criteria to arrive at a classification decision for each individual.

It is shown that the algorithm is highly reliable and has been able to detect correctly all members of the first group. It is also able to detect all eleven patients in each of the next two groups correctly. Only one of the healthy members in the second and third was classified as high-risk.

In this study, four different criteria were proposed and used in the proposed algorithm in order to detect the abnormalities in testing subjects. From each testing subject, various features were extracted and used as input for the criteria, and based on the results of all four criteria, a decision was made about the type of subject, as to whether he/she is normal, a high risk subject or a patient.

The proposed algorithm gave reliable results in detecting the ICU needed patients, but still needs to be improved. The difference between the proposed method in this study and other similar research in this field of study is that by using the presented algorithm in this study, existence of any abnormality in a patient will be found, while in most of similar studies in this area, a specific abnormality is found in a patient or among a database of subjects. Therefore, our

results are more general and more useful from the point of view of clinical applications.

This method tends to be more detective rather than predictive, and this could be one drawback of the algorithm. Further investigations need to be carried out to improve its predicting ability.

References

[1] H. Cao, L. Eshelman, N. Chbat, L. Nielsen, B. Gross, and M. Saeed, "Predicting ICU hemodynamic instability using continuous multiparameter trends," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '08)*, pp. 3803–3806, August 2008.

[2] C. B. Laramee, L. Lesperance, D. Gause, and K. McLeod, "Intelligent alarm processing into clinical knowledge," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6657–6659, 2006.

[3] W. Friesdorf, B. Buss, and M. Gobel, "Monitoring alarms—the key to patient's safety in the ICU?" *Intensive Care Medicine*, vol. 25, no. 12, pp. 1350–1352, 1999.

[4] L. J. Eshelman, K. P. Lee, J. J. Frassica, W. Zong, L. Nielsen, and M. Saeed, "Development and evaluation of predictive alerts for hemodynamic instability in ICU patients," *Proceedings of the Annual Symposium proceedings/AMIA Symposium*, pp. 379–383, 2008.

[5] M. Pagani, V. Somers, R. Furlan et al., "Changes in autonomic regulation induced by physical training in mild hypertension," *Hypertension*, vol. 12, no. 6, pp. 600–610, 1988.

[6] A. Blasi, J. Jo, E. Valladares, B. J. Morgan, J. B. Skatrud, and M. C. K. Khoo, "Cardiovascular variability after arousal from sleep: time-varying spectral analysis," *Journal of Applied Physiology*, vol. 95, no. 4, pp. 1394–1404, 2003.

[7] D. R. Westenskow, J. A. Orr, F. H. Simon, H. J. Bender, and H. Frankenberger, "Intelligent alarms reduce anesthesiologist's response time to critical faults," *Anesthesiology*, vol. 77, no. 6, pp. 1074–1079, 1992.

[8] C. L. Tsien, I. S. Kohane, and N. McIntosh, "Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit," *Artificial Intelligence in Medicine*, vol. 19, no. 3, pp. 189–202, 2000.

[9] B. Müller, A. Hasman, and J. A. Blom, "Evaluation of automatically learned intelligent alarm systems," *Computer Methods and Programs in Biomedicine*, vol. 54, no. 3, pp. 209–226, 1997.

[10] K. Becker, B. Thull, H. Käsmacher-Leidinger et al., "Design and validation of an intelligent patient monitoring and alarm system based on a fuzzy logic process model," *Artificial Intelligence in Medicine*, vol. 11, no. 1, pp. 33–53, 1997.

[11] A. Singh and J. V. Guttag, "A Comparison of Non-symmetric Entropy-based Classification trees and Support Vector Machine for Cardiovascular Risk Stratification," in *Proceedings of the Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC '11)*, pp. 79–82, 2011.

[12] D. L. Timms, S. D. Gregory, M. C. Stevens et al., "Hemodynamic modeling of cardiovascular system using mock circulation loops to test cardiovascular devices," in *Proceedings of the 33rd Annual International IEEE EMBS Conference*, pp. 4301–4304, Boston, Mass, USA, August—September 2011.

[13] P. Ghorbanian, A. Jalali, A. Ghaffari et al., "An improved procedure for detection of heartarrhythmias with novel preprocessing techniques," *Expert Systems*. In press.

[14] U. Rajendra Acharya, P. Subbanna Bhat, S. S. Iyengar, A. Rao, and S. Dua, "Classification of heart rate data using artificial neural network and fuzzy equivalence relation," *Pattern Recognition*, vol. 36, no. 1, pp. 61–68, 2003.

[15] Y. Özbay, R. Ceylan, and B. Karlik, "A fuzzy clustering neural network architecture for classification of ECG arrhythmias," *Computers in Biology and Medicine*, vol. 36, no. 4, pp. 376–388, 2006.

[16] J. Lacroix and J. Cotting, "Severity of illness and organ dysfunction scoring in children," *Pediatric Critical Care Medicine*, vol. 6, no. 3, pp. S126–S134, 2005.

[17] A. Jalali, A. Ghaffari, P. Ghorbanian, and C. Nataraj, "Identification of sympathetic and parasympathetic nerves function in cardiovascular regulation using ANFIS approximation," *Artificial Intelligence in Medicine*, vol. 52, no. 1, pp. 27–32, 2011.

[18] M. Javorka, Z. Lazarova, I. Tonhajzerova et al., "Baroreflex analysis in diabetes mellitus: linear and nonlinear approaches," *Medical and Biological Engineering and Computing*, vol. 49, no. 3, pp. 279–288, 2011.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2012, Article ID 272570, 7 pages
doi:10.1155/2012/272570

Research Article

An Intelligent Dynamic MRI System for Automatic Nasal Tumor Detection

Wen-Chen Huang and Chun-Liang Liu

Department of Information Management, National Kaohsiung First University of Science and Technology, Kaohsiung 811, Taiwan

Correspondence should be addressed to Wen-Chen Huang, wenh@nkfust.edu.tw

Received 30 July 2011; Revised 14 November 2011; Accepted 16 November 2011

Academic Editor: Jiann-Shing Shieh

Dynamic magnetic resonance images (DMRIs) are one of the major tools for diagnosing nasal tumors in recent years. The purpose of this research is to propose a new method to be able to automatically detect tumor region and compare three classifiers' tumor detection performance for DMRI. These three classifiers are AdaBoost, SVM, and Bayes-Gaussian classifier. Three measurable metrics, sensitivity, specificity, accuracy values, match percent, and correspondence ratio, are used for evaluation of each specific classifiers. The experimental results show that SVM has the best sensitivity value, and Bayesian classifier has the best specificity and accuracy values. Moreover, the detected tumor regions that are marked with red color are shown by using each of these three classifiers.

1. Introduction

MRI provides a variety of different cross-section digital images, showing the structure of the nasopharynx and the source of the disease. Nasopharyngeal carcinoma (NPC) is a malignant tumor, located in the skull base and often occurs in South Asia. The capacity of the tumor for the NPC is a major harbinger of indicators. Therefore, it is necessary to estimate the capacity of the tumor. DMRI is a major nasal tumor detection tool, which has been widely used for radiation research [1]. Huang et al. [2] apply an RSI (relative signal increase) curve to identify these recurrent tumors. Hsu et al. [3] use pharmacokinetic analysis to identify NPC region. Zhou et al. [4] propose a fuzzy clustering method to filter out normal tissue region in T1-weighted (T1W) and contrast-enhanced T1W (CET1W) images. Lee et al. [5] uses Bayesian probability calculation and local histogram in T2-weighted images, compared to T1W's intensity variety to detect NPC region. Huang and Chang [6] use Fuzzy C-means (FCM) [7] and grey prediction to separate the tumor and normal tissues. Zhou et al. [8] use knowledge-based fuzzy clustering (KBFC), maximum likelihood, and seed growing to identify tumor region and analyze each method's accuracy.

Support Vector Machines (SVMs) [9, 10] have been widely used in tumor segmentation. Zhang et al. [11] compare the tumor segmentation results with one-class SVM and two-class SVM. They use MP (Match Percent) and CR (Correspondence Ratio) to evaluate their performance. The results showed that one-class SVM is superior to two-class SVM. Zhou et al. [12] use two-class SVM and kernel trick [13] to derive a new algorithm called query-based two-class SVM classifier, which is better than traditional MLP-based classifier [14]. It is available for radiologist to use as a preoperative diagnostic tool. Ritthipravat et al. [15] use region growing method and probabilistic map to find some candidate tumor region. Zhou et al. [12] also mentioned that so far, in the current diagnostic imaging or radiation therapy, a radiologist or radiation therapists need to manually describe the scope of the tumor. Therefore, we propose this new algorithm to be able to detect tumor and draw the candidate regions automatically for radiologists.

