International Journal of Innovative Computing, Information and Control
Volume **9**, Number **7**, July **2013**

ICIC International ©2013 ISSN 1349-4198
pp. 3015–3037

GENETIC ALGORITHM AND ROUGH SETS BASED HYBRID APPROACH FOR REDUCTION OF THE INPUT ATTRIBUTES IN MEDICAL SYSTEMS

Kursat Zuhtuogullari[1,*], Novruz Allahverdi[2] and Nihat Arikan[3]

[1]Department of Electronics and Computer Education
Technical Education Faculty
[2]Department of Computer Engineering
Technology Faculty
Selcuk University
Selcuklu, Konya 42003, Turkey
[*]Corresponding author: zuhtuoglu@selcuk.edu.tr; noval@selcuk.edu.tr

[3]Department of Urology
Faculty of Medicine
Ankara University
Ankara, Turkey
narikan@hotmail.com

Received May 2012; revised October 2012

Abstract. *The information based systems with large input spaces require high processing times and memory when soft computing methods are used. Attribute reduction mechanisms are very important to overcome these problems by representing the data with the significant attributes. In this study, the genetic algorithm and rough sets based software is designed to realize feature reduction mechanism for the medical databases. The number of the input attributes can be reduced by the GARSBS and the output of the reduction mechanism is combined with variable input neural network classification software. Successful and faster reducing software is developed and a new modified selection mechanism consisting of variable number of generations in the genepool is constructed for obtaining performance. Extreme information storage, time consumption and input number restriction problems of the feature reduction algorithms are solved and classification processing times are reduced by using the generated approach. In the study, urological measurements are accepted as the input variables of the system. The number of the input variables (twenty) has been reduced to the reducts with eight, seven, six and five elements and the classification accuracy has been tested by the artificial neural network part of the software.*

Keywords: Rough sets theory, Hybrid rough sets and genetic algorithm based approach, Attribute reduction mechanism, Urological measurements

1. **Introduction.** In this study, genetic algorithm and rough sets based feature reduction software (GARSBS) is developed for finding the significant and dominant features (attributes) for making classification according to the urological measurements. The developed system supports not only the medical systems but also the information systems with large number of inputs. Most of reduction algorithms have the input number restrictions and high memory demand and high processing time problems that give rises to the memory errors. These problems are eliminated by the designed GARSBS. In addition, a new modified selection algorithm that is based on artificial selection method is proposed for faster processing and eliminating the genetic algorithm system to be locked

into the local solution candidates. The modified selection algorithm with variable generation genepool based on the artificial selection algorithm supports 4, 6, 8, 10 ..., $2*n$ generations and modified candidate selection system. The software is developed by using Delphi programming language.

The using of rough sets theory for input data reduction is a significant method in data mining and feature reduction mechanisms. In this study, a new modified version of artificial selection method with variable input generation genepool and improved solution candidate selection system is proposed and constructed for getting the optimum performance from genetic algorithm and rough sets based system.

The classification accuracies of GARSBS are compared with different approaches in the literature. Softwares using the Quick Reduct, Reverse Reduct, Decision Relative Discernility Function Based Reduction and Relative Dependency method are used for testing the performance and making comparison with the urological database. Delphi programming language is used for the developping test software interfaces.

Software using Decision Relative Discernibility based approach is accelerated by optimizing the software. The fitness function of the developed hybrid system software is determined by the rough sets based attribute reduction methodology. The software system is developed for reducing the input variables of the medical system with high accuracy and performance. A successful hybrid system software using proposed modified selection is developed for getting maximum performance from the classifier for the urological input data. The systems with high input variables need tremendous memory and time consumption when processed with artificial intelligence based systems.

In recent years different versions of rough sets theory have been applied for knowledge based systems for clustering and data mining approaches. The aim of the rough sets attribute reduction mechanisms is for making the analysis of the hidden data in the soft computing and knowledge based systems [1-3]. The reducing of number the input attributes and selecting dominant features that represents the database are made for processing the data efficiently by soft computing based methods. This procedure is significant for information systems because processing databases that consist of high inputs takes longer processing times or causes memory errors in software systems. The rough sets theory has a significant role in the feature reduction mechanisms of the knowledge based systems. The facts in the hidden data are analyzed and attribute reduction of input variables are made for calculating the significant input attributes of the systems. The rough sets based methodology aims to select most significant attributes of the data set without transforming the data during selection process. The methodology helps to select the significant features in the data bases for finding the reducts (reduced input variables) to represent the complete data set. The used approach is very efficient for finding the optimal reducts of the input database of the medical system by decreasing computation times and preventing memory errors in high input numbered databases.

In the knowledge based systems the input data is represented by the table where each row represents a case of an event, an object or pattern with input variables and the decision variable. There are many benefits of feature selection [1]. The hidden trends within the data can be recognized by representing the data with the most significant attributes. Using the reducts of the feature selection algorithms causes the soft computing based learning algorithms to be efficient [2,3]. The problems with high dimension data can be solved by rough sets based mechanism. The runtimes of learning algorithms are improved by the efficient selection mechanism. Usage of data mining based featured reduction mechanisms supports the artificial intelligent based learning algorithms for solving the data with large dimensions and reduces the utilization times [4-6].

The genetic algorithms are the family of computational models that encode solutions for a specific area by using data structures named as chromosomes. Implementations of genetic algorithms preserve critical information and search for better solutions by applying selection, reproduction and mutation operators. Genetic algorithm techniques are often accepted as optimizers for solving the problems in the knowledge based systems [7-10]. The genetic algorithm methods evaluate the potential solutions and produce new solutions for finding the optimal solution. In the developed system a new selection method based on the artificial selection is constructed in the genetic algorithm part of the developed rough-genetic hybrid system. The stopping criteria of rough genetic hybrid system for finding the optimum solution generation is determined by the attribute dependency value of GARSBS interface. Hybrid attribute reduction software with new modified selection mechanism based on artificial selection is constructed with the adaptation of large databases and the system has the capability of reading the system input values from the text files. The backpropagation based artificial neural network classifier software with variable input is developed and added to output of the rough-genetic hybrid system software. The neural network part of the software is constructed with the adaptation of variable input data. Successful results were obtained from the developed hybrid medical system software. The artificial neural network part of the developed software is developed compatible with multiple input variables. In the part 2, the problems in the high dimensionality data processing techniques and reducing algorithms are expressed. In the part 3, the general features of rough sets theory and genetic algorithms are summarized. In the part 4, the general structure of GARSBS is explained and in the part 5, the constructed variable input Artificial Neural Network system software is expressed. In the part 6, the obtained results are given and interpretations of the results are realized.

2. **The Problems of Processing Techniques Using Soft Computing Methods for High Dimensional Data.** In the information based systems, the processing times and high memory demand for processing data is a significant problem [2,3]. The systems with large inputs require more computation time and memory for processing with the soft computing methods. Large memory demands and high processing times give rise to the memory errors by requiring more memory than the allocated part determined by the operating system. For overcoming the problems, data reduction algorithms are developed for reducing the number of the input attributes [4-6]. Most of the data reduction algorithms aim to explore the significant attributes for reduction the number of the input variables. The input dimensionality reduction algorithms aim to find significant features to represent the information system. However, in the data mining systems, most of the reduction algorithms based on rough sets based methodologies require high memory and the computation times [3-5]. The decision relative discernibility matrix and function based reducing procedures and most of rough sets based reduction algorithms require high memory usages that give rise to memory errors and also the long computation times are required. These problems are eliminated by the developed hybrid system software by faster processing times and effective memory usage.

Another software using decision relative discernibility is constructed for comparing with the developed GARSBS. In the genetic algorithm systems, locking to the local solutions in the selection algorithms is also a problem. In the used modified artificial selection algorithm, this problem is eliminated and results are obtained faster and effectively.

The GARSBS have effective memory usage property and reduct derivation with high classification accuracy, adaptation of large input data bases and prevention of locking to local solutions in the genetic algorithm part and searching the solution points in a larger search space by the variable gene pool mechanism with modified candidate selection. The

proposed modified selection mechanism based on artificial selection mechanism helps the algorithm to find better solution points quickly and effectively by enlarging the qualified solution space.

GARSBS determines the reducts by a user defined search parameters named as "attribute dependency" used in rough sets methodology and this specification gives the opportunity to find the reducts with higher accuracies and representation capability for the whole database when the test data are used. The system is tested in the neural network part of the designed GARSBS by using the test data and the advantage of exploring a larger search space by finding more reducts with higher classification accuracies more effectively.

Most of the reduction algorithms in the literature have some disadvantages. In the Relative Dependency Method and Reverse Reduct Algorithm, backward elimination of the inputs are made and this property prevent the system for searching most of other reducts in the systems with large input variables. In the Quick reduct algorithm forward processing of reducts are made. However, Quick Reduct, Reverse Reduct and Relative Dependency methods require more processing times when the systems with high input number and high input spaces. These algorithms have input number restrictions when large input numbered data sets are evaluated when processed with softwares. Quick Reduct and Reverse reduct algorithms have disadvantage of not scanning most of the reducts.

