An Intelligent Human-Computer Interaction System for Decision Support

Chee Siong Teh and Chee Peng Lim

***Abstract*—**This paper proposes a novel architecture for developing decision support systems. Unlike conventional decision support systems, the proposed architecture endeavors to reveal the decision-making process such that humans' subjectivity can be incorporated into a computerized system and, at the same time, to preserve the capability of the computerized system in processing information objectively. A number of techniques used in developing the decision support system are elaborated to make the decision-marking process transparent. These include procedures for high dimensional data visualization, pattern classification, prediction, and evolutionary computational search. An artificial data set is first employed to compare the proposed approach with other methods. A simulated handwritten data set and a real data set on liver disease diagnosis are then employed to evaluate the efficacy of the proposed approach. The results are analyzed and discussed. The potentials of the proposed architecture as a useful decision support system are demonstrated.

***Keywords*—**Interactive evolutionary computation, multivariate data projection, pattern classification, topographic map

I. INTRODUCTION AND MOTIVATION

COMPUTERIZED decision support systems, an inclusive term of many types of information systems that support decision-making [1], evolve from two main research fields: interactive computer systems and theoretical organizational decision-making [2]. Both classification and prediction techniques are important elements in most decision support systems. However, to date, classification and prediction often do not provide any ability that reveals the underlying structure and information of the data set. In general, the process of classification and prediction is conducted autonomously without much involvement from domain users. Such an approach does not take advantage of the remarkable perceptive and associative abilities of human observers to perceive clusters and correlations, which usually can lead to better understanding of structures in data (see Figure 1). This concealed approach is somehow disapproved with the Gestalt theory, in which an important aspect of the theory is that humans gain most understanding by comprehending the meaning of the underlying progression of parts [3].

Fig. 1 Concealed design in the classification process

According to [2], many researches on decision support systems focus on well-defined or highly structured problems, in which the systems are given a clearly specified problem statement that explains the 'goal state', the 'current state', and the 'permissible operations' they may use to get to the goal state from the current state. In addition, [4][5] pointed out the difficulties to understand how such systems produce recommendation based on a given set of input. To the end-user, the reasoning process of these systems is a "black box". Indeed, many current decision support systems [6]-[9] do not acknowledge the role and importance of human behavior in the decision making process. Decision making in the real world, however, often deals with unstructured or ill-defined problems in which decision makers have only a vague idea about the goal state and the current state as well as that the permissible operations may be undefined. In such cases, the performance of a decision support system that cannot take into account the undefined operations will be compromised.

More than forty years ago, Simon [10] introduced the notion of 'bounded rationality' of human decision-making abilities. This notion argued that people are limited in the amount of information they can process and the methods they use to integrate information. Simon also believed that if these limitations could be transcended, then the decision-making effectiveness would be enhanced. The integration of human and the computer was foreseen to be the best method for achieving these new heights for decision makers.

Computers are known for their amazing power to organize, store, process, and retrieve vast volume of information. One of the major contributions of computerized decision support systems is their ability to deal with large amount of information. The way of automated information processing of such systems also means all procedures to process information are predefined. As the computer cannot be told to do what humans cannot define, Phelps et al [2] pointed the deficiencies of developing decision support systems that do not support decision-making that operates within bounded rationality. In

Manuscript received December 25, 2005. This work was sponsored by University Malaysia Sarawak, Malaysia.

C. S. Teh is a lecturer at Faculty of Cognitive Sciences and Human Development, University Malaysia Sarawak, 94300, Kota Samarahan, Malaysia.(e-mail: csteh@fcs.unimas.my).

C. P. Lim is an associate professor at School of Electrical and Electronic Engineering, University of Science Malaysia, Malaysia. (e-mail: cplim@eng.usm.my).

other words, such systems take the rational approach that consider a decision in a logical, sequential, and in-depth analysis of alternatives on the basis of information without considering emotion or social pressure. Although these systems overcome humans' limited ability to process information of the alternatives, it is crucial to acknowledge that not all alternatives led themselves to quantification that allows easy comparison [11].

Instead of producing a system that automates decision-making, this work aims to utilize a more pragmatic approach that allows participation of users in the decision making process. With such integration, the deficiencies of the system can be mitigated by the subjectivity (qualitative) evaluation of humans and, at the same time, the deficiencies of humans can be mitigated by the objectivity (quantitative) evaluation of the system. To enable a smooth integration of human into the decision support system, this research places emphasis on the human cognitive or thinking processes that are fundamental to decision-making, comprehends the deficiencies or errors that are associated with these processes, and devises methods of compensating for these deficiencies in processing.

The organization of the paper is as follows. The following section elaborates a novel architecture for integrating human and the computer into a co-operative platform. Various methods used to incorporate support human cognitive processes in decision-making are presented. An artificial data set is used to illustrate clearly the proposed approach. A pen-based handwritten digits recognition dataset and a real medical application on liver diseases diagnosis are included to demonstrate the efficacy of the system.

II. The Proposed System Architecture

Figure 2 shows the system architecture of the proposed interactive decision support system. In contrast to most decision support systems (as shown in Figure 1) which place sole reliance on the computer system to perform the decision-making process, the proposed system architecture employs a number of approaches, such as data visualization, classification, prediction, and evolutionary computational search, to uncover the decision-making process in an effort to allow humans to utilize their perceptive and associative abilities in the process. Revealing the underlying data structure and providing opportunities for human intervention during the decision-making process enables the integration of humans' subjective intelligence into the system and thus, reduces the limitation of conventional decision support systems. The following sub-sections describe the techniques used in the proposed approach.

A. Multi-dimensional data scaling and visualization

Advanced methods of pattern recognition, data analysis, and visualization are becoming crucial to uncover important structures and interesting correlations in data in order to generate useful, meaningful, and even unpredictable information from flood of data. Consequently, a large number of artificial neural networks and machine learning algorithms, particularly for feature extraction and data projection, have been proposed [12]-[16].