2. System Framework

In the training process, a tumor ROI is extracted by hand as a ground truth. Therefore, we have two parts of data—normal and tumor. Three classifiers, AdaBoost, SVM, and Bayes-Gaussian, are used to classify these two groups of

Table 1: The number of slices of each DMRI group.

DMRI	E1322	E1563	E1971	E2074	E6618	E6632
Number of slices	15	7	6	8	8	12

data. After training process, the proposed system is capable of distinguishing tumor region automatically and the performance of each classifier is provided.

2.1. System Framework. The dynamic nuclear magnetism radiography (DMRI) [16] is a sequence of MRI by injecting Gadolinium developing agent into patient's nasal region after every 0, 5, 30, 60, 120, 300, ... seconds. After injecting Gadolinium developing agent, the gray level intensity of patient's tumor spot will be gradually increased, finally reaching a stable state. This continuous MRI-scanned technology, making the tumor region and normal tissues have different grayscale value changes. The changes in the tumor region than normal tissue are larger as well as faster. According to this characteristic we can do a preliminary observation of tumor's size and region. However, it is very difficult to identify the tumor region by the naked eye. Therefore, the development of an effective detection system to help doctors make the diagnosis is necessary. Each MRI image has $256 * 256$ pixels. This research used 6 groups of patient's materials to make the training and the testing. The numbers of slices of each MRI group are not all the same. Table 1 shows these 6 groups of materials, respectively, by E1322, E1563, E1971, E2074, E6618, and E6632.

As the reagent, after injection, tumor location at different times with the grayscale value of image intensity will be different. Therefore, we will first have to act according to these DMRI intensity difference. Our method definition is as follows:

$$D_i = A_i - A_0, \tag{1}$$

A_i is the MRI which is obtained in the different time spots, but A_0 is in the DMRI's first chart (before injecting agents).

As the tumor over time made some of grayscale intensity values increased, while the distribution of grayscale value is between the ranges of 0 and 255 where 0 represents the black and 255 is white. A threshold value is set up according to histogram distribution. The purpose of setting up threshold is to remove background and some unlikely tumor regions, therefore, to increase the detection accuracy. After computing the intensity difference between the ith DMRI and the first DMRI, N consecutive stacks of images are shown as

$$\text{stack}(d_1, d_2, d_3, d_4, d_5, \ldots d_N). \tag{2}$$

Then ROI is chosen. ROI is the tumor area. Otherwise, the region belongs to a normal tissue area. These two distinct data are used for training and testing by several classification tools based on the ground truth [16]. Some evaluation parameters are used for the accuracy confirmation. Overall, the flowchart of the system framework is shown in Figure 1.

2.2. Evaluation Parameters. This paper has used sensitivity, specificity, accuracy, MP (Match Percent), and CR (correspondence ratio) as five kinds of evaluation parameters. They are defined as follows:

$$\begin{aligned}
\text{TP} + \text{FP} &= 1,\\
\text{FN} + \text{TN} &= 1,\\
\text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}},\\
\text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}},\\
\text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},\\
\text{Match Percent (MP)} &= \frac{\text{TP}}{\text{GT}}(\%),\\
\text{Correspondence Ratio (CR)} &= \frac{\text{TP} - 0.5 * \text{FP}}{\text{GT}}(\%),
\end{aligned} \tag{3}$$

where positive represents the tumor region and negative represents the normal tissue. TP, FP, TN, FN, and GT definitions are as follows: TP: true positive when system determines a tumor region in fact as tumor region, FP: false positive when system determines what is not tumor region actually as a tumor region), TN: true negative when system determines what is normal tissue in fact as normal tissue), FN: false negative system when determines what is not normal tissue in fact as normal tissue), and GT: ground truth (standard tumor region, doctor thought there is a tumor region).

2.3. AdaBoost. AdaBoost is evolved by Boosting. Boosting means the promotion and is proposed by Valiant [17]. It belongs to PAC (Probably Approximately Correct) architecture of machine learning domain. Kearns and Valiant [18] proposed a weak learning ability concept and has improved the concept of Valiant. The weak learning ability concept is that, after several iterations, a weak learner with the voting mechanism (majority vote) is better than a strong learner. Such a concept is confirmed by Schapire [19]. Freund and Schapire [20] unify the above concept to propose the AdaBoost method. AdaBoost is that in the training process after several iterations, a weak learner becomes a strong learner by constantly upgrading.

In the training process, given an equal weight, first result is obtained through a weak learner. Then after voting mechanism, the misclassified data are given bigger weight and the correct classified data are given lower weight. In this way, those data which have worse classifier are forced to have more training. Therefore, a training error is gradually decreased through several iterations. Eventually, a weak learner becomes a strong learner. Many recent AdaBoost researches focus on the choice of weak learner. According to many research reports, the accuracy of weak learner cannot be too high, about 50% is the maximum. Leshem and Ritov [21] proposed a new AdaBoost method was treated as the weak learner with Random Forests [22], such algorithm

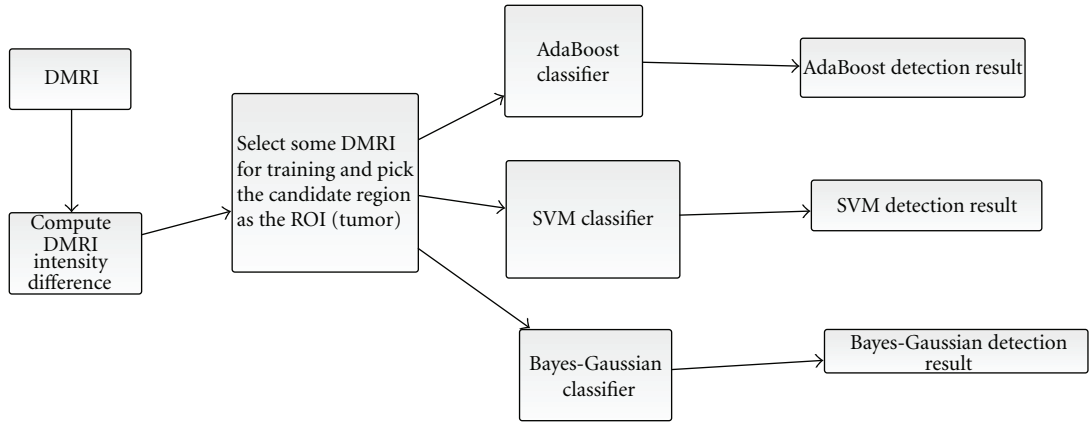


Figure 1: The proposed system framework.

achieves very good classified effect and reduces classified error rate. We adopt Leshem's method [21] as the AdaBoost classifier in this paper.

2.4. Support Vector Machines (SVMs). Support vector machines are proposed by Cortes and Vapnik [10] and are based on the statistical learning theory, mainly with binary classification. It is also widely used in a variety of practical problems, such as pattern recognition, document classification, and biological information. The concept of SVM is through the data to construct an optimal hyper plane. This hyper plane is used as an interface to classify two different groups of data. In recent years the related SVM research focuses on searching for the parameter of cost and gamma, because parameters with the correct classification rate have a great relationship. Lin et al. [23] proposed new SVM classified tool (http://www.csie.ntu.edu.tw/~cjlin/ libsvmtools/#4). Libsvm improves the original SVM concept and provides the tool to search for the best SVM's parameter automatically. Libsvm achieves the extremely good classified accuracy. Therefore, we use Libsvm to carry on the classification to work and to analyze the result.

2.5. Bayes Classifier for Gaussian Pattern Classes. In medical image segmentation, Bayes classifier is widely used. It is originally from Bayesian theorem and is used to judge the unknown category through the probability of statistical analysis to minimize a classification error. The use of supervised learning, classification must be prior knowledge of classification of patterns, and training through the training sample study, to effectively deal with the future data classification. It is based on Bayesian theorem that the exchange prior and posteriori probabilities with the decision of classification characteristics among the various attributes are conditional independence assumptions, to predict the outcome of classification. The principle of Bayes classifier is to use the attribute's relationship, through the training sample, to study the classification mechanism. Bayesian classification theory is based on the statistical principle of the classification; each type of sample comes to its average value and standard deviation. Under normal distribution, the variable x with variable probability of emergence of $P(x)$ is defined as follows [24]:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-u)^2}{\sigma^2}\right], \tag{4}$$

where σ is the standard deviation and u is the mean value.

