

A hybrid intelligent system for medical data classification



Manjeevan Seera^a, Chee Peng Lim^{b,*}

^a Intel Technology, Penang, Malaysia

^b Centre for Intelligent Systems Research, Deakin University, Australia

ARTICLE INFO

Keywords:

Fuzzy Min–Max neural network
Classification and regression tree
Random forest
Hybrid intelligent systems
Medical decision support

ABSTRACT

In this paper, a hybrid intelligent system that consists of the Fuzzy Min–Max neural network, the Classification and Regression Tree, and the Random Forest model is proposed, and its efficacy as a decision support tool for medical data classification is examined. The hybrid intelligent system aims to exploit the advantages of the constituent models and, at the same time, alleviate their limitations. It is able to learn incrementally from data samples (owing to Fuzzy Min–Max neural network), explain its predicted outputs (owing to the Classification and Regression Tree), and achieve high classification performances (owing to Random Forest). To evaluate the effectiveness of the hybrid intelligent system, three benchmark medical data sets, viz., Breast Cancer Wisconsin, Pima Indians Diabetes, and Liver Disorders from the UCI Repository of Machine Learning, are used for evaluation. A number of useful performance metrics in medical applications which include accuracy, sensitivity, specificity, as well as the area under the Receiver Operating Characteristic curve are computed. The results are analyzed and compared with those from other methods published in the literature. The experimental outcomes positively demonstrate that the hybrid intelligent system is effective in undertaking medical data classification tasks. More importantly, the hybrid intelligent system not only is able to produce good results but also to elucidate its knowledge base with a decision tree. As a result, domain users (i.e., medical practitioners) are able to comprehend the prediction given by the hybrid intelligent system; hence accepting its role as a useful medical decision support tool.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Research in computerized intelligent systems for medical applications is an important and exciting domain. In general, a physician typically accumulates his/her knowledge based on patients' symptoms and the confirmed diagnoses. In other words, prognostic relevance of symptoms towards certain diseases and diagnostic accuracy of a patient are highly dependent on a physician's experience (Meesad & Yen, 2003). As medical knowledge and treatment therapy progress rapidly, e.g. the occurrence of new diseases and the availability of new drugs, it is challenging for a physician to keep up-to-date with all recent knowledge and development in clinical practice (Meesad & Yen, 2003). On the other hand, with the advent of computing technologies, it is now relatively easy to acquire and store a lot of information digitally, e.g. in dedicated databases of electronic patient records (Pavlopoulos & Delopoulos, 1999). As such, the deployment of computerized medical decision support systems becomes a viable approach to assisting physicians to swiftly and accurately diagnose patients (Chabat, Hansell, & Yang, 2000). Nevertheless, numerous

issues have to be overcome before a useful medical decision support system can be developed and deployed, which include decision making in the presence of uncertainty and imprecision (Tsipouras, Voglis, & Fotiadis, 2007). While medical experts' knowledge and experience is important, ranging from assessing a patient's condition to making a diagnosis, advances in machine learning (Kwiatkowska, Atkins, Ayas, & Ryan, 2007) techniques have opened up the way for medical practitioners to exploit computerized intelligent systems for decision support in their workplace, e.g. surgical imagery and X-ray photography (Isola, Carvalho, & Tripathy, 2012). When treating a patient, a physician first needs to narrow down the suspected disease to the root cause (out of a list of probable causes with similar symptoms) using his/her knowledge and experience, and then confirms the diagnosis by performing a number of tests (Isola et al., 2012). Concomitantly, computerized intelligent systems can be useful in assisting the physician to arrive at an informed decision quickly, e.g. by learning from similar past cases in a large database of electronic patient records and inferring the diagnosis for the current patient with proper justifications. The advantages of using such intelligent systems include increasing diagnosis accuracy and, at the same time, reducing time and costs associated with patient treatment (Çomak, Polat, Güneş, & Arslan 2007).

* Corresponding author. Tel.: +61 3 5227 3307.

E-mail address: chee.lim@deakin.edu.au (C.P. Lim).

Machine learning models have been developed to support various medical decision making tasks. As an example, intelligent classifiers have been used for prognosis, diagnosis, and screening of diabetes, breast cancer and Parkinsons disease (Luukka, 2011). A number of neural-fuzzy models have been used as classifiers for heart disease, because they are capable of learning from data samples (i.e., patient records) and generalizing beyond the training samples (Kahramanli & Allahverdi, 2009). These include fuzzy neural networks, fuzzy probabilistic neural networks, and fuzzy learning vector quantization networks (Sekar, Dong, Shi, & Hu, 2012). However, one key limitation of these models is the lack of ability to explain their predictions (Markowska-Kaczmar & Matkowski, 2006). This is the motivation of this research, whereby we attempt to devise a machine learning-based system that is able to reveal its reasoning in dealing with an input case, and to provide justification for its predictions.

In machine learning, neural networks have significant advantages for medical decision support applications (Downs, Harrison, Kennedy, & Cross, 1996). Compared with expert systems, neural networks avoid the time-consuming and demanding knowledge acquisition process by directly learning complex association between input symptoms and target diseases from data samples i.e., patient records (Hayes-Roth, Waterman, & Lenat, 1983). In addition to learning, neural networks possess other useful properties, which include handling incomplete or missing data as well as filtering noise, uncertainty or imprecision (Downs et al., 1996). In view of the salient features of neural networks, the Fuzzy Min-Max (FMM) neural network is investigated for developing a usable and useful medical decision support tool in this paper. In order to further strengthen the FMM network for medical applications, the Classification and Regression Tree (CART) and Random Forest (RF) models are incorporated to produce a hybrid intelligent system. The proposed hybrid model serves as an extension of our previous work (Seera, Lim, Ishak, & Singh, 2012), which was focused on an offline model (i.e. FMM-CART) for undertaking fault detection and diagnosis problems. Here, the motivation is to extend the hybrid model such that it is equipped with the necessary characteristics for undertaking medical decision support tasks. While CART has the advantage of rule extracting in the form of a tree structure, it is less flexible in performing incremental learning from data samples. While FMM has the advantage of one-pass training with incremental learning properties, it lacks the capability of producing rules to explain its predictions. On the other hand, RF has the benefit of forming an ensemble of CART whereby the best tree can be identified to produce high prediction accuracy. Therefore, the hybrid model, i.e. FMM-CART-RF, has three distinctive capabilities, viz, learning incrementally from data samples (owing to FMM), explaining its predicted outputs (owing to CART), and achieving high classification performances (owing to RF). This is the key contribution of this research.

From the perspective of decision making, we commonly seek a second opinion (and sometimes even more) before making important decisions, especially one that has medical implications (Polikar, 2006). Then, different opinions are weighed and combined based on a thought process before the final decision is made. In machine learning, such ensemble concept is utilized too in developing a highly accurate model. In particular, bagging algorithms are useful for constructing an ensemble of decision trees, and one such variant is RF. RF is an effective ensemble method for data mining (Zhang, Zulkernine, & Haque, 2008). It has shown good results in many applications, which include automatic intrusion detection systems (Zhang et al., 2008) and medical data classification (Wu, Ye, Liu, & Ng, 2012). RF has some important characteristics such as providing useful internal estimates of strength, correlation, and variable importance, while producing high accuracy in comparison with many standard classification models (Wu et al., 2012). As a

result, RF is exploited in this paper to form a group of diverse CART models so that the best tree can be chosen for classification and rule extraction purposes.

The hybrid model proposed in this research has two important practical implications in the domain of medical decision support. Firstly, the ability to provide explanation and justification for the prediction is of paramount importance, in order to convince domain users (i.e., medical practitioners) with the outcome given by a computerized decision support system. This ability is essential in safety critical applications, such as medical diagnosis and prognosis, whereby domain users need to understand, and be convinced of, how the computerized system arrives at such a prediction (Economou, Goumas, & Spiropoulos, 1996). The elucidated rules in the form of a decision tree from the hybrid model is, therefore, important in practice, whereby the rules could serve as a source of second opinions in medical diagnostic situations (Kovalerchuk, Vityaev, & Ruiz, 2000). Secondly, accuracy of a decision support system is very crucial in medical applications. As stated in Luukka (2011), a high false negative rate of a screening system would increase the risk of patients by depriving them from getting the necessary medical attention, while a high false alarm rate would cause unnecessary worry and stress in patients as well as increase the demand on medical resources. On the other hand, as indicated in Kinney (2003), a decision support system with high specificity and variable sensitivity could save medical costs and improve scheduling of vestibular patients in an otolaryngology clinic. Besides that, Huang, Yang, King, and Lyu (2006) also recognized the usefulness of machine learning models in reducing cost and saving time for undertaking medical diagnostic tasks. As shown in the experimental study, the proposed hybrid model not only is able to achieve high accuracy, sensitivity, and specificity rates, but also to provide explanation for its predictions in the form of a decision tree; hence demonstrating its usefulness as a decision support system in practical environments.