The Johnson algorithm of Rosetta Software reducer algorithms of the Rosetta software has also been used for comparing the results. The working principles of the developed system and the algorithms are expressed in the part 4, and test results and conclusions are explained in the part 6. The Johnson algorithm using full discernibility aims to find a single reduct. The object related discernibility based part of Johnson has generated the attribute combinations with low classification accuracies when tested with urological test database.

The relative dependency method in the literature also makes the backward elimination of the attributes using the "relative dependence" parameter and the proportion of the number of the indiscernible subsets are evaluated. One drawback of this algorithm is in the condition that when the number of the subsets in the numerator and the denumerator part of the equation is equal and when the relative dependency is measured in the lower nodes are below 1, the algorithm is stopped.

3. **The General Features of the Rough Sets Theory and Genetic Algorithms.**

The rough sets based methodology manipulates the significant knowledge and groups of these methodologies and the extensions of them support many theoretical developments in the computational intelligence. Rough sets theory calculations have a significant place in the areas of medical applications of machine learning, data mining, and decision analysis. In addition to these areas, rough sets methods are used in intelligent control and pattern recognition systems [4-6]. The rough sets theory and its applications are often computationally efficient techniques for addressing problems such as hidden data discovery, data reduction, data significance evaluation and decision rule generation. An information system can be regarded as the table of data consisting of objects (rows) and attributes (columns). The conditional attributes of the system indicate the input variables and the decision attribute of the system indicates the classification indexes. In the developed approach the input values are the uroflowmetric measurements and residual urine volume and the output value of the developed genetic algorithm and rough sets based system is the classification according to the risk factor according to the measurements. An advanced software is constructed with the adaptation of large data bases. In

the constructed system a new genetic selection algorithm is constructed by modifying the artificial selection algorithm for generating optimal reducts for the developed system. In the developed software, attribute dependency is used as the fitness value for genetic algorithm based feature reduction mechanism. A variable input neural network software is added to the GARSBS for the classification purposes of the test data.

The rough sets based methods for the knowledge based systems consist of data selection or pre-processing and data reduction for the data mining processes. Data selection part of the rough sets based systems a target dataset is selected or created. The data pre-processing part of these systems includes noise removal and reduction. The aim of data selection phase is to improve the overall quality of information for software based systems. In the data reduction part, the useful features to represent the data are determined for representing the data set for efficient processing of data. In this procedure, data-mining method is selected depending on the goals of the knowledge discovery task for the extraction of information to represent input data for soft computing methods. The choice of algorithm used depends on many factors including the source of the dataset and the efficient processing. In the interpreting part of the data mining systems, knowledge that will be used for the processing are explored and evaluation systems use the prior knowledge. There are often many features in data mining algorithms related with knowledge discovery.

Genetic Algorithms are a family of computational soft computing models for finding the optimal solutions by exploring the solution space. These algorithms produces the optimal solutions to the systems and produce solution models for exploring the best solution candidates by using the soft computing methods inspired by genetics. The genetic algorithm based models proposes the advanced solution techniques for calculating the optimal results by producing new solution candidates. These algorithms encode a potential solution candidates named as chromosomes and the group of potential solution candidates that is named as generation [8-10]. Genetic algorithm based soft computing methods aims to find the better by applying genetic algorithm operators like selection, crossover and mutation. These operators are the computational mathematical models for finding the optimal solutions. An implementation of a genetic algorithm begins with a population of chromosomes. Then the genetic algorithm based systems uses the genetic algorithm operators for finding better solutions. The generation term symbolises the solution space produced by algorithm and includes potential solution candidates (chromosomes). In a broader usage of the term, a genetic algorithm is any population based model that uses selection and recombination operators to generate new sample solution points in a search space [9,10].

4. **The General Structure of the Developed Hybrid System.** In this study, a hybrid system software using new modified artificial selection strategy consisting variable generations in the genepool and improved solution candidate selections system is constructed. The system integrates the genetic algorithm methods with the rough sets attribute reduction. Proposed new modified variable generation genepool selection mechanism depends on artificial selection algorithm. In the developed GARSBS, genetic algorithm system is combined with rough sets attribute reduction system for calculating the optimal reducts of the medical system.

The modified and improved selection mechanism enables the system to find the reducts more effectively by reducing the computation time and increasing the system performance. The genetic algorithm based strategy enables the reducing algorithm to be adapted to the high input variable information based system when compared with different reduction

algorithms. The GARSBS has also the specification of finding different reducts by generating different solution candidates. However, the most of the existing algorithms can found limited number of the reducts because of their algorithmic structure.

In the genetic algorithm based feature selection system, attribute dependency value is accepted as the fitness value and the stopping criteria is defined by the fitness value of each chromosome (solution candidate). A new modified selection mechanism based on artificial selection algorithm is developed for the genetic algorithm selection mechanism. This selection mechanism approach is the modified version of the artificial selection method. The developed approach supports the classical artificial selection mechanism and the new constructed modified version of the algorithm. In the GARSBS, the urological database is used for testing the system with high input attributes instead of open access databases because most of open access databases consists of limited number of input attributes and transactions and they are not suitable for testing the advanced software systems.

Rough set theory is an extension of conventional set theory that supports approximations in decision making. A rough set is itself the approximation of a set by a pair of precise concepts that are called lower and upper approximations [3,4]. The lower approximation is a description of the domain objects that are known with certainty to belong to the subset of interest, whereas the upper approximation is a description of the objects that possibly belong to the subset. $I = (U, A)$ represents an is an information system, where U is a the nonempty set of finite objects and A is a nonempty finite set of attributes such that $a : U \to V_a$ for every $a \in A$. V_a represents the set of values that attribute a may take. For decision systems, $A = \{C \cup D\}$, where C is the set of input features and D is the set of class indexes for classification purposes. The Indiscernibility (*IND*) feature of the rough sets can be expressed by Equations (1)-(3). The relation given in Equation (1) shows that two objects are equivalent if and only if they have the same vectors of attribute values for the attributes in P. With any $P \subseteq A$ there is an associated equivalence relation *IND*(P): [3-6].

$$IND\,(P) = \left\{(x, y) \in U^2 | \forall a \in P, a\,(x) = a\,(y)\right\} \tag{1}$$

The partition of U, determined by *IND* (P), is expressed as $U/IND(P)$ or U/P, which is the set of equivalence classes generated by *IND* (P).

$$U/IND(P) = \otimes\{U/IND(\{a\})|a \in P\} \tag{2}$$

where

$$A \otimes B = \{X \cap Y | X \in A,\ Y \in B,\ X \cap Y \neq \varnothing\} \tag{3}$$

The lower approximation and upper approximation concepts are expressed by Equations (4) and (5). X can be approximated using the information in P by constructing the P-lower and P-upper approximations of the classical crisp set X. Equations (4) and (5) show the mathematical expressions of the upper and the lower approximation concepts respectively.

$$\underline{P}X = \{x|[x]_P \subseteq x\} \tag{4}$$

$$\overline{P}X = \{x|[x]_P \cap x \neq \phi\} \tag{5}$$

The positive, negative and boundary regions of the rough sets are expressed by Equations (6) and (8) respectively.

$$POS_P(Q) = \bigcup_{x \in U/Q} \underline{P}X \tag{6}$$

$$NEG_P(Q) = U - \bigcup_{x \in U/Q} \overline{P}X \tag{7}$$

$$BNP(Q) = \bigcup_{x \in U/Q} \overline{P}X - \bigcup_{x \in U/Q} \underline{P}X \tag{8}$$

Feature dependency values are calculated for the fitness value of the generated candidates in GARSBS. The genetic algorithm based system uses the feature dependency values of rough sets methodologies for each chromosome for finding the optimal reducts with high performance. In the rough sets theory feature dependency value is the ratio of the positive region to the solution space and expressed in Equation (9). A set of attributes Q, depends on a set of attributes P and for P, $Q \subset A$, Q depends on P in a degree k and denoted by $P \Rightarrow kQ$. In this study, attribute (feature) dependency value of rough sets methodology is used as the fitness function value for each solution candidate and stopping criteria in the developed hybrid system software. Stopping criterion of the system is denoted by α and this threshold value can be determined by the user in the developed system. The stopping criterion for the GARSBS is expressed in Equation (10). The solution candidates that are equal or higher than α threshold level are accepted as the results (reducts) in the GARSBS. In the modified artificial selection algorithm, the solution addition type added to the algorithm is calculated by Equation (11). The percentage values are expressed for the gene pool. The number of the random selected chromosomes used in modified artificial selection algorithm is determined by the percentage value expressed in Equation (11). The "*RSC*" term is used for the abbreviation of "random selected chromosomes" and "*BSC*" and "*WSC*" are used as the abbreviations named as best solution candidates and worst solution candidates respectively.