Table 2: The sensitivity, specificity, and accuracy values after AdaBoost classifier.

Data	1322	1563	1971	2074	6618	6632	average
Sensitivity	0.9885	0.8958	0.9807	0.9985	0.9705	0.9839	0.9697
Specificity	0.7283	0.5725	0.6546	0.9440	0.6615	0.6896	0.7084
Accuracy	0.8112	0.6225	0.7340	0.9696	0.7404	0.7725	0.7750

3. Experimental Results

This section shows experimental results based on three classifiers: AdaBoost, SVM, and Bayes-Gaussian classifier. Table 2 shows the sensitivity, specificity and accuracy values after AdaBoost classifier. Table 3 shows the MP and CR values after AdaBoost classifier. Figure 2 shows the detection result images after AdaBoost classifier, where (a) is for ground truth, and in (b) color blue is for true positive and color red is for false negative. Table 4 shows the sensitivity, specificity and accuracy values after SVM classifier. Table 5 shows the MP and CR values after SVM classifier. Table 6 shows the sensitivity, specificity, and accuracy values after Bayes-Gaussian classifier. Figure 3 show the detection result images after SVM classifier, where (a) is for ground truth, and in (b) color blue is for true positive and color red is for false negative. Table 7 shows the MP and CR values after Bayes-Gaussian classifier. Figure 4 shows the detection images after Bayes-Gaussian classifier, where (a) is for ground truth (b) shows the detection result by color white, and (c) color blue is for true positive and color red is for false negative.

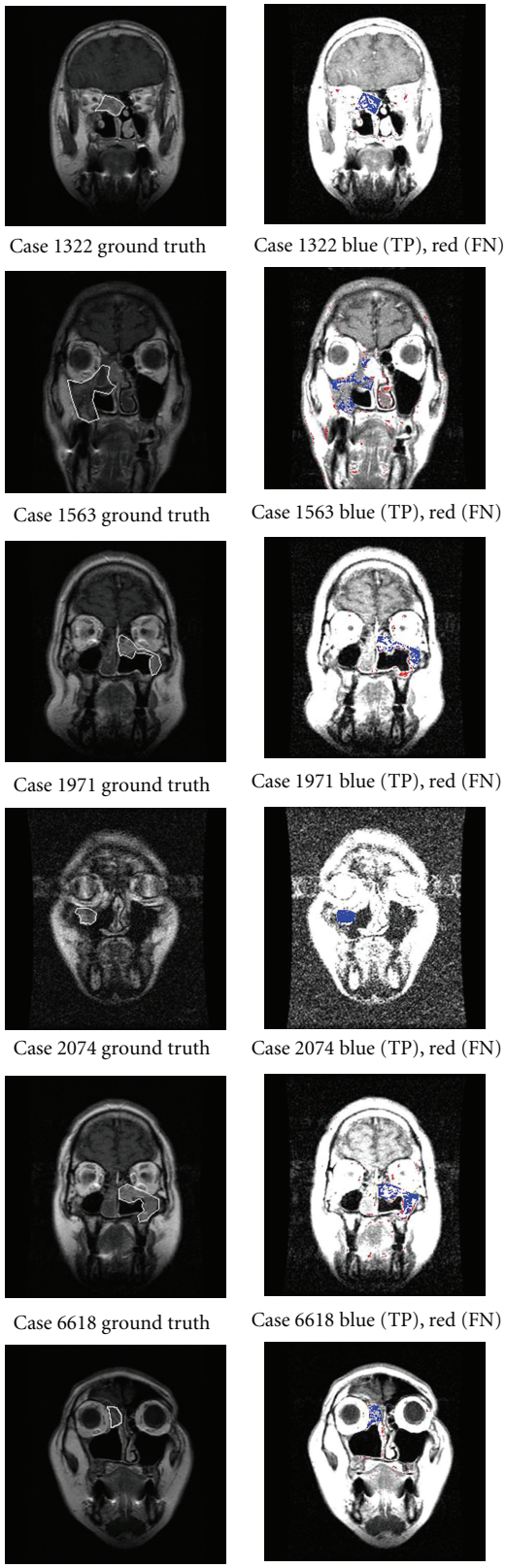


Figure 2: AdaBoost detection results. (a) Ground truth; (b) color blue is for true positive and color red is for false negative.

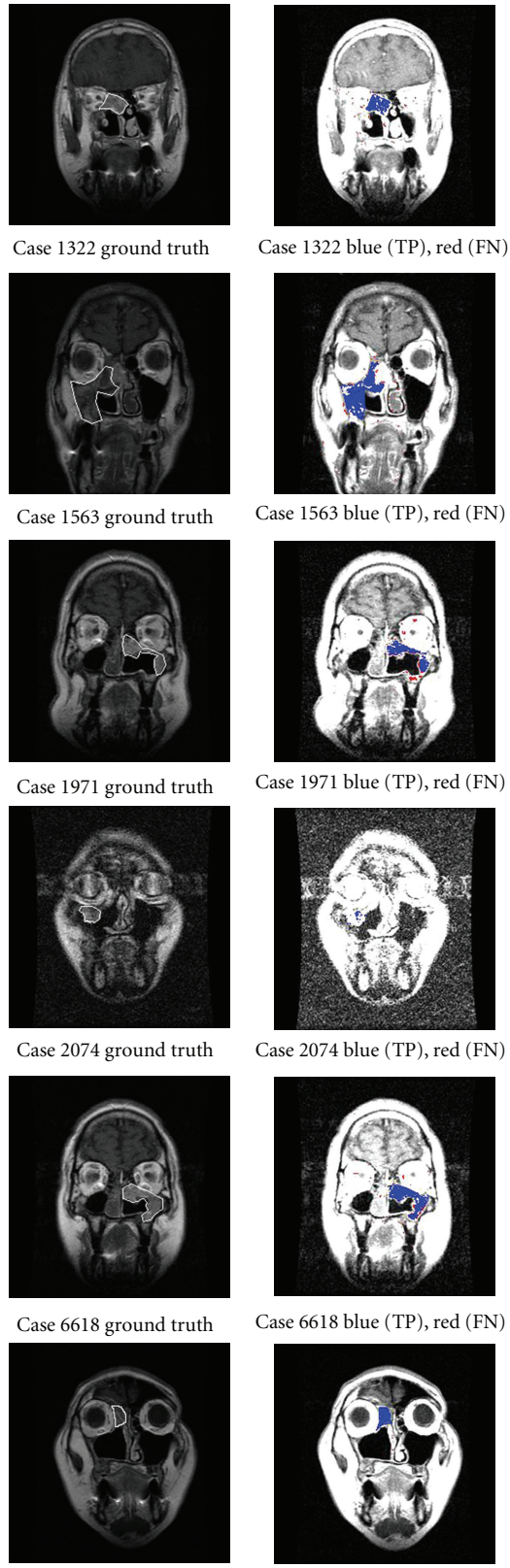


Figure 3: SVM detection results. (a) Ground truth; (b) color blue is for true positive and color red is for false negative.

