Staging of Cervical Cancer with Soft Computing

Pabitra Mitra, Sushmita Mitra*, *Member, IEEE*, and Sankar K. Pal, *Fellow, IEEE*

Abstract—**This paper describes a way of designing a hybrid decision support system in soft computing paradigm for detecting the different stages of cervical cancer. Hybridization includes the evolution of knowledge-based subnetwork modules with genetic algorithms (GA's) using rough set theory and the Interactive Dichotomizer 3 (ID3) algorithm. Crude subnetworks obtained via rough set theory and the ID3 algorithm are evolved using GA's. The evolution uses a restricted mutation operator which utilizes the knowledge of the modular structure, already generated, for faster convergence. The GA tunes the network weights and structure simultaneously. The aforesaid integration enhances the performance in terms of classification score, network size and training time, as compared to the conventional multilayer perceptron. This methodology also helps in imposing a structure on the weights, which results in a network more suitable for extraction of logical rules and human interpretation of the inferencing procedure.**

Index Terms—**Classification, genetic algorithms, ID3, knowledge-based networks, medical decision support system, modular neural networks, rough sets, rule extraction.**

I. Introduction

THE WORLD-WIDE occurrence of the cancer cervix cases show [1] that only 20% of these cases occur in the developed nations while 80% of the cases are found in the developing countries that include India [2]. In India, the cancer of the uterine cervix is the most frequent malignancy observed in females, as per the reports of the different cancer registries published in the Indian Council of Medical Research (ICMR) biennial report [3].

The medical records of the Chittaranjan National Cancer Institute (CNCI), Calcutta, India, for the last five years show that the commonest malignancy observed in the females is cancer of the cervix, which comprises about 40% of the total female cases diagnosed. Cervical cancer has a more or less well-defined treatment modality. Treatment modalities available for this particular type of malignancy are surgery, radiotherapy and chemotherapy or a combination of all these, depending on the stage of the disease at the time of diagnosis and also the physical condition of the patient during treatment.

Staging is a process that uses information learned about cancer through diagnostic processes, such as the size of the tumor, how deeply the tumor has invaded tissues at the site of the origin, the extent of any invasion into surrounding organs, and the extent of metastasis (spread) to lymph nodes or distant organs. This is a very important process because proper staging is the most important factor in selecting the right treatment plan. Cervical cancer is most frequently staged using the International Federation of Gynecology and Obstetrics (FIGO) system of staging. This system classifies the disease in Stages I–IV.

With the advent of various new modalities of treatment in the field of cancer, the decision making toward a particular treatment regime to be adopted for each individual patient has become a complex process and should keep pace with the advancement of medical sciences. More often, there is a large amount of information to be processed, much of which is quantifiable. Intuitive thought processes involve rapid unconscious data processing and combine available information by law of average and consequently have a low intraperson and interperson consistency. So from the point of intuitive decision making, the clinician of today should move toward analytic decision making which, though typically slow, is conscious and consistent and clearly spells out the basis of the decisions.

In the field of oncology, the decision making is not only restricted to finding out the correct diagnosis and planning of the proper treatment modality but also to take cognizance of the factors like the patient's socio-economic background, his or her ability to pursue a prolonged and expensive treatment program, and also whether the ill effects of the treatment modalities will outweigh its efficacy. Unlike other diseases, the comprehensive cancer treatment involves not only the above mentioned treatment interventions at the onset of the disease, but a lifetime follow-up information on each patient is essential to prevent recurrence and to calculate the disease-free survival necessary to evaluate any treatment efficacy. If a computerized program could be developed taking all these factors into consideration, that would help in the analytical decision making toward treatment and other related parameters, it will go a long way to make the task of a practicing oncologist much easier.

The objective of this article is to design a medical decision support system, for cancer management, using a knowledge-based network in combination with rough set theory and genetic algorithms (GA's) in soft computing paradigm. The proposed system is able to exploit the parallelism, self-learning, and fault tolerance characteristics of artificial neural network (ANN) models, knowledge encoding capabilities of rough set theory, and the adaptive, parallel and robust searching characteristics of GA's. The model is built on the data of cancer of uterine cervix for detecting its various stages.

Relevance of Soft Computing: Many of the early efforts to apply artificial intelligence to medical reasoning problems, have primarily used rule-based systems [4]. Such programs are typically easy to create, because their knowledge is cataloged in the form of *if-then* rules. In relatively well-constrained domain such programs show skilled behavior. But in real-life situations, there is considerable degradation of performance due to both presence of ambiguity and incomplete information as well as inadequate modeling of the diseases by the rules. Other conventional methods like Bayes classifier and flow charts are also unable to deal with most complex clinical decision making problems. It is necessary to provide health care researchers from different domains with software tools combining different machine learning models, that can extract information from large imperfect data sets, without manual intervention. There are currently large repositories of data in several domains of health care. These data may support clinical, bibliographic, administrative, or epidemiological studies.

Soft Computing is a consortium of methodologies which works synergetically and provides, in one form or another, flexible information processing capability for handling real life ambiguous situations [5]. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness and low-cost solutions. There are ongoing efforts to integrate ANN's with fuzzy set theory, rough set theory, and GA's and other methodologies in soft computing paradigm [5], [6].

