International Journal of Scientific Research in Computer Science and Engineering
Research Paper Vol-1, Issue-5 ISSN: 2320-7639

Applying Rough Set Theory for Medical Informatics Data Analysis

M.Durairaj[1*], T.Sathyavathi[2]

[1*,2] *Department of Computer Science, Engineering and Technology, Bharathidasan University,Tamilnadu, India.durairaj_m@bdu.ac.in*

Available online at www.isroset.org

Received: 12 September 2013 Revised: 24 September 2013 Accepted: 18 October 2013 Published: 30 October 2013

Abstract- In the medical field each and every data is important, because these data are very essential for human life. Medical data analysis is a very big and complex task. The medical data consists of imprecise, (or) uncertainty, (or) incomplete data. Therefore the medical data analysis process requires excellent techniques for processing, storing and accessing the datasets. Some of the traditional techniques are available to process the incomplete data and these techniques requires additional information to process the imprecise dataset. In this paper, we propose an intelligent technique of rough set theory for analyzing the imprecise medical data, which could be used for extracting knowledge without changing the knowledge of the original. In comparison to traditional techniques, rough set theory gives the optimal result from the analysis process without loss of information. ROSETTA is a toolkit for analyzing tabular data within the framework of rough set theory that could be applied in the original dataset to compute the reduced set without the loss of the knowledge of the original set. In this paper, the medical data set of recorded information from IVF (in-vitro fertilization) tests are used for data analysis, in which the influential parameters (tests) are identified using Rough Set Theory. The identified influential parameters display the determining impact on the result of IVF treatment (Test tube baby treatment). ROSETTA toolkit used to predict the influential parameters in the IVF treatment.

Keywords: Rough Sets Theory, Medical Data Analysis, ROSETTA tool kit, in-vitro Fertilization**.**

I. INTRODUCTION

Medical data is a very informative and consist of essential knowledge of medical field but data analysis and extracting required knowledge from the medical data set is a difficult task. The medical data analysis process requires advanced techniques for processing, storing and accessing information of the data. Traditional techniques are not capable enough to produce optimal results from incomplete or redundant data through the analysis process. In this work, we applied an intelligent technique called rough set theory that is accustomed for its data processing and delivering a reduced set of the tabular data. Rough set theory is proposed by Z.Pawlak[1] [2] in the late nineteen eighties. Rough set theory produces the optimal result for without loss of information from the original set. The medical data sets of infertilitycouples who undergo infertility treatment are collected from various infertility treatment centers, used in this analysis process. The collected clinical data set consists of various test results of IVF treatment and other relevant factors which affects the success rate of infertility treatment. Infertility is on the rise across the globe and cost of treatment on the rise, which makes the researchers to predict the success rate of IVF treatment prior to the beginning of treatment. This could help the Gynecologist to recommend their patients the types of treatments and procedures. Knowing the success rate of IVF treatment prior to the start of treatment will reduce psychological stress of patients which increase their chances of getting pregnant.

The proposed rough set theory reduction technique produces the optimal reduct se of large data.This reduct set gives the information in which factors affect the fertilization success rate. The ROSETTA software package contains different processing algorithms; in this work we use a Johnson reduction algorithm which produces optimal results without loss of information.

II. REVIEW OF LITERATURE

The objective of this research work was rough set theory used in medical data (In-vitro fertilization data) analysis. The review of literature contents consists of In-vitro fertilization data analysis information. More number of authors published papers on related to fertilization data analysis. Some of the few papers are discussed in this work. These papers are helpful in analysis process and reduced the complexity of analysis task. This review gives the idea of data collection, how to handle the data in analysis process, how to predict the fertility rate for collected data and what are the techniques are used in analysis process.

M.Durairaj and K.Meena presented a novel technique in their paper for predicting fertility rate of animal sperm. In this paper used the key terms Artificial Neural Network (ANN), in-vitro fertilization (IVF) and Artificial

Corresponding Author: M.Durairaj

insemination (AI). The Artificial insemination (AI) is one of the most successful reproductive technologies developed to improve reproductive efficiency. The percentage of Pre-freeze motility the assays of sperm function such as acrosome reaction, zona binding ability, in-vitro fertilization, and in-vitro embryo production are used to predict fertility in the field. In this paper used an optimized Artificial Neural Network (ANN) technique and proper selection algorithm. These techniques are used to predict the fertilization potential rate of animal sperm. The output of the Artificial Neural Network (ANN) results are compared to non-return rates (NRR) obtained from animal reproduction laboratory. The ANN produces the better and valid results compare to non-return rates (NRR) and traditional statistical analysis techniques. Therefore ANN can effectively replace the traditional statistical analysis method for predicting semen fertility rate [3].