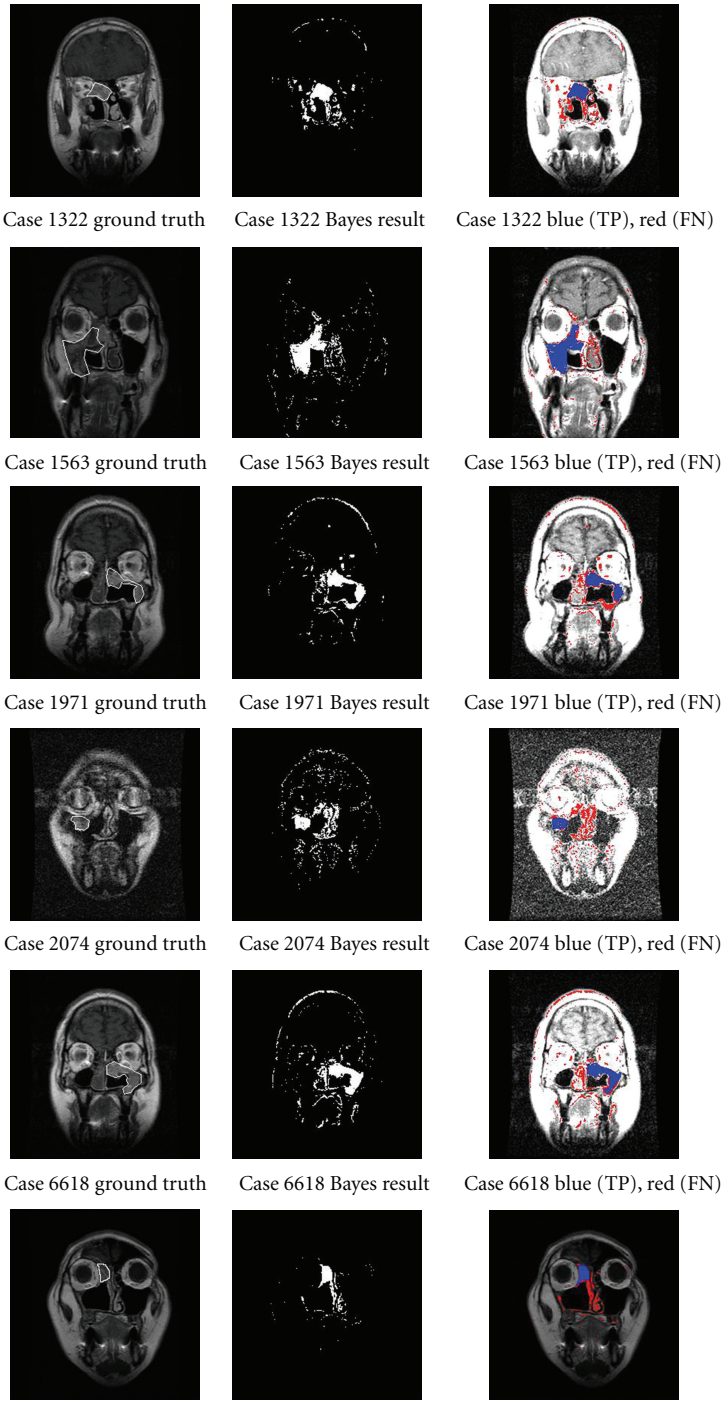


Figure 4: Bayes-Gaussian detection results. (a) Ground truth; (b) detection result shown by color white; (c) color blue is for true positive and color red is for false negative.

4. Conclusions

The probability of tumor recurrence in nasopharyngeal carcinoma is high, and relapse after the effects of the use of radiation therapy is far less than that used for the first time. Through the use of surgical resection, preoperative diagnosis becomes extremely important. This study proposed a new method and used this method in the data of multiple operations and training, testing, then to improve the tumor identification rate. It is based on this foundation to build

Table 3: The MP and CR value after AdaBoost classifier.

No.	#TP	#FP	#FN	#GT	MP	CR
E1322	255	150	82	405	0.63	0.44
E1563	365	951	257	1316	0.28	−0.08
E1971	285	312	120	597	0.48	0.22
E2074	254	16	12	270	0.94	0.91
E6618	361	367	172	728	0.50	0.24
E6632	154	124	81	278	0.55	0.33
Average					0.56	0.34

Table 4: The sensitivity, specificity, and accuracy values after SVM classifier.

Data	1322	1563	1971	2074	6618	6632	Average
Sensitivity	0.9938	0.9855	0.9870	1	0.9916	0.9987	0.9927
Specificity	0.8484	0.8328	0.7839	0.5466	0.8947	0.9788	0.8142
Accuracy	0.9086	0.8949	0.8587	0.5852	0.9379	0.9886	0.8623

Table 5: The MP and CR value after SVM classifier.

No.	#TP	#FP	#FN	#GT	MP	CR
E1322	333	72	57	405	0.82	0.73
E1563	1055	261	94	1316	0.80	0.70
E1971	434	163	122	597	0.73	0.59
E2074	46	224	0	270	0.17	−0.24
E6618	643	85	86	728	0.88	0.82
E6632	272	6	11	278	0.98	0.97
Average					0.73	0.59

Table 6: The sensitivity, specificity, and accuracy values after Bayes-Gaussian classifier.

Data	1322	1563	1971	2074	6618	6632	Average
Sensitivity	0.9310	0.9050	0.9207	0.8551	0.9050	0.9581	0.9125
Specificity	0.9791	0.9435	0.9481	0.9825	0.9507	0.9888	0.9655
Accuracy	0.9538	0.9234	0.9340	0.9091	0.9266	0.9730	0.9367

Table 7: The MP and CR value after Bayes-Gaussian classifier.

No.	#TP	#FP	#FN	#GT	MP	CR
E1322	397	8	813	405	0.98	0.97
E1563	1245	71	791	1316	0.95	0.92
E1971	567	30	1044	597	0.95	0.92
E2074	266	4	1382	270	0.98	0.98
E6618	694	34	1144	728	0.95	0.93
E6632	275	3	387	278	0.99	0.98
Average					0.97	0.95

the medical use of nasopharyngeal tumor recognition system, to ensure compatibility with the physician for the disease diagnosis, treatment, and pathology monitoring. This study proposes AdaBoost algorithm, support vector machines, and Bayesian classification classifiers to distinguish the tumor region with the normal region. The tumor detecting validation comparisons on the three classifiers are through MP value, CR value, sensitivity, specificity, and accuracy of five parameters. Final results show that SVM has the best sensitivity value, and Bayesian classifier has the best specificity and accuracy values. Some morphology techniques, such as open and close, are used to make a more complete regional cancer identification. The proposed system has no intention to replace the physician's diagnosis job. By using visual aids of this recognition system the physicians' misjudgment to enhance the overall diagnosis accuracy reduces.

References

[1] G. Brix, G. Layer, P. Schlag et al., "Recurrent rectal cancer: diagnosis with dynamic MR imaging," *Radiology*, vol. 189, no. 3, pp. 881–889, 1993.

[2] W. C. Huang, C. C. Hsu, C. Lee, and P. H. Lai, "Recurrent nasal tumor detection by dynamic MRI," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 4, pp. 100–105, 1999.

[3] C. C. Hsu, P. H. Lai, C. Lee, and W. C. Huang, "Automated nasopharyngeal carcinoma detection with dynamic gadolinium-enhanced MR imaging," *Methods of Information in Medicine*, vol. 40, no. 4, pp. 331–335, 2001.

[4] J. Zhou, T. K. Lim, V. Chong, and J. Huang, "Segmentation and visualization of nasopharyngeal carcinoma," *Computers in Biology and Medicine*, vol. 33, no. 5, pp. 407–424, 2003.

[5] F. K. H. Lee, D. K. W. Yeung, A. D. King, S. F. Leung, and A. Ahuja, "Segmentation of nasopharyngeal carcinoma (NPC) lesions in MR images," *International Journal of Radiation Oncology Biology Physics*, vol. 61, no. 2, pp. 608–620, 2005.

[6] W. C. Huang and C. P. Chang, "Automatic nasal tumor detection by grey prediction and fuzzy C-means clustering," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 542–547, October 2006.

[7] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 672–682, 1992.

[8] J. Zhou, V. Chong, T. K. Lim, and J. Houng, "MRI tumor segmentation for nasopharyngeal carcinoma using knowledge-based fuzzy clustering," *International Journal of Information Technology*, vol. 8, no. 2, 2002.

[9] K.-P. Lin and M.-S. Chen, "On the design and analysis of the privacy-preserving SVM classifier," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1704–1717, 2011.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[11] J. Zhang, K. Ma, and V. Chong, "Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine," in *Proceedings of the International Workshop on Advanced Image Technology (IWAIT '04)*, 2004.

[12] J. Zhou, K. L. Chan, P. Xu, and V. Chong, "Nasopharyngeal carcinoma lesion segmentation from MR images by support vector machine," in *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pp. 1364–1367, April 2006.

[13] B. Scholkopf and A. J. Smola, *Learning with Kernels Support Vector Machines: Regularization, Optimization and Beyond*, MIT Press, Cambridge, Mass, USA, 2002.

[14] M. G. Penedo, A. Mosquera, M. J. Carreira, and D. Cabello, "Lung nodule detection in curvature space with multilayer perceptron network," in *Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, pp. 1130–1132, Amsterdam, The Netherlands, November 1996.