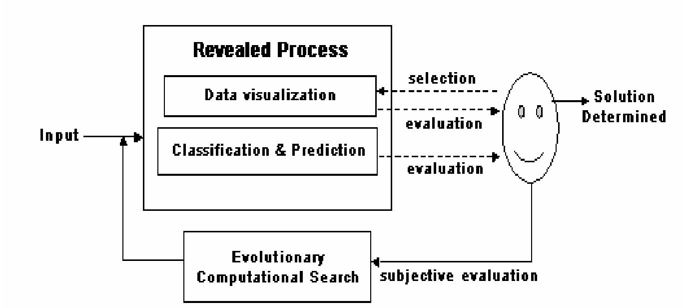

Fig. 2 Architecture of the proposed interactive decision support system

Feature extraction and projection of multivariate data enables the visualization of high dimensional data in order to better understand the underlying structure, explore the intrinsic dimensionality, and analyze the clustering tendency of multivariate data [17]. In addition, visualization has played an important part as it takes advantage of human's perceptive and associative abilities to perceive clusters and correlations, which usually lead to better understanding of the underlying data structure. This work utilizes the kernel-based Maximum Entropy learning Rule (kMER) [18] as a multivariate data projection and visualization technique for mapping high dimensional data vectors to a lower dimensional space.

The kMER model is an unsupervised competitive learning scheme that produces an equiprobabilistic topology-preserving mapping from the n-dimensional input space to a discrete lattice, with N formal neurons in a fixed topological 2D space. The lattice topology is rectangular and its dimensionality is the same as that of the input space in which the lattice is based upon. kMER is able to extract the maximum amount of information available about the input distribution during its learning process. It has been successfully implemented in modeling sensory coding, non-parametric Blind Source Separation, and non-parametric density estimation, e.g. for cluster or classification purposes [19].

Consider a lattice A with a regular and fixed topology, with dimension d_A in the d-dimensional space $V \subseteq \Re^d$. Let A be a lattice with N neurons, labeled i = 1,2,…,N, and corresponding kernel $K(\mathbf{v}-\mathbf{w}_i,\sigma_i)$, $\mathbf{v}\in V$, with a radially-symmetric receptive field (RF) around its center, for example Gaussian, which has center $\mathbf{w}_i$ and radius σ_i. The cross-section τ_i of the kernel, as shown in Figure 3, defines the RF region S_i with radius σ_i in the V-space.

When the current input $\mathbf{v}$ falls in S_i, then supra-threshold activation occurs and the threshold is elevated, else sub-threshold activation occurs and the threshold is lowered. These events are formalized by associating S_i with a code membership function:

$$\xi_i(\mathbf{v}) = \begin{cases} 1, & if \quad \mathbf{v} \in S_i \\ 0, & if \quad \mathbf{v} \notin S_i \end{cases} \tag{1}$$

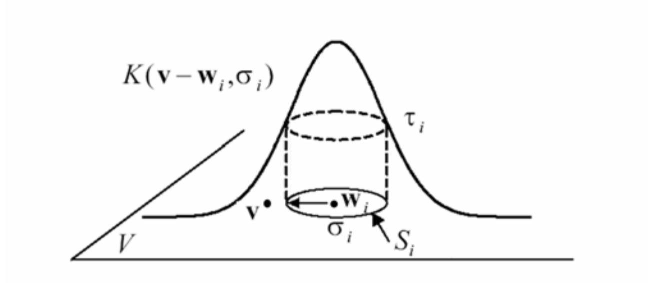

Fig. 3 A receptive field (RF) kernel $K(\mathbf{v}-\mathbf{w}_i,\sigma_i)$ and a RF region S (adapted from [19])

Since the RF definition is not lattice-based, a different type of competition at the learning stage is used. A fuzzy code membership function, $\Xi_i(\mathbf{v})$, is introduced to update the RF center $\mathbf{w}_i$ proportion to the function and in the general direction of **v,** as follows:

$$\Xi_i(\mathbf{v}) = \frac{\xi_i(\mathbf{v})}{\sum_{k\in A}\xi_k(\mathbf{v})}, \forall i \in A \tag{2}$$

The rationale is to ensure that $0 \le \Xi_i(\mathbf{v}) \le 1$ and $\sum_i \Xi_i(\mathbf{v}) = 1$.

Depending on the activation state of the neurons, the kernel centers $\mathbf{w}_i$ and radii σ_i are adapted according to the following two learning rules. In the "batch" mode learning, given a training set $M = \{\mathbf{v}^\mu\}$ of M input samples, the RF centers are updated as follows.

$$\Delta w_i = \eta \sum_{\mathbf{v}^\mu \in M} \sum_{j\in A} \Lambda(i,j,\sigma_\Lambda(t))\Xi_i(\mathbf{v}^\mu) Sgn(\mathbf{v}^\mu - \mathbf{w}_i), \quad \forall i \in A \tag{3}$$

where Sgn(.) is a sign function applied component wise for each neighborhood range, $\sigma_\Lambda(t)$; t is the present time step; and η is a learning rate. The Gaussian neighborhood function is used as follow:

$$\Lambda(i,j,\sigma_\Lambda(t)) = \exp\left(\frac{(\mathbf{r}_i - \mathbf{r}_j)^2}{2\sigma_\Lambda(t)^2}\right) \tag{4}$$

where $\mathbf{r}_i$ and $\mathbf{r}_j$ represent the lattice coordinates of neurons i and j, and decrease the range in the following way:

$$\sigma_\Lambda(t) = \sigma_{\Lambda 0}\exp\left(-2\sigma_{\Lambda 0}\frac{t}{t_{\max}}\right) \tag{5}$$

with t the maximum number of time step, and $\sigma_{\Lambda 0}$ the range spanned by the neighborhood function at $t = 0$.

The kernel radii σ_i are updated by using equation (6) so that the activation probability for unit i converges to $P(\xi_i(\mathbf{v}) \ne 0) = \rho / N, \forall i$, with ρ a scale factor, and $\rho_r \doteq \rho N/(N-\rho)$.

$$\Delta\sigma_i = \eta \sum_{\mathbf{v}^\mu \in M}\left(\frac{\rho_r}{N}(1-\xi_i(\mathbf{v}^\mu)) - \xi_i(\mathbf{v}^\mu)\right), \forall i \in A \tag{6}$$

The mathematical details and proof of convergence can be found in [18].

B. Statistical Pattern Classification and Prediction

This section describes a neural network model that implements a classical non-parametric density estimation procedure. In addition, it elaborates the integration of kMER into the network to form a hybrid system. The performance of the network is then compared with Bayes' optimal results and other classification algorithms.

1) The PNN Model: The Probabilistic Neural Network (PNN) is a simple but powerful non-parametric classifier, which was proposed by [20] and originated from Parzen's [21] kernel-based probability density function (pdf). The training process of the network is one-pass, without iteration for weight adaptation. The PNN algorithm is as follows.