The organization of this paper is as follows. Literature reviews related to intelligent models for medical applications as well as rule extraction methods are detailed in Section 2. In Section 3, the dynamics of the proposed FMM-CART-RF model is presented. The experimental study, results, and discussion using a number of benchmark medical data sets are presented in Section 4. Finally, conclusions and suggestions for further work are presented in Section 5.

2. Literature Review

The literature review presented in this section is divided into two main parts: (i) intelligent systems for medical applications; (ii) rule extraction methods the emphasis on medical applications. In machine learning, supervised learning is a commonly used method for tackling medical problems. The task of identifying the smallest sets of genes and constructing a highly accurate classification model of cancers from microarray data was attempted using a fuzzy neural network and the Support Vector Machine (SVM) (Wang, Chu, & Xie, 2007). Important genes were chosen using a feature importance ranking scheme. The identified genes were then fed to the learning algorithm for classification. Cascio et al. (2006) utilized a supervised neural network for classification of breast cancer. A set of representative features was extracted and used for detecting masses in mammographic images. Among various supervised learning models, the Multi-Layer Perceptron (MLP)-based neural networks are commonly used in different application domains. In the medical area, MLP was used to classify patterns of scoliosis spinal deformity using features from the central axis curve of spinal deformity (Lin, 2008). Osareh, Shadgar, and Markham (2009) used an MLP neural network with a genetic-based algorithm for

automated identification of exudate pathologies in retinopathy images. An MLP neural network with a fast learning algorithm was used for the prediction of post-dialysis blood urea concentration (Azar, 2013). Ma and Wang (2012) developed a multilevel feed-forward neural network-based hypertension model for forecasting meteorological factors that could remarkably affect hypertension.

On the other hand, Kondo, Kondo, and Ueno (2009) used a group method of data handling (GMDH)-based neural network for three-dimensional analysis of heart images. The GMDH-based network operated using a heuristic self-organization method. It was applied successfully to liver cancer problems (Kondo, Kondo, Takao, & Ueno, 2010). The algorithm identified characteristics of the cancer images using some feedback-loop calculations. The radial basis function-type of neurons were adopted in the network structure. Coyle, Prasad, and McGinnity (2009) used a self-organizing fuzzy neural network for sensitivity analysis of an Electroencephalogram (EEG)-based brain-computer interface. Walczak (2005) deployed the backpropagation neural network to predict blood transfusion requirements of trauma patients, in order to enable its efficient utilization in hospitals. While the backpropagation algorithm is the most widely used technique for training neural network models, it has some limitations which include the possibility of being trapped in local solutions and producing low classification rates in some cases (Örkcü & Bal, 2011).

Successful applications of intelligent systems to many other medical problems have also been reported. A constructive algorithm that creates a compact neural network architecture was employed for classification of early breast cancer patients (Franco, Subirats, Molina, Alba, & Jerez, 2007). A neural network with the weighted fuzzy membership functions was employed to detect premature ventricular contractions (Lim, 2009). Besides that, a nonlinear dynamical neural network was used in a brain-machine interface model to interpret temporal and spatial neural activities (Sanchez, Erdogmus, Nicoletis, Wessberg, & Principe, 2005). A multi-resolution massively trained neural network for suppressing the contrast of ribs and clavicles in chest radiographs was described in Suzuki, Abe, MacMahon, and Doi (2006). A fused hierarchical neural network for cardiovascular disease using hemodynamic parameters derived from non-invasive sphygmogram was presented in Sekar et al. (2012). Srinivasan, Eswaran, and Sriraam (2007) used the Elman and Probabilistic neural networks for EEG detection. A principal component analysis-enhanced cosine radial basis function neural network classifier was employed for classification of EEG signals into healthy, ictal, and interictal categories (Ghosh-Dastidar, Adeli, & Dadmehr, 2008). The SVM for both prediction and detection of Cyclosporine A blood concentrations in therapeutic drug monitoring of kidney recipients was reported in Camps-Valls et al. (2007). It was reported that SVM method would likely to provide poor performance if the number of features was higher than that of the data samples.

In addition to producing accurate predictions, rule extraction is another important requirement for intelligent systems to be useful in medical applications. This is because the extracted rules provide domains users with the explanation and justification of predictions. A survey by Andrews, Diederich, and Tickle (1995) identified three categories of rule extraction methods, viz, decompositional, pedagogical, and eclectic. The decompositional category aims to extract individual hidden and output nodes by analyzing the weight vector associated with each hidden node in the network. By treating each hidden node independently, a local rule aggregation process is accomplished to obtain a composite rule base for the network. Decision trees belong to this decompositional category (Wang & Fu, 2005). The pedagogical category aims to extract symbolic rules which map the input–output relationship as closely as possible to the way the network understands the relationship.

The number of rules and their form do not directly correspond to the number of weights or the network structure (Saad & Wunsch, 2007). The electric category combines both decompositional and pedagogical methods. Knowledge from the internal structure and/or weight vectors in the trained network is used to complement a symbolic learning algorithm for rule extraction.

Fuzzy rules are commonly expressed in simple if-then format, and have been used in various applications. Fuzzy-based diagnostic techniques were used in a drowsiness detection system (Picot, Charbonnier, & Caplier, 2012). Cascading decision rules according to a medical scale for drowsiness evaluation was made by monitoring a single EEG channel. A neuro-fuzzy classifier was used for classification and extraction of knowledge from the training set of sleep-wake states, and a number of stages were developed for healthy infants of 6 months and onwards (Held et al., 2006). A fuzzy rule-based decision support system was used for the diagnosis of coronary artery disease, where the system produced a set of rules from a decision tree (Tsipouras et al., 2008). Fuzzy rules generated from rule-based knowledge discovery methods were also used to provide decision support in the medical informatics domain (Papageorgiou, 2011). The method elicited knowledge in form of fuzzy rules, and fed them into a fuzzy cognitive map to form a dynamic decision support system.

Association rule mining is able to discover relationships among different attributes in a database and to produce if-then rules. Association rule mining was applied to automated detection and classification of transient events in EEG recordings (Exarchos, Tzallas, Fotiadis, Konitsiotis, & Giannopoulos, 2006). The decision was interpreted by using a set of association rules. Ribeiro, Traina, Traina, and Azevedo-Marques (2008) used the association rule-mining technique to examine medical images of mammograms. Based on the analyzed outcomes, suggestions were generated by employing association rules. Time-series data mining pertaining to blood and urine biochemical test data from chronic hepatitis patients was conducted, and knowledge was mined from the data samples (Abe, Ohsaki, Yokoi, & Yamaguchi, 2006).

Decision trees are useful as they provide predictions in a comprehensible manner that resembles human reasoning. A probabilistic boosting tree algorithm was used for processing information from magnetic resonance imaging volumes and cortical surfaces (Tu et al., 2007). The approach selected and combined hundreds of features at different scales, such as curvatures, gradients, and shape index. A decision tree that employed boosting was used for dengue fever diagnosis (Rao & Kumar, 2012). Feature selection was performed using a wrapper-based method with genetic search, with the aim to extract a subset of influential symptoms that could provide diagnosis of the illness. A rule-based algorithm cascaded with a neural network was introduced to detect epileptic seizure segments in EEG signals, where the algorithm screened out segments of EEG patterns as epileptic based on features in the power spectrum (Karayiannis et al., 2006).

Logistic regression and multilayer perceptron models with automatic relevance determination were applied to analyze malignant or benign ovarian tumors (Aung et al., 2007). Explanatory rules were extracted using the orthogonal search-based algorithm. A rough set theory was deployed for extracting diagnostic rules from medical databases (Fakih & Das, 2006). The SVM was used for learning while the evolutionary algorithm was deployed for generating rules in a comprehensible and concise form for modeling breast cancer instances (Stoean & Stoean, 2013). However, the SVM results are not easy to interpret, as the explanation is non-intuitive even to human experts. Falco (2013) used differential evolution, where explicit knowledge was extracted for automatic classification of items in a medical database. Carmona, González, Del Jesus, Navío-Acosta, and Jiménez-Trevino (2010) used a multi-objective evolutionary algorithm for rule extraction in a

study pertaining to different categories of patients visiting the psychiatric emergency department in a given period of time of the day. The extracted rules provided useful information regarding the rate of admission to the psychiatric emergency department.