ANN's are highly parallel connectionist systems modeled on biological neurons. ANN's are now widely used for medical decision support as they have the capacity to model highly nonlinear data distributions [7], [8]. ANN's generally consider a fixed topology of neurons connected by links in a pre-defined manner. Recently, there have been some attempts in improving the efficiency of neural computation by using knowledge-based nets. These [9] constitute a special class of ANN's that consider crude domain knowledge to generate the initial network architecture, which is later refined in the presence of training data. Recently, the theory of rough sets has been used to generate knowledge-based networks.

The theory of *rough sets* [10] has emerged as a major mathematical tool for managing uncertainty that arises from granularity in the domain of discourse, i.e., from the indiscernibility between objects in a set. The primary role of rough sets here is in managing uncertainty and extracting domain knowledge [11].

The Interactive Dichotomizer 3 (ID3) approach [12] to pattern recognition and classification consists of a procedure for synthesizing an efficient discrimination tree for classifying patterns that have nonnumeric attributes or feature values. The discrimination tree can also be expressed in the form of a body of rules and, because of this, ID3 is also often thought of as an inductive inference procedure for machine learning or rule acquisition. These rules can also be encoded to generate a knowledge-based network.

A recent trend in neural network design for large scale problems is to split the original task into simpler subtasks, and use a subnetwork module for each of the subtasks [13].

GA's are randomized search and optimization techniques guided by the principles of evolution and natural genetics [14]. They are efficient, adaptive, and robust search processes, producing near-optimal solutions and having a large amount of implicit parallelism. Many researchers have combined GA's with neural networks for building more powerful adaptive systems.

Design Approach: In the present article an evolutionary strategy is used in soft computing framework for designing a modular knowledge-based network using both rough set theory and ID3 algorithm for staging of cervical cancer. The evolutionary training algorithm generates the weight values for a parsimonious network. Rough set theory and the ID3 algorithm are used to obtain the sets of probable knowledge-based subnetworks which form the initial population of the GA. These modules are then integrated and evolved with a *restricted* mutation operator that helps preserve extracted localized rule structures as potential solutions.

The connection weights of the evolved modular knowledge-based network are used for extracting refined rules for the problem domain. This helps in minimizing human interaction and associated inherent bias during the phase of knowledge-base formation and also reduces the possibility of generating contradictory rules. The extracted rules help in alleviating the knowledge acquisition *bottleneck*, refining the initial domain knowledge, and providing reasoning and explanation facilities. One realizes that, specially in the medical domain, the final responsibility for any diagnostic decision always has to be accepted by the medical practitioner. So the doctor may want to verify the justification behind the decision reached, based on his/her expertise. This requires the system to be able to explain its mode of reasoning for any inferred decision/recommendation, preferably in rule form, to convince the user that its reasoning is correct. All these aspects serve to demonstrate the utility of automated rule extraction in a medical decision support system.

Encoding of domain knowledge reduces the search space of the network and speeds up convergence. Use of GA's helps in optimizing the network topology. This sort of automatic generation of network architecture is novel to our approach. It is unlike previously reported works using ANN's and related techniques [7], [8]. Classification performance of the models for different stages is compared with that of the conventional multilayer perceptron (MLP). The rules extracted are also verified by oncologists.

II. Knowledge-Based MLP

Knowledge is extracted using two methods: 1) Rough set theoretic concepts and 2) ID3 algorithm. The extracted crude domain knowledge is encoded among the connection weights of an MLP. This helps one to automatically generate an appropriate network architecture in terms of hidden nodes and links. The methods model arbitrary decision regions with multiple object representatives.

A. Rough MLP [15]

The feature space gives the condition attributes and the output classes the decision attributes, so as to result in a decision table. Rules are then generated from the table by computing relative reducts. The dependency factors of these rules are used to encode the initial connection weights of the resultant knowledge-based network. For further theoretical details, one may refer to the Appendix.

While designing the initial structure of the network, the union of the rules of the l classes is considered. The input layer consists of n attribute values while the output layer is represented by l classes. The hidden layer nodes model the conjuncts in the antecedent part of a rule. The output layer nodes model the disjuncts. For each conjunct, corresponding to one output class (one dependency rule), one hidden node is dedicated. Only those input attributes that appear in this conjunct are connected to the appropriate hidden node, which in turn is connected to the corresponding output node. Each disjunct is modeled at the output layer by joining the corresponding hidden nodes.

B. ID3 Algorithm

ID3 is effective when there is a body of data consisting of a large number of patterns, each of which is made up of a long list of nonnumeric feature/attribute values. The class membership of some of these patterns are known.

In the ID3 approach [12], we make use of labeled examples and determine how features might be examined in sequence until all the labeled examples have been classified correctly. In addition, the class membership depends on certain combinations of feature values, as discovered by the discrimination tree. These might also provide insight into the basic mechanism that determine the processes being examined. For further algorithmic details, see the Appendix.