For a given data set **v,** the Parzen density function is

$$p(\mathbf{v}) = \frac{1}{N\sigma^d}\sum_{n=1}^{N} G\left(\frac{\|\mathbf{v}-\mathbf{v}_n\|}{\sigma}\right) \tag{7}$$

where $\mathbf{v} \in R^d$, G is the kernel function and σ is the smoothing parameter of the kernel function. The kernel function often takes the Gaussian type as follows.

$$G(\mathbf{v}) = \frac{1}{(2\pi)^{d/2}}\exp\left(-\frac{\mathbf{v}^2}{2}\right) \tag{8}$$

The Bayesian posterior probabilities are then computed,

$$P(C_{i*} \mid \mathbf{v}) = \frac{p(\mathbf{v} \mid C_i)P(C_i)}{\sum_{j=1}^{k} p(\mathbf{v} \mid C_j)P(C_j)} \tag{9}$$

with C_i the class labels, k the number of classes $P(C_i)$ the a priori probability of class C_i, $p(\mathbf{v} \mid C_i)$ the conditional density of class C_i and **v** a given sample. Once the neurons are labeled, we can determine the class-conditional densities and classify a given sample into class $P(C_{i*})$ when

$$p(\mathbf{v} \mid C_{i*})P(C_{i*}) > p(\mathbf{v} \mid C_j)P(C_j), \quad \forall j \ne i* \tag{10}$$

or when its posterior probability satisfies:

$$P(C_{i*} \mid \mathbf{v}) > P(C_j \mid \mathbf{v}), \quad \forall j \ne i* \tag{11}$$

2) The Hybrid kMER-PNN Model: The PNN algorithm described above uses all the samples in the training set to estimate the pdf and to perform classification. If the data samples in the training set are corrupted by noise, the classification performance may be affected. In the proposed hybrid kMER-PNN model, the prototype vectors from each class of the trained kMER map, instead of the original training samples (often with a large size), are used to estimate the pdf. The advantages of the integration are two-fold. First, kMER is used as the underlying clustering algorithm to reduce the number of pattern nodes required in the PNN; second, the PNN is used as the probability estimation algorithm to provide probabilistic prediction from kMER. Suppose that the trained

kMER map has N_{C_i} (the number of RF regions of class C_i) nodes with label C_i and the corresponding prototype vectors are $\mathbf{w}_j^{C_i}$, $j = 1,2,3.., N_{C_i}$, the pdf of class C_i is estimated using

$$p(\mathbf{v} \mid C_i) = \frac{1}{N_{C_i} (2\pi\sigma_j^2)^{d/2}} \sum_{j=1}^{N_{C_i}} \exp\left(-\frac{\left\| \mathbf{v} - \mathbf{w}_j^{C_i} \right\|}{2\sigma_j^2} \right), \quad (12)$$

$$\forall j = 1,2,3..., N_{C_i}$$

After obtaining the estimated pdf of each class, the posterior probabilities are determined by using equation (9). Note that $P(C_i) = N_{C_i} / \sum_j N_{C_i}$ since with kMER, each RF region is equiprobabilistic and σ_j are the kernel values located on each neuron weight. The proposed hybrid approach has the following advantages:

a. kMER reduces the number of pattern nodes as required in the PNN
b. kMER provides a good representative of the training samples; it makes the PNN classifier more robust in the presence of noise in data.
c. the selection of the “smoothing parameter” of the original PNN is automated; in contrast to other optimization techniques, which need to be tuned using the trial-and-error [22] or cross-validation methods [23].

Since kMER allocates a kernel at each neuron weight, instead of at each input sample such as the GTM algorithms [24], it can go beyond the Parzen-window technique that utilizes fixed radii kernels. This leads to variable kernel estimation for non-parametric density estimation. The process of the hybrid algorithm is shown in Figure 4.

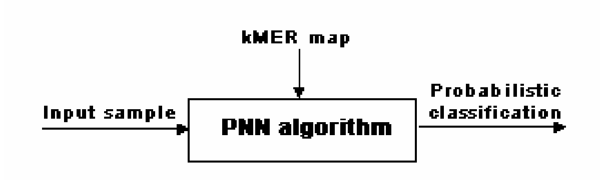

Fig. 4 The proposed kMER-PNN model

C. An Example

The following example aims to demonstrate the capability of kMER to produce topographic maps and its applicability in the proposed hybrid architecture. The study followed the procedure used in Herbin and co-workers [25]. A data set of three equally probable Gaussians with M = 900 and centered at (-0.4,-0.3), (0.4,-0.3) and (0.0,0.3) in the unit square $[1,1]^2$, with the standard deviations all equal to 0.2 was first generated. The distribution of the generated data samples is shown in Figure 5.

In the experiment, we employed kMER with an N = 24x24 planar lattice, and the neighborhood function and random weight and radius initialization procedures as suggested before. All samples M are used to train the map. The RF region and the weight centers obtained for ρ_r = 1 are shown in Figure 6. For comparison, we generated another map by utilizing the standard Self Organizing Map (SOM) using the SOM Toolbox [26][27]. The SOM and kMER projected maps are shown in Figure 7.

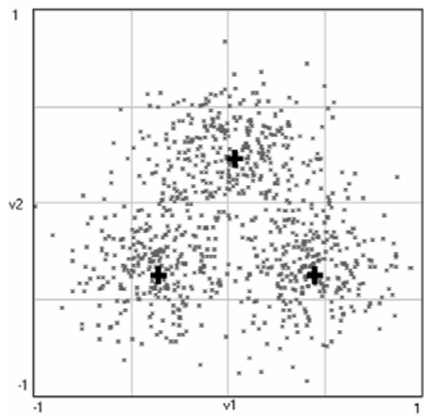

Fig. 5 Scatter plot of three times 300 samples drawn from three Gaussian distribution with “+” as the center of each Gaussian

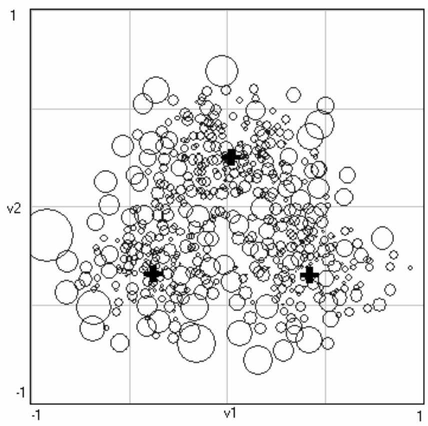

Fig. 6 Prototype vectors and RF regions distribution obtained with kMER for ρ_r = 1 and after 2000 training epochs