M.Durairaj and K.Meena represented an intelligent technique for predicting the semen quality. In this paper used the key terms Artificial Neural Network, Rough sets theory (RST), Fertility rate prediction, IRNNs. This paper used a hybrid prediction system consists of Rough set theory (RST) and Artificial Neural Network (ANN). The Rough set theory (RST) is useful tool for reduce the input to ANN to improve the classification and prediction. ANN technique used to predict the fertility ability. The RST technique used to reduce the input of ANN.These combinations of hybrid system improve the prediction ability, accuracy and reduce the training time. These combinations of hybrid system called IRNNs (Intelligent Rough Neural Network system). This hybrid system predicts the semen quality. The hybrid approach based on the rough sets feature selection mechanism and neural network efficient classification. The combination for this approach is to build more powerful systems that can reduce drawbacks of implementing a single machine learning techniques[4].

M.Durairaj, K.Meena and S.Selvaraju presented a novel technique for predicting the fertility of animal sperm. The objective of this paper is to apply a relatively new data mining approach of rough set theory to analyze in vitro functional parameters to select most significant parameter that can be used to predict cleavage rate of given sperm. The traditional statistical models are not tools for knowledge discovery because of their model assumption of representativeness of the sample and their sensitivity to irrelevant features[5].

The prediction of sperm fertilizing ability has great economic importance for breeding animals when artificial insemination is used. When evaluating semen the ultimate goal is to accurately predict its fertilizing potential. Even after much progress has been made the ability to predict the fertility of semen with laboratory test is still limited mainly due to complexity of the spermatozoon and fertilization process. Since the basic purpose of semen evaluation procedures is to ensure that only good quality and highly fertile semen is used for artificial insemination purposes it is very essential to properly analyze the data and carefully select a parameter or combination of in vitro functional tests parameters that can be used to accurate prediction of fertility of animals.

M.Durairaj and K.Meena presented a novel technique for predicting the fertility rate of animal sperm. The objective of this paper is to apply a relatively new data mining approach of rough set theory and Artificial Neural Network (hybrid) can be directly applied to classification and regression without additional transformation mechanisms in the dataset. In this paper, a new Rough Neural Network (RNN) algorithm for the proposed hybrid approach in order to predict semen fertility rate is described[6].

III. METHODOLOGY

3.1 Rough Set Theory

Rough sets theory methodology is concerned with the classification and analysis of imprecise, uncertain orincomplete information and knowledge and it is considered one of the first non-statistical approaches in data analysis [2].

3.2 Basic Notations Of Rough Set Theory

The basic notations of rough set theory are information system, approximation, and reduction of attributes.

3.2.1 Information System

An information system or information table can be viewed as a table, consisting of objects (rows) and attributes (columns) as shown in Fig. 1. Knowledge representation in rough sets is done via information system which is a tabular form of an OBJECT → ATTRIBUTE VALUE relationship.

More precisely an information system $I = < U, \Omega, V_q, f_q > qc\Omega$, where $U \rightarrow$ is a finite set of objects, $U=\{x_1, x2, x3, \ldots\ldots x_n\}$;

$\Omega \rightarrow$ is a finite set of attributes (features) the attributes in Ω are further classified into disjoint condition attributes A and decision attributes D, $\Omega = AUD$;

For each $q \in \Omega$,

- V_q is a set of attribute values for q,
- Each f_q: $U \rightarrow V_q$ is an information function which assigns particular values from the domains of attributes to objects such that $f_q(x_i) \in V_q$ for all $x_i \in U$ and $q \in \Omega$.

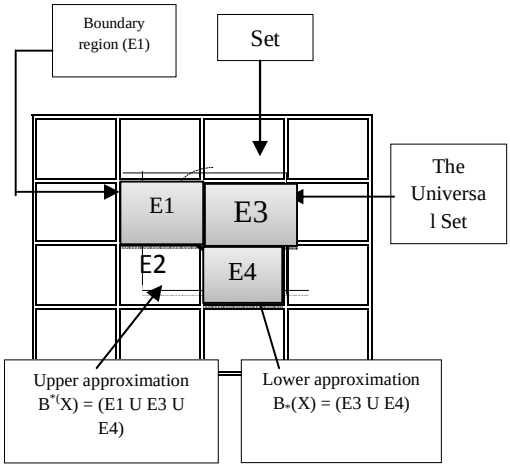


Fig 1. Information system

3.2.2 Indiscernibility Relation

Indiscernibility relation is a central concept in rough set theory and is considered as a relation between two objects or more where all the values are identical in relation to a subset of considered attributes.

3.2.3approximations

An approximation space is an ordered pair A= (U, R), where U is a finite and non-empty set of elements called attributes R is an equivalence relation about U. Any set B⊆ A there is an associated equivalence relation called B-indiscernibility relation.

$$IND_A(B) = \{(x, y) \in U^2 \mid \forall a \in B, a(x) = a(y)\}$$

If $(x, y) \in IND_A(B)$ then x and y are indiscernible from each other by attributes from b. The Indiscernibility is an equivalence relation.