A variety of machine learning and rule extraction methods with respect to medical applications have been reviewed. While standard learning models such as MLP, RBF, or SVM, can be used as classifiers, no explanations are given to justify their predictions. As such, rule extraction from intelligent systems is important for explaining the predicted outcome, especially in medical applications. Examples of useful rule-based models include association rule mining and decision trees. From the literature review, it is clear that computational models that possess both incremental learning and rule extraction capabilities are rare. As a result, our research focuses on the design and development of a computational model that encompasses both capabilities. In the next section, the proposed hybrid intelligent system, i.e., FMM-CART-RF, that is able to perform incremental learning and provide high classification accuracy, as well as to allow the extraction of explanatory rules in the form of a decision tree is described.

3. The hybrid intelligent system

In this paper, a hybrid intelligent model, i.e., FMM-CART-RF, is developed to undertake medical data classification problems. The hybrid intelligent system possesses two important properties, i.e., incremental learning with high performance and rule extraction with justifiable predictions. Fig. 1 shows the procedure of FMM-CART-RF. In order to allow FMM, CART, and RF to operate efficiently as a hybrid intelligent system, a number of modifications are needed, as explained in the following sub-sections.

3.1. The FMM network

The FMM network has three layers of nodes, i.e., F_A is the input layer, F_B is the hidden layer, and F_C is the output layer. The input and output layers contain nodes equal in number to the number of dimensions of the input pattern and the number of the target classes, respectively. The hidden layer, known as the hyperbox layer, contains nodes that can be created incrementally. Each F_B node represents a hyperbox fuzzy set. F_A to F_B connections contain the minimum and maximum points of the hyperboxes, denoted by matrices \mathbf{V} and \mathbf{W} , respectively. F_B and F_C connections are binary values, and are stored in matrix \mathbf{U} . Equation (1) is used to assign the values between F_B and F_C connections, i.e.,

$$u_{jk} = \begin{cases} 1 & \text{if } b_j \text{ is a hyperbox for class } C_k, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where b_j is the j th node and C_k is the k th node. Each F_C node represents a class. The output of the F_C node represents the degree to which the h th input pattern, $\mathbf{A}_h = (a_{h1}, a_{h2}, \dots, a_{hn}) \in I^n$, fits within class k . The transfer function for each F_C nodes performs the fuzzy

union of the appropriate hyperbox fuzzy set values, and is defined as:

$$c_k = \max_{j=1}^m b_j u_{jk}, \quad (2)$$

where the membership function for the j th hyperbox, $b_j(\mathbf{A}_h)$, $0 \leq b_j(\mathbf{A}_h) \leq 1$, is used to measure the extent to which the h th input pattern, \mathbf{A}_h , falls outside hyperbox B_j . The output of the F_C class nodes can be used in two different ways. In the case of a soft decision, the outputs are used directly. In the case of a hard decision, the F_C node with the highest value is chosen, and its node value is set to 1 to indicate that it is the closest pattern class, while other F_C node values are set to 0, i.e., the principle of winner-takes-all.

The hyperbox fuzzy sets are the key component of the FMM network. A parameter called the expansion coefficient, $\theta \in [0, 1]$, is used to control the hyperbox size. A small value of θ leads to the creation of a large number of hyperboxes, and vice versa. For an n -dimension input pattern, a unit cube, I^n , is defined, and the definition of each hyperbox fuzzy set B_j is:

$$B_j = \{X, V_j, W_j, f(X, V_j, W_j)\} \quad \forall X \in I^n, \quad (3)$$

where $\mathbf{V}_j = (v_{j1}, v_{j2}, \dots, v_{jn})$ is the minimum point of B_j and $\mathbf{W}_j = (w_{j1}, w_{j2}, \dots, w_{jn})$ is the maximum point of B_j . In essence, the membership function is computed with respect to the minimum and maximum points of the hyperbox, and to the extent to which the input pattern fits into the hyperbox. The combined fuzzy set that classifies the k th pattern class, C_k , is:

$$C_k = \bigcup_{j \in K} B_j, \quad (4)$$

where K is the set of hyperboxes associated with class k . The FMM training procedure is focused on establishing and fine-tuning the class boundaries.

In FMM, hyperboxes from the same class are allowed to overlap one another. However, the overlapped regions of hyperboxes from different classes need to be eliminated. The membership function for the j th hyperbox, $b_j(\mathbf{A}_h)$, is used to measure the extent each component of the input pattern is larger (or smaller) than the maximum (or minimum) point along each dimension that falls outside the minimum and maximum boundaries of the hyperbox. As $b_j(\mathbf{A}_h)$ becomes closer to 1, the point is said to be “more” contained in the respective hyperbox. The membership function criterion is the sum of two complements, viz. the average of the maximum and minimum point violations. The resulting membership function is:

$$b_j(\mathbf{A}_h) = \frac{1}{2n} \sum_{i=1}^n [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - w_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, v_{ji} - a_{hi})))], \quad (5)$$

where γ is the sensitivity parameter that controls the speed the membership value decreases when the distance between \mathbf{A}_h and B_j increases.

1. Provide the input data samples
2. Start the FMM learning procedure
 - a) Initialize the hyperbox minimum and maximum points
 - b) Identify the closest hyperbox to the input pattern for expansion, if not, add a new hyperbox
 - c) Check whether the expansion causes any overlaps among hyperboxes
 - d) If overlaps exist, contract the hyperboxes to eliminate the overlapped regions
 - e) Calculate the centroid point of the hyperbox
3. Use the generated hyperboxes as the input samples to CART
 - a) Split the nodes recursively, with each node assigned to a predicted class
 - b) Use the centroid point and the confidence factor for tree building
 - c) Conduct bagging for RF, with the capability of random attribute selection
4. Use the final tree for data classification

Fig. 1. The procedure for developing the hybrid intelligent system.

To ensure a smooth synergy with the CART model, two modifications to FMM are needed. The first modification, adopted from Quteishat and Lim (2008), is concerned with the centroid computation of FMM hyperboxes. As the FMM hyperbox are constructed using the minimum and maximum points, they do not represent the centroid of data samples contained within the hyperbox. The centroid, which represents the most populated data samples within the hyperbox, can be computed as:

$$C_{ji}^{new} = C_{ji} + \frac{|a_{hi} - C_{ji}|}{N_j}, \quad (6)$$

where C_{ji} is the centroid of the j th hyperbox in the i th dimension, a_{hi} is the i -dimensional element of the h th input pattern, and N_j is the number of patterns included in the j th hyperbox.

The second modification is concerned with the hyperbox confidence factor. A confidence factor, CF_j , for each hyperbox B_j is introduced. The confidence factor measures the usage frequency and predictive accuracy of the hyperbox. It can be computed using (Carpenter & Tan, 1995):

$$CF_j = (1 - \lambda)U_j + \lambda A_j, \quad (7)$$

where U_j is the usage of hyperbox j , A_j is the accuracy of hyperbox j , and $\lambda \in [0, 1]$ is a weighting factor. The usage measure of a hyperbox, U_j , is defined as the number of prediction patterns classified by any hyperbox B_j divided by the maximum number of prediction patterns classified by any hyperbox with the same class. The accuracy measure of a hyperbox, A_j , is defined as the number of correctly classified prediction patterns by any hyperbox B_j divided by the maximum number of correctly classified patterns with the same class. In essence, the confidence factor provides a measure for hyperboxes that are frequently used and that are generally accurate in prediction, as well as hyperboxes that are rarely used, but yet highly accurate.

3.2. The CART and RF models

To build a tree classifier, the training samples, which comprise a set of features (attributes) and their class labels, are needed. The tree building procedure starts with a set of training data and their associated class labels. The training set is recursively partitioned into smaller subsets as the tree is built. Based on the decision matrix from the distribution of classes in the training set, each resulting node is assigned a predicted class. The CART tree building procedure begins at the root node where the entire training set is used. No further decision to the partition is made if all the data samples belong to the same class. Otherwise, the best possible variable to split the node into two leaf nodes is identified. This process is repeated for each new leaf node until a completely distinguishable CART tree is obtained.