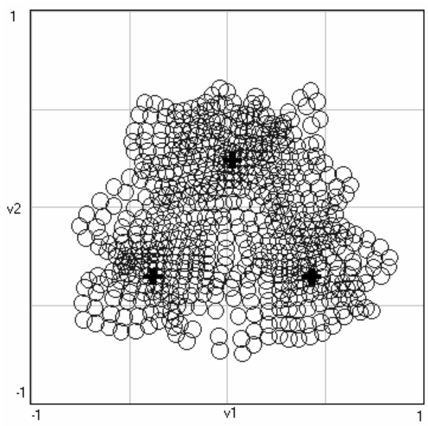

Fig. 7 Prototype vectors distribution obtained with SOM. Since SOM does not provide RF regions, all circles shown are of equal radii (σ = 0.02) and serve only for visualization purpose

1) Visualization of clusters and borders: Humans tend to notice phenomena in visually displayed data precisely [28]. By inspecting Figure 6 and Figure 7, for instance, one can

observe three distinct clusters in the prototype vectors distribution, which are also similar with the original distribution of the sample data as depicted in Figure 5. However, when comparing both Figures 6 and 7, in addition to cluster information, the circles of different radii shown in the distribution obtained with kMER provide additional insight into the input sample density information.

Displaying the best match unit (BMU) of each data sample on the map provided another visualization method. The minimum Euclidean distance (minEuC) between each data and the neurons determines the BMU (winning neuron). For every BMU of each data, the class of the BMU is labeled based on the data label. Figures 8 and 9 show the labels obtained from the training data set and these labels provided another obvious distinction of the various regions or cluster borders on the map. Comparing both figures, it is found that SOM (Figure 9) produces much more "dead" units than the kMER model (Figure 8). These dead units do not sufficiently contribute to the minimization of the overall distortion of the map and, hence, resulting in a less 'optimal' usage of the map resources.

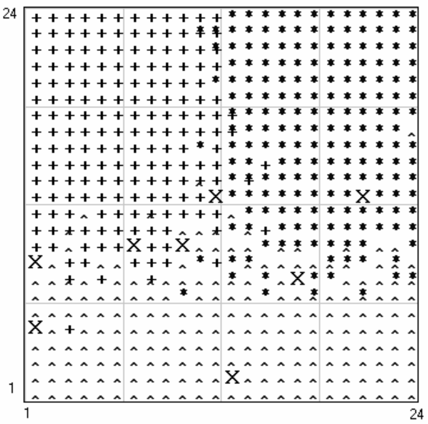

Fig. 8 The neurons trained by kMER algorithm (t = 2000, ρ_r = 1) and labeled with "+", "*" and "^" using minEuC method. "X" indicates dead unit that is inactive/ unlabeled

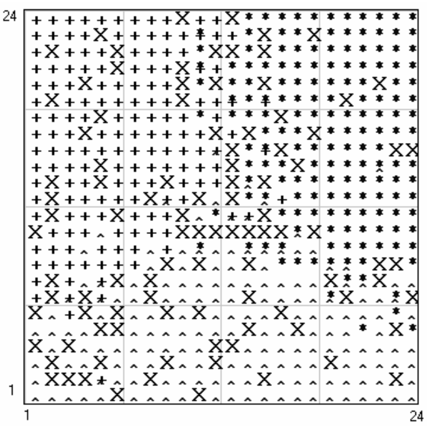

Fig. 9 The neurons trained by SOM algorithm and labeled with "+", "*" and "^" using minEuC method. "X" indicates inactive/unlabeled dead units

2) Classification performance*:* Classification errors occur due to the overlapping of the Gaussian distributions. As for the 900 input samples, the error rate of kMER-PNN is found to be the lowest (3.45% or 31 misclassifications). Herbin's method produced an error rate of 5.89% (53 misclassifications), and for the Bayesian approach, it was 5.43% (49 misclassifications) [25]. The error rate for the hybrid SOM-PNN algorithm is 6.80% (61 misclassifications) with $\sigma = 0.1$.

III. Evolutionary Computational Search

Humans, by and large, do not behave like intuitive statisticians [29], that is, they do not follow the principles of probability theory in judging the likelihood of uncertain events. Instead, we use judgmental heuristics (mental shortcuts) to compute subjective probabilities. Decision makers evaluate the probability of an event according to how similar that event is to its "parent population" and/or according to the degree to which the event reflects the salient features of the process by which it was generated [30]. In other words, we think events are more "likely" when they are similar to our expectations or preferences. Based on this argument, a more meaningful decision support system should therefore be able to provide opportunities for humans to incorporate their subjective judgments in searching alternative solutions through the decision-making process.

The Evolutionary Computation (EC) is a biologically inspired general computational concept that uses population-based searching algorithm and outputs multiple candidates, as system outputs. The Genetic Algorithm (GA) [31] is one of the typical paradigms uses the concept of EC to perform the search and optimization based on multiple searching points or solutions in the problem space. GA usually represents solutions using the chromosomes with bit coding (genotype) and searches for the better solution candidates in the genotype space using GA operators. Figure 10 shows the flow of data representation in the GA process. The possible solutions are expressed in bit code (genotype) and decoded to the values (phenotype) used in finding solutions for the application task. The population of possible solutions is applied to the application task and evaluated the results. These results (fitness values) are then used to select the parent solutions that determine the next searching points. The GA operators (crossover and mutation) are applied to generate the offsprings in the next generation, and this process iterates until the GA search converges to the required searching level.

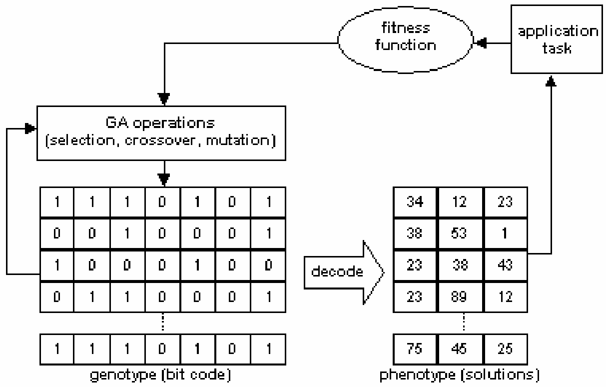

Fig. 10 The flow of data in the GA process (adapted from [32])

In Interactive Evolutionary Computation (IEC) [33], the roles of EC/GA and a human user are integrated; the GA performs the searches while the human user evaluates the result. In other words, human user replaces the fitness function. However, to reduce human psychological and physical fatigue during the searching process, the concept of active user intervention has been incorporated. Visualized IEC [34] is proposed in this case to accelerate convergence when searching for a global optimum by combining the different capabilities of the GA and humans. The GA using its operators (e.g. selection, mutation and crossover) performs sound systematic search in the *n*-dimensional space while human visualizes and performs global search in a mapped 2D space.