3.2.3.1 Lower Approximation $B_*(X)$

Lower approximation (B_*) is a description of the domain objects that areknown with certainty to belong to the subset of interest.

$$B_*(X) = \cup\{Y \in U \mid IND(P): Y \subseteq X\}$$

3.2.3.2 Upper Approximation $B^*(X)$

Upper approximation is a description of the objects that possibly belong to the subset of interest. The upper approximation of a set X regarding R is the set of all objects which can be possibly classified with X regarding R.

$$B^*(X) = \cup\{Y \in U \mid IND(P): Y \cap X \neq \varphi\}$$

3.2.3.3 Boundary Region

The difference between the upper and lower approximation referred to as boundary region. The B-boundary of X in the information system I, is defined as:

$$BND(X) = B^*(X) - B_*(X)$$

3.2.4 Decision Tables & Decision algorithms

A decision table constrains two types of attributes designated as the condition attribute and decision attribute. Each row of the table determines a decision rule which specifies the decisions (actions) that must be taken when conditions are indicated by condition attributes. The number of consistency rules contained in the decision table known as a factor of consistence, which can be denoted by $\gamma(C, D)$, where C is the condition attribute and D is the decision attributes. If $\gamma(C, D) = 1$, the decision table is consistent but if $\gamma(C, D) \neq 1$ the table of decision is inconsistent.

3.2.5 Dependency Of Attributes

In the analysis of data it is important discover the dependence between attributes. A set of attributes D depends totally on a set of attributes C, denoted as $C \Rightarrow D$ if all values of attributes from D are uniquely determined by values of attributes from C then D depends totally on C. The partial dependency means that only some values of D are determined by values of C. If D and C is subsets of A can be affirmed that D depends on C in degree K ($0 \leq k \leq 1$) denoted $\Rightarrow$kD.And if $k = \gamma(C, D)$. If K=1 D depends totally on C, this dependency denoted byI(C) ⊆ I(D) if $k < 1$ it is said that D depends partially on C. If k =0 the decision attribute D does not depends on condition attributes C.

3.2.6 Reduction Of Attributes

Reduct is a minimum attributes subset that retains the decision attributes a dependence degree to conditional attributes. The subset R⊆B ⊆ A such that $Y_B(Y) = Y_R(Y)$ is called Y-reduct of B and denoted as $Red_Y(B)$. The core is possessed by every legitimate reduct and cannot be removed from the information system without deteriorating basic knowledge of the system. The set of all indispensable attributes of B is called the Y - core. Formally,

$$Core_Y(B) = \cap Red_Y(B)$$

3.2.7 Accuracy

Accuracy measures how much a set is rough. If a set has B(X) = B(X) = X, the set is precise called crisp and for every element $x \in X \in U$. This is expressed by the formula.

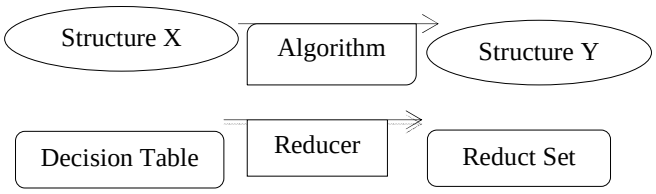
$$\alpha_B(X) = |B_*(X)| / |B^*(X)|$$

When $0 \leq \alpha_B(X) \leq 1$, and if $\alpha_B(X) = 1$ X is crisp with respect to B.

3.1.1ROSETTA (A Rough Set Toolkit For Analysis The Data)

ROSETTA is a toolkit for analyzing tabular data within the framework of rough set theory. ROSETTA is designed to support the overall data mining and knowledgediscovery process: From initial browsing and preprocessing of the data, via computation of minimal attribute sets and generation of if-then rules or descriptive patterns, validation and analysis of the induced rules or patterns. ROSETTA offers a highly intuitive GUI environment where data-navigational abilities are emphasized. In addition to the core features described there, the system includes a large number of algorithms for discretization, reduct computation, and rule pruning and classifier evaluation [7]. The system has two main components: structures and algorithms. Structures

are different data sets such as decision systems, reducts, rules, etc. Algorithms are applied to structures to produce new structures. For example, algorithms for reduct computation are applied to decision tables to produce reducts:
Example:



Steps involved in processing data

- Import/export
- Preprocessing
- Computation
- Post processing
- Validation and analysis

IV. EXPERIMENTAL RESULTS

4.1 Experimental data

The IVF test datasets used in this analysis process contain 23 fields and 114 objects (or) records. The fields are Age (F), Duration of infertility, previous pregnancy, Medical disorders, BMI(F), Endometriosis, Tubal infertility, Ovulatory factor, Hormonal Factor, Cervical Factor, Semen Ejaculate Volume, Liquefaction Time, Sperm Concentration, Sperm Motility, Sperm Vitality, Sperm Morphology, No.of Oocytes Retrieved, No.of Embryos Transferred, Male Factor Only, Female Factor Only and Combined Factor. The rough set methodology used in this analysis process for predicting the IVF success rate potential.