The CART tree uses the Gini impurity as a measure to determine how often a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset. An error occurs when a learning sample takes the wrong branch while being classified by the tree, as normally a limited number of learning samples are used to build the tree. This problem can be mitigated by introducing a modification whereby each class in the tree is tagged with a confidence factor from FMM. In the original Gini computation, the sum of the Gini index is measured over m classes. With the modification, a new way for computing the Gini index is adopted, as follows:

$$Gini(D) = 1 - \sum_i w^2(i), \quad (8)$$

where D is a data partition or the set of training data and $w(i)$ is the normalized weights. This modification serves as a useful approach

to improve the performance of CART, as shown in the experimental studies.

Next, the goodness-of-split is measured by an impurity function defined for each node. Ideally, each leaf node should be “pure”, in which only one class dominates. When a split at node t occurs, the goodness-of-split is computed as follows.

$$\Delta i(s, t) = i(t) - p_L[i(t_L)] - p_R[i(t_R)], \quad (9)$$

where s is a particular split, p_L is the proportion of the cases at node t that go into the left child node (t_L), p_R is the proportion of cases at node t that go into the right child node (t_R), $i(t_L)$ is impurity of the left child node, and $i(t_R)$ is impurity of the right child node (Yohannes, 1999). Given a number of splits generated by a particular variable at a node, the rule is to select the split that results in the maximum reduction in the impurity measure of the parent node. On the other hand, RF is an ensemble method that offers several advantages, i.e., the abilities to operate efficiently with large data sets and to perform robustly against outliers and noise (Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012). In this study, the RF ensemble is established by a combination of CART. The RF is built using the bagging method in tandem with random attribute selection. Given a training set, D , the procedure for generating k decision trees is as follows. For each iteration, i ($i = 1, 2, \dots, k$), a training set, D_i , is established by sampling with replacement from D . In other words, each D_i is a bootstrapped sample of D , whereby some training data may occur more than once in D_i , while others may be excluded (Han, Kamber, & Pei, 2012). Let F be the number of attributes used to determine the split at each node, and F is smaller than the number of available attributes. To construct a decision tree, M_i , randomly select, at each node, F attributes as the candidates for the split at the node (Han et al., 2012). In this study, the trees in an RF ensemble are grown using the CART procedure to its maximum size. However, the pruning step in the CART algorithm is omitted. At each node of an RF tree, a subset of the available attributes is randomly selected, and the best split within the attributes is selected for that node. The tree with the highest classification accuracy rate is then selected and is used for classification.

4. Experimental results and discussion

4.1. Experimental setup

In this section, three publicly available data sets from the UCI machine learning data repository (Newman, Hettich, Blake, & Merz, 2007) were used for evaluation, i.e., Breast Cancer Wisconsin, Pima Indians Diabetes, and Liver Disorders. Before the experiments, all variables in the data set were first normalized between 0 and 1. They were then used by FMM-CART-RF for learning and prediction. A decision tree was produced to provide explanation for the predictions. The k -fold cross-validation was used, as it could minimize bias associated with random sampling of training samples (Delen, Walker, & Kadam, 2005). With k -fold cross validation, the whole data set was randomly divided into k mutually exclusive subsets with approximately equal number of samples. FMM-CART-RF was trained and tested k times, whereby each time, one of the k subsets was used as the test set and the remaining as the training set. The average results from k different runs were computed. In order to facilitate performance comparison, three different experimental train-to-test ratios were adopted, i.e., 1:1 (2-fold cross-validation), 4:1 (5-fold cross-validation), and 9:1 (10-fold cross-validation). Each experiment was repeated 30 times. The common performance metrics in medical applications, i.e., accuracy, sensitivity, specificity, and area under the Receiver Operating Characteristic (ROC) curve, were computed to quantify the

performance (Loo, 2005). The bootstrap method (Efron, 1979) was employed to compute the results. A total of 1,000 bootstrap samples generally could provide accurate results while using 2,000 bootstrap replications could provide useful and valid results (Efron & Tibshirani, 1993). In this study, the performance metrics were computed with 5,000 re-samplings, in order to produce reliable estimate of the results. In addition to FMM-CART-RF, FMM-CART and original FMM were implemented for performance comparison purposes.

4.2. Breast Cancer Wisconsin data set

This breast cancer data set was created by Wolberg from the University of Wisconsin. The aim was to diagnose breast masses based on Fine Needle Aspiration (FNA). Nine visually assessed features were identified from an FNA sample, which was considered most relevant to the diagnosis. The features used were: (1) clump thickness, (2) uniformity of cell size, (3) uniformity of cell shape, (4) marginal adhesion, (5) single epithelial cell size, (6) bare nuclei, (7) bland chromatin, (8) normal nucleoli, and (9) mitoses.

Table 1 shows the overall results. Table 1 shows the overall results. The best result was produced by FMM-CART-RF (98.84%) using the 9:1 train-to-test ratio. The results from five publications, which were based on different train-to-test ratios, were used for comparison, as shown in Table 1. The Fuzzy-AIRS (Polat, Şahan, Kodaz, & Güneş, 2007) method produced the best accuracy of 98.51%, followed by BC FRPCA1 method (Luukka, 2009) at 98.19%, Sim (Luukka, 2011) at 97.49%, Pedagogical (Stoean & Stoean, 2013) at 97.07% and Real-coded GA (Örkcü & Bal, 2011) at 96.50%. Nevertheless, this performance was inferior to that of FMM-CART-RF.

In addition to accuracy, both sensitivity and specificity measures were computed too. Sensitivity measures the proportion of positive cases that are correctly identified while specificity mea-

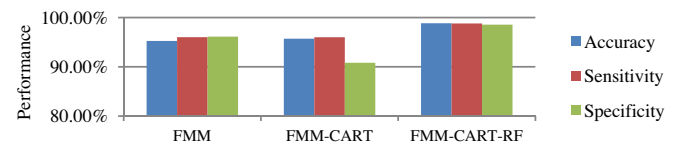


Fig. 2. Comparison of results for FMM, FMM-CART, FMM-CART-RF for Breast Cancer Wisconsin data set.

Table 2
Results of ROC for Breast Cancer Wisconsin data set.

Source	Method	ROC
Luukka (2011)	Sim	0.956
	Sim + F1	0.971
	Sim + F2	0.971
This study	FMM	0.961
	FMM-CART	0.973
	FMM-CART-RF	0.987

asures the proportion of negative cases that are correctly identified. While the sensitivity and specificity rates were not provided in the five papers, the sensitivity and specificity rates for FMM, FMM-CART, and FMM-CART-RF are calculated, as shown in Fig. 2. It can be noticed that the accuracy, sensitivity, and specificity rates for FMM-CART-RF are the highest, as compared with those from FMM and FMM-CART.

Another important performance metric in medical application, i.e. area under the ROC curve, was also computed. In Table 2, a comparison of the areas under the ROC curve is listed. The results from all models reported in Luukka (2011), i.e., Sim, Sim + F1, and Sim + F2, were lower than those from FMM-CART-RF and FMM-CART. FMM-CART-RF managed to produce the best result; hence ascertaining its usefulness in this application.

In addition to achieving good classification performances, another key advantage of FMM-CART-RF is its capability of producing a decision tree for explaining its predictions. The decision tree for the Breast Cancer Wisconsin data set is shown in Fig. 3. Its interpretation is as follows. Based on the “Uniformity of Cell Shape” feature, the decision tree starts to split. When the corresponding feature is smaller than 0.0708, splitting occurs based on the “Bare Nuclei” feature. When the corresponding feature is larger than or equal to 0.1787, FMM-CART-RF classifies the input as “malignant”, otherwise the input is predicted as “benign”. The other branch of the tree checks the “Uniformity of Cell Shape” feature. When the corresponding feature is smaller than 0.0755, the “Bland Chromatin” feature is checked. When the corresponding feature is larger than or equal to 0.1788, FMM-CART-RF classifies the input as “malignant”, otherwise the input is predicted as “benign”. When the “Uniformity of Cell Shape” feature is larger than or equal to 0.0755, the “Uniformity of Cell Shape” feature is scrutinized again. If the corresponding feature is larger than or equal to 0.1240, FMM-CART-RF classifies the input as “malignant”. Otherwise, the tree splits based on the “Clump Thickness” feature. If the corresponding feature is smaller than 0.1714, FMM-CART-RF classifies the input as “benign”. Otherwise, the tree splits again based on the “Bare Nuclei” feature. If the corresponding feature is larger than or equal to 0.0564, FMM-CART-RF classifies the input as “malignant”, otherwise the input is predicted as “benign”.