In this work, kMER produces a topological map that serves as the visualized 2D search space. This 2D space helps users to estimate a rough global optimum position in the space and uses the position as a new elite individual [34]. In addition, use of hybrid kMER-PNN provides prediction of target class probabilities that can further give clue on the possibility of achieving convergence. This clue is important since in a decision-making process, humans continue to collect additional information until they reduce their subjective uncertainty below some acceptable point [35]. Indeed, the inclusion of active user interaction in EC together with other supports (i.e. visualization, classification, and prediction) has helped reveal the decision-making process, which enables the incorporation of humans' subjective judgments into the decision-making process.

IV. Applications

The following section illustrates the applicability of the proposed architecture in two case studies: (i) a pen-based handwritten digits recognition dataset (to demonstrate the feasibility of the proposed architecture); (ii) a real medical decision support task of diagnosing liver diseases (to demonstrate the applicability of the proposed architecture to real-world problems).

A. Pen-based handwritten digits recognition

The simulated 8 points (16 attributes comprising x and y coordinates of digits) pen-based handwritten digits recognition dataset [36], obtained from the UCI repository of Machine Learning Databases, was employed to evaluate the visualization capability of the kMER algorithm. The algorithm used a 24x24 planar lattice, and the initial weights with 16 attributes were randomly chosen from the uniform distribution of (0, 1]. The learning rate was fixed at $\eta = 0.001$ with $\rho = 1$, and the maximum batch learning time was set to 1000 epochs. Figure 11 shows the topology-preserved maps generated by the algorithms. The class of the BMU was used to label the grid units belonging to digits 0 to 9 (10 classes). It is obvious that kMER is able to provide visualization of the cluster boundaries.

Figure 13 shows an example of the output for 3 generations of IEC searches. The input sample of this example was displayed on the top left-hand side of the figure. The hybrid kMER-PNN provided the probabilistic prediction of the input sample class, which was 94.83% as digit '5' and 3.92% as digit '8'. The results show how visualization and probability prediction were embedded into the system.

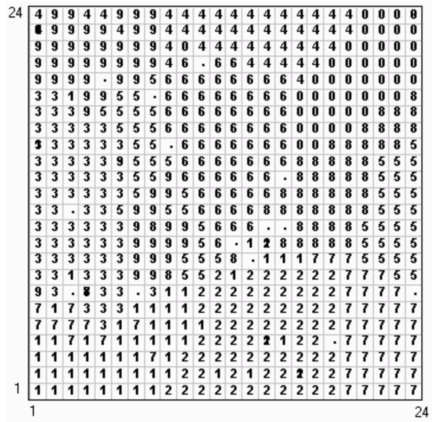

Fig. 11 Visualization of the pen-based handwritten digits data set generated using kMER

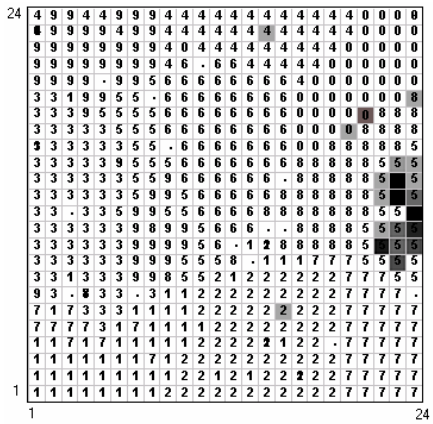

Fig. 12 Topographic map with shaded grids to reveal the interactive EC search process

Eight cases were retrieved from the dataset in the first generation of EC. Based on the probability prediction values, the class with the highest percentage was determined. Three most related cases from this class and five related cases from other classes were retrieved. All these cases were selected based on the shortest Euclidean distance from the input sample. Based on these eight displayed cases, the user could use the vertical scroll bar at the side of each window to subjectively evaluate each case. The higher a bar was adjusted, the higher the fitness value of the particular case.