[15] P. Ritthipravat, C. Tatanun, T. Bhongmakapat, and L. Tuntiyatorn, "Automatic segmentation of nasopharyngeal carcinoma from CT images," in *Proceedings of the 1st International Conference on BioMedical Engineering and Informatics (BMEI '08)*, pp. 18–22, May 2008.

[16] L. Wang, Z. Yuemin, and L.; Hongying, "Multiscale modeling and simulation of the cardiac fiber architecture for DMRI," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 16–19, 2012.

[17] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.

[18] M. Kearns and L. Valiant, "Learning boolean formulae or finite automata is as hard as factoring," Tech. Rep. TR-14-88, Aiken Computation Laboratory, Karrard University, Cambridge, Mass, USA, 1988.

[19] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.

[20] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[21] G. Leshem and Y. Ritov, "Traffic flow prediction using adaboost algorithm with random forests as a weak learner," in *Proceedings of the International Conference on Computer, Information, and Systems Science, and Engineering*, January 2007.

[22] B. Leo, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] C. J. Lin, R. C. Weng, and S. S. Keerthi, "Trust region Newton method for large-scale logistic regression," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, 2007.

[24] Q. M. Chang and T.-T. Lin, "Neural network and bayesian classification applied to image segmentation a comparative study," *Journal of Agricultural Machinery*, vol. 8, no. 3, 1999.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2012, Article ID 863549, 6 pages
doi:10.1155/2012/863549

Research Article

Medical Pattern Recognition: Applying an Improved Intuitionistic Fuzzy Cross-Entropy Approach

Kuo-Chen Hung

Department of Logistics Management, National Defense University, No. 70, Section 2, Jhongyang N. Road, Beitou District, Taipei City 112, Taiwan

Correspondence should be addressed to Kuo-Chen Hung, kuochen.hung@msa.hinet.net

Received 14 June 2011; Accepted 22 August 2011

Academic Editor: Maysam Abbod

One of the toughest challenges in medical diagnosis is the handling of uncertainty. Since medical diagnosis with respect to the symptoms uncertain, they will be assumed to have an intuitive nature. Thus, to obtain the uncertain optimism degree of the doctor, fuzzy linguistic quantifiers will be used. The aim of this article is to provide an improved nonprobabilistic entropy approach to support doctors examining the work of the preliminary diagnosing. The proposed entropy measure is based on intuitionistic fuzzy sets, extrainformation regarding hesitation degree, and an intuitive and mathematical connection between the notions of entropy in terms of fuzziness and intuitionism has been revealed. An illustrative example for medical pattern recognition demonstrates the usefulness of this study. Furthermore, in order to make computing and ranking results easier and to increase the recruiting productivity, a computer-based interface system has been developed to support doctors in making more efficient judgments.

1. Introduction

Medical diagnostic investigations are very important and complex. Uncertainty is an inseparable aspect of medical diagnosis problems. A symptom is an uncertain indication of a disease as it may or may not occur with or as a result of the disease. Uncertainty characterizes a relation between symptoms and diseases [1, 2]. Hence, coping efficiently with uncertainty leads us to more accurate decision making, and this is considered as a fundamental challenge in medicine.

The fuzzy set framework has been utilized in several different approaches to model the diagnostic process. In the approach formulated by Sanchez [3] in 1979, he adopted the compositional rule of inference by Zadeh [4] as an inference mechanism. It accepts fuzzy descriptions of a patient's symptoms and infers fuzzy descriptions of patient's diseases by means of the fuzzy relationships described before.

With S, a set of symptoms, and a fuzzy relation R from S to D, where D is a set of diseases. And with P, a set of patients, and a fuzzy relation Q from P to S, and by "max-min composition," we get the fuzzy relation $T = Q \times R$ with the membership function

$$\mu_T(p,d) = \max_{s\in S} \left(\min\{\mu_Q(p,s), \mu_R(s,d)\}\right), \quad s \in S,\ d \in D,\ p \in P. \tag{1}$$

Furthermore, based on the concepts of fuzzy sets (FSs) theory, numerous fuzzy approaches to medical diagnosis have been addressed, and the readers can refer to [3, 5–13].

In addition, several investigations in medical diagnosis have addressed these issues based on intuitionistic fuzzy sets (IFSs), such as Shannon et al. [14] developed an intuitionistic fuzzy logic approach for decision making in medicine; De et al. [1] used intuitionistic fuzzy sets to apply to medical diagnosis; Ahn et al. [15] have also presented a medical diagnostic method by applying intuitionistic fuzzy set to perform classification of solution sets and linear regression approach; Innocent and John [16] presented a new method for computing a diagnostic support index which uses vague

symptom and temporal information in a clinical diagnosis context; Szmidt and Kacprzyk [17] provided a measure method for intuitionistic fuzzy sets and showed its usefulness in medical diagnostic reasoning; Xu [18] proposed a new method for deriving the correlation coefficients to the interval-valued intuitionistic fuzzy set theory and showed its application in medical diagnosis; Todorova et al. [19, 20] provided two algorithms by intuitionistic fuzzy Voronoi diagrams (IFVDs) mechanism to estimate the classification in medicine; Khatibi and Montazer [21] take an intelligent approach towards the bacteria classification problem by using five similarity measures of FSs and IFSs to examine their capabilities in encountering uncertainty in the medical pattern recognition. Currently, Ye [22] also presents a cosine similarity measure and a weighted cosine similarity measure based on the concept of the cosine similarity measure for IFSs.

Furthermore, Vlachos and Sergiadis [23] presented an intuitionistic fuzzy cross-entropy (IFCE) approach to discrimination measures for IFSs and their application in medical diagnosis. Zhang and Jiang [24] have also proposed two kinds of new vague information entropy measures which are vague entropy and vague cross-entropy to solve medical diagnosis problem.

Among the IFCE approach, this paper also provides an improvement to examine their capabilities to cope with uncertainty in the medical pattern recognition.

The remainder of this study is organized as follows. In Section 2, the relative basic concepts including IFSs, fuzzy cross-entropy, and IFCE methods are described. Section 3 presents an improved IFCE approach, illustrates an example for medical diagnosis, and provides some discussions in Section 4. A computer-based interface has developed in Section 5. Finally, a conclusion is drawn in Section 6.

2. Related Works

2.1. IFSs Theory. Fuzzy sets theory, proposed by Zadeh [4] in 1965, has shown successful applications in various fields. In this theory, the membership of an element to a fuzzy set is a single value between zero and one. But in reality, it may not always be certain that the degree of nonmembership of an element to a fuzzy set is just equal to 1 minus the degree of membership, that is, there may be some hesitation degree. Thus, as a generalization of fuzzy sets, the concept of IFSs was introduced by Atanassov [25] in 1983. Burillo and Bustince [26] showed that this notion coincides with the notion of vague sets (VSs).

The IFSs is as an extension of fuzzy sets. An IFSs A in a fixed set E is an objective with the expression

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle \mid x \in E\}, \tag{2}$$

where the functions $\mu_{\mathrm{A}} : \mathrm{E} \to [0,1]$ and $\nu_{\mathrm{A}} : \mathrm{E} \to [0,1]$ denote the degree of membership and the degree of nonmembership of the element $x \in E$, respectively. For every $x \in E$,

$$0 \le \mu_A(x) + \nu_A(x) \le 1. \tag{3}$$

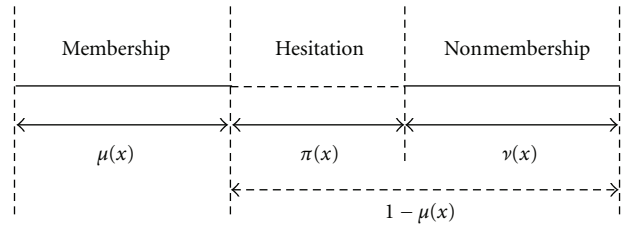


Figure 1: The descriptions of membership, nonmembership and hesitation degree.

When $\mu_A(x) + \nu_A(x) = 1$, for every $x \in E$, then the IFSs will degenerate to a fuzzy set. Hence, we can consider a fuzzy set with its membership function $\mu_A(x)$, having the IFSs expression as

$$A = \{\langle x, \mu_A(x), 1 - \mu_A(x)\rangle \mid x \in E\}, \tag{4}$$

under the condition of $\nu_A(x) = 1 - \mu_A(x)$, for every $x \in E$.