Equation (12) shows the number of generations in the genepool and the number of starting generations constructed randomly. "n" is a user defined parameter of the software. The software supports the new modified artificial selection algorithm that supports $2{*}n$ generations $(4, 6, 8, 10 \ldots,$ etc.) in the genepool whereas the classical version only supports two generation in the genepool. In the proposed modified version, starting $2 * n$ generations are constructed randomly in the modified version when the genetic algorithm part of the software is initialized. "n" is a user defined parameter and can be changed by the user of the software for getting optimal performance and expresses the number of the generations collected in the genepool and starting generations. The starting generations are used for generating the following generation. And last $2{*}n$ $(4, 6, 8, \ldots)$ generations are used for constructing the genepool used for determining the intermediate region used for crossover and mutation operator. This specification also enlarges the search space for obtaining the reducts. The modification mentioned above decreases the computation time and prevents the algorithm to be locked in the local solution points. The percentage values of the random selected chromosomes are named as solution addition percentage in the developed software interface.

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \tag{9}$$

$$\gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \geq \alpha \tag{10}$$

$$RSC\% = 100\% - (BSC\% + WSC\%) \tag{11}$$

$$\text{The number of generations in the Gene Pool} = 2 * n \tag{12}$$

In this study, a new selection algorithm that is based on artificial selection algorithm is constructed for getting optimum performance for the developed system. The new selection algorithm approach based on artificial selection is proposed and better results are obtained by reducing the computation time of the GARSBS. The constructed system also supports the classical and the new modified artificial selection algorithm versions developed for this study. In the classical artificial selection method, the chromosomes are ordered according to the fitness function. In the system that uses the attribute dependency

value, the chromosomes of the last two generations are ordered in the descending order. In the classical method, best and worst valued chromosomes are used for generating the gene pool and the best and worst chromosome percentages are determined by the user. Selection strategies are significant for the combinational operators [7,8]. The generated system supports the classical version and also the new constructed version that is the modified version of the artificial selection. In the classical version, the chromosomes of the last two generations are collected in the gene pool for the selection. In the developed modified selection system, software chooses the best valued chromosomes and the worst valued chromosomes and the random the chosen chromosomes in the desired percentage values. The best and the worst percentage values and the percentage of the random chosen chromosomes can be selected by the user of the software. The first $2*n$ generations are generated randomly by the modified system. In the modified system, "n" is a parameter determined by the user of the software and the genepool can consists of 4, 6, 8, 10, 12 ... generations when the n is specified as 2, 3, 4, 5 and 6, etc. The proposed new modified selection mechanism supports $2*n$ generations in the genepool. And the last system uses last $2*n$ generations iteratively in the following steps. Random chosen chromosomes can be selected from the interior region determined by boundary between the best selected and the worst selected chromosomes regions. This property enables the software to determine the random selected chromosomes except the best and the worst region.

The best, the random chosen from interior region and the worst fitness valued chromosomes are used for generating the optimal and successful solutions with less computation time. The reduction system also decreases the training times of the artificial neural network classifier system. Software interface GARSBS and variable input artificial neural network software that uses back propagation algorithm are developed by using Delphi programming language.

The constructed software is generated with the adaptation of the multiple input databases and the selection method is used for determining the gene pool for the crossover and mutation operators of the genetic algorithm. Order based crossover and partially matched crossover methods are integrated to the developed software and two mutation operators named as inversion and adjacent two change mutation methods are used. Adjacent two change mutation method is in the GARSBS and determines adjacent two input change by using permutation coding. Mutation operators are used for obtaining new input combinations when crossover rate is selected as high.

The high performance system with high reliability and accuracy is obtained with the reduced processing times. The general structure of the generated software is shown in Figure 1. The general structure of the genetic algorithm part and modified selection mechanism of the software is shown in Figure 2. The Artificial Neural Network (ANN) classifier software is added to the outputs of the system. The ANN uses backpropagation based classifier. The number of inputs, hidden neurons and learning rate can be adjusted by the developed advanced ANN software. The developed ANN software has the property to train the selected columns determined by the attribute reduction mechanism of GARSBS.

Another software using decision relative discernibility matrix and function based reducing mechanism is constructed for comparing the performance with the GARSBS. The decision relative discerniblility based reduction software is accelerated and optimized for comparing the performance with GARSBS.

A discernibility matrix of a decision table $(U, C \cup D)$ is a symmetric $|U| \times |U|$ matrix and the discernibility functions can be calculated by using this matrix and approach. A discernibility function f_d is a Boolean function of m Boolean variables $a_1^*, \ldots, a_n^*$ (corresponding to the attributes named as $a_1, \ldots, a_n$ from a given entry of the discernibility

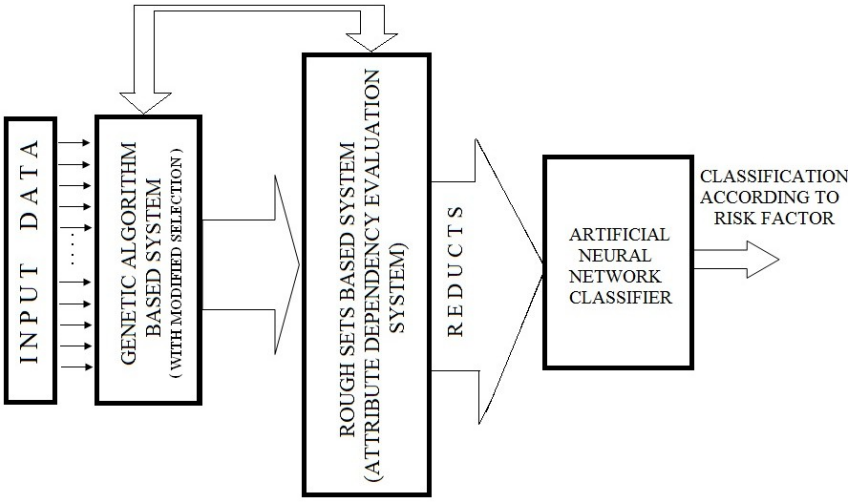


FIGURE 1. The general structure of the developed hybrid system software

matrix). The discernibility function is expressed in Equation (13). f_d represents the discernibility function and i and j are the indexes used for the matrix cells.

$$f_D(a_1^*, \ldots, a_m^*) = \wedge\{\vee C_{ij} | 1 \leq j \leq i \leq |U|, C_{ij} \neq \varnothing\} \tag{13}$$

The results obtained from the developed system are compared with the Johnson algorithm based reducer used in Rosetta software. The Johnson based reducer finds the reducts by using the a variation of Greedy algorithm. This algorithm has a natural bias towards finding a single prime implicant of minimal length. The reduct named as "B" is found by running the algorithm expressed below. The S denotes the set of sets corresponding to the discernibility function and $w(S)$ shows a weight for set S in S that automatically computed from the data. Support for computing approximate solutions is provided by aborting the loop when "enough" sets have been removed from S, instead of requiring that Shas to be fully emptied [11-13].

1. $B = \varnothing$
2. a shows the attribute that maximizes $\sum w(S)$ where the sum is taken over all sets S in S that contain a. Currently, ties are resolved arbitrarily.
3. Add a to B.
4. Remove all sets S from S that contain a.
5. If $S = \varnothing$; return B, otherwise, go to Step 2.

For the test procedures the algorithms in the literature are used. The Quick Reduct algorithm used for the test procedure is expressed below. In the Quick Reduct algorithm, the dependency of each attribute is calculated, and the best candidate chosen and added to the first node. And the algorithm iteratively continues by adding new nodes (inputs). In Reverse Reduct algorithm, backward elimination of the attributes are made [14]. Elimination of the attributes are made with starting from the least informative ones. Attribute dependency value of each candidate is important in the elimination procedure. The least significant attributes are eliminated initially. And the algorithm continues until finding the reduct. Large feature subsets have to be taken into account in this algorithm. This procedure is not always used in large databases because the algorithm must calculate large variety of subsets.

The relative dependence method in the literature uses the proportional values of the numbes of indiscernibity subsets. The algorithm aims to find the numbers of the indiscernible subsets and uses the proportional value [15]. The algorithm aims to make

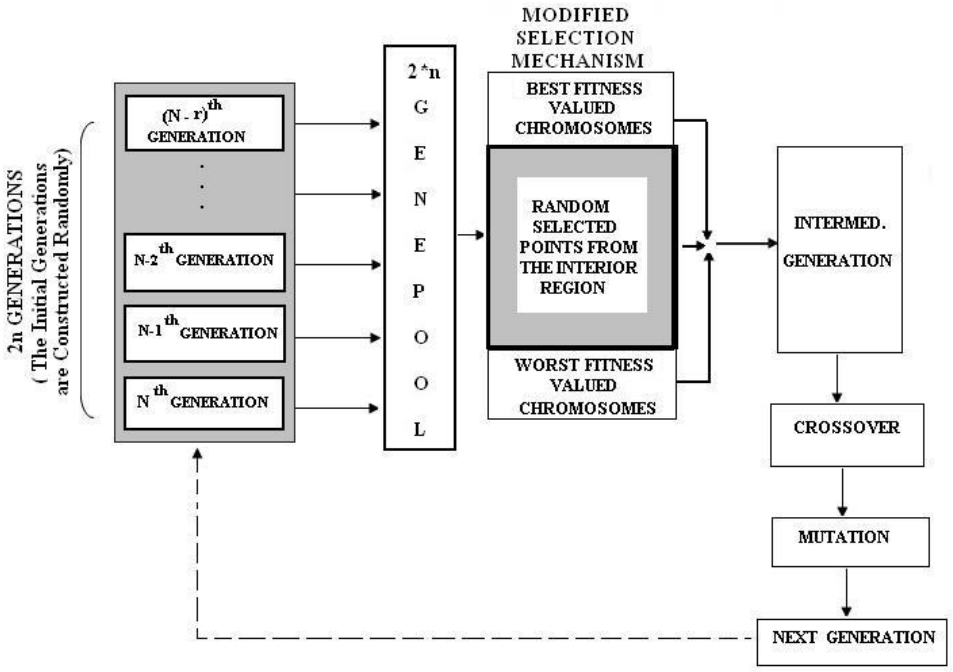


FIGURE 2. The general structure of the genetic algorithm based hybrid system, the proposed new variable generation gene pool generation system and new modified selection mechanism based on artificial selection