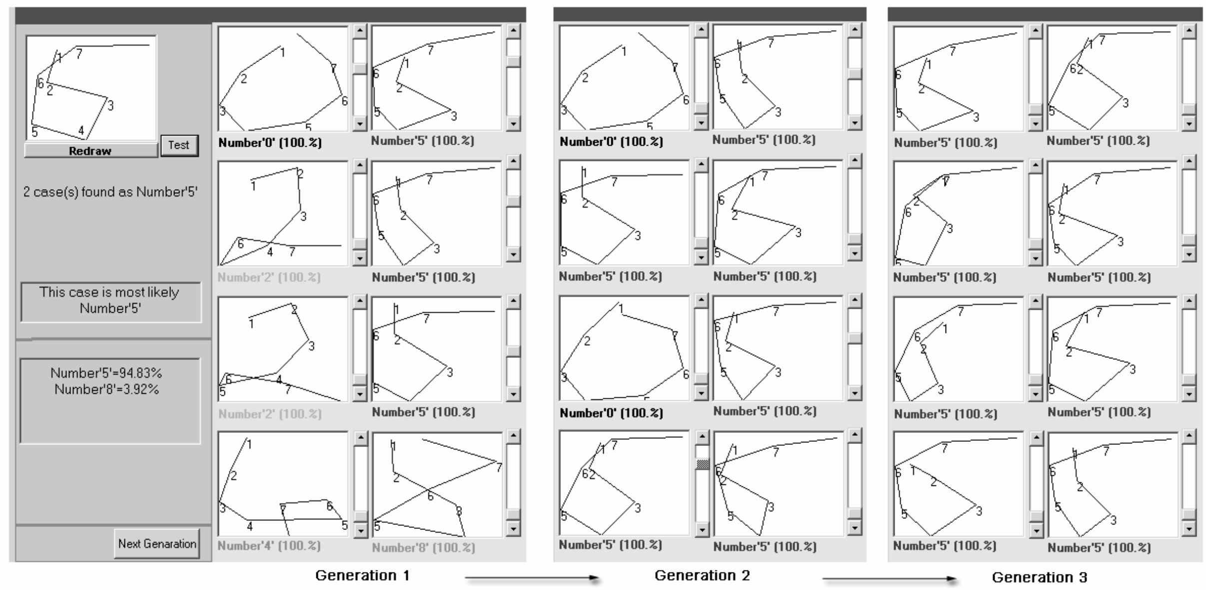

Fig. 13 The output of the IEC searches from generation 1 to generation 3

After providing the fitness values, the system would then display the next generation results based on EC. This generation would retrieve cases that were of even a closer match to the user's judgments. In the above example, output of the third generation could further verify the earlier probabilistic prediction that the digit was '5'. This is because in the first two generations, the user had provided higher preference for cases that showed similarity to digit '5'. However, through the interactive search, similar input sample could also generate different outputs in later generations if the user had provided a higher preference for cases other than digit '5'. Such outputs had revealed the related cases from the dataset to the user.

Visualization generated in Figure 12 had also revealed the interactive EC search process. The shaded grids represented the distribution of cases retrieved from each generation of EC searches. The shading intensity of these grids showed the fitness values, in which the darkest grids represented cases of the highest fitness values. Such visualization enabled the user to perceive their preference throughout the interactive process. Existence of convergence areas, such as the one shown in Figure 12, could most probably increase the certainty of the decision made.

B. Liver Diseases Diagnosis

The liver is a complex organ with many vital roles [37]. The liver stores fuel for the body (produced from sugars), and it is involved in the processing of fats and proteins. Bile produced by the liver is involved in the digestion and absorption of fat in the intestines. The liver also makes proteins that are essential for blood clotting, and it helps remove poisons and toxins from the body. Although the liver is an important organ, liver diseases are one of the commonest causes of morbidity and mortality among hospital patients [38]. These patients suffer from the risk of progressive liver damage and neoplastic transformation. Thus, regular follow-up of these patients, which usually include performing Liver Function Tests (LFTs), is essential.

LFTs are a group of blood tests that measure substances in the blood that reflect whether the liver has been injured and the extent of the injuries [39][37]. LFTs include tests for bilirubin, a breakdown product of hemoglobin, and ammonia, a protein by-product that is normally converted into urea by the liver before being excreted by the kidneys. LFTs also commonly include tests to measure levels of several enzymes, which are special proteins that help the body breaks down and metabolises other substances. Enzymes that are often measured in LFTs include alanine aminotransferase (ALT), aspartate aminotransferase (AST) and alkaline phosphataste (ALP).

The process of identifying the type of liver diseases in patients requires knowledge and experience. Indeed, diagnostic problem solving in a medical domain is a primary example of decision making under uncertainty. Diagnostic problems are often based on more than one item of data because it is highly exceptional for a single symptom or measurement to be pathologically significant to one disease without any associations with others. In the case of diagnosing liver diseases, besides considering historical data, symptoms and other signs through examination, the levels of the range of substances revealed through the LFTs also aid in the differential diagnosis of liver disease and injury, and to help monitor response to treatment. In other words, such diagnosis involves both objective and subjective considerations.

Clinicians and doctors who make such decisions often proceed by assigning, probably unconsciously, different weights to different data items, and then choose the most probably prediction [40]. However, many different outcomes may correspond to identical set of data or, conversely, distinct sets of data may point to the same disease. Moreover, some data items may be noisy or imprecise, making the whole diagnostic process probabilistic in nature. These factors can often result in an overall degradation of diagnostic accuracy by humans. An automated decision support system will potentially improve the objective judgments in such a diagnostic process. However, as both objective and subjective considerations are involved in the diagnosis of liver diseases, an interactive human-computer decision support system is foreseen to be more advantageous because both considerations can be incorporated into the decision-making process.

1) Experiments: The study included a total number of 139 patients of Timberland Medical Centre, Kuching, Sarawak, Malaysia. Information on the LFTs was retrieved from the patients' record. The identification of different types of liver disease categories (either Normal, Hepatitis Carrier, Chronic Hepatitis, or Cirrhosis) for each of the patients was made by the gastroenterologist involved in the patients' care. Seven variables were used, i.e., total bilirubin (T.Bil), total protein (T.Pro), alanine aminotransferase (ALT), albumin (Alb), aspartate aminotransferase (AST), globulin (Glb), and alkaline phosphatase (ALP). Besides, another 400 (100 cases for each category) simulated records were added into the dataset. These simulated records were generated using Gaussian white noises based on the means and variances for each variable of each category.

Figure 14 shows a screenshot of the interface of the decision support system. Basically, the interface consists of three main components. The first component enables a user to input the values of the seven variables obtained from the results of the LFT of a particular patient. Once the input values are submitted, the system provides visualization of the position of the input data (winning neuron) on the 2D map, which is projected using the kMER algorithm. This position provides valuable information on the category of liver disease the patient might suffer from. If the input fell on the border of two different clusters, the patient might be diagnosed as either normal or a hepatitis carrier. The system also provides information on the number of related cases for each category that are within the neuron (4 Normal cases and 1 Hepatitis carrier case in this example).

Use of the hybrid kMER-PNN model also enabled the user to estimate the probability density function from the historical dataset. The probabilistic prediction of the category the patient might fall into was shown in percentage. The high percentage for both Normal (45.9%) and Hepatitis carrier (30.2%) once again proved the overlapping of these two categories. So far, this system has demonstrated the benefits of embedding visualization and probability prediction into the decision-making process.

Figure 15 illustrates another interface of the system that provides visualization for active user interaction and intervention in EC. User intervention in EC search aimed to accelerate convergence, which subsequently would lead to the decision or solution of problem by providing similar or related cases based on humans' judgments. Eight cases were retrieved in the first generation of EC. Based on the probability prediction values, the category with the highest percentage was determined. Five most related cases from this category and three related cases from each of the other categories were retrieved. All these cases were selected based on the shortest Euclidean distance from the patient's data, which was input earlier by the user.