III. Modular Knowledge-Based Network: Evolutionary Design

It is believed that the use of modular neural network (MNN) enables a wider use of ANN's for large scale systems. Embedding modularity (i.e., to perform local and encapsulated computation) into neural networks leads to many advantages compared to the use of a single network. For instance, constraining the network connectivity increases its learning capacity and permits its application to large scale problems [13]. It is easier to encode *a priori* knowledge in modular neural networks.

GA's are highly parallel and adaptive search processes based on the principles of natural selection [14]. Here we use GA's for evolving the weight values as well as the structure of the MLP used in the framework of modular neural networks. Unlike other theory refinement systems which train only the *best* network approximation obtained from the domain theories, the initial population here consists of all possible networks generated from rough set theoretic as well as ID3 rules. This is an advantage because potentially valuable information may be wasted by discarding the contribution of less successful networks at the initial level itself.

We use two phases. First an l-class classification problem is split into l two-class problems. Let there be l sets of subnetworks, with n inputs and one output node each. Rough set theoretic concepts and ID3 algorithm are used to encode domain knowledge into each of the subnetworks. The number of hidden nodes and connectivity of the knowledge-based subnetworks is automatically determined.

At the end of training, the modules/subnetworks corresponding to each two-class problem are concatenated to form an initial network for the second phase. The intermodule links are initialized to small random values as depicted in Fig. 1. A set of such concatenated networks forms the initial population of the GA. Note that the individual modules cooperate, rather than compete, with each other while evolving toward the final solution. The mutation probability for the intermodule links is set to a high value, while that of intra-module links is set to a relatively lower value. This sort of *restricted* mutation helps preserve some of the localized rule structures, already extracted and evolved, as potential solutions. The initial population for the GA of the entire network is formed from all possible combinations of these individual network modules and random perturbations about them.

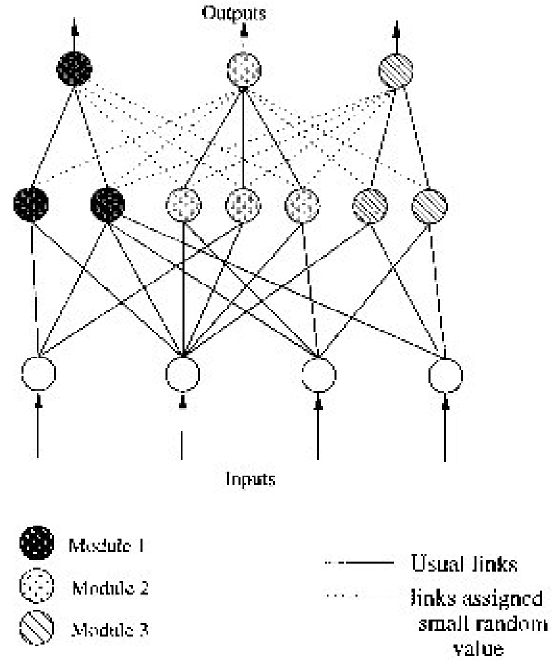

Fig. 1. Intramodule and intermodule links.

The GA used involves three procedures—encoding of the problem parameters in the form of binary strings, application of genetic operators like crossover and mutation, selection of individuals based on some objective function to create a new population. These are discussed as follows:

1) Chromosomal Representation: The problem variables consist of the weight values and the input/output fuzzification parameters. Each of the weights is encoded into a binary word of 16 bit length. An additional bit is assigned to each weight to indicate the presence or absence of the link. Thus a total of 17 bits are assigned for each weight.

2) Crossover and Mutation Operators: It is obvious that due to the large string length, single point crossover would have little effectiveness. Multiple point crossover is adopted to ensure a high probability for only one crossover point occurring within a word encoding a single weight. The crossover probability is fixed at 0.7. The search string being very large, the influence of mutation is more on the search. Each of the bits in the string is chosen to have some mutation probability (*pmut*). This mutation probability however has a spatio-temporal variation. This is done to ensure least alterations in the structure of the individual modules already evolved.

3) Choice of Fitness Function: An objective function of the form: $F = \alpha_1 f_1 + \alpha_2 f_2$, is chosen, where

$$f_1 = \frac{\text{no. of Correctly Classified Sample in Training Set}}{\text{Total no. of Samples in Training Set}}$$

and

$$f_2 = 1 - \frac{\text{no. of links present}}{\text{Total no. of links possible}}.$$

Here α_1 and α_2 determine the relative weightage of each of the factors.

4) Selection: Selection is done by the *roulette wheel* method. The probabilities are calculated on the basis of ranking of the individuals in terms of the objective function, instead of the objective function itself.