backward elimination of the features where attributes are considered to be removed if the relative dependency equals 1, upon their removal. Relative dependency is calculated by Equation (14) in the literature and expressed by κ. The relative dependence compares the number of the indiscernible subsets of the numerator and denumerator of Equation (14). The drawback of the algorithm is the comparison of the number of the subsets calculated in the numerator and the denumerator part. Reverse Reduct, Quick Reduct and Relative dependency and decision relative discernibility based methods require more computation time and memory and they are not suitable for high dimensional input data. If they are tested with high dimensional data these methods give memory errors because of high memory demand which is higher than the allocated memory of the operating system. They have the restrictions of the input data and so another urological database with 10 input variable and 115 transaction is used for comparing the results.

$$\kappa = \frac{|U/IND|}{|U/IND(R \cup D)|} \tag{14}$$

Order based crossover and partially matched crossover methods are used in the GARS BS. In order based crossover method, random numbers of the solution points are selected from the parent chromosomes. In the first chromosome the selected genes will stay in the same places but the corresponding genes in the second chromosome will be beside the genes of the first chromosome that occupy the same places [16-18]. The order based crossover method is expressed in Figure 3. The numbers in the figure represents the inputs of the system. Figure 3 shows the selected chromosomes and Figure 4 shows the resultant chromosomes after crossover method is used.

Partially matched crossover operator is used as the second crossover operator in the system. The partially matched crossover method was suggested by Goldberg and Lingle [16-18]. The partially matched crossover method can be implemented in the following

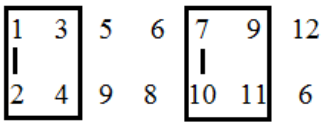


FIGURE 3. The selected chromosomes for the order based crossover method

1 2 5 6 7 10 12

2 1 9 8 10 7 6

FIGURE 4. The resultant chromosomes after the order based crossover method is applied

2 3 | 5 7 9 | 1 6 12 15

5 9 | 12 6 10 | 8 4 2 17

FIGURE 5. The chromosomes selected for partially matched crossover

2 3 | 12 6 10 | 1 6 12 15

5 9 | 5 7 9 | 8 4 2 17

FIGURE 6. The candidate chromosomes after crossover operator is applied

algorithm. In the partially matched crossover method, two random places are selected. Exchange the subtours formed by these two places and get the temporary solutions named as offspring '1' and '2'. The reputations of the genes in the solution candidate chromosome are prevented by changing by the genes in the second chromosome located between selected places in this method. The partially matched crossover method is shown in Figures 5-7. In Figure 5, the selected chromosomes for the crossover and the selected places for the crossover are shown. In Figure 6, the candidate offspring chromosomes are shown. In Figure 7, the repetition of the genes are prevented by transferring the matched the genes in the corresponding chromosome in the selected region.

Inversion and adjacent two change mutation operators are used in the system. In the inversion mutation method, a subtour is randomly selected by determining two points in the chromosome and the genes between the selected points are inverted [16-19]. In the adjacent two input change mutation method, adjacent two genes are selected and the places of the genes are inverted. In the adjacent two change method the places of the genes are interchanged [18,19]. In the GARSBS, the input variables are expressed as the permutation coding mode, and this method is used for interchanging the place of the two input variables. Mutation operators are used for the getting new solution candidates in the following generations when the crossover probability is accepted as high.

Figure 8 shows the inversion mutation method and Figure 9 shows the adjacent two change mutation method. This mutation method is adapted to the inputs of the hybrid system.

In the knowledge based systems, a data set is represented as a table where each row represents a case an event, an object or pattern with input variables and the decision variable. Every column represents an attribute a variable of an observation, a property,

2 3 | 12 6 10 | 1 7 5 15

12 10 | 5 7 9 | 8 4 2 17

FIGURE 7. The resultant chromosomes when the repetitions are prevented

2 5 6 | 8 10 12 14 | 15 17

2 5 6 | 14 12 10 8 | 15 17

FIGURE 8. The inversion mutation method

2 5 6 [8 10] 12 14 15 17

2 5 6 [10 8] 12 14 15 17

FIGURE 9. Adjacent two change mutation method

etc. that can be measured for each pattern. In the knowledge based systems that uses the rough sets approach, the outcome of the classification in the input data base table is called the decision attribute.

Uroflowmetry is a diagnostic test that is made for checking for abnormalities in the flow rate of a patient's urine. Uroflowmetric measurements are significant for determining the urological illnesses like urethral obstructions, urethral strictures, abnormal bladder activities, prostatic diseases and bladder obstructions and also the kidney problems. After uroflowmetry test, the amount of urine left in the bladder is measured. The residual urine volume shows the volume of urine left in the bladder after the uroflowmetry test. Voiding time show the time passed for voiding the bladder and measured by the uroflowmetry devices [24-26].

The constructed software interface is shown in Figure 9. The software is constructed with the adaptation of generating solution candidates with multiple input variables and can produce solution candidates with desired input number range. The system has 20 input variable and 1 classification variable (decision variable). The input variables of the system are uroflowmetric measurements named as maximum flowrate (mlt./s), average flowrate (mlt./s) and residual urine volume (mlt.) and the sampled flowrate values (mlt./s) from the uroflowmetry graph in the period of $T/4$ and $3T/4$. T represents the voiding time and maximum flowrate is expressed by the input variable named as 'a_1', average flowrate and residual urine volume are represented by the a_2 and a_3 respectively.

The input variables $a_4, a_5, a_6, a_7, a_8, a_9, \ldots, a_{20}$ (17 sampled uroflowmetric values) represent the flowrate measurements in the period of $T/4$ and $3T/4$. The period of $T/4$ and $3T/4$ are divided into 17 parts in the uroflowmetry test for getting the sapled flowrate values. In the database, maximum flowrate is expressed by 4 linguistic variables named as very low, low medium and high and denoted by 1, 2, 3 and 4 in the table. The average flowrate is expressed by 4 linguistic variables very low, low, medium and high and encoded by 1, 2, 3 and 4 respectively. The maximum flowrate and the average flowrate and residual urine volume values are expressed in the list below. The same threshold levels for the average flowrate values are accepted for the threshold values of the sampled flowrate values. The residual urine volume is expressed by none, medium, high and very high and denoted by the numbers expressed above. The sampled flowrate values

$(a_4, a_5, a_{6,}a_7, a_8, a_9, \ldots, a_{20})$ are denoted by four linguistic variables named as very low, low, medium and high and denoted by 1, 2, 3 and 4 respectively.

The input data base consists of 120 transactions and some of the transactions are shown in Table 1. The constructed software has the capability of processing high dimensional input data with fast computation times and optimized for using memory efficiently. The classification attribute is denoted by the letter 'd' and three linguistic variables named as very risky, risky and healthy and symbolised by the numbers 1, 2 and 3 respectively. The database consists of 120 transactions. Each transaction (rows) denotes the patients and each column represents the urological measurements. The database is taken from the patient database and constructed by the help of urology expert. The classification attribute determines the very risky, risky and healthy groups according to the uroflowmetric measurements and residual urine volume.

Maximum Flowrate (mlt./s)

Very Low.....	0 mlt./s $\leq$ x $<$ 10 mlt./s	1
Low.....	10 mlt./s $\leq$ x $<$ 19 mlt./s	2
Med.....	19 mlt./s $\leq$ x $<$ 30 mlt./s.	3
High.....	30 mlt./s $\leq$ x $\leq$ 40 mlt./s	4

Average Flowrate (mlt./s)

Very Low.....	0 mlt./s $\leq$ x $<$ 7 mlt./s	1
Low.....	7 mlt./s $\leq$ x $<$14 mlt./s	2
Medium.....	14 mlt./s $\leq$ x $<$ 25 mlt./s	3
High.....	25 mlt./s $\leq$ x $\leq$ 40 mlt./s	4

Residual urine volume (mlt.)

None.....	0 mlt.	1
Medium.....	0 mlt. $<$ x $<$ 50 mlt.	2
High.....	50 mlt. $\leq$ x $<$ 150 mlt.	3
Very High	150 mlt. $\leq$ x $\leq$ 500 mlt.	4

The software has the capability of scanning different number of inputs if desired and has the capability of processing high input numbered systems without out of memory error. The problem of the reduction mechanisms with high input variables are very long processing times and out of memory errors. The developed system with modified selection mechanism (GARSBS) eliminates these problems. The large input numbered systems can be processed by the developed system with high processing speed and efficiency.