For each IFSs A in E, "hesitation degree" has consideration. We call the hesitation degree of an element $x \in E$ in A by the following expression:

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x), \tag{5}$$

and we can consider $\pi_A(x)$ as a hesitancy degree of x to A. From (5), it is evident that

$$0 \le \pi_A(x) \le 1, \quad \forall x \in E. \tag{6}$$

The illustration of these degrees has been exhibited in Figure 1.

Therefore, to describe an intuitionistic fuzzy set completely, we need at least two functions from the triplet [27, 28]: (1) membership function, (2) nonmembership function, and (3) hesitation degree.

2.2. FCE Method. Let us recall the definition of the cross-entropy by Kullback [29]. Let p and q be two probability distributions of the discrete random variable X. Then, the cross-entropy measure of p from q can be obtained as

$$I(p,q) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)}, \tag{7}$$

which measures the amount of discrimination of p from q. Lin [30] pointed out that (7) is undefined if $q(x) = 0$ and $p(x) = 0$ for any $x \in E$. To overcome this drawback, He proposed a modified cross-entropy measure described

$$K(p,q) = \sum_{x \in X} p(x) \ln \frac{2p(x)}{p(x) + q(x)}, \tag{8}$$

which is well defined and independent of the values of $p(x)$ and $q(x)$, $x \in E$.

Based on (8), Shang and Jiang [31] defined the FCE between two sets $\widetilde{A}, \widetilde{B} \in U(X)$, where hereinafter X denotes the finite universe of discourse. Thus, they provided the definition of FCE approach as the following.

Definition 1 (see [31]). Let $\widetilde{A}$ and $\widetilde{B}$ be two FSs defined on X. Then,

$$E_{\mathrm{FS}}\left(\widetilde{A},\widetilde{B}\right) = K\left(\mu_{\widetilde{A}},\mu_{\widetilde{B}}\right) + K\left(1-\mu_{\widetilde{A}},1-\mu_{\widetilde{B}}\right) \tag{9}$$

is called fuzzy cross-entropy, where n is the cardinality of the finite universe X. The equation (9) is the degree of discrimination of A from B. However, $E_{\mathrm{FS}}(\widetilde{A},\widetilde{B})$ is not symmetric with respect to its arguments. Thus, Shang and Jiang [31] also proposed a symmetric discrimination information measure based on $E_{\mathrm{FS}}(\widetilde{A},\widetilde{B})$, given by

$$D_{\mathrm{FS}}\left(\widetilde{A},\widetilde{B}\right) = E_{\mathrm{FS}}\left(\widetilde{A},\widetilde{B}\right) + E_{\mathrm{FS}}\left(\widetilde{B},\widetilde{A}\right). \tag{10}$$

Moreover, they showed that $D_{\mathrm{FS}}(\widetilde{A},\widetilde{B}) \geq 0$ and $D_{\mathrm{FS}}(\widetilde{\mathrm{A}},\widetilde{\mathrm{B}}) = 0$ if and only if $\widetilde{\mathrm{A}} = \widetilde{\mathrm{B}}$.

2.3. IFCE Method. Let us consider two sets $A, B \in U(X)$. In order to derive a cross-entropy measure for IFSs, we have to exploit the information carried by both the membership and the nonmembership function. In an analogous manner to Bhandari and Pal [32], we consider the following quantity:

$$I^{\mu}(A,B;x_i) = \ln\frac{\mu_A(x_i)}{\mu_B(x_i)}, \tag{11}$$

as the amount of information for discrimination of $\mu_A(x_i)$ from $\mu_B(x_i)$. Therefore, the expected information for discrimination of A against B, based solely on the membership function, is given by

$$I^{\mu}(A,B) = \sum_{i=1}^{n}\mu_A(x_i)\ln\frac{\mu_A(x_i)}{\mu_B(x_i)}. \tag{12}$$

Similarly, considering the nonmembership function that we have,

$$I^{\nu}(A,B) = \sum_{i=1}^{n}\nu_A(x_i)\ln\frac{\nu_A(x_i)}{\nu_B(x_i)}. \tag{13}$$

Hence, the information for discrimination in favor of A against B is obtained as the sum of the quantities $I^{\mu}(A,B)$ and $I^{\nu}(A,B)$, that is,

$$I''_{\mathrm{IFS}}(A,B) = I^{\mu}(A,B) + I^{\nu}(A,B). \tag{14}$$

In order to overcome the drawback of (14) being undefined if $\mu_B(x_i) = 0$ or $\nu_B(x_i) = 0$ for any $x_i \in X$, a modified version of (14) is introduced based on (8) and is given by

$$I_{\mathrm{IFS}}(A,B) = K(\mu_A,\mu_B) + K(\nu_A,\nu_B). \tag{15}$$

According to Shannon's inequality, Lin [30] pointed out that one can easily prove that $I_{\mathrm{IFS}}(A,B) \geq 0$ and $I_{\mathrm{IFS}}(A,B) = 0$ if and only if $A = B$. A similar measure for FSs, denoted by $\mathrm{I}_{\mathrm{FS}}(\widetilde{\mathrm{A}},\widetilde{\mathrm{B}})$, was defined by Bhandari and Pal [32]. For FSs in general $I_{\mathrm{FS}}(\widetilde{A},\widetilde{B}) \neq I_{\mathrm{FS}}(\widetilde{A}^c,\widetilde{B}^c)$. However, for IFSs, the following equality holds:

$$I_{\mathrm{IFS}}(A,B) = I_{\mathrm{IFS}}(A^c,B^c), \tag{16}$$

where A^c and B^c are the complementary sets of A and B, respectively. Therefore, in analogy with (9), we proceed to the following definition of the intuitionistic fuzzy cross-entropy.

Definition 2. For two sets $A, B \in U(X)$, $I_{\mathrm{IFS}}(A,B)$ is the intuitionistic fuzzy cross-entropy between A and B. $I_{\mathrm{IFS}}(A,B)$ can also be called discrimination information for IFSs.

However, one can observe that $I_{\mathrm{IFS}}(A,B)$ is not symmetric with respect to its arguments. Therefore, a symmetric measure is defined as follows.

Definition 3. For two sets $A, B \in U(X)$,

$$D_{\mathrm{IFS}}(A,B) = I_{\mathrm{IFS}}(A,B) + I_{\mathrm{IFS}}(B,A) \tag{17}$$

is called a symmetric discrimination information measure for IFSs.

It can easily be verified that $D_{\mathrm{IFS}}(A,B) \geq 0$ and $D_{\mathrm{IFS}}(A,B) = 0$ if and only if $A = B$. Moreover, $I_{\mathrm{IFS}}(A,B)$ and $D_{\mathrm{IFS}}(A,B)$ degenerate to their fuzzy counterparts when A and B are FSs.

3. An Improved IFCE Approach

It is easy to see that cross-entropy proposed by Vlachos and Sergiadis [23] only involve the first two parameters, that is, the membership degree and nonmembership degree of IFSs. Furthermore, it does not take the third parameter (hesitation) into account in the above derivation and discrimination measure of IFSs.

Furthermore, Szmidt and Kacprzyk [33] have showed that omitting one of the three parameters may lead to incorrect results, and thus all the three parameters should be taken into account when calculating association or distance between two IFSs.

Motivated by the idea of Szmidt and Kacprzyk [33], we improve the feasibility of original IFCE by adding the third parameter (hesitation degree) consideration. Therefore, a modified formulation is improved to reach better discrimination measure of IFCE approach.

Following (12) and (13), we may keep considering hesitation function for discrimination measure similarly. Thus, we can obtain

$$I^{\pi}(A,B) = \sum_{i=1}^{n}\pi_A(x_i)\ln\frac{\pi_A(x_i)}{\pi_B(x_i)}. \tag{18}$$

Next, the information for discrimination in favor of A against B can be calculated as the sum of the quantities $I^{\mu}(A,B)$, $I^{\nu}(A,B)$, and $I^{\pi}(A,B)$. The equation (14) can be revised as

$$I'_{\mathrm{IFS}}(A,B) = I^{\mu}(A,B) + I^{\nu}(A,B) + I^{\pi}(A,B). \tag{19}$$

To avoid the drawback of (19) being undefined if $\mu_B(x_i) = 0$, $\nu_B(x_i) = 0$ or $\pi_B(x_i) = 0$ for any $x_i \in X$, an improved version of (19) is introduced based on (15) and is given as

$$I_{\mathrm{IFS}}(A,B) = K(\mu_A,\mu_B) + K(\nu_A,\nu_B) + K(\pi_A,\pi_B). \tag{20}$$

Table 1: Symptoms characteristic for the diagnoses considered.