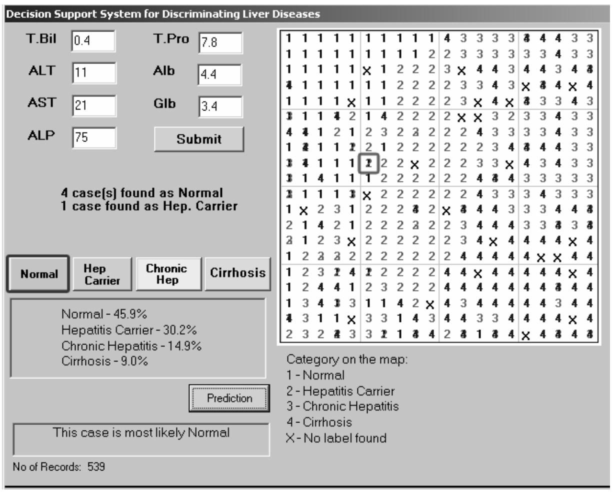

Fig. 14 An interface that provides visualization and probability prediction

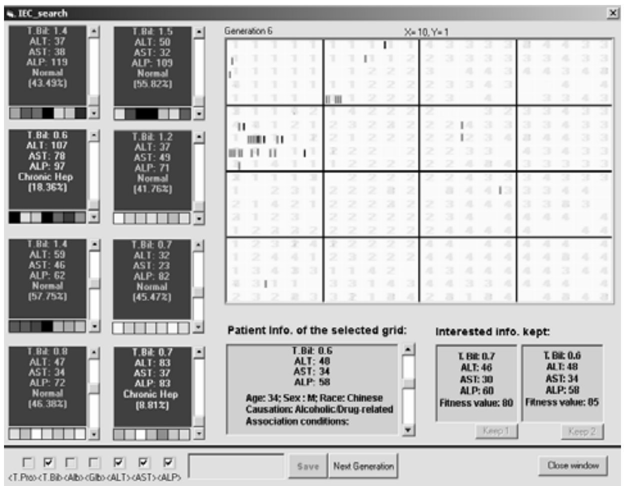

Fig. 15 An interface that provides visualization for active user interaction and intervention in EC

In addition to the information (values of 4 variables, probability prediction) displayed in each window, the user could retrieve other patient's information for each case, which would be useful in assisting his/her decision. Based on this information, the user could then use the vertical scroll bar at the side of each window to subjectively evaluate each case. The higher a bar was adjusted, the higher the fitness value of the particular case. After providing the fitness values, the system would then display the next generation results based on EC. This generation would retrieve cases that are of even closer match to the user's judgments.

The right side of the interface displayed the kMER map. The bars represent the distribution of cases retrieved from the previous and current generation of EC searches, in which these cases were mapped from n-D space to a 2D space. The shading of the bars also showed the fitness values where the darkest bars represented cases of the highest fitness values. Such visualization helped the user to roughly estimate a global optimum position in the 2D space.

The user could also select any other grids in the 2D space, in which he/she would most probably select the one with a high fitness value. The chosen grid would then become a new elite case and the case with the lowest fitness value in the current generation of EC population would be replaced with this chosen case. The EC mated and created an offspring using the parent or current generation population to include the newly added case. Indeed, the opportunity for the user to select case to be incorporated into the EC would accelerate the EC convergence, as user would often include cases that were most related to the solution of the problem based on his/her subjective judgments [34].

V. Summary

In summary, this paper has proposed a new architecture for developing interactive decision support systems that integrate human and the computer in a cooperative platform. The proposed architecture attempts to reveal the decision-making process by embedding various techniques, such as data visualization, classification, prediction, and evolutionary computational search to enable humans' subjectivity to be incorporated into a decision support system while preserving the remarkable ability of computer system to process information objectively. The characteristics of the proposed architecture have been compared and illustrated using two simulated datasets. The practicality of the architecture is demonstrated through the successful implementation of the system for a real-world medical diagnostic problem, i.e., liver diseases diagnosis. Indeed, we believe the proposed architecture can also be employed by decision support systems of other application types. Further work will focus on evaluating the effectiveness of the system for decision support in various domains. In addition, an incremental learning system is being devised to allow the system to adapt to changing environments.

Acknowledgment

Special thanks to Dr Simon Huang from Timberland Medical Centre, Kuching, Sarawak, Malaysia who has assisted in data collection and experiment management. This research was sponsored by University Malaysia Sarawak. The views and conclusions presented in this paper are those of the authors and do not represent an official opinion, expressed or implied, of University Malaysia Sarawak or the acknowledged individuals.

References

[1] D.J. Power, "What Is A DSS?," *The On-Line Executive Journal for Data-Intensive Decision Support*, vol. 1, no. 3, 1997, pp. 223-232.

[2] R.H. Phelps, R.M. Pliske and S.A. Mutter, "Improving decision making: a cognitive approach," in *Human Productivity Enhancement-Organizations, Personnel, and Decision Making*, vol. 2, J. Zeidner, ed., Praeger: USA ,1987.

[3] G.H. Bower and E. Hilgard, *Theories of learning*, 5th ed., Englewood Cliffs, NJ:Prentice Hall, 1981.

[4] F. Hayes-Roth and N. Jacobstein, "The state of knowledge-based systems," *Communications of the ACM*, vol. 37, no. 3, 1994, pp. 27-39.

[5] S. Franklin, *Artificial Minds*, Cambridge, MA: MIT Press, 1997.

[6] H. Yan, J. Zheng, Y. Jiang, C. Peng and Q. Li, "Development of a decision support system for heart disease diagnosis using multilayer perceptron," *Proc. of the Int. Symposium on Circuits and Systems*, ISCAS '03, vol. 5, pp. 709-712, May 2003.

[7] C.E. Vasios, G.K. Matsopoulos, K.S. Nikita, N. Uzunoglu and C. Papageorgiou, "A decision support system for the classification of event-related potentials," 6th Seminar on *Neural Network Applications in Electrical Engineering*, NEUREL '02, 2002, pp. 159-164.

[8] G. Bowker and S.L. Star, "Situations vs. standards in long-term, wide-scale decision-making: the case of the international classification of diseases" *Intl. conference on System Sciences*, vol. 4, 1991, pp. 8-11.

[9] G.D. Tattersall, K. Chichlowski and R. Limb, "Pre-processing and visualization of decision support data for enhanced machine classification" *1st Intl. Conf. on Intelligent Systems Engineering*, pp. 275-280, 19-21 Aug 1992.

[10] H.A. Simon, *The new science of management decision*, NY:Harper and Row, 1960.

[11] G. Moorhead and R. W. Griffin, *Organizational Behavior–Managing People and Organizations*, 4th ed., Houghton Mifflin:USA, 1995.