Most of reduction algorithms requires much memory demand and time consumption when used in computer softwares and are unsuitable for processing high input numbered databases. A second urological database that consists of 115 transactions and 10 input variables are used for comparing the performance of some reducing algorithms the with GARSBS. Test softwares using the algorithms decision relative discernibility, Quick Reduct, Reverse Reduct and Relative Dependency method are constructed for comparison. Some of the transactions of the second database is shown in Table 2. In Table 2, x_1 is used for maximum flowrate, x_2 is used for average flowrate, x_3 is used for residual urine volume and x_4 ... x_{10} represent the sampled uroflowmetric flowrate values in the period of $T/4$ and $3T/4$. The the sampled uroflowmetric flowrate values uses the threshold values specified for the average flowrate (mlt/s) values.

The artificial intelligence part of the software is developed with the adaptation of processing variable numbered of input variables for the flexibility to train the reducts. The artificial intelligence software part is optimized and also developed with high processing speed property. The number of hidden neurons can be determined by the user for getting

TABLE 1. Some of the transactions in the urological database 1 with 20 inputs

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}	a_{19}	a_{20}	d
1	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	1	3	1	2	1	1	1	2	1	2	2	1	1	1	2	1	1	1	1	1
3	1	1	2	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1
4	2	1	1	1	2	1	1	1	2	1	1	1	1	1	1	2	1	1	2	1	1
5	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	2	1	2	1	2	1	1	2	1	1	1	1	2	1	1	2	1	1	2	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	2	2	3	1	2	1	2	1	1	1	2	2	2	1	1	1	1	2	1	1	1
9	3	2	2	2	2	2	2	1	2	2	2	3	2	3	2	1	2	2	2	2	2
10	2	2	3	2	2	2	1	2	2	2	2	2	2	1	2	1	2	1	2	1	1
11	2	2	2	1	2	1	2	1	2	1	2	2	2	1	1	1	2	1	2	1	1
12	2	2	4	1	1	1	1	1	1	2	1	2	2	1	2	1	1	1	1	1	1
13	3	2	1	1	2	2	1	2	2	2	3	2	2	2	2	2	2	2	2	2	2
14	3	2	2	1	3	1	2	2	2	2	2	2	2	2	2	3	1	2	2	2	2
15	3	2	2	2	2	2	1	2	3	2	2	3	1	2	3	2	2	1	2	2	2
16	3	2	2	2	2	2	2	2	2	2	2	2	3	1	2	2	2	2	3	1	2
17	2	2	3	1	2	2	1	2	1	1	2	2	1	1	2	1	1	1	1	1	1
18	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3
19	3	2	2	2	2	3	2	3	2	2	3	2	1	2	2	2	2	2	2	2	2
20	3	2	1	2	2	2	2	2	3	2	2	3	2	1	2	2	2	2	2	2	2

TABLE 2. Some of the transactions in the urological database 2 with 10 inputs

	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x7$	$x8$	$x9$	$x10$	d
1	1	1	2	1	1	1	1	1	2	1	1
2	2	1	1	1	2	1	2	1	1	1	1
3	2	1	1	1	1	1	1	2	1	1	1
4	2	1	1	2	1	1	1	2	1	1	1
5	2	1	1	1	2	2	2	1	1	2	1
6	1	1	3	1	1	1	1	1	1	1	1
7	3	3	1	3	3	3	3	3	4	3	3
8	3	1	1	1	1	1	1	3	1	1	1
9	3	3	1	3	3	3	4	3	3	3	3
10	2	1	1	1	1	2	1	1	1	1	1

the optimal performance from the training system. Figure 11 shows the constructed artificial neural network software interface. The learning speed and number of input variables can also be determined by the user for obtaining the flexibility. The back propagation algorithm is used in the developed artificial neural network training system attached to the reduct calculation part of the GARSBS. All the data in the system are normalized between 0 and 1. Figure 12 shows the normalization part of neural network part of GARSBS. The modified artificial selection algorithm used in the developed software prevents the system to be locked into the local points.

5. **Artificial Neural Network Classifier System.** The artificial neural network classifier system software is constructed so flexible with the adaptation of variable input data and test interface. The output of the GARSBS is attached to developed flexible artificial