4.3. Pima Indians Diabetes data set

Owned by the National Institute of Diabetes and Digestive and Kidney Diseases, the Pima Indian Diabetes data set was concerned with the presence or absence of diabetes among Pima-Indian

Table 1
Results of CART, FMM, FMM-CART and comparison with other methods for the Breast Cancer Wisconsin data set.

Source	Method	Accuracy (%)	Ratio	Runs
Luukka (2009)	BC FRPCA1	98.19	1:1	30
	BC FRPCA2	98.13	1:1	30
	BC FRPCA3	98.16	1:1	30
	BC Original	97.49	1:1	30
	BC PCA	97.72	1:1	30
Luukka (2011)	Sim	97.49	1:1	30
	Sim + F1	97.10	1:1	30
	Sim + F2	97.18	1:1	30
Örkcü and Bal (2011)	Binary-coded GA	94.00	9:1	10
	BP	93.10	9:1	10
	Real-coded GA	96.50	9:1	10
Polat et al. (2007)	AIRS	97.20	9:1	30
	Fuzzy-AIRS	98.51	9:1	10
Stoean and Stoean (2013)	Cooperative coevolution	96.69	2:1	30
	Decompositional	95.93	2:1	30
	Pedagogical	97.07	2:1	30
	SVMs	96.50	2:1	30
This study	FMM	95.14	4:1	30
	FMM	95.21	1:1	30
	FMM	95.26	9:1	30
	FMM-CART	93.14	4:1	30
	FMM-CART	94.86	1:1	30
	FMM-CART	95.71	9:1	30
	FMM-CART-RF	97.29	1:1	30
	FMM-CART-RF	97.57	4:1	30
	FMM-CART-RF	98.84	9:1	30

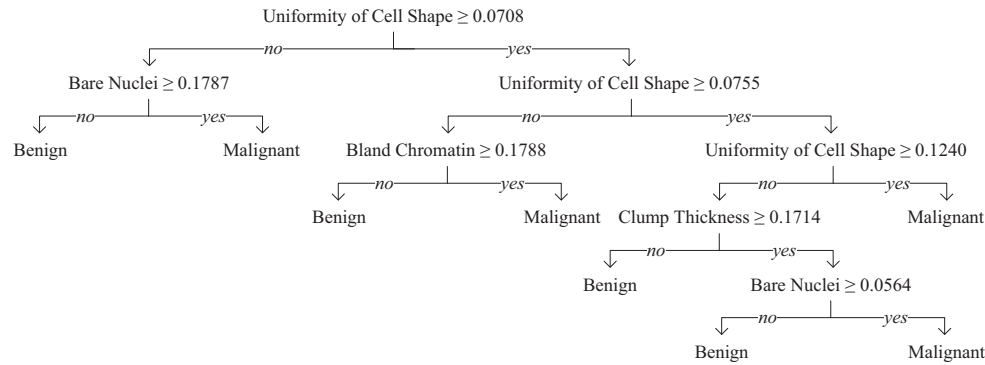


Fig. 3. The FMM-CART-RF decision tree for the Breast Cancer Wisconsin data set.

Table 3

Results of CART, FMM, FMM-CART and comparison with other methods for the Pima Indians Diabetes data set.

Source	Method	Accuracy (%)	Ratio	Runs
Luukka (2011)	Sim	75.29	1:1	30
	Sim + F1	75.84	1:1	30
	Sim + F2	75.97	1:1	30
Örkücü & Bal (2011)	Binary-coded GA	74.80	9:1	10
	BP	73.80	9:1	10
	Real-coded GA	77.60	9:1	10
This study	FMM	68.10	4:1	30
	FMM	68.49	1:1	30
	FMM	69.28	9:1	30
	FMM-CART	69.13	1:1	30
	FMM-CART	69.53	9:1	30
	FMM-CART	71.35	4:1	30
	FMM-CART-RF	76.56	4:1	30
	FMM-CART-RF	76.87	1:1	30
	FMM-CART-RF	78.39	9:1	30

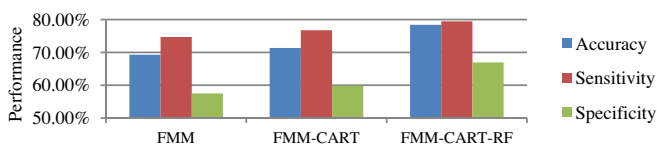


Fig. 4. Comparison of results for FMM, FMM-CART, FMM-CART-RF for Pima Indians Diabetes data set.

women living near Phoenix, Arizona. Some constraints were placed on the selection of patients, e.g., all patients were females with at least 21 years old of Pima Indian heritage. There were eight features in each data sample: (1) number of times pregnant, (2) plasma glucose concentration a 2 h in an oral glucose tolerance test, (3) diastolic blood pressure, (4) triceps skin fold thickness, (5) 2-h serum insulin, (6) body mass index, (7) diabetes pedigree function, (8) age.

The results of FMM-CART-RF and other methods are summarized in Table 3. Again, FMM-CART-RF produced the highest accuracy using the train-to-test ratio of 9:1. The results from two publications were compared, as shown in Table 3. The best reported accuracy rates were 75.97% from the Sim + F2 method in Luukka (2011) and 77.60% from the Real-coded GA method in Örkücü and Bal (2011). Nevertheless, they were inferior to that of FMM-CART-RF (78.39% accuracy).

Similar to the previous experiment, no sensitivity and specificity rates were reported in the respective publications. The computed accuracy, sensitivity, and specificity rates for FMM,

Table 4

Results of ROC for Pima Indians Diabetes data set.

Source	Method	ROC
Luukka (2011)	Sim	0.762
	Sim + F1	0.703
	Sim + F2	0.667
This study	FMM	0.661
	FMM-CART	0.683
	FMM-CART-RF	0.732

FMM-CART, and FMM-CART-RF are shown in Fig. 4. Again, the results from FMM-CART-RF were the highest for all three indicators as compared with those from FMM and FMM-CART. The comparison of the areas under the ROC curve is listed in Table 4. FMM-CART-RF ranked second in terms of the area of the ROC curve, i.e., better than those from Sim + F1 and Sim + F2, but inferior to that of Sim from Luukka (2011).

The decision tree for the Pima Indians Diabetes data set is shown in Fig. 5, and the interpretation is as follows. Based on the “Plasma glucose concentration” feature, the decision tree starts to split. When the corresponding feature is larger than or equal to 0.1494, the feature of “Body mass index” is checked. When the corresponding feature is larger than or equal to 0.1306, FMM-CART-RF classifies the input as “positive”, otherwise the input is predicted as “negative”. On the other hand, if the “Plasma glucose concentration” feature is smaller than 0.1494, the “Age” feature is checked to determine the next branch. When the “Age” feature is smaller than 0.1258, FMM-CART-RF classifies the input as “negative”. Otherwise, the “Body mass index” feature is examined. When the corresponding feature is smaller than 0.1811, FMM-CART-RF classifies the input as “negative”. When the corresponding feature value is larger than or equal to 0.1811, the tree branches to the “Diabetes pedigree” feature. When the corresponding feature is larger than or equal to 0.1565, FMM-CART-RF classifies the input as “positive”. Otherwise, splitting occurs at “Plasma glucose concentration”. When the corresponding feature is smaller than 0.1229, FMM-CART-RF classifies the input as “negative”. Otherwise, splitting occurs again at “Diastolic blood pressure”. When the corresponding feature is larger than or equal to 0.1498, FMM-CART-RF classifies the input as “negative”, otherwise the input is predicted as “positive”.