IV. Implementation and Results

The data consists of a set of 221 cervical cancer patient cases obtained from the database of the CNCI. Cross-validation of results is made with oncologists. There are four classes corresponding to the Stages I–IV of the cancer, each containing 19, 41, 139, and 19 patient cases, respectively. The features of the proposed model represents the presence or absence of the symptoms, and the signs observed upon physical examination. The 21 Boolean input features refer to *Vulva: healthy* $(Vu(h))$, *Vulva: lesioned* $(Vu(l))$, *Vagina: healthy* $(Va(h))$, *Vagina: spread to upper part* $(Va(u))$, *Vagina: spread to middle part* $(Va(m))$, *Vagina: spread to lower part* $(Va(l))$, *Cervix: healthy* $(Cx(h))$, *Cervix: eroded* $(Cx(e))$, *Cervix: small ulcer* $(Cx(su))$, *Cervix: ulcerative growth* $(Cx(u))$, *Cervix: proliferative growth* $(Cx(p))$, *Cervix: ulcero-proliferative growth* $(Cx(l))$, *Paracervix: free* $(PCx(f))$, *Paracervix: infiltrated* $(PCx(i))$, *Urinary bladder base: soft* $(BB(s))$, *Urinary bladder base: hard* $(BB(h))$, *Rectrovaginal septum: free* $(RVS(f))$, *Rectrovaginal septum: infiltrated* $(RVS(i))$, *Parametrium: free* $(Para(f))$, *Parametrium: spread, but not upto* $(Para(nu))$ and *Parametrium: spread upto* $(Para(u))$, respectively. Let the proposed methodology be termed Model S. The dependency rules used for knowledge encoding are obtained by two methods—using rough set theory and using the ID3 algorithm. The performance of the methodologies is compared with that of an ordinary MLP (termed Model O) trained using backpropagation (BP) with weight decay.

A. Knowledge Encoding and Classification

The dependency rules generated via rough set theory and ID3 algorithm, and used in the encoding scheme are provided in Tables I and II. Recognition scores obtained for each of the data by the proposed modular network (Model S) are then presented and compared. Here, 50% of the samples are used as training set and the network is tested on the remaining samples.

These extracted rules are encoded to generate the knowledge-based MLP (Model S). Table III demonstrates the performance of Model S, using knowledge encoding by both rough sets and ID3 algorithm. It is observed to be superior, in both cases, to that of Model O. This can be corroborated from Fig. 2.

B. Rule Extraction

Consider a simple heuristic for rule extraction. Let us define the following quantities: *Thres*$_1$ = mean of the weights > 0, *Thres*$_2$ = mean of the weights $>$ *Thres*$_1$, *Thres*$_3$ = mean of the weights $>$ *Thres*$_2$. We consider weights having value greater than *Thres*$_3$ as strong connections (plotted as thick lines in Fig. 3), weights having value between *Thres*$_2$ and *Thres*$_3$ as moderate links (plotted as normal lines in Fig. 3). We obtained *Thres*$_1$ $= 24.98$, *Thres*$_2$ $= 76.64$ and *Thres*$_3$ $= 85.09$. If the same set of threshold values are applied to Model O, no strong links are obtained. Hence, it is not possible to extract any crisp rules from it. On the other hand, the network obtained using the proposed Model S contains a number of strong links which can be used in extracting meaningful logical rules.

A sample set of refined rules extracted from the network, considering only the strong and moderate links, is presented below.

TABLE I
CRUDE RULES OBTAINED VIA ROUGH SET THEORY

I	Cx(su) ∨ Para(f), Cx(p) ∨ Para(f), Cx(su) ∨ Para(nu)
II	Va(h) ∨ Cx(u), Va(h) ∨ Cx(l),
	Va(u) ∨ Cx(u), Para(nu), Pcx(f)
III	Para(nu), Para(u), Va(u)
	(Va(u) ∧ Cx(u)) ∨ Cx(l) ∨ Va(m)
	(Va(h) ∧ Cx(u)) ∨ (Va(u) ∧ Cx(u)) ∨ Cx(l)
	(Va(u) ∧ Cx(p)) ∨ Va(m) ∨ Cx(l)
IV	(Va(l) ∧ Cx(u)) ∨ (Cx(u) ∧ Va(u)) ∨ (Va(l) ∧ Para(u))
	(Va(l) ∧ Cx(p)) ∨ Va(m).

TABLE II
CRUDE RULES EXTRACTED VIA THE ID3 ALGORITHM

I	Para(f) ∧ Va(h) ∧ Cx(u)
	Para(nu) ∧ Va(h) ∧ Cx(u) ∧ PCx(f) ∧ BB(s)
II	Va(u) ∧ Para(f) ∧ Cx(h)
	Va(h) ∧ Pcx(i) ∧ Cx(u) ∧ BB(s) ∧ Para(nu)
	Va(u) ∧ Para(nu) ∧ Cx(l) ∧ BB(s)
III	Va(h) ∧ Cx(l) ∧ Para(u)
	Para(u) ∧ Cx(u) ∧ PCx(i) ∧ BB(s)
	Va(u) ∧ Cx(u) ∧ Para(u)
IV	Va(l) ∧ Cx(u) ∧ Para(u)
	Va(m) ∧ BB(h) ∧ Cx(u) ∧ Para(u)
	Va(m) ∧ Cx(p) ∧ BB(h) ∧ Para(u).