4.2 Analysis steps

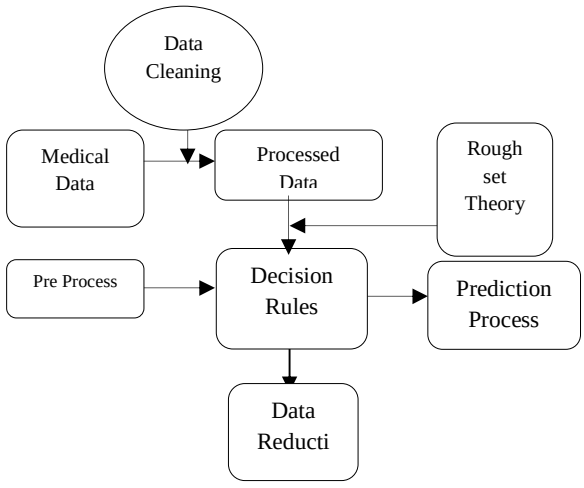


Fig. 2. Analysis steps

The data analysis steps involved in the medical data processing are as illustrated in Fig. 2.

4.3Analysis procedure

This ROSETTA toolkit consists of inbuilt rough set algorithms to process data analysis. The toolkit follows the some important procedures for producing the accurate result. The steps are importing data from any valid data source (excel) format, applying the binary splitting algorithm in the imported data to split the original dataset into training and test data, remove the missing values, finally applying the reduction and classification algorithms. The reduction algorithm is used to compute the reduct set and the classification algorithm is used to reduct rule and compute the classification result.

4.3.1 Reduction process

Johnson reduction algorithm is used in this reduction process since this algorithm produces more accurate results. Chosen Johnson reduction algorithm applied to predict the influential parameters which can be used to estimate the success rate of treatment. Genetic algorithm produces the 86 combination of reduct set where the Johnson reduction algorithm produces 17 combinations of reduct set. Since the Johnson reduction algorithm produces a minimal set of combinations efficiently, this algorithm chosen for reduction process.

Johnson algorithm

Johnson(C,fD)

C, the set of conditional attributes

fD, the discernibility function.

(1) R ←∅; bestc=0;

(2) while(fDnot empty)

(3) for each a∈ Cthat appears in fD

(4) c= heuristic(a)

(5) if(c >bestc)

(6) bestc=c; bestAttr_ a

(7) R ← R ∪a

(8) fD_ removeClauses(fD, a)

(9) returnR

This is a simple greedy heuristic algorithm that is often applied to discernibility functions to find a single reduct. The algorithm begins by setting the current reduct candidate, R, to the empty set. Then, each conditional attribute appearing in the discernibility function is evaluated according to the heuristic measure. For the standard

Johnson algorithm, this is typically a count of the number of appearances an attribute makes within clauses; attributes that appear more frequently are considered to be more significant. The attribute with the highest heuristic value is added to the reduct candidate and all clauses in the discernibility function containing this attribute are removed. As soon as all clauses have been removed, the algorithm terminates and returns the reductR. R is assured to be a reduct as all clauses contained within the discernibility function have been addressed. Variations of the algorithm involve alternative heuristic functions in an attempt to guide the search down better path.

The reduction algorithm results

In Fig. 3, numbers of reduct sets produced through the application of Johnson reduction algorithm are illustrated. The Johnson reduction algorithm produced 17 combinations of reduct sets.

Rosetta - [ivf reduction set]

File Edit View Window Help

	Reduct	Support	Length
1	{AGE(F), NO#OF OOCYTES RETRIEVED}	100	2
2	{DURATION OF INFERTILITY (YEARS)}	100	1
3	{DURATION OF INFERTILITY (YEARS), HORMONAL FACTOR}	100	2
4	{DURATION OF INFERTILITY (YEARS), SPERM CONCENTRATION}	100	2
5	{DURATION OF INFERTILITY (YEARS), TUBAL INFERTILITY}	100	2
6	{AGE(F), MEDICAL DISORDERS}	100	2
7	{DURATION OF INFERTILITY (YEARS), NO#OF EMBRYOS TRANSFERRED}	100	2
8	{AGE(F), STAGES}	100	2
9	{AGE(F), DURATION OF INFERTILITY (YEARS)}	100	2
10	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS}	100	2
11	{DURATION OF INFERTILITY (YEARS), NO#OF OOCYTES RETRIEVED}	100	2
12	{DURATION OF INFERTILITY (YEARS), OVULATORY FACTOR}	100	2
13	{AGE(F), BMI(F)}	100	2
14	{DURATION OF INFERTILITY (YEARS), MALE FACTOR ONLY}	100	2
15	{AGE(F), COMBINED FACTOR}	100	2
16	{SEMEN EJACULATE VOLUME, NO#OF OOCYTES RETRIEVED}	100	2
17	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS}	100	2