4.4. Liver Disorders data set

The Liver Disorders data set was prepared by the BUPA Medical Research Company. The features were obtained from blood tests (the first five features) and daily alcohol consumption (the last feature), as follows: (1) mean corpuscular volume, (2) alkaline

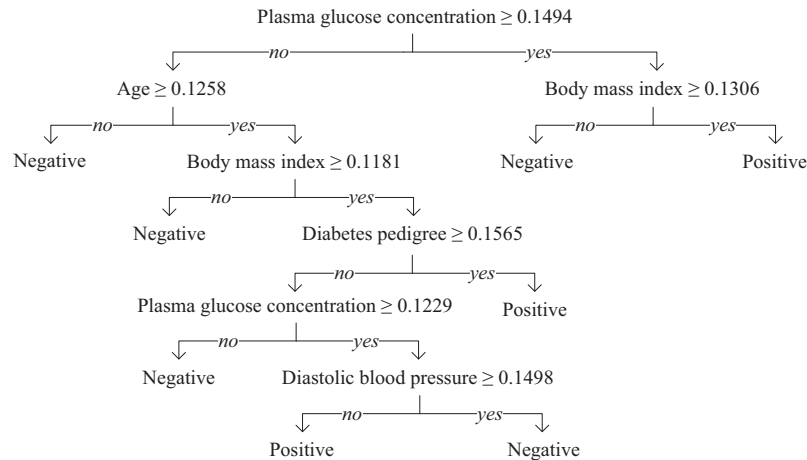


Fig. 5. The FMM-CART-RF decision tree for the Pima Indians Diabetes data set.

Table 5

Results of CART, FMM, FMM-CART and comparison of other methods for Liver Disorders data set.

Source	Method	Accuracy (%)	Ratio	Runs
Çomak et al. (2007)	LSSVM	60.00	9:1	–
	LSSVM with fuzzy weighting	94.29	9:1	–
Luukka (2009)	Liver FRPCA1	68.25	1:1	30
	Liver FRPCA2	67.88	1:1	30
	Liver FRPCA3	70.25	1:1	30
	Liver Original	63.09	1:1	30
	Liver PCA	66.50	1:1	30
Özşen & Güneş (2009)	AWAIS	70.17	9:1	10
	GA-AWAIS	85.21	9:1	10
Polat et al. (2007)	AIRS	81.00	9:1	30
	Fuzzy-AIRS	83.38	9:1	30
This study	FMM	63.78	1:1	30
	FMM	66.13	9:1	30
	FMM	67.25	4:1	30
	FMM-CART	91.00	1:1	30
	FMM-CART	91.71	9:1	30
	FMM-CART	92.61	4:1	30
	FMM-CART-RF	93.61	1:1	30
	FMM-CART-RF	93.62	4:1	30
	FMM-CART-RF	95.01	9:1	30

Table 6

Results of ROC for Liver Disorders data set.

Source	Method	ROC
Çomak et al. (2007)	LSSVM	0.336
	LSSVM with fuzzy weighting	0.950
This study	FMM	0.671
	FMM-CART	0.917
	FMM-CART-RF	0.955

that of FMM-CART-RF with 9:1 train-to-test ratio, but better than those from FMM-CART-RF with 1:1 and 4:1 train-to-test ratios.

The comparison of accuracy, sensitivity, and specificity rates for LSSVM, LSSVM with Fuzzy Weighting (Çomak et al., 2007) with FMM, FMM-CART, and FMM-CART-RF is shown in Fig. 6. The LSSVM with fuzzy weighting in Çomak et al. (2007) produced the best sensitivity and specificity results, i.e., 95.00% and 93.33%, respectively. However, both results were still slightly inferior to those from FMM-CART-RF, i.e., 96.54% sensitivity and 94.44% specificity. Notice that the sensitivity rate of LSSVM (Çomak et al., 2007) was 100%, but at the expense of a very low specificity rate. The comparison of the areas under the ROC curve is summarized in Table 6. The result in Çomak et al. (2007) for LSSVM with Fuzzy Weighting was slightly lower than that of FMM-CART-RF. The decision tree for the Liver Disorders data set is shown in Fig. 7.

The interpretation of the decision tree is as follows. Based on the “gammagt” feature, the tree starts to split. When the corresponding feature is smaller than 0.0612, splitting occurs by referring to the “sgpt” feature. When the corresponding feature is larger than or equal to 0.1017, FMM-CART-RF classifies the input as “Class 1”, otherwise splitting occurs. When the “alkphos” feature is larger than or equal to 0.1755, FMM-CART-RF classifies the input as “Class 1”, otherwise the “gammagt” feature is scrutinized. If the corresponding feature is larger than or equal to

phosphatase, (3) alamine aminotransferase, (4) aspartate aminotransferase, (5) gamma-glutamyl transpeptidase, and (6) number of half-pint equivalents of alcoholic beverages drunk per day.

Table 5 shows the overall the results. Again, the best result was yielded by FMM-CART-RF. The results in four publications with different train-to-test ratios are included in Table 5 for performance comparison. The LSSVM with fuzzy weighting in Çomak et al. (2007) ranked second, i.e., 94.29% accuracy, which was inferior to

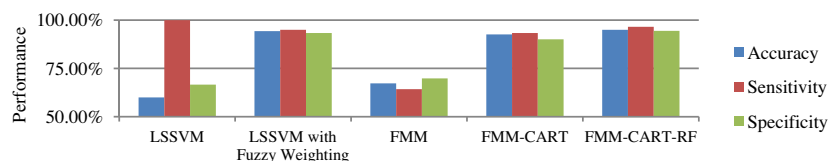


Fig. 6. Comparison of results for LSSVM (Çomak et al., 2007), LSSVM with Fuzzy Weighting (Çomak et al., 2007), FMM, FMM-CART, FMM-CART-RF for Liver Disorders data set.

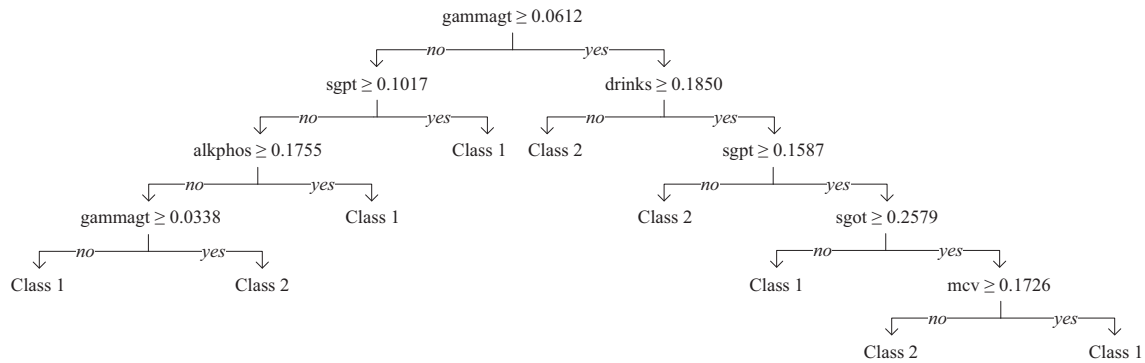


Fig. 7. The FMM-CART-RF decision tree for Liver Disorders data set.

0.0338, FMM-CART-RF classifies the input as “Class 2”, otherwise the input is predicted as “Class 1”. The other branch from the top of the tree (whereby the “gammagt” feature is larger than or equal to 0.0612) examines the “drinks” feature. When the corresponding feature is smaller than 0.1850, FMM-CART-RF classifies the input as “Class 2”, otherwise the “sgpt” feature is checked. When the corresponding feature is smaller than 0.1587, FMM-CART-RF classifies the input as “Class 2”, otherwise the tree splits again. When the “sgot” feature is smaller than 0.2579, FMM-CART-RF classifies the input as “Class 1”, otherwise the tree comes to another split. When the “mcv” feature is larger than or equal to 0.1726, FMM-CART-RF classifies the input as “Class 1”, otherwise the input is predicted as “Class 2”.

5. Conclusions

In this paper, a hybrid intelligent model, i.e., FMM-CART-RF, has been proposed for undertaking medical decision support tasks. A series of empirical studies using three benchmark medical data sets from the UCI Machine Learning Repository, namely Breast Cancer Wisconsin, Pima Indians Diabetes, and Liver Disorders, has been conducted to evaluate the efficacy of the hybrid model. Different experimental configurations have been adopted in order to provide a fair performance comparison with different models reported in the literature. The findings reveal that FMM-CART-RF is able to yield better performances in comparison with FMM and FMM-CART. Its results are also comparable with those from other methods published in the literature. From the empirical evaluation, it is evident that FMM-CART-RF is able to make accurate predictions with explanatory rules in the form of a decision tree for handling medical decision support problems.