TABLE III
COMPARATIVE CLASSIFICATION SCORES FOR DIFFERENT MODELS

Stage	Model O		Model S with knowledge encoding via			
			Rough set theory		ID3 algorithm	
	Train	Test	Train	Test	Train	Test
I(%)	65.00	64.70	65.00	64.70	90.00	89.71
II(%)	69.05	67.73	69.05	68.13	73.81	72.04
III(%)	93.66	93.01	94.13	90.02	90.14	90.02
IV(%)	42.11	40.09	44.21	41.87	42.11	40.19
Net(%)	80.97	79.23	81.02	79.52	82.74	80.02
# links	175		118		82	
Cycles	90		50		50	

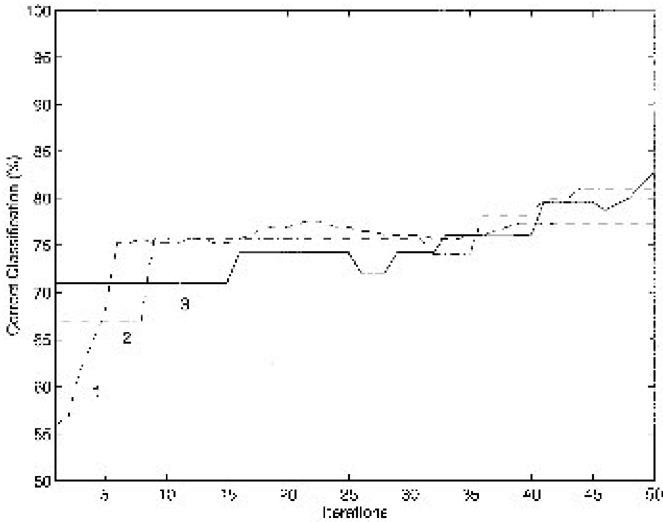

Fig. 2. Evolution of overall correct classification percentage with training sweeps. 1. Network trained with BP (Model O). 2. Model S with initial rule encoding via rough set theory, 3. Model S with initial rule encoding via ID3 algorithm.

For a network with initial weight encoding from the crude rules obtained by rough set theory

$$
\begin{aligned}
I &\leftarrow (Va(h) \wedge Para(f)) \\
&\quad \vee (Cx(h) \wedge Cx(u) \wedge BB(s)) \\
II &\leftarrow (PCx(f) \wedge PCx(i)) \vee Para(f) \vee Para(nu) \\
III &\leftarrow Va(h) \wedge Cx(u) \wedge Cx(l) \wedge Para(u) \\
IV &\leftarrow Va(m) \vee (Cx(u) \wedge Cx(p)) \\
&\quad \vee (Para(nu) \wedge Para(u)).
\end{aligned}
$$

For a network with initial weight encoding from the crude rules obtained by ID3 algorithm

$$
\begin{aligned}
I &\leftarrow Cx(u) \wedge Cx(l) \wedge PCx(f) \wedge BB(s) \\
&\quad \wedge Para(f) \wedge Para(nu) \\
II &\leftarrow Va(h) \wedge Va(u) \wedge Cx(p) \wedge Cx(l) \\
&\quad \wedge Para(nu) \wedge Para(u) \\
III &\leftarrow Va(h) \wedge Cx(u) \wedge Cx(l) \wedge PCx(i) \wedge BB(s) \\
IV &\leftarrow Va(u) \wedge Va(m) \wedge Va(l) \wedge Cx(u) \\
&\quad \wedge Cx(p) \wedge PCx(f) \wedge BB(h) \\
&\quad \wedge Para(nu) \wedge Para(u).
\end{aligned}
$$

Here, we provide the expertise obtained from oncologists. In Stage I, the cancer has spread from the lining of the cervix into the deeper connective tissue of the cervix. But it is still confined within the cervix. Stage II signifies the spread of cancer beyond the cervix to nearby areas like parametrial tissue, that are still inside the pelvic area. In Stage III, the cancer has spread to the lower part of the vagina or the pelvic wall. It may be blocking the ureters (tubes that carry urine from the kidneys to the bladder). Stage IV is the most advanced stage of cervical cancer. Now the cancer has spread to other parts of the body, such as rectum, bladder, or lungs. It may be mentioned that the rules generated by the proposed algorithm are validated by the experts' opinion.

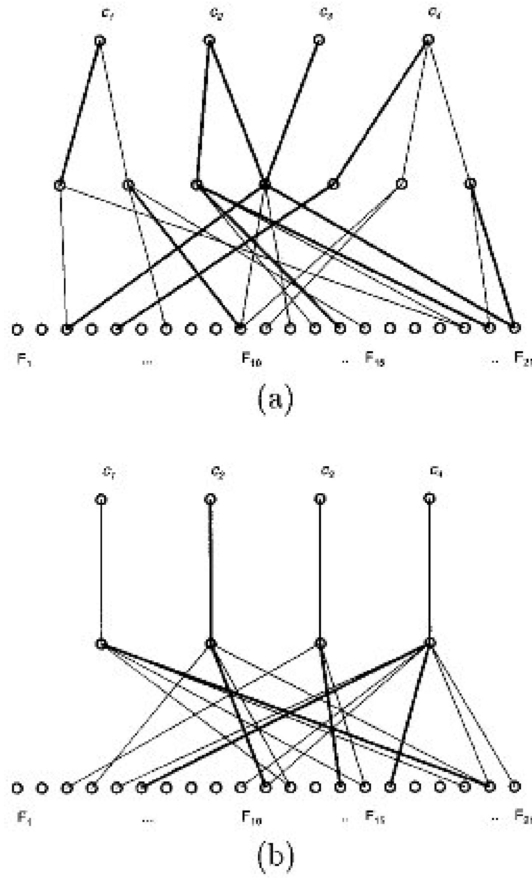

Fig. 3. Connectivity of the network obtained for the data, using Model S, with initial rule encoding via (a) rough set theory, and (b) ID3 algorithm.