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Temperature	(0.4, 0.0, 0.6)	(0.7, 0.0, 0.3)	(0.3, 0.3, 0.4)	(0.1, 0.7, 0.2)	(0.1, 0.8, 0.1)
Headache	(0.3, 0.5, 0.2)	(0.2, 0.6, 0.2)	(0.6, 0.1, 0.3)	(0.2, 0.4, 0.4)	(0.0, 0.8, 0.2)
Stomach pain	(0.1, 0.7, 0.2)	(0.0, 0.9, 0.1)	(0.2, 0.7, 0.1)	(0.8, 0.0, 0.2)	(0.2, 0.8, 0.0)
Cough	(0.4, 0.3, 0.3)	(0.7, 0.0, 0.3)	(0.2, 0.6, 0.2)	(0.2, 0.7, 0.1)	(0.2, 0.8, 0.0)
Chest pain	(0.1, 0.7, 0.2)	(0.1, 0.8, 0.1)	(0.1, 0.9, 0.0)	(0.2, 0.7, 0.1)	(0.8, 0.1, 0.1)

Table 2: Symptoms characteristic for the patients considered.

	Temperature	Headache	Stomach pain	Cough	Chest pain
Al	(0.8, 0.1, 0.1)	(0.6, 0.1, 0.3)	(0.2, 0.8, 0.0)	(0.6, 0.1, 0.3)	(0.1, 0.6, 0.3)
Bob	(0.0, 0.8, 0.2)	(0.4, 0.4, 0.2)	(0.6, 0.1, 0.3)	(0.1, 0.7, 0.2)	(0.1, 0.8, 0.1)
Joe	(0.8, 0.1, 0.1)	(0.8, 0.1, 0.1)	(0.0, 0.6, 0.4)	(0.2, 0.7, 0.1)	(0.0, 0.5, 0.5)
Ted	(0.6, 0.1, 0.3)	(0.5, 0.4, 0.1)	(0.3, 0.4, 0.3)	(0.7, 0.2, 0.1)	(0.3, 0.4, 0.3)

One can also prove that $I_{\mathrm{IFS}}(A,B) \geq 0$ and $I_{\mathrm{IFS}}(A,B) = 0$ if and only if $A = B$, according to Shannon's inequality. A similar measure for IFSs, the following equality also holds:

$$I_{\mathrm{IFS}}(A,B) = I_{\mathrm{IFS}}(A^c, B^c), \tag{21}$$

where A^c and B^c are the complementary sets of A and B, respectively. Therefore, in analogy with (20), we can also proceed to the symmetric discrimination definition of the proposed intuitionistic fuzzy cross-entropy

$$D_{\mathrm{IFS}}(A,B) = I_{\mathrm{IFS}}(A,B) + I_{\mathrm{IFS}}(B,A). \tag{22}$$

4. Medical Diagnosis Example and Discussions

4.1. Illustrative Example for Medical Diagnosis. We will illustrate an application using the improved IFCE approach for medical pattern recognition problem.

Let us consider the same example in De et al. [1] and Szmidt and Kacprzyk [34, 35]. They consist of a set of patients P = Al, Bob, Joe, Ted, a set of diagnoses D = viral fever, malaria, typhoid, stomach problem, chest pain, and a set of symptoms S = temperature, headache, stomach pain, cough, chest pain. Table 1 presents the characteristic symptoms for the diagnoses considered. The symptoms for each patient are given in Table 2. Each element of the tables is given in the form of a pair of numbers corresponding to the membership, nonmembership, and hesitation values, respectively, for example, the temperature for viral fever is described by $(\mu, \nu, \pi) = (0.4, 0, 0.6)$ in Table 1.

In order to find a proper diagnosis, we calculate for each patient $p_i \in P$, where $i \in \{1, \ldots, 4\}$, the symmetric discrimination information measure for IFSs $D_{\mathrm{IFS}}(s(p_i), d_k)$ between patient symptoms and the set of symptoms that are characteristic for each diagnosis $d_k \in D$, with $k \in \{1, \ldots, 5\}$. Based on (20) and (22), we assign to the ith patient the diagnosis whose symptoms have the lowest symmetric discrimination information measure from patient's symptoms. The diagnosed results for the considered patients are given in Table 3. We assign to the ith patient the diagnosis whose symptoms have the lowest symmetric discrimination information measure from patient's symptoms. The diagnosed results for the considered patients are given in Table 3.

According to Table 3, we know that Al, Joe, and Ted maybe suffer viral fever whereas Bob faces a stomach pain problem.

4.2. Discussions. Through this study, some issues can be raised and depicted as follows.

(1) From the practical case of medical diagnosis study, the proposed method can provide a useful way to help doctors perform preliminary diagnosis. The proposed method differs from previous methods for medical diagnosis decision making due to the fact that the proposed method considers the degree of hesitation compared to other existing approaches. In the future, the proposed method may be a merit for the preliminary diagnosed models to solve the medical diagnosis problem using IFSs.

(2) Sometimes, the difference between two adjacent ranking scores is very close. Observing Table 3, the suffered outcomes of AL, the scores of viral fever and malaria are close. Joe suffered from viral fever and typhoid, of which the scores are close. If it appears in the first and second rank, doctors require making an advanced diagnosis with their expert knowledge to evaluate the values of two IFSs because the doctor process allows little deficiency caused by the practical case study among uncertain environment and usage to achieve the optimal judgment. Therefore, the purpose of advanced medical diagnosis is to reduce the risk which may occur in medical diagnosis.

5. A Computer-Based Interface

As more and more decisions in real organizational settings are made, applying IFSs into medical diagnosis analysis to deal with imprecision, uncertainty, and fuzziness in decision making may become a popular research topic in

Table 3: The diagnosed results by the improved IFCE approach.

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Al	**0.7873**	0.7885	0.8902	2.1573	2.5384
Bob	1.6195	2.6886	0.9896	**0.2886**	1.8327
Joe	**1.0441**	1.5938	1.0668	2.2708	2.8497
Ted	**0.4997**	0.8992	0.9952	1.4273	1.9536

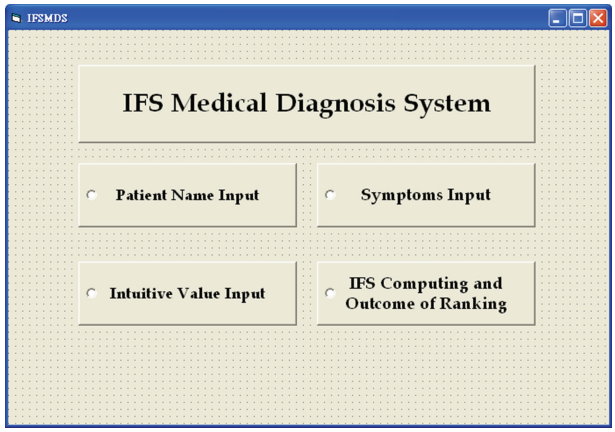


Figure 2: The functional interface of IFSMDS.

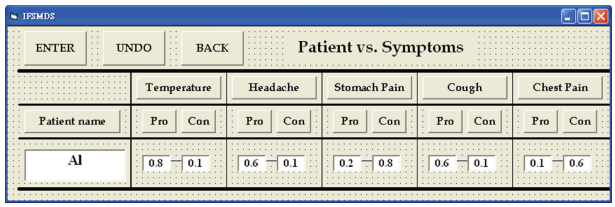


Figure 3: Input intuitionistic diagnosed value on each symptom for patient.

the current uncertain environment. The application of IFSs in supporting doctors can provide a useful way to help the decision analyzer make his/her decisions efficiently.

In this paper, in order to make computing and ranking the results much easier and to increase the recruiting productivity, we have developed an information system called intuitionistic fuzzy sets medical diagnosis system (IFSMDS) as shown in Figure 2. This prototype system is developed by Visual Basic 6 and ACCESS on a N-tier client server architecture. In IFCEDSS, doctors need to key in the patient's name and symptoms data, respectively. The intuitive values of each patient on each relation between symptoms and diseases have showed, as illustrated in Figure 3. The system can calculate the assessment value of each patient on each symptom. The diagnosed results are shown as Figure 4. If ranking score is the lowest, then the patient is more likely to suffer from the disease.