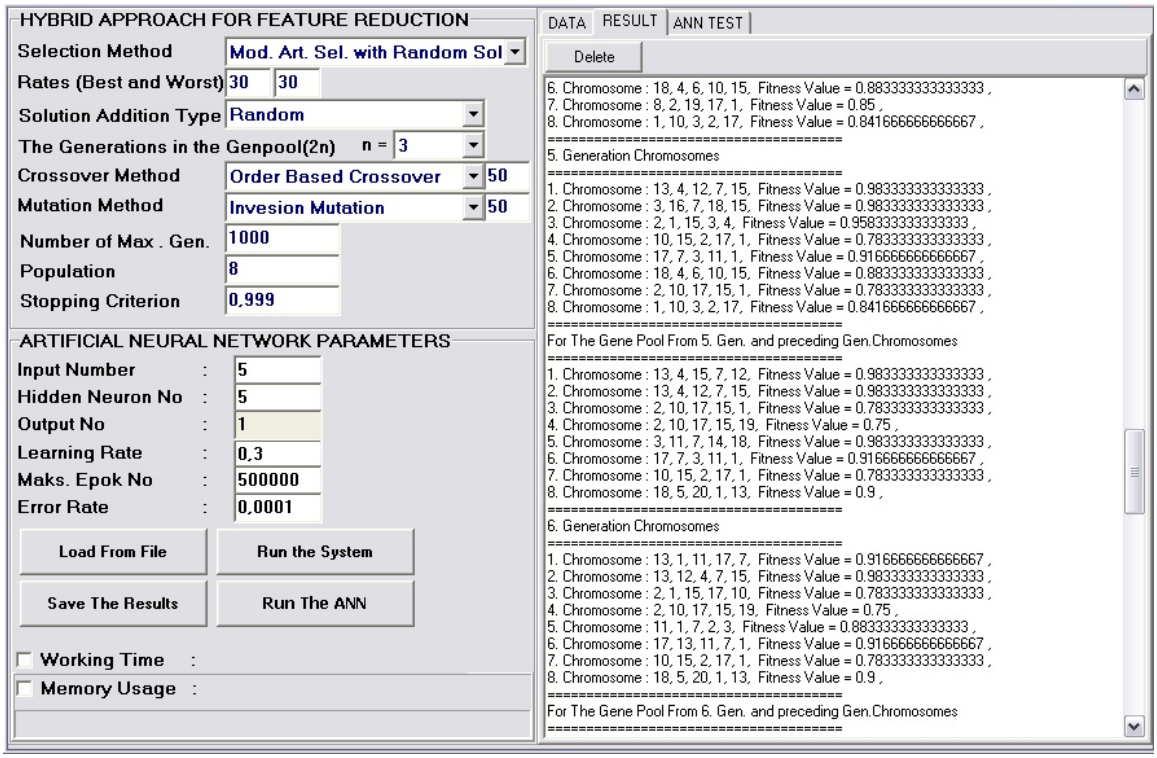


FIGURE 10. The general appearance of developed software

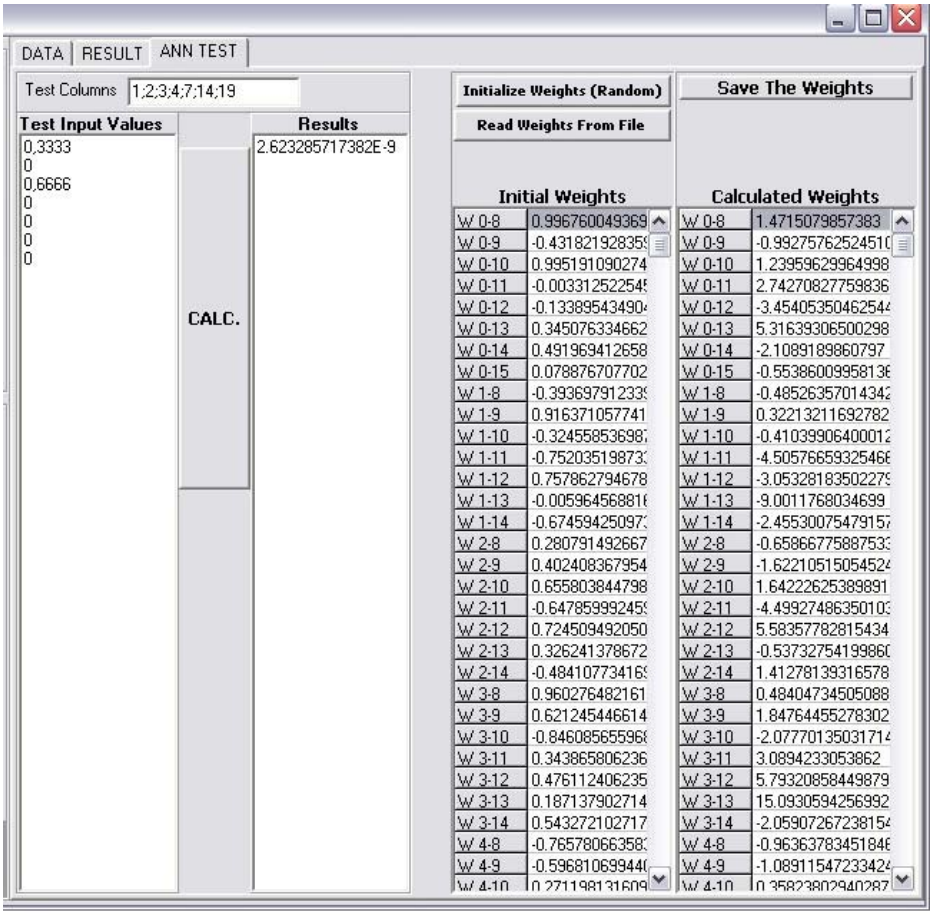
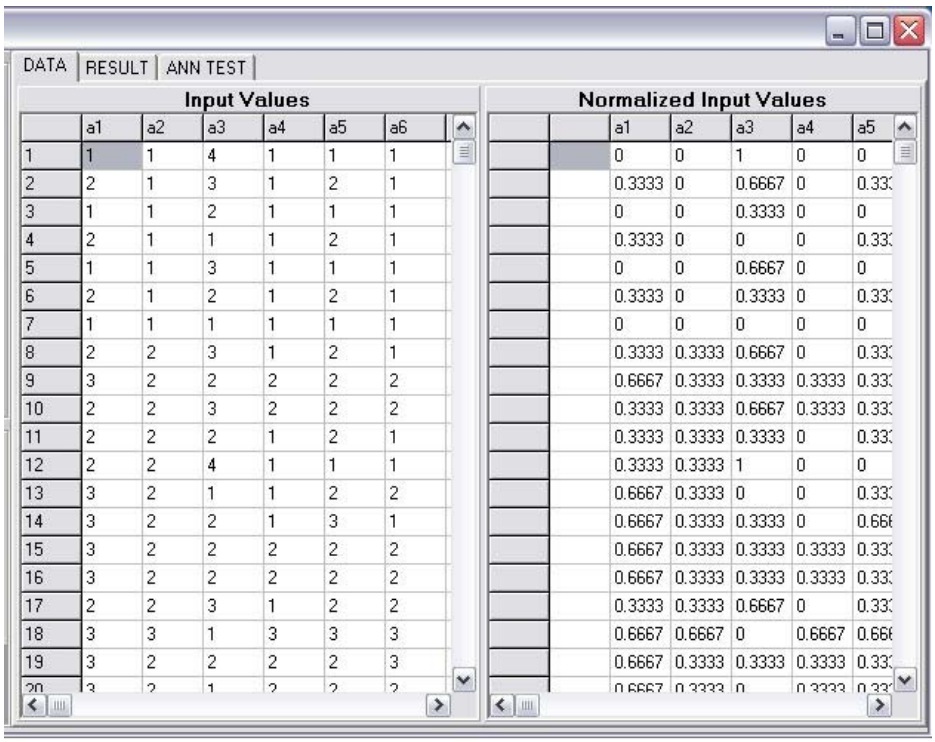


FIGURE 11. Artificial neural network software interface of the developed software



	a1	a2	a3	a4	a5	a6
1	1	1	4	1	1	1
2	2	1	3	1	2	1
3	1	1	2	1	1	1
4	2	1	1	1	2	1
5	1	1	3	1	1	1
6	2	1	2	1	2	1
7	1	1	1	1	1	1
8	2	2	3	1	2	1
9	3	2	2	2	2	2
10	2	2	3	2	2	2
11	2	2	2	1	2	1
12	2	2	4	1	1	1
13	3	2	1	1	2	2
14	3	2	2	1	3	1
15	3	2	2	2	2	2
16	3	2	2	2	2	2
17	2	2	3	1	2	2
18	3	3	1	3	3	3
19	3	2	2	2	2	3

	a1	a2	a3	a4	a5
	0	0	1	0	0
	0.3333	0	0.6667	0	0.33
	0	0	0.3333	0	0
	0.3333	0	0	0	0.33
	0	0	0.6667	0	0
	0.3333	0	0.3333	0	0.33
	0	0	0	0	0
	0.3333	0.3333	0.6667	0	0.33
	0.6667	0.3333	0.3333	0.3333	0.33
	0.3333	0.3333	0.6667	0.3333	0.33
	0.3333	0.3333	0.3333	0	0.33
	0.3333	0.3333	1	0	0
	0.6667	0.3333	0	0	0.33
	0.6667	0.3333	0.3333	0	0.66
	0.6667	0.3333	0.3333	0.3333	0.33
	0.6667	0.3333	0.3333	0.3333	0.33
	0.3333	0.3333	0.6667	0	0.33
	0.6667	0.6667	0	0.6667	0.66
	0.6667	0.3333	0.3333	0.3333	0.33

FIGURE 12. The normalization part of the developed software

neural network classifier software. The number of input variables and the hidden neurons, the error rate and the learning rate variables can be changed by the user for the increasing system performance. Back propagation method is used in the developed system test software. Calculated weights can be recorded and read to the text files for faster processing purposes. The used system, forward propagation is used for calculating the output value of the network. In the backward propagation phase the updating the weights are made by ANN. Net and output values for middle layer neurons are calculated by Equations (15) and (16). C_j represents the output value of middle neuron [27-29].

$$NET_j = \sum_{i=1}^{i=n} X_i.W_{ij} \tag{15}$$

$$C_j = \frac{1}{1 + e^{-(NET_j^a + \beta_j^a)}} \tag{16}$$

In the backward propagation algorithm the initial weights are updated according to the position of the neurons. The updated weights are applied to the next iteration. Updating of the weights between the middle and output layer are made by using Equations (17)-(19). In Equation (17), λ is the learning constant and α is the momentum coefficient. β represents the bias weights and the $\Delta\beta$ represents the change of the weights of the biases. C_m represents the output value of the output neuron and C_j represents the output value of the middle neuron [27-29].

In Equations (17) and (18), ΔA_{jm}^a represents the change in the weight between middle and output neuron.

$$\Delta A_{jm}^a(t) = \lambda.\delta_m.C_j^a + \alpha\Delta A_{jm}^a(t-1) \tag{17}$$

$$\Delta A_{jm}^a(t) = \lambda.\delta_m.C_j^a + \alpha\Delta A_{jm}^a(t-1) \tag{18}$$

$$\delta_m = C_m.(1 - C_m).E_m \tag{19}$$

The new values of the weights are calculated by Equations (20)-(22). Weights of the bias neurons are updated using Equations (21) and (22). A_{jm} represents the weights between the middle layer and the output layer and ΔA^a_{jm} represents the change in the weight of A_{jm}. In the equations used for updating the weights of backpropagation network, k is an index used for representing the input layer, j denotes the middle layer and m represents the output layer.

$$A^a_{jm}(t) = A^a_{jm}(t-1) + \Delta.A^a_{jm}(t) \tag{20}$$

$$\Delta\beta^{\varsigma}_m(t) = \lambda.\delta_m + \alpha.\Delta\beta^{\varsigma}_m(t-1) \tag{21}$$

$$\beta^{\varsigma}_m(t) = \beta^{\varsigma}_m(t-1) + \Delta\beta^{\varsigma}_m(t) \tag{22}$$

In the update phase of the weights between the middle layer and the input layer, Equations (23)-(27) are used [27-29].

$$\Delta A^i_{kj}(t) = \lambda.\delta^a_j.\mathrm{C}^i_k + \alpha\Delta A^i_{kj}(t-1) \tag{23}$$

$$\delta^a_j = f'(NET).\sum_m \delta_m.A^a_{jm} \tag{24}$$

$$A^i_{kj}(t) = A^i_{kj}(t-1) + \Delta A^i_{kj}(t) \tag{25}$$

$$\Delta\beta^a_j(t) = \lambda.\delta^a_j + \alpha.\Delta\beta^a_j(t-1) \tag{26}$$

$$\beta^a_j(t) = \beta^a_j(t-1) + \Delta\beta^a_j(t) \tag{27}$$

In the neural network classifier system, normalization procedure is made according to the columns. The Normalization equation is used in the procedure is expressed in Equation (28). The values of the input attributes and the output attribute are normalized between 0 and 1. In the equation $a_{\min}$ represents the minimum value for a in the column and $a_{\max}$ represents the maximum value in the column and 'i' symbolizes the column number.

$$\text{The Normalized } (a_i) = \frac{(a_i - a_{\min})}{(a_{\max} - a_{\min})} \tag{28}$$

In Equation (29), classification accuracy is expressed. The classification accuracy is the proportion of the correctly classified objects to the number of all classification objects in the test data. "*C. Acc.*" is used as the abbreviation for "classification accuracy in Equation (27).

$$C.Acc.\% = \left(\frac{\text{no of correctly class. samples}}{\text{no of all test samples.}}\right) * 100 \tag{29}$$

6. **Results, Discussions and Conclusions.** In the developed system software, high accuracy results were obtained during the classification procedure. The modified version of artificial selection algorithm with modified solution candidate selection mechanism and user defined multiple generation system prevent the developed algorithm to be locked in the local solution candidates and this property of the developed system enables the system for exploring the reducts of the system with less computation time with high efficiency.