Classification performances of the different methodologies suggested do not show significant difference as is evident from Table III. However, the process of knowledge encoding and structured training was successful in imposing a structure on the weight values. It can be seen from Fig. 4(a) that a large number of connections are absent in the proposed knowledge-based model, whereas the links present are very strong. On the other hand, Fig. 4(b) shows that network without knowledge encoding has a larger number of links with nonzero values, that are almost uniformly distributed. This can be interpreted as follows: the proposed methodology results in a sparse network having stronger links, whereas ordinary MLP results in a dense network with moderate and weak links. Hence, the knowledge-based network is better suited for the extraction of crisp and more interpretable rules. This is an advantage in the medical domain, where explanation of the results obtained is required to be available for examination by clinical practioners.

The performance of the popular C4.5 machine learning system [16] (based on the ID3 algorithm) on the data set was also studied as a benchmark. The program gave classification scores of 81.5% on training data and 80.2% on test data. These are comparable with the results presented in Table III. Sample rules generated by C4.5 are

$$
\begin{aligned}
I &\leftarrow Va(h) \wedge PCx(f) \wedge Para(f) \\
II &\leftarrow Para(f)
\end{aligned}
$$

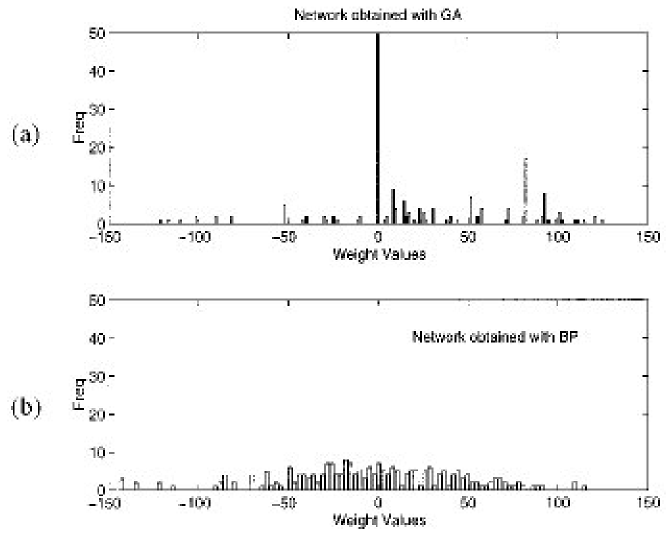

Fig. 4. Distribution of the weight values for a network (a) with knowledge encoding via rough set and (b) without knowledge encoding.

$$II \leftarrow BB(s)$$
$$III \leftarrow BB(s) \wedge Para(u).$$

Note that the rules obtained using C4.5 are significantly poorer than those obtained by the proposed methodology for the knowledge-based network. This is due to the fact that only statistically significant instances of the stages are represented in the rules by C4.5. On the other hand, in the proposed model the rare patient cases are also preserved and incorporated into the network in the process of knowledge encoding and structured training. This leads to a more complete rule base.

V. Conclusion and Discussion

The steady expansion of medical knowledge has made it difficult for the physician to remain abreast of medicine outside a narrow field. Consultation with the specialist is a solution when the clinical problem lies beyond the physician's competence. However, frequently expert opinion is either unavailable or not available in a timely fashion. This is an acute problem in medical management, specially when the number of patients is very high. With increasing cost of medical resources exhaustive clinical tests like PAP smear are becoming exorbitantly costly, and doctors often have to rely on less costly examinations for clinical assessment and subsequent management. Moreover, if the patients come from widely varying socio-economic background, medical practitioners often have access to only incomplete or ambiguous information about the patient's history. This introduces a large amount of uncertainty in the inferencing procedure.

In recent years, a number of sophisticated machine learning methodologies have been developed using ANN's, fuzzy sets, and rough sets, and successfully used to solve several complex real life problems. Such soft computing tools, when used in conjunction, have the properties of intolerance to uncertainties, robustness, and low cost. We suggest one such integrated methodology for detecting the different stages of cervical cancer using a modular knowledge-based network with GA's.

The algorithm presented here involves synthesis of several MLP modules, encoding rules generated by 1) rough set theory and 2) ID3 algorithm, for each class. These knowledge-based modules are refined using a GA. The genetic operators are implemented in such a way that they help preserve the modular structure already evolved. It is seen that this methodology along with modular network decomposition results in superior performance in terms of classification score, training time, and network sparseness; thereby enabling easier extraction of rules which are more interpretable to practitioners.