Fig 3. Johnson reductset(influential parameters)

Rosetta - [ivf reduct set using genetic]

File Edit View Window Help

	Reduct	Support	Length
11	{DURATION OF INFERTILITY (YEARS), BMI(F), SPERM CONCENTRATION, SPERM VITALITY, NO#OF EMBRYOS TRANSFERRED}	100	5
12	{DURATION OF INFERTILITY (YEARS), OVULATORY FACTOR, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED}	100	5
13	{DURATION OF INFERTILITY (YEARS), BMI(F), SPERM CONCENTRATION, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED}	100	5
14	{DURATION OF INFERTILITY (YEARS), STAGES, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED}	100	5
15	{DURATION OF INFERTILITY (YEARS), SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, MALE FACTOR ONLY}	100	5
16	{DURATION OF INFERTILITY (YEARS), BMI(F), SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED}	100	5
17	{AGE(F), DURATION OF INFERTILITY (YEARS), SPERM CONCENTRATION, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, MALE FACTOR ONLY}	100	6
18	{AGE(F), DURATION OF INFERTILITY (YEARS), SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, COMBINED FACTOR}	100	6
19	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, BMI(F), ENDOMETRIOSIS, HORMONAL FACTOR, NO#OF OOCYTES RETRIEVED}	100	6
20	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, STAGES, HORMONAL FACTOR, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED}	100	6
21	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, BMI(F), SPERM VITALITY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	6
22	{DURATION OF INFERTILITY (YEARS), BMI(F), OVULATORY FACTOR, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED}	100	6
23	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, CERVICAL FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF EMBRYOS TRANSFERRED}	100	6
24	{DURATION OF INFERTILITY (YEARS), OVULATORY FACTOR, HORMONAL FACTOR, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED, MALE FACTOR ONLY}	100	6
25	{DURATION OF INFERTILITY (YEARS), OVULATORY FACTOR, HORMONAL FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF EMBRYOS TRANSFERRED}	100	6
26	{DURATION OF INFERTILITY (YEARS), BMI(F), STAGES, SPERM CONCENTRATION, SPERM VITALITY, NO#OF OOCYTES RETRIEVED}	100	6
27	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, HORMONAL FACTOR, SPERM CONCENTRATION, NO#OF OOCYTES RETRIEVED, COMBINED FACTOR}	100	6
28	{DURATION OF INFERTILITY (YEARS), BMI(F), HORMONAL FACTOR, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED}	100	6
29	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, TUBAL INFERTILITY, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, COMBINED FACTOR}	100	6
30	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, COMBINED FACTOR}	100	6
31	{DURATION OF INFERTILITY (YEARS), BMI(F), TUBAL INFERTILITY, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED}	100	6
32	{DURATION OF INFERTILITY (YEARS), CERVICAL FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF EMBRYOS TRANSFERRED, FEMALE FACTOR ONLY}	100	6
33	{DURATION OF INFERTILITY (YEARS), BMI(F), ENDOMETRIOSIS, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED}	100	6
34	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY, COMBINED FACTOR}	100	6
35	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, HORMONAL FACTOR, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, COMBINED FACTOR}	100	6
36	{DURATION OF INFERTILITY (YEARS), HORMONAL FACTOR, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, COMBINED FACTOR}	100	6
37	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, HORMONAL FACTOR, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED}	100	6
38	{DURATION OF INFERTILITY (YEARS), CERVICAL FACTOR, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED, COMBINED FACTOR}	100	6
39	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	6
40	{DURATION OF INFERTILITY (YEARS), BMI(F), OVULATORY FACTOR, HORMONAL FACTOR, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED}	100	6
41	{DURATION OF INFERTILITY (YEARS), BMI(F), ENDOMETRIOSIS, SPERM CONCENTRATION, SPERM VITALITY, NO#OF OOCYTES RETRIEVED}	100	6
42	{DURATION OF INFERTILITY (YEARS), HORMONAL FACTOR, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	6
43	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, HORMONAL FACTOR, SPERM MOTILITY, SPERM VITALITY, NO#OF OOCYTES RETRIEVED}	100	6
44	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, HORMONAL FACTOR, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	6
45	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, SPERM MOTILITY, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY}	100	6
46	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, STAGES, HORMONAL FACTOR, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	6
47	{DURATION OF INFERTILITY (YEARS), CERVICAL FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSF	100	7
48	{DURATION OF INFERTILITY (YEARS), CERVICAL FACTOR, LIQUEFACTION TIME, SPERM MOTILITY, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRE	100	7
49	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, HORMONAL FACTOR, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY, FEMALE FACTOR	100	7
50	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, ENDOMETRIOSIS, HORMONAL FACTOR, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	7
51	{DURATION OF INFERTILITY (YEARS), HORMONAL FACTOR, CERVICAL FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRA	100	7