In the classical artificial selection algorithm the generations from the last two generations are used and the chromosomes are ordered according to their fitness values and best and worst valued chromosomes are selected for constructing the genepool. In the proposed new modified artificial selection, the initial "$2*n$" generations are constructed randomly and the the previos $2*n$ generations (4, 6, 8, ...) are used for constructing the genepool. The genepool is constructed from the best, worst and random selected chromosomes. The random selected chromosomes are selected from the interior region determined by the bottom border of the best chromosomes and upper border of worst selected chromosomes. "n" is a user defined parameter for determining the wideness of the genepool.

Constructing a larger genepool from the previous "$2*n$" generations enables the system to derive better solution candidates (chromosomes) with better fitness values. In the test procedures, the number of the generations in the genepool increases a better fitness values chromosomes are obtained in the genepool. During the test operations the number of the generations in the genepool were increased when using the proposed modified artificial selection, the average computation times were decreased. The memory usages when using a larger genepool were also decreased because of reducing in the computation time.

Attribute dependency values of rough sets methodology has been used as the fitness values for chromosomes and stopping criteria of the GARSBS and user of the software can determine the attribute dependency threshold value. By using the threshold value to desired ranges and genetic algorithm based search system, large number of reducts can be calculated and also can be adapted to the consistent and inconsistent databases. The developed system with modified selection prevented "out of memory" errors and decreased the training times in the medical test systems. The software can also adaptable to different information based systems with high input spaces.

Successful and satisfactory results were obtained during the reduction process. The developed software has the capability of scanning different numbered input spaces and searching property of scanning for the reducts between the desired ranges.

The modified artificial selection algorithm version used in the software decreased computation times and prevented the developed algorithm to be locked to the local solution candidates during test operations. The GARSBS can be run for different threshold values (attribute dependency values) and different number of attribute ranges. The number of the input variables (twenty) has been reduced to the reducts with eight, seven, six and five elements. Some of the reducts found by GARSBS are listed with the attribute dependency values in Table 3 when the urological database with 20 input variables were used.

TABLE 3. Some of the reducts found by the GARSBS when urological input data with 20 input variables have been used

Element No	The Reducts	Attribute Dependency Value
8	a_1, a_2, a_3, a_5, a_7, a_9, a_{12}, a_{15}	1,0
	a_1, a_2, a_3, a_4, a_5, a_8, a_{10}, a_{12}	0,983
	a_2, a_3, a_4, a_5, a_6, a_8, a_{10}, a_{12}	0,983
7	a_1, a_2, a_3, a_4, a_7, a_{14}, a_{19}	1
	a_1, a_2, a_3, a_4, a_5, a_{12}, a_{15}	0,983
6	a_1, a_2, a_3, a_5, a_8, a_{12}	0,958
5	a_1, a_2, a_3, a_4, a_{20}	1

Table 4 shows the average fitness values of the chromosomes when 50 tests were made and stopped in the specified generation number. The results were obtained from the specified generation and show the attribute dependencies of the chromosomes in the intermediate generation used for crossover and mutation. The enlarged genepool with 6 and 8 generations using modified artificial selection enabled the system to work faster and effectively and the intermediate generation used for the crossover and mutation was constructed with the chromosomes with higher fitness (attribute dependency) values when compared with artificial selection strategy as shown in Table 4. The addition of random selected chromosomes to best and worst selected chromosomes in the modified selection system also increased the overall fitness value of the intermediate generation.

TABLE 4. The average fitness values of the chromosomes when 50 tests were made and stopped in the same generation before reaching the final reducts

Chromosome No in the Int. Reg.	Mod. Art. Sel. with 6 generations in the genepool	Mod. Art. Sel. with 8 generations in the genepool	Artificial Selection
1	0,82	0,85	0,6
2	0,78	0,84	0,55
3	0,7	0,75	0,5
4	0,65	0,68	0,4
5	0,55	0,59	0,38
6	0,45	0,47	0,36

GARSBS calculates the significant attributes of the risk determination system for the urological illnesses like urethral obstructions, urethral strictures and the illnesses, and determines the risk factor according to the urological measurements (uroflowmetric measurements and residual urine volume). In the urological database with 20 input variables, the reducts named as a_1, a_2 and a_3 represent maximum flowrate, average flowrate and residual urine volume respectively. Most of the reducts with attribute (feature) dependency value of 1, contain the reducts named as a_1, a_2 and a_3. This calculation shows the significance of the maximum flowrate, average flowrate and residual urine volume measurements and calculated reducts contains the some of the sampled flowrate values. The input variables $a_4, a_5, a_6, a_7, a_8, a_9, \ldots, a_{20}$ (17 sampled uroflowmetric values) represent the flowrate measurements in the period of $T/4$ and $3T/4$. The input medical data is classified according to the calculated reducts with high performance. The higher attribute dependency values guarantee no information loss or minimum information loss when reducing the high input data. The reduct containing the elements a_1, a_2, a_3, a_5, a_7, a_9, a_{12}, a_{15} shows that, for faster processing of the medical data base with 20 input variables, we can use the significant (dominant) attributes named as maximum flowrate, average flowrate, residual urine volume and the sampled uroflowmetric measurements named as a_5, a_7, a_9, a_{12}, a_{15}. During the test operations of the ANN part, the average classification accuracy about %90 is reached when the data that are not in the training set are tested. The risk degree determination can be done by the input measurements calculated by the hybrid reducing system.

The processing times of the artificial neural network system for training procedure has been decreased averagely above %50 during the test operations made with the full and reduced data sets. The test operations are made with Core2Quad 3.0 processor with 8 GB RAM.

The decision relative discernibility matrix and function based reducing procedures and most of rough sets based reduction algorithms require high memory usages that give rise to memory errors and also the long computation times. The discernibility matrix and function based reducing software is constructed for the test procedure for comparing performance with the developed hybrid system. The discernibility matrix and function based attribute reducing software constructed for the testing the system, supports maximum 12 inputs when the database with 120 transactions (rows) is used. When the test data with 20 input variables is tested with the decision relative discernibility matrix and function based system, the processing time exceeds 4 hours and exceeds the memory allocated by the operating system and causes memory errors. When testing the decision relative

discernibility approach, in the task manager of the operating system the, Memory – Peak Working Set exceeds 800 MB (in the task manager of operating system) that give rise to memory error.

The GARSBS supports high dimensional inputs and works efficiently without memory errors and has high processing speed within the range of 2 minutes and 30 minutes when artificial selection algorithm in the genetic algorithm part has been used and this calculation time depends upon the population number and generated chromosomes. The modified artificial selection method has been tested and reduced the computation time %30-%45 averagely when compared with artificial selection algorithm and also the modified selection mechanism prevents the genetic algorithm based system to locked into the local solutions and find the reducts more effectively. Modified artificial selection with 6 and 8 generations in the Genepool have decreased the computation time more than %50 and %60 when compared with classical artificial selection algorithm in the test operations. The allocated memory values by the Operating System to the software (The Memory Peak Working Sets) are shown in Table 5. The average computation times and memory usage values of the tested systems are shown for the different urological databases consisting 10 and 20 inputs respectively in Table 5. The average computation time and memory usage values are better than the tested systems.

The Quick Reduct, The Reverse Reduct and Relative Dependency based algorithms demand high memory and time usage and the constructed test softwares using these algorithms do not support 20 input variables with 120 transactions for the reason that test algorithms are not suitable for high dimensional data and require high memory than

TABLE 5. The average time consumption and peak memory working set values of GARSBS and test softwares

	Tested System	Time (average)	Number of Inputs	Time (average)	Allocated Memory Peak Working Set (MB)
1	GARSBS	Art. Sel.	20	2 min-30 min	90-350 MB
		Mod. Art. Sel. with 2 Gen.	20	1,5 min-20 min	80-250 MB
		Mod. Art. Sel with 6 Gen.	20	1 min-8 min	60-150 MB
		Mod. Art. Sel with 8 Gen.	20	40 sec-7 min	50-120 MB
		Art. Sel.	10	1 min-6 min	40 MB-80 MB
		Mod. Art. Sel with 6 Gen.	10	30 sec-4 min	35 MB-70 MB
2	Decision Relative Discernibility Fuction Based Reducer		20	Exceeds 4 Hours	Exceeds 800 MB causing Memory Error (Does Not Support)
			10	10 min-25 min	150-500 MB
3	Quick Reduct		20	Does Not Support (Extreme Memory Usage causing Memory Error)	
			10	5-20 min	380-500 MB
4	Reverse Reduct		20	Does Not Support (Extreme Memory Usage causing Memory Error)	
			10	6-20 min	370-550 MB
5	Relative Dependency Method		20	Does Not Support (Extreme Memory Usage causing Memory Error)	
			10	10-20 min	200-350 MB

allocated by the operation system. Therefore, the classification accuracy tests of these algorithms are made by using another urological database with 10 input variables. In addition, these algorithms are not suitable for scanning most of reduct combinations.