[12] W. Siedlecki, K. Siedlecka and J. Sklansky, "An overview of mapping techniques for exploratory pattern analysis," *Pattern Recognition*, vol. 21, no. 5, 1988, pp. 411-429.

[13] A.K. Jain and J. Mao, "Artificial neural network for nonlinear projection of multivariate data," *in Proc. IEEE Int. Joint Conf. Neural Networks*, vol. 3, 1992, pp. 335-340.

[14] J. Mao and A.K. Jain, "Discriminant analysis neural networks," *in Proc IEEE Int. Conf. Neural Networks*, vol. 1, 1993, pp. 300-305.

[15] A. König, "A survey of methods for multivariate data projection, visualization and interactive analysis," *in Proc. of the 5th Int. Conf. on Soft Computing and Information/Intelligent Systems*, IIZUKA'98, 55-59, October 16-20, 1998.

[16] A. König and T. Michel, "DIPOL-SOM - a distance preserving enhancement of the self-organizing map for dimensionality reduction and multivariate data visualization," *Workshop on Self-Organizing Maps - Intelligent Systems and innovational Computing* (WSOM&acute'03); Sept. 11-14, 2003.

[17] J. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *IEEE Trans. On Neural Networks*, vol. 6, no. 2, 1995, pp. 296-317.

[18] M.M. Van Hulle, "Kernel-based equiprobabilistic topographic map formation," *Neural Computation*, vol. 10, no. 7, 1998, pp.1847-1871.

[19] M.M. Van Hulle, "Faithful representations with topographic maps," *Neural Networks*, vol. 12, 1999, pp. 803-823.

[20] D.F. Specht, "Probabilistic neural networks", *Neural Networks*, vol.3, 1990, pp.109-118.

[21] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, 1962, pp. 1065-1076.

[22] J. Chen, H. Li, S. Tang and J. Sun, “A SOM-based probabilistic neural network for classification of ship noises,” *IEEE Int. Conf. On Communication, Circuits and Systems and West Sino Expositions*, vol. 2, 2002, pp. 1209-1212.
[23] F. Ma, W. Wang, W. W. Tsang, Z. Tang, S. Xia and X. Tong, “Probabilistic segmentation of volume data for visualization using SOM-PNN classifier,” *IEEE Symposium on Volume Visualization*, vol. 169, 1998, pp. 71-78.
[24] C.M. Bishop, M. Svensén and C.K.I. Williams, “GTM: A principled alternative to the self-organizing map,” *Proc. ICANN’96*, 1996, pp. 165-170.
[25] M. Herbin, N. Bonnet and P. Vautrot, “A clustering method based on the estimation of the probability density function and on the skeleton by influence zones: Application to image processing,” *Pattern Recognition Letters*, vol. 17, 1996, pp. 1141-1150.
[26] J. Vesanto (2000, April), “Neural network tool for data mining: SOM toolbox," SOM Toolbox Team, Neural Network Research Centre, Helsinki University of Technology, Finland.
[Online]. Available: http://www.cis.hut.fi/projects/somtoolbox
[27] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological Cybernetics*, vol. 43, 1982, pp. 59-69.
[28] S. Goonatilake and S. Khebbal, *Intelligent hybrid systems*, John Wiley & Sons, England, 1995.
[29] R. Nisbett and L. Ross, *Human inference: strategies and shortcomings of social judgment*, Englewood Cliffs, Prentice-Hall:NJ, 1980.
[30] D. Kahneman and A. Tversky, “Subjective probability: a judgment of representative,” *Cognitive Psychology*, vol. 3, 1972, pp. 430-454.
[31] J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI,1975; MIT Press, Cambridge, MA, 1992.
[32] H. Takagi, “Introduction to fuzzy systems, neural networks and genetic algorithms,” *in Intelligent Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms*, D. Ruan (ed.), Kluwer Academic Publishers:USA, 1997, pp 1-33.
[33] H. Takagi, Interactive evolutionary computation: system optimization based on human subjective evaluation,” *IEEE Int. Conf. on Intelligent Engineering Systems* (INES’98), Vienna, Austria, pp. 1-6 Sept, 1998.
[34] H. Takagi, Visualized IEC: Interactive evolutionary computation with multidimensional data visualization” *IECON Proceedings (Industrial Electronics Conference)*, vol. 4, 2000, pp. 2738-2743.
[35] R.M. Corbin, “Decisions that might not get made,” *in Cognitive Process In Choice and Decision Behavior*, T.S. Wallstem (ed.), Hillsdale, NJ, Erlbaum Associates, 1980.
[36] F. Alimoglu, *Combining multiple classifiers for pen-based handwritten digit recognition*, M.Sc Thesis, Institute of Graduate Studies in Science and Engineering, Bogazici University, 1996
[37] F. Meier (2002, June), “Liver function tests,” [Online]. Available: http://kidshealth.org/parent/
[38] O.L. Wong, *The distribution of c-reactive protein and alpha-fetoprotein in liver disease*, (Unpublished Master Dissertation), Faculty of Medicine, Universiti Kebangsaan Malaysia, 1996.
[39] J.O. Flores (2004, June), “Liver function tests,” Gale Encyclopedia of Medicine, [Online]. Available: http://www.findarticles.com/p/articles/
[40] S.S. Cross, R.F. Harrison and R.L. Kennedy, “Introduction to neural networks,” The Lancet, vol. 346, 1995, pp. 1075-1079.

CS Teh was born in the city of Malacca, Malaysia in 1971. He obtained his Master degree in Computer Graphics and Virtual Environments from The University of Hull, UK in 1997. Currently he is a PhD candidate at School of Electrical and Electronic Engineering, University of Science Malaysia. His major field of studies includes artificial neural network modeling, pattern classification, interactive decision support, and virtual reality.

CP Lim received the BEng (Elect.) degree with first class honors from the University of Technology Malaysia in 1992, the MSc (Eng) and PhD degrees from the University of Sheffield, UK in 1993 and 1997. He is currently Associate Professor at School of Electrical & Electronic Engineering, University of Science Malaysia. His research interests include soft computing, pattern classification, medical diagnosis, and fault detection and diagnosis.
Dr. Lim has published more than 100 technical papers, and received six best paper awards at national and international conferences. He is recipient of the Japan Society for the Promotion of Science Research Fellowship (2002), Fulbright Scholarship (2002), and Commonwealth Fellowship (2003), as well as The Outstanding Young Malaysians Award (2001) and National Young Scientist Award (2002) of Malaysia.