52	{DURATION OF INFERTILITY (YEARS), BMI(F), STAGES, LIQUEFACTION TIME, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY}	100	7
53	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, CERVICAL FACTOR, SPERM CONCENTRATION, SPERM VITALITY, NO#OF EMBRYOS TRANSFERRED, COMBINED FACT	100	7
54	{DURATION OF INFERTILITY (YEARS), TUBAL INFERTILITY, LIQUEFACTION TIME, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, CO	100	7
55	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, BMI(F), SPERM CONCENTRATION, SPERM MOTILITY, NO#OF EMBRYOS TRANSFERRED, COMBINED FACTOR}	100	7
56	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, BMI(F), ENDOMETRIOSIS, CERVICAL FACTOR, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED}	100	7
57	{DURATION OF INFERTILITY (YEARS), STAGES, OVULATORY FACTOR, SPERM CONCENTRATION, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, COMBINED FACTOR}	100	7
58	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, STAGES, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY, FEMALE FACTOR ONLY}	100	7
59	{DURATION OF INFERTILITY (YEARS), TUBAL INFERTILITY, CERVICAL FACTOR, SPERM CONCENTRATION, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED,	100	7
60	{DURATION OF INFERTILITY (YEARS), BMI(F), STAGES, SPERM CONCENTRATION, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY}	100	7
61	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, ENDOMETRIOSIS, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, FEMALE FACTOR ONLY, CO	100	7
62	{AGE(F), DURATION OF INFERTILITY (YEARS), STAGES, SPERM CONCENTRATION, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY, COMBINED FACTOR}	100	7
63	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, ENDOMETRIOSIS, SEMEN EJACULATE VOLUME, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED, FEMALE FACT	100	7
64	{DURATION OF INFERTILITY (YEARS), TUBAL INFERTILITY, OVULATORY FACTOR, SEMEN EJACULATE VOLUME, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF OOCYT	100	7
65	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, OVULATORY FACTOR, HORMONAL FACTOR, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED, FEMALE FACTO	100	7
66	{DURATION OF INFERTILITY (YEARS), BMI(F), SPERM MOTILITY, SPERM VITALITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED, FEMALE FACTOR ONLY}	100	7
67	{MEDICAL DISORDERS, BMI(F), SPERM MOTILITY, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, MALE FACTOR ONLY}	100	7
68	{DURATION OF INFERTILITY (YEARS), BMI(F), STAGES, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF EMBRYOS TRANSFERRED, COMBINED FACTOR}	100	7
69	{DURATION OF INFERTILITY (YEARS), BMI(F), TUBAL INFERTILITY, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF EMBRYOS TRANSFERRED, FEMALE FACTOR ONLY}	100	7
70	{DURATION OF INFERTILITY (YEARS), TUBAL INFERTILITY, OVULATORY FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED, MALE FACTOR	100	7
71	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, CERVICAL FACTOR, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, NO#OF EMBRYOS TRANSFERRED, COMBI	100	7
72	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, STAGES, OVULATORY FACTOR, HORMONAL FACTOR, NO#OF OOCYTES RETRIEVED, COMBINED FACTOR}	100	7
73	{AGE(F), DURATION OF INFERTILITY (YEARS), OVULATORY FACTOR, HORMONAL FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, NO#OF OOCYTES RETRIEVED}	100	7
74	{AGE(F), DURATION OF INFERTILITY (YEARS), BMI(F), HORMONAL FACTOR, SPERM CONCENTRATION, SPERM VITALITY, NO#OF OOCYTES RETRIEVED}	100	7
75	{DURATION OF INFERTILITY (YEARS), OVULATORY FACTOR, SPERM CONCENTRATION, SPERM MOTILITY, SPERM VITALITY, NO#OF EMBRYOS TRANSFERRED, COMBINED F	100	7
76	{DURATION OF INFERTILITY (YEARS), TUBAL INFERTILITY, OVULATORY FACTOR, HORMONAL FACTOR, SPERM CONCENTRATION, NO#OF OOCYTES RETRIEVED, MALE FAC	100	7
77	{AGE(F), DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, BMI(F), SPERM VITALITY, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY}	100	7
78	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, BMI(F), ENDOMETRIOSIS, TUBAL INFERTILITY, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY}	100	7
79	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, OVULATORY FACTOR, HORMONAL FACTOR, LIQUEFACTION TIME, NO#OF OOCYTES RETRIEVED, MALE FACTO	100	7
80	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, LIQUEFACTION TIME, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	7
81	{DURATION OF INFERTILITY (YEARS), ENDOMETRIOSIS, OVULATORY FACTOR, HORMONAL FACTOR, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED, COMBINED	100	7
82	{DURATION OF INFERTILITY (YEARS), MEDICAL DISORDERS, TUBAL INFERTILITY, SPERM MOTILITY, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, MALE FACTOR ONLY}	100	7
83	{DURATION OF INFERTILITY (YEARS), STAGES, LIQUEFACTION TIME, SPERM CONCENTRATION, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, COMBINED FACTOR}	100	7
84	{DURATION OF INFERTILITY (YEARS), LIQUEFACTION TIME, SPERM CONCENTRATION, SPERM VITALITY, NO#OF OOCYTES RETRIEVED, FEMALE FACTOR ONLY, COMBINED F	100	7
85	{DURATION OF INFERTILITY (YEARS), BMI(F), SPERM CONCENTRATION, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED, COMBINED FACTOR}	100	7
86	{DURATION OF INFERTILITY (YEARS), BMI(F), SPERM CONCENTRATION, SPERM MOTILITY, SPERM MORPHOLOGY, NO#OF EMBRYOS TRANSFERRED, FEMALE FACTOR ONLY	100	7