The novelty of our model is as follows: a) We have used rough sets and ID3 algorithm to encode initial domain knowledge. This helps in reducing the search space and speeding convergence. A crude network topology can be automatically determined. b) Use of GA's helps in refining the network architecture while preserving the initially encoded localized structures. c) The modular approach helps in speeding up the whole procedure.

Note that the data for cervical cancer considered here is rather simple. This makes the classification performance of the proposed method comparable to that of the conventional MLP. However in the case of more complicated data with linearly nonseparable decision regions or overlapping classes the superiority of the proposed method becomes more evident. The present investigation not only provides a decision support system for Cervical Cancer management, but also demonstrates a way how different *soft computing* tools like neural networks, GA's and rough set theory can be integrated in order to build an efficient decision making system for pattern classification and rule generation.

Appendix

Rule Generation with Rough Sets: Let $\mathcal{S} = \langle U, A \rangle$ be a decision table, with C and $D = \{d_1, \ldots, d_l\}$ its sets of condition and decision attributes, respectively. Divide the decision table $\mathcal{S} = \langle U, A \rangle$ into l tables $\mathcal{S}_i = \langle U_i, A_i \rangle$, $i = 1, \ldots, l$ corresponding to the l decision attributes $d_1, \ldots, d_l$, where, $U = U_1 \cup \cdots \cup U_l$ and $A_i = C \cup \{d_i\}$.The size of each $\mathcal{S}_i$ $(i = 1, \ldots, l)$ is first reduced with the help of a threshold on the number of occurrences of the same pattern of attribute values.

Consider the case of class c_k in the l-class problem domain. Let there be m sets $O_1, \ldots, O_m$ of objects in the table having identical attribute values, and *card* $(O_i) = n_{k_i}$, $i = 1, \ldots, m$, such that $n_{k_1} \geq \ldots \geq n_{k_m}$ and $\sum_{i=\_}^{m} n_{k_i} = n_k$. The attribute-value table can now be represented as an $m \times n$ array. Let $n_{k'_1}, n_{k'_2}, \ldots, n_{k'_m}$ denote the distinct elements among $n_{k_1}, \ldots, n_{k_m}$ such that $n_{k'_1} > n_{k'_2} > \cdots > n_{k'_m}$. Let a heuristic threshold function Tr be defined so that all entries having frequency less than Tr are eliminated, resulting in a reduced attribute-value table.

ID3 Model: ID3 uses an information-theoretic approach. The procedure is that at any point we examine the feature that provides the greatest gain in information or, equivalently, the greatest decrease in entropy. Entropy is defined as $-p \log_2 p$, where probability p is determined on the basis of frequency of occurrence.

The general case is that of N labeled patterns partitioned into sets of patterns belonging to classes c_i, $i = 1, 2, 3, \ldots, l$. The population in class c_i is n_i. Each pattern has n features and each feature has J (≥ 2) values. The ID3 prescription for synthesizing an efficient decision tree can be stated as follows [12]:

Step 1) Calculate initial value of entropy

$$Entropy(I) = \sum_{i=1}^{l} -(n_i/N)\,\log_2(n_i/N) = \sum_{i=1}^{l} -p_i\,\log_2 p_i.$$

Step 2) Select that feature which results in the maximum decrease in entropy (gain in information), to serve as the root node of the decision tree.

Step 3) Build the next level of the decision tree providing the greatest decrease in entropy.

Step 4) Repeat Steps 1–3. Continue the procedure until all subpopulations are of a single class and the system entropy is zero.

At this stage, one obtains a set of leaf nodes (subpopulations) of the decision tree, where the patterns are of a single class. There can be some nodes which cannot be resolved any further. Now we describe the knowledge encoding algorithm using the decision tree generated by ID3. Let us consider the leaf nodes only. The path from the root to a leaf can be traversed to generate the rule corresponding to a pattern from that class. In this manner one obtains a set of rules for all the pattern classes, in the form of intersection of the features/attributes encountered along the traversal paths.

References

[1] L. A. Brinton, "Epidemiology of cervical cancer—Overview, the epidemiology of cervical cancer and human papilloma virus," *IACR*, vol. 119, pp. 4–23, 1992.

[2] K. Jayant, R. S. Rao, B. M. Nene, and P. S. Dale, "Improved stage at diagnosis of cervical cancer with increased cancer awareness in a rural Indian population," *Int. J. Cancer*, vol. 63, pp. 161–163, 1995.

[3] "National Cancer registry programme," ICMR, Chennai, India, Biennial Rep., 1988–89.

[4] P. Szolovits, R. S. Patil, and W. B. Schwatrz, "Artificial intelligence in medical diagnosis," *Ann. Internal Med.*, vol. 108, pp. 80–87, 1988.

[5] L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Commun. ACM*, vol. 37, pp. 77–84, 1994.

[6] S. K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*. New York: Wiley, 1999.

[7] C. Perneel, J. M. Themlin, J. M. Renders, and M. Acheroy, "Optimization of fuzzy expert systems using genetic algorithms and neural networks," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 300–312, **<month?>** 1995.