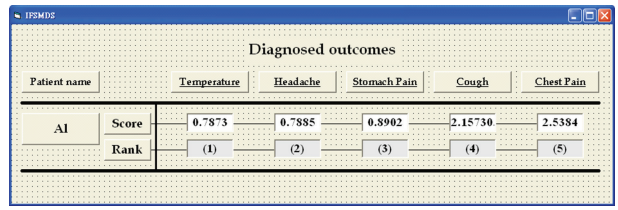


Figure 4: The outcomes and ranking of medical diagnosis.

6. Concluding and Remarks

In this paper, an improved IFCE approach is presented based on IFSs. The idea of improvement is to add the hesitation degree and reveal an intuitive and mathematical connection between the notions of entropy for IFSs in terms of fuzziness and intuitionism based on this entropy measure. Illustrative examples have demonstrated the usefulness of the proposed discrimination information measure for medical diagnosis. In addition, in order to prevent misjudgment, we have suggested that when the diagnosed results between the diseases are very close, an advanced diagnosis is necessary to avoid some error risks of the medical diagnosis.

Acknowledgments

The author would like to express appreciation to the anonymous reviewers for their very helpful comments on improving this paper. This research is partially supported by the National Science Council of the Republic of China, Taiwan with Grant no. NSC 100-2410-H-606-002-MY2.

References

[1] S. K. De, R. Biswas, and A. R. Roy, "An application of intuitionistic fuzzy sets in medical diagnosis," *Fuzzy Sets and Systems*, vol. 117, no. 2, pp. 209–213, 2001.

[2] E. Straszecka, "Combining uncertainty and imprecision in models of medical diagnosis," *Information Sciences*, vol. 176, no. 20, pp. 3026–3059, 2006.

[3] E. Sanchez, "Medical diagnosis and composite fuzzy relations," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds., North-Holland, Amsterdam, The Netherlands, 1979.

[4] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[5] F. Steimann and K. P. Adlassnig, "Fuzzy Medical Diagnosis," http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.47.9705.

[6] F. F. Yao and J. S. Yao, "Fuzzy decision making for medical diagnosis based on fuzzy number and compositional rule of inference," *Fuzzy Sets and Systems*, vol. 120, no. 3, pp. 351–366, 2001.

[7] J. Palma, J. M. Juarez, M. Campos, and R. Marin, "Fuzzy theory approach for temporal model-based diagnosis: an application to medical domains," *Artificial Intelligence in Medicine*, vol. 38, no. 2, pp. 197–218, 2006.

[8] R. Seising, "From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis," *Artificial Intelligence in Medicine*, vol. 38, no. 3, pp. 237–256, 2006.

[9] A. Quteishat and C. P. Lim, "Application of the fuzzy min-max neural networks to medical diagnosis," *Lecture Notes in Computer Science*, vol. 5179, part 3, pp. 548–555, 2008.

[10] I. Gadaras and L. Mikhailov, "An interpretable fuzzy rule-based classification methodology for medical diagnosis," *Artificial Intelligence in Medicine*, vol. 47, no. 1, pp. 25–41, 2009.

[11] C. M. Own, "Switching between type-2 fuzzy sets and intuitionistic fuzzy sets: an application in medical diagnosis," *Applied Intelligence*, vol. 31, no. 3, pp. 283–291, 2009.

[12] S. Badaloni and M. Falda, "Temporal-based medical diagnoses using a fuzzy temporal reasoning system," *Journal of Intelligent Manufacturing*, vol. 21, no. 1, pp. 145–153, 2010.

[13] S. Lekkas and L. Mikhailov, "Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases," *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 117–126, 2010.

[14] A. Shannon, S. Kim, Y. Kim, J. Sorsich, K. Atanassov, and P. Georgiev, "A possibility for implementation of elements of the intuitionistic fuzzy logic in decision making in medicine," in *Proceedings of the First International Conference on Intuitionistic Fuzzy Sets*, J. Kacprzyk and K. Atanassov, Eds., vol. 3 of *Notes on intuitionistic fuzzy sets*, no. 4, pp. 40–43, 1997.

[15] J. Y. Ahn, Y. H. Kim, and S. K. Kim, "A fuzzy differential diagnosis of headache applying linear regression method and fuzzy classification," *IEICE Transactions on Information and Systems*, vol. E86-D, no. 12, pp. 2790–2793, 2003.

[16] P. R. Innocent and R. I. John, "Computer aided fuzzy medical diagnosis," *Information Sciences*, vol. 162, no. 2, pp. 81–104, 2004.

[17] E. Szmidt and J. Kacprzyk, "A similarity measure for intuitionistic fuzzy sets and its application in supporting medical diagnostic reasoning," in *Proceedings of the Lecture Notes in Computer Science (ICAISC '04)*, L. Rutkowski, J. H. Siekmann, R. Tadeusiewicz, and L. A. Zadeh, Eds., vol. 3070, pp. 388–393, Springer, Berlin, Germany, 2004.

[18] Z. S. Xu, "On correlation measures of intuitionistic fuzzy sets," *Lecture Notes in Computer Science*, vol. 4224, pp. 16–24, 2006.

[19] L. Todorova, K. Atanassov, S. Hadjitodorov, and P. Vassilev, "On an intuitionistic fuzzy approach for decision making in medicine: part 1," *International Journal Bioautomation*, vol. 6, pp. 92–101, 2007.

[20] L. Todorova, K. Atanassov, S. Hadjitodorov, and P. Vassilev, "On an intuitionistic fuzzy approach for decision making in medicine: part 2," *International Journal Bioautomation*, vol. 7, pp. 64–69, 2007.

[21] V. Khatibi and G. A. Montazer, "Intuitionistic fuzzy set vs. fuzzy set application in medical pattern recognition," *Artificial Intelligence in Medicine*, vol. 47, no. 1, pp. 43–52, 2009.

[22] J. Ye, "Cosine similarity measures for intuitionistic fuzzy sets and their applications," *Mathematical and Computer Modelling*, vol. 53, no. 1-2, pp. 91–97, 2011.

[23] I. K. Vlachos and G. D. Sergiadis, "Intuitionistic fuzzy information—applications to pattern recognition," *Pattern Recognition Letters*, vol. 28, no. 2, pp. 197–206, 2007.

[24] Q. S. Zhang and S. Y. Jiang, "A note on information entropy measures for vague sets and its applications," *Information Sciences*, vol. 178, no. 21, pp. 4184–4191, 2008.

[25] K. Atanassov, "Intuitionistic fuzzy sets," VII ITKR Session, Sofia (Deposed in Central Library Of The Bulgarian Academy Of Sciences, 1697/84), 1983, (in Bulgarian).

[26] P. Burillo and H. Bustince, "Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 78, no. 3, pp. 305–316, 1996.

[27] K. T. Atanassov, "More on intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 33, no. 1, pp. 37–45, 1989.

[28] K. Atanassov, *Intuitionistic Fuzzy Sets: Theory and Applications*, Physica, Heidelberg, Germany, 1999.

[29] S. Kullback, *Information Theory and Statistics*, Dover, New York, NY, USA, 1968.

[30] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[31] X. G. Shang and W. S. Jiang, "A note on fuzzy information measures," *Pattern Recognition Letters*, vol. 18, no. 5, pp. 425–432, 1997.

[32] D. Bhandari and N. R. Pal, "Some new information measures for fuzzy sets," *Information Sciences*, vol. 67, no. 3, pp. 209–228, 1993.

[33] E. Szmidt and J. Kacprzyk, "Distances between intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 114, no. 3, pp. 505–518, 2000.

[34] E. Szmidt and J. Kacprzyk, "Intuitionistic fuzzy sets in intelligent data analysis for medical diagnosis," in *Proceedings of the Computational Science (ICCS '01)*, V. N. Alexandrov, J. J. Dongarra, B. A. Juliano, R. S. Renner, and C. J. K. Tan, Eds., vol. 2074, pp. 263–271, Springer, Berlin, Germany, 2001.

[35] E. Szmidt and J. Kacprzyk, "Intuitionistic fuzzy sets in some medical applications," in *Proceedings of the 7th Fuzzy Days Dortmund*, B. Reusch, Ed., vol. 2206 of *Computational intelligence: theory and applications*, pp. 148–151, Springer, Berlin, Germany, 2001.