Fig 4.Reduct sets produced by Genetic algorithm

The reduct sets produced by Genetic algorithm are depicted in Fig.3, where the genetic algorithm produces 86 combinations of reduct sets.

4.3.2 Results and Discussion

The reduction rule explains the rule support, stability, length, coverage and accuracy. Each row of the reduction rule is called descriptors (Attribute→ value). The left hand side of the rule is called the antecedent and right hand side of the rule is called consequent. This reduction rule result used in the classification process. This rule is used to make the confusion matrix. The reduction rules generated are given in Fig. 5 and confusion matrix generated is given in Fig.6.

Rosetta - [ivf reduction set rule]

File Edit View Window Help

	Rule	LHS Support	RHS Support	RHS Accuracy	LHS Coverage	RHS Coverage	RHS Stability	LHS Length	RHS Length
1	AGE(F)(25) AND NO#OF OOCYTES RETRIEVED(4) => IVF TREATMENT(SUCCESS)	2	2	1.0	0.035088	0.074074	1.0	2	1
2	AGE(F)(25) AND NO#OF OOCYTES RETRIEVED(14) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1
3	AGE(F)(25) AND NO#OF OOCYTES RETRIEVED(10) => IVF TREATMENT(UNSUCCESS)	2	2	1.0	0.035088	0.066667	1.0	2	1
4	DURATION OF INFERTILITY (YEARS)(11) => IVF TREATMENT(SUCCESS)	2	2	1.0	0.035088	0.074074	1.0	1	1
5	DURATION OF INFERTILITY (YEARS)(13) => IVF TREATMENT(UNSUCCESS)	2	2	1.0	0.035088	0.066667	1.0	1	1
6	DURATION OF INFERTILITY (YEARS)(10) => IVF TREATMENT(UNSUCCESS)	7	7	1.0	0.122807	0.233333	1.0	1	1
7	DURATION OF INFERTILITY (YEARS)(24) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	1	1
8	DURATION OF INFERTILITY (YEARS)(6) => IVF TREATMENT(SUCCESS)	3	3	1.0	0.052632	0.111111	1.0	1	1
9	DURATION OF INFERTILITY (YEARS)(15) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	1	1
10	DURATION OF INFERTILITY (YEARS)(1) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	1	1
11	DURATION OF INFERTILITY (YEARS)(4) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	1	1
12	DURATION OF INFERTILITY (YEARS)(9) => IVF TREATMENT(UNSUCCESS)	2	2	1.0	0.035088	0.066667	1.0	1	1
13	DURATION OF INFERTILITY (YEARS)(30) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	1	1
14	DURATION OF INFERTILITY (YEARS)(18) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	1	1
15	DURATION OF INFERTILITY (YEARS)(8) AND HORMONAL FACTOR(H1) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1
16	DURATION OF INFERTILITY (YEARS)(3) AND HORMONAL FACTOR(H1) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	2	1
17	DURATION OF INFERTILITY (YEARS)(14) AND SPERM CONCENTRATION(17) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1
18	DURATION OF INFERTILITY (YEARS)(7) AND SPERM CONCENTRATION(17) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	2	1
19	DURATION OF INFERTILITY (YEARS)(12) AND TUBAL INFERTILITY(T0) => IVF TREATMENT(SUCCESS)	2	2	1.0	0.035088	0.074074	1.0	2	1
20	AGE(F)(25) AND MEDICAL DISORDERS(M3) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	2	1
21	AGE(F)(35) AND MEDICAL DISORDERS(M3) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1
22	DURATION OF INFERTILITY (YEARS)(12) AND NO#OF EMBRYOS TRANSFERRED(4) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	2	1