[8] K. Tumer, N. Ramanujam, J. Ghosh, and R. Richards-Kortum, "Spectroscopic detection of cervical pre-cancer through radial basis function networks," *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 953–962, Aug. 1998.

[9] G. G. Towell and J. W. Shavlik, "Knowledge-based artificial neural networks," *Artif. Intell.*, vol. 70, pp. 119–165, 1994.

[10] Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning About Data*. Dordrecht: Kluwer Academic, 1991.

[11] S. K. Pal and A. Skowron, Eds., *Rough Fuzzy Hybridization: New Trends in Decision Making*. Singapore: Springer Verlag, 1999.

[12] Y. H. Pao, *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley, 1989.

[13] B. M. Happel and J. J. Murre, "Design and evolution of modular neural network architectures," *Neural Networks*, vol. 7, pp. 985–1004, 1994.

[14] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.

[15] M. Banerjee, S. Mitra, and S. K. Pal, "Rough fuzzy MLP: Knowledge encoding and classification," *IEEE Trans. Neural Networks*, vol. 9, pp. 1203–1216, June 1998.

[16] J. R. Quinlan, *C4.5, Programs for Machine Learning*. San Diego, CA: Morgan Kauffman, 1993.

Pabitra Mitra received the B.Tech. degree in electrical engg. in 1996, from the Indian Institute of Technology, Kharagpur, India.

He worked as a Scientist at Centre for Artificial Intelligence and Robotics, Bangalore, India, and is currently a Junior Research Fellow at the Machine Intelligence Unit, Indian Statistical Institute, Calcutta, India. His research interests are in the fields of data mining, knowledge discovery, medical expert systems, and learning.

Sushmita Mitra (M'99) received the B.Sc. degree (with honors) in physics and the B.Tech. and M.Tech. degrees in computer science from the University of Calcutta, Calcutta, India, in 1984, 1987, and 1989, respectively, and Ph.D. degree in computer science from Indian Statistical Institute, Calcutta, India, in 1995.

During 1992–1994, she was with the European Laboratory for Intelligent Techniques Engineering, Aachen, Germany, as a German Academic Exchange Service (DAAD) Fellow. Since 1995, she is an Associate Professor of the Indian Statistical Institute, where she joined in 1989. She was a Visiting Professor at Meiji University, Japan in 1999. She has co-authored *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing Paradigm* (New York: Wiley, 1999). Her research interests include pattern recognition, fuzzy sets, artificial intelligence, neural networks, and soft computing.

Dr. Mitra was a recipient of the National Talent Search Scholarship (1978–83) from the National Council for Educational Research and Training, India, the IEEE TNN Outstanding Paper Award in 1994 and CIMPA-INRIA-UNESCO Fellowship in 1996.

Sankar K. Pal (M'81–SM'84–F'93) received the M.Tech. and Ph.D. degrees in radio physics and electronics in 1974 and 1979, respectively, from the University of Calcutta, Calcutta, India. In 1982, he also received the Ph.D. degree in electrical engineering along with the DIC, from Imperial College, University of London, London, U.K.

He is a Distinguished Scientist, and Founding Head of Machine Intelligence Unit, at the Indian Statistical Institute, Calcutta, India. He worked at the University of California, Berkeley, and the University of Maryland, College Park, during 1986–1987 as a Fulbright Post-doctoral Visiting Fellow; at the NASA Johnson Space Center, Houston, TX during 1990–1992 and 1994 as a Guest Investigator under the NRC-NASA Senior Research Associateship Program; and at the Hong Kong Polytechnic University, Hong Kong in 1999 as a Visiting Professor. He served as a Distinguished Visitor of IEEE Computer Society (USA) for the Asia-Pacific Region during 1997–1999. He is a co-author of six books including *Fuzzy Mathematical Approach to Pattern Recognition* (New York: Wiley, 1986), and *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*, (New York: Wiley, 1999), and more than 200 research papers. His research interests include pattern recognition, image processing, soft computing, neural nets, GA's, and fuzzy systems.

Prof. Pal is a Fellow of the Third World Academy of Sciences, Italy, and all the four National Academies for Science/Engineering in India. He has received the 1990 S. S. Bhatnagar Prize (which is the most coveted award for a scientist in India), 1993 Jawaharlal Nehru Fellowship, 1993 Vikram Sarabhai Research Award, 1993 NASA Tech Brief Award, 1994 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award, 1995 NASA Patent Application Award, 1997 IETE - Ram Lal Wadhwa Gold Medal, 1998 Om Bhasin Foundation Award, and the 1999 G. D. Birla Award for Scientific Research. Prof. Pal is an Associate Editor, IEEE TRANSACTIONS ON NEURAL NETWORKS (1994–1998), Pattern Recognition Letters, Neurocomputing, Applied Intelligence, Information Sciences, Fuzzy Sets and Systems, and Fundamenta Informaticae; a Member, Executive Advisory Editorial Board, IEEE TRANSACTIONS ON FUZZY SYSTEMS and Int. Journal of Approximate Reasoning, and a Guest Editor of many journals including the IEEE Computer.