The main contribution of this paper is the hybrid model that possesses three important characteristics for tackling medical decision support tasks, as follows:

- (i) Online learning whereby the FMM model is used to perform incremental learning from data samples with a one-pass training procedure. Learning in FMM is also able to avoid the catastrophic forgetting issue, i.e. current knowledge can be absorbed and added into the network structure (as hyperboxes) without corrupting or destroying previously learned knowledge (i.e. solving the stability-plasticity dilemma in learning systems). This is important to ensure a stable knowledge base can be established in the hybrid model for solving complex learning tasks, such as in medical diagnosis and prognosis applications.
- (ii) High performance whereby the RF method is applied to form an ensemble of CART models. Using the bagging principle, RF is able to formulate a diverse group of trees based on ran-

dom attribute selection. It is also useful for mitigating the effects of outliers and/or noise. This is important to ensure a robust and accurate hybrid model can be established and provide high performance in medical decision support tasks, whereby positive (through the sensitivity measure) patients could receive the necessary medical treatment and attention timely, while negative patients (through the specificity measure) could avoid going through the unnecessary tests that incur additional load on medical resources.

- (iii) Rule extraction whereby the CART model is employed to reveal knowledge captured in FMM into the decision tree format. The rules elucidated from CART are useful for explaining the prediction given by the hybrid model. This is particularly important in medical decision support tasks in where domain users (i.e. medical practitioners) not just would like to have an accurate prognostic or diagnostic prediction, but also would like to be convinced that the prediction is based on reasonable justifications; hence accepting the use of computerized intelligent systems in clinical practice.

Further work will focus on improving the robustness of FMM-CART-RF for real-world implementation. In addition to providing explanatory rules, the capability of handling missing data is another important aspect in medical applications (Lim, Leong, & Kuan, 2005). As such, it is beneficial to equip FMM-CART-RF with a robust method to overcome missing data problems in real-world environments. On the other hand, to facilitate the usage in a stand-alone mode in various locations, a portable hardware device that has interfaces with different medical sensors for data acquisition can be developed, with FMM-CART-RF incorporated into the device for real-time medical diagnosis. In this case, FMM-CART-RF will not be limited to clinics or hospitals, but can be used in field studies. The provision of a portable decision support tool is useful for medical practitioners to use the tool to treat patients in different locations, including remote areas. Besides that, to further realize widespread applicability of the hybrid model as a decision support tool, the benefits of telemedicine and telecare should be exploited (Coates, Clarke, Davison, & Patterson, 2012). As such, a fully functional web-based platform to house FMM-CART-RF can be developed. The web-based platform will allow diagnosis or prognosis to be conducted without limiting the physical location of the patient (subject to the availability of the Internet). In addition, the principle of transfer learning (Pan & Yang, 2010) can be investigated to ensure portability of the hybrid model. Transfer learning aims to train a machine learning model in a source domain, and use it to solve problems in a different, but related target domain by adapting knowledge acquired from the source domain (Pan & Yang, 2010). With transfer learning, medical knowledge learned

from one site can be transferred and utilized in another site; hence saving time and cost in developing medical decision support systems for different sites. Finally, as FMM-CART-RF is a generic model, its application is not restricted to medical problems. As such, it would be interesting to examine the efficacy of FMM-CART-RF in other domains whereby the rule extraction capability also plays an important role (Yap, Lim, & Au, 2011), e.g. to detect incipient faults of complex processes in a power generation plant.

References

- Abe, H., Ohsaki, M., Yokoi, H., & Yamaguchi, T. (2006). Implementing an integrated time-series data mining environment based on temporal pattern extraction methods: A case study of an interferon therapy risk mining for chronic Hepatitis. *New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science*, 4012, 425–435.
- Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems*, 8(6), 373–389.
- Aung, M. S. H., Lisboa, P. J. G., Etchells, T. A., Testa, A. C., Van Calster, B., Van Huffel, S., et al. (2007). Comparing analytical decision support models through boolean rule extraction: A case study of ovarian tumour malignancy. *Advances in Neural Networks, Lecture Notes in Computer Science*, 4492, 1177–1186.
- Azar, A. T. (2013). Fast neural network learning algorithms for medical applications. *Neural Computing and Applications*, 23(3–4), 1019–1034.
- Camps-Valls, G., Soria-Olivas, E., Perez-Ruixo, J. J., Perez-Cruz, F., Artes-Rodriguez, A., & Jimenez-Torres, N. V. (2007). Therapeutic drug monitoring of kidney transplant recipients using profiled support vector machines. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3), 359–372.
- Carmona, C. J., González, P., Del Jesus, M. J., Navío-Acosta, M., & Jiménez-Trevino, L. (2010). Evolutionary fuzzy rule extraction for subgroup discovery in a psychiatric emergency department. *Soft Computing*, 15(12), 2435–2448.
- Carpenter, G. A., & Tan, A. (1995). Rule Extraction: from neural architecture to symbolic representation. *Connection Science*, 7, 3–27.
- Cascio, D., Fauci, F., Magro, R., Raso, G., Bellotti, R., De Carlo, F., et al. (2006). Mammogram segmentation by contour searching and mass lesions classification with neural network. *IEEE Transactions on Nuclear Science*, 53(5), 2827–2833.
- Chabat, F., Hansell, D. M., & Yang, G. (2000). Computerized decision support in medical imaging. *IEEE Engineering in Medicine and Biology Magazine*, 19(5), 89–96.
- Coates, S., Clarke, A., Davison, G., & Patterson, V. (2012). Tele-EEG in the UK: A report of over 1000 patients. *Journal of Telemedicine and Telecare*, 18(5), 243–246.
- Çomak, E., Polat, K., Güneş, S., & Arslan, A. (2007). A new medical decision making system: Least square support vector machine (LSSVM) with fuzzy weighting pre-processing. *Expert Systems with Applications*, 32(2), 409–414.
- Coyle, D., Prasad, G., & McGinnity, T. M. (2009). Faster self-organizing fuzzy neural network training and a hyperparameter analysis for a brain–Computer Interface. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(6), 1458–1471.
- Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2), 113–127.
- Downs, J., Harrison, R. F., Kennedy, R. L., & Cross, S. S. (1996). Application of the fuzzy ARTMAP neural network model to medical pattern classification tasks. *Artificial Intelligence in Medicine*, 8(4), 403–428.
- Economou, G. P. K., Goumas, P. D., & Spiropoulos, K. (1996). A novel medical decision support system. *Computing & Control Engineering Journal*, 7(4), 177–183.
- Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, 7(1), 1–26.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Exarchos, T. P., Tzallas, A. T., Fotiadis, D. I., Konitsiotis, S., & Giannopoulos, S. (2006). EEG transient event detection and classification using association rules. *IEEE Transactions on Information Technology in Biomedicine*, 10(3), 451–457.
- Fakih, S. J., & Das, T. K. (2006). LEAD: A methodology for learning efficient approaches to medical diagnosis. *IEEE Transactions on Information Technology in Biomedicine*, 10(2), 220–228.
- Falco, I. D. (2013). Differential evolution for automatic rule extraction from medical databases. *Applied Soft Computing*, 13(2), 1265–1283.
- Franco, L., Subirats, J. L., Molina, I., Alba, E., & Jerez, J. M. (2007). Early breast cancer prognosis prediction and rule extraction using a new constructive neural network algorithm. *Computational and Ambient Intelligence, Lecture Notes in Computer Science*, 4507, 1004–1011.
- Ghosh-Dastidar, S., Adeli, H., & Dadmehr, N. (2008). Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering*, 55(2), 512–518.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). MA, USA: Morgan Kaufmann.
- Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). *Building expert systems*. London: Addison-Wesley.
- Held, C. M., Heiss, J. E., Estevez, P. A., Perez, C. A., Garrido, M., Algarin, C., et al. (2006). Extracting fuzzy rules from polysomnographic recordings for infant sleep classification. *IEEE Transactions on Biomedical Engineering*, 53(10), 1954–1962.
- Huang, K., Yang, H., King, I., & Lyu, M. R. (2006). Maximizing sensitivity in medical diagnosis using biased minimax probability machine. *IEEE Transactions on Biomedical Engineering*, 53(5), 821–831.
- Isola, R., Carvalho, R., & Tripathy, A. K. (2012). Knowledge discovery in medical systems using differential diagnosis, LAMSTAR, and k-NN. *IEEE Transactions on Information Technology in Biomedicine*, 16(6), 1287–1295.
- Kahramanli, H., & Allahverdi, N. (2009). Extracting rules for classification problems: AIS based approach. *Expert Systems with Applications*, 36(7), 10494–10502.
- Karayannis, N. B., Mukherjee, A., Glover, J. R., Ktonas, P. Y., Frost, J. D., Hrachovy, R. A., et al. (2006). Detection of pseudosinusoidal epileptic seizure segments in the neonatal EEG by cascading a rule-based algorithm with a neural network. *IEEE Transactions on Biomedical Engineering*, 53(4), 633–641.
- Kinney, W. C. (2003). Web-based clinical decision support system for triage of vestibular patients. *Otolaryngology-Head and Neck Surgery*, 128(1), 48–53.
- Kondo, C., Kondo, T., & Ueno, J. (2009). Three-dimensional medical image analysis of the heart by the revised GMDH-type neural network self-selecting optimum neural network architecture. *Artificial Life and Robotics*, 14(2), 123–128.
- Kondo, T., Kondo, C., Takao, S., & Ueno, J. (2010). Feedback GMDH-type neural network algorithm and its application to medical image analysis of cancer of the liver. *Artificial Life and Robotics*, 15(3), 264–269.
- Kovalerchuk, B., Vityaev, E., & Ruiz, J. F. (2000). Consistent knowledge discovery in medical diagnosis. *IEEE Engineering in Medicine and Biology Magazine*, 19(4), 26–37.
- Kwiatkowska, M., Atkins, M. S., Ayas, N. T., & Ryan, C. F. (2007). Knowledge-based data analysis: First step toward the creation of clinical prediction rules using a new typicality measure. *IEEE Transactions on Information Technology in Biomedicine*, 11(6), 651–660.
- Lim, C. P., Leong, J. H., & Kuan, M. M. (2005). A hybrid neural network system for pattern classification tasks with missing features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4), 648–653.
- Lim, J. S. (2009). Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system. *IEEE Transactions on Neural Networks*, 20(3), 522–527.
- Lin, H. (2008). Identification of spinal deformity classification with total curvature analysis and artificial neural network. *IEEE Transactions on Biomedical Engineering*, 55(1), 376–382.
- Loo, C. K. (2005). Accurate and reliable diagnosis and classification using probabilistic ensemble simplified fuzzy ARTMAP. *IEEE Transactions on Knowledge and Data Engineering*, 17(11), 1589–1593.
- Luukka, P. (2009). Classification based on fuzzy robust PCA algorithms and similarity classifier. *Expert Systems with Applications*, 36(4), 7463–7468.
- Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, 38(4), 4600–4607.
- Ma, Y. x., & Wang, S.-g. (2012). The application of artificial neural network in medical meteorology. *Future Control and Automation, Lecture Notes in Electrical Engineering*, 172, 237–242.
- Markowska-Kaczmar, U., & Matkowski, R. (2006). Experimental study of evolutionary based method of rule extraction from neural networks in medical data. *Applications in Medicine, Web Mining, Marketing, Image and Signal Mining, Lecture Notes in Computer Science*, 4065, 76–90.
- Meesad, P., & Yen, G. G. (2003). Combined numerical and linguistic knowledge representation and its application to medical diagnosis. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 33(2), 206–222.
- Newman, D. J., Hettich, S., Blake, C. L., & Merz, C. J. (2007). UCI Repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science. <http://www.ics.uci.edu/mllearn/MLRepository.html>.
- Örkücü, H. H., & Bal, H. (2011). Comparing performances of backpropagation and genetic algorithms in the data classification. *Expert Systems with Applications*, 38(4), 3703–3709.
- Osareh, A., Shadgar, B., & Markham, R. (2009). A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. *IEEE Transactions on Information Technology in Biomedicine*, 13(4), 535–545.
- Özgen, S., & Güneş, S. (2009). Attribute weighting via genetic algorithms for attribute weighted artificial immune system (AWAIS) and its application to heart disease and liver disorders problems. *Expert Systems with Applications*, 36(1), 386–392.
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.
- Papageorgiou, E. I. (2011). A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. *Applied Soft Computing*, 11(1), 500–513.
- Pavlopoulos, S. A., & Delopoulos, A. (1999). Designing and implementing the transition to a fully digital hospital. *IEEE Transactions on Information Technology in Biomedicine*, 3(1), 6–19.
- Picot, A., Charbonnier, S., & Caplier, A. (2012). On-line detection of drowsiness using brain and visual information. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 42(3), 764–775.
- Polat, K., Şahan, S., Kodaz, H., & Güneş, S. (2007). Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism. *Expert Systems with Applications*, 32(1), 172–183.

- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3), 21–45.
- Quteishat, A., & Lim, C. P. (2008). A modified fuzzy min–max neural network with rule extraction and its application to fault detection and classification. *Applied Soft Computing*, 8(2), 985–995.
- Rao, V. S. H., & Kumar, M. N. (2012). A new intelligence-based approach for computer-aided diagnosis of dengue fever. *IEEE Transactions on Information Technology in Biomedicine*, 16(1), 112–118.
- Ribeiro, M. X., Traina, C., Traina, C., & Azevedo-Marques, P. M. (2008). An association rule-based method to support medical image diagnosis with efficiency. *IEEE Transactions on Multimedia*, 10(2), 277–285.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), 93–104.
- Saad, E. W., & Wunsch, D. C. (2007). Neural network explanation using inversion. *Neural Networks*, 20, 78–93.
- Sanchez, J. C., Erdogmus, D., Nicoletis, M. A. L., Wessberg, J., & Principe, J. C. (2005). Interpreting spatial and temporal neural activity through a recurrent neural network brain-machine interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(2), 213–219.
- Seera, M., Lim, C. P., Ishak, D., & Singh, H. (2012). Fault detection and diagnosis of induction motors using motor current signature analysis and a hybrid FMM-CART model. *IEEE Transactions on Neural Networks and Learning Systems*, 23(1), 97–108.
- Sekar, B. D., Dong, M.-C., Shi, J., & Hu, X. Y. (2012). Fused hierarchical neural networks for cardiovascular disease diagnosis. *IEEE Sensors Journal*, 12(3), 644–650.
- Srinivasan, V., Eswaran, C., & Sriraam, N. (2007). Approximate entropy-based epileptic eeg detection using artificial neural networks. *IEEE Transactions on Information Technology in Biomedicine*, 11(3), 288–295.
- Stoean, R., & Stoean, C. (2013). Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. *Expert Systems with Applications*, 40(7), 2677–2686.
- Suzuki, K., Abe, H., MacMahon, H., & Doi, K. (2006). Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). *IEEE Transactions on Medical Imaging*, 25(4), 406–416.
- Tsipouras, M. G., Exarchos, T. P., Fotiadis, D. I., Kotsia, A. P., Vakalis, K. V., Naka, K. K., et al. (2008). Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling. *IEEE Transactions on Information Technology in Biomedicine*, 12(4), 447–458.
- Tsipouras, M. G., Voglis, C., & Fotiadis, D. I. (2007). A framework for fuzzy expert system creation—Application to cardiovascular diseases. *IEEE Transactions on Biomedical Engineering*, 54(11), 2089–2105.
- Tu, Z., S. Zheng, Yuille, A.L., Reiss, A.L., Dutton, R.A., Lee, A.D., Galaburda, A.M., Dinov, I., Thompson, P.M., & Toga, A.W. (2007). Automated extraction of the cortical sulci based on a supervised learning approach. In *IEEE transactions on medical imaging* (Vol. 26, No. 4, pp. 541–552).
- Walczak, S. (2005). Artificial neural network medical decision support tool: predicting transfusion requirements of ER patients. *IEEE Transactions on Information Technology in Biomedicine*, 9(3), 468–474.
- Wang, L., & Fu, X. (2005). *Data mining with computational intelligence*. Berlin: Springer-Verlag.
- Wang, L., Chu, F., & Xie, W. (2007). Accurate cancer classification using expressions of very few genes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(1), 40–53.
- Wu, Q., Ye, Y., Liu, Y., & Ng, M. K. (2012). SNP selection and classification of genome-wide SNP data using stratified sampling random forests. *IEEE Transactions on NanoBioscience*, 11(3), 216–227.
- Yap, K. S., Lim, C. P., & Au, M. T. (2011). Improved GART neural network model for pattern classification and rule extraction with application to power systems. *IEEE Transactions on Neural Networks*, 22(12), 2310–2323.
- Yohannes, Y., & Webb, P. (1999). *Classification and regression trees*. USA: International Food Policy Research Institute.
- Zhang, J., Zulkernine, M., & Haque, A. (2008). Random-forests-based network intrusion detection systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(5), 649–659.