For the test procedure the test operations were done with the reducts obtained by the Rosetta software Johnson Reducer algorithm [11,12]. Urological database with 20 input variables were used during the test operations. The reducts found by Johnson reducer of Rosetta had an average classification accuracy value of %60,1 when sampled points from the unreduced test data is used and tested in the ANN part of GARSBS. When the tests with 10 urological database with 10 input variables an average classification accuracy value of %60 was obtained.

During the test operations, the found reducts with 2 or higher were evaluated in Johnson algorithm. The average value of the test results of the reducts obtained by full discernibility and object related discernibiliy are used by the Johnson Algorithm. The accuracy value of GARSBS is above %90 and this classification accuracy is higher than the tested Johnson reducer algorithm. Some of the reducts that are found by the Johnson reducer system from the urological database with 20 inputs are expressed below.

a_1, a_3, a_4, a_{15}, a_{20}
a_1, a_5, a_{20}
a_4, a_5
a_2, a_{17}
a_1, a_3 ... etc.

The average classification accuracies calculated by the ANN part and shown in Table 6 when the urological database with 20 and 10 inputs were used. Table 7 shows some

TABLE 6. Average classification accuracies of the tested softwares

	The Tested Software	Average Classification Accuracies (Urological Database 1 with 20 inputs)	Average Classification Accuracies (Urological Database 2 with 10 inputs)
1	GARSBS with Mod. Sel.	%90	%90
2	Johnson Reducer	%60	%55
3	Dec. Relative. Disc.	– (not supported)	%75
4	Quick Reduct	– (not supported)	%70
5	Reverse Reduct	– (not supported)	%70
6	Relative Dependency Method	– (not supported)	%65

TABLE 7. Some of the reducts calculated by the constructed test softwares and GARSBS when urological database with 10 input variables is used

	The Tested Software	Some of the Reducts Calculated
1	GARSBS	$x_1x_2x_3x_4x_5$, $x_1x_2x_3x_6$, $x_1x_2x_3x_5x_6, \ldots, x_1x_3x_5x_6x_8$
2	Decision Relative. Disc.	$f_d = x_1x_3x_4x_5x_6x_9 + \ldots + x_1x_3x_4x_6x_7x_9$ $+x_3x_4x_7x_8x_9x_{10}$
3	Quick Reduct	$x_1x_3x_4x_5x_6x_9$
4	Reverse Reduct	$x_2x_3x_4x_5x_7x_9x_{10}$
5	Relative Dependency Method	$x_3x_4x_7x_8x_9x_{10}$
6	Johnson Reducer of Rosetta (Full and Object Related Dis.)	$x_1x_3x_4x_5x_6x_9$, $x_1x_7x_{10}$, $x_6x_9, \ldots, x_2x_7$, $x_3x_5x_8x_{10}$

of the reducts calculated by the constructed test softwares, GARSBS and also Johnson Reducer of Rosetta software when urological database with 10 input variables was used.

GARSBS with modified selection had higher classification accuracies when compared with the Johnson Reducer, Decision Relative Discernibility, Quick, Reverse and Relative Dependency based constructed test softwares. GARSBS has the scanning capability more number of different successful reducts in high dimensional input spaces when compared with the tested systems. Quick Reduct and Reverse Reduct and Decision Relative Discernibility based algorithms have disadvantage of input data restrictions. Quick Reduct and Reverse reduct algorithms have disadvantages of not scanning of most the reducts in the information based systems.

GARSBS has the advantage of effective memory usage and fast processing when large databases are used. The advantages of GARSBS are obtaining the reducts with high classification accuracies, adaptation to high numbered input spaces and deriving larger number of reducts and also effective processing time and memory usages.

Acknowledgement. This project is supported by the Scientific Research Projects Unit of Selcuk University.

REFERENCES

[1] D. Pei, On definable concepts of rough set models, *Information Sciences*, vol.177, no.19, pp.4230-4239, 2007.

[2] J. P. Herbert and J. Yao, Criteria for choosing a rough set model, *Computers & Mathematics with Applications*, vol.57, no.6, pp.908-918, 2009.

[3] Z. Pawlak and A. Skowron, Rudiments of rough sets, *Information Sciences*, vol.177, no.1, pp.3-27, 2007.

[4] L. Shen and H. T. Loh, Applying rough sets to market timing decisions, *Decision Support Systems*, vol.37, no.4, pp.583-597, 2004.

[5] P. Pattaraintakorn and N. Cercone, Integrating rough set theory and medical applications, *Applied Mathematics Letters*, vol.21, no.4, pp.400-403, 2008.

[6] B. Mak and T. Munakata, Rule extraction from expert heuristics: A comparative study of rough sets with neural networks and ID3, *European Journal of Operational Research*, vol.136, no.1, pp.212-229, 2002.

[7] *http://www.obitko.com/tutorials/genetic-algorithms/*, 2012.

[8] K. Miettinen, M. M. Makela, P. Neittaanmaki and J. Periaux, Evolutionary algorithms in engineering and computer science Part 1, in *An Introduction to Evolutionary Computation and Some Applications*, D. B. Fogel (ed.), New York, John Wiley & Sons, 1999.

[9] C. L. Karr and L. M. Freeman, Industrial applications of genetic algorithms, *Image Calibration Transformation Matrix Solution Using Genetic Algortihm*, T. P. Dickens (ed.), CRC Press, New York, 1999.

[10] L. M. Schmitt, C. L. Nehaniv and R. H. Fujii, Linear analysis of genetic algorithms, *Theoretical Computer Science*, vol.200, no.1-2, pp.101-134, 1998.

[11] D. S. Johnson, Approximation algorithms for combinatorial problems, *Journal of Computer and System Sciences*, vol.9, pp.256-278, 1974.

[12] A. Øhrn, J. Komorowski, A. Skowron and P. Synak, The design and implementation of a knowledge discovery toolkit based on rough sets: The ROSETTA system, in *Rough Sets in Knowledge Discovery 1: Methodology and Applications, Studies in Fuzziness and Soft Computing*, L. Polkowski and A. Skowron (eds.), 1998.

[13] A. Øhrn, J. Komorowski, A. Skowron and P. Synak, The ROSETTA software system, in *Rough Sets in Knowledge Discovery 2: Applications, Case Studies and Software Systems, Studies in Fuzziness and Soft Computing*, L. Polkowski and A. Skowron (eds.), 1998.

[14] R. Jensen and Q. Shen, *Computational Intelligence and Feature Selection, Rough and Fuzzy Approaches*, IEEE Press, John Wiley and Sons, 2008.

[15] J. Han, R. Sanches and X. Hu, Feature selection based on relative attribute dependency: An experimental study, *RSFDGrC'05 Proc. of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, vol.1, pp.214-223, 2005.

[16] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Massachusetts, USA, 1989.
[17] C. L. Karr and L. M. Freeman, Industrial applications of genetic algorithms, Chapter 2, *Image Calibration Transformation Matrix Solution Using Genetic Algorithm*, T. P. Dickens (ed.), CRC Press, New York, 1999.
[18] T. Muarata and H. Ishibuchi, Performance evaluation of genetic algorithms for flowshop scheduling problems, *Proc. of the 1st IEEE International Conference on Evolutionary Computation*, pp.812-817, 1994.
[19] T. Murata and H. Ishibuchi, Positive and negative combination effects of crossover and mutation operators in sequencing problems, *Proc. of the IEEE International Conference on Evolutionary Computation*, pp.812-817, 1996.
[20] G. Renner and A. Ekárt, Genetic algorithms in computer aided design, *Computer-Aided Design*, vol.35, no.8, pp.709-726, 2003.
[21] P. Bommel, T. Weide and C. Lucasius, Genetic algorithms for optimal database design, *Information and Software Technology*, vol.36, no.12, pp.725-732, 1994.
[22] D. Whitley, An overview of evolutionary algorithms: Practical issues and common pitfalls, *Information and Software Technology*, vol.43, no.14, pp.817-831, 2001.
[23] Ø. Braaten, O. K. Rødningen, I. Nordal and T. P. Leren, The genetic algorithm applied to haplotype data at the LDL receptor locus, *Computer Methods and Programs in Biomedicine*, vol.61, no.1, pp.1-9, 2000.
[24] Z. Vesna, L. Milica, V. Marina, S. Andjelka and D. Lidija, Correlation between uroflowmetry parameters and treatment outcome in children with dysfunctional voiding, *Journal of Pediatric Urology*, vol.6, no.4, pp.396-402, 2010.
[25] B. A. Erickson, B. N. Breyer and J. W. McAninch, Changes in Uroflowmetry maximum flow rates after urethral reconstructive surgery as a means to predict for stricture recurrence, *The Journal of Urology*, vol.186, no.5, pp.1934-1937, 2011.
[26] E. A. Tanagho and D. Deng, *Smith's General Urology*, 17th Edition, 2008.
[27] E. Oztemel, *Yapay Sinir Aglari*, Papatya Publishers, Istanbul, 2006.
[28] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, New Jersey, 1999.
[29] A. Zaknich, *Neural Networks for Intelligent Signal Processing*, World Scientific, River Edge, NJ, 2003.