23	AGE(F)(25) AND STAGES(15) => IVF TREATMENT(SUCCESS)	2	2	1.0	0.035088	0.074074	1.0	2	1
24	AGE(F)(25) AND DURATION OF INFERTILITY (YEARS)(3) => IVF TREATMENT(SUCCESS)	4	4	1.0	0.070175	0.148148	1.0	2	1
25	DURATION OF INFERTILITY (YEARS)(8) AND ENDOMETRIOSIS(1) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	2	1
26	DURATION OF INFERTILITY (YEARS)(12) AND ENDOMETRIOSIS(1) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1
27	DURATION OF INFERTILITY (YEARS)(2) AND NO#OF OOCYTES RETRIEVED(2) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1
28	DURATION OF INFERTILITY (YEARS)(5) AND NO#OF OOCYTES RETRIEVED(2) => IVF TREATMENT(UNSUCCESS)	1	1	1.0	0.017544	0.033333	1.0	2	1
29	DURATION OF INFERTILITY (YEARS)(7) AND OVULATORY FACTOR(OV1) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1
30	AGE(F)(25) AND BMI(F)(29.9) => IVF TREATMENT(SUCCESS)	2	2	1.0	0.035088	0.074074	1.0	2	1
31	DURATION OF INFERTILITY (YEARS)(5) AND MALE FACTOR ONLY(Y) => IVF TREATMENT(UNSUCCESS)	2	2	1.0	0.035088	0.066667	1.0	2	1
32	AGE(F)(35) AND COMBINED FACTOR(C) => IVF TREATMENT(UNSUCCESS)	5	5	1.0	0.087719	0.166667	1.0	2	1
33	SEMEN EJACULATE VOLUME(2) AND NO#OF OOCYTES RETRIEVED(5) => IVF TREATMENT(UNSUCCESS)	2	2	1.0	0.035088	0.066667	1.0	2	1
34	DURATION OF INFERTILITY (YEARS)(12) AND MEDICAL DISORDERS(M1) => IVF TREATMENT(SUCCESS)	1	1	1.0	0.017544	0.037037	1.0	2	1

Fig 5. Reduction rule generated

4.3.3 Classification process

In ROSETTA classification algorithm, the reduction rule is used for the classification process.The reduction rule helps to produce the classification result.

Rosetta - [ivf data classification result]

File Edit View Window Help

		Predicted		
		SUCCESS	UNSUCCESS	
Actual	SUCCESS	17	4	0.809524
	UNSUCCESS	26	10	0.277778
		0.395349	0.714286	0.473684
ROC	Class	UNSUCCESS		
	Area	0.511905		
	Std. error	0.079766		
	Thr. (0, 1)	0.604		
	Thr. acc.	0.0		

Fig 6. Confusion matrix generated

Confusion matrix (or) contingency table (or) error matrix:

A confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes (commonly mislabeling one as another). The confusion matrix is often called the contingency table or the error matrix (see Fig. 7).

A confusion matrix is a table with two rows and two columns that reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN). This allows more detailed analysis than the mere proportion of correct guesses (accuracy). Accuracy is not a reliable metric for the real performance of a classifier, because it will yield misleading results if the data set is unbalanced.

		Predicted	
		-	+
Actual	-	TN	FP
	+	FN	TP

Fig. 7. Confusion Matrix

Sensitivity (true positive rate (TPR)) = TP / (TP + FN)
1-Specificity (false positive rate (TNR)) = TN / (TN + FP)
PPV = TP / (TP + FP)
NPV = TN / (TN + FN)
Accuracy = (TP + TN) / (TP + FP + TN + FN)
The classification result TPR and FPR calculation
Sensitivity for actual=TP/(TP+FN)=17/(17+4)= 0.809524
Specificity for actual =TN/(TN+FP)=10/(10+26)= 0.277778
Sensivity for predicted =17/(17+26)= 0.395349
Specificity for predicted =10/(10+4)= 0.714286
Accuracy=(TP+TN)/TP+FP+TN+FN=0.473684

V. RESULT &DISCUSSIONS

Influential parameters obtained through applying rough set theory are AGE (F), BMI (F), DURATION OFFERTILITY (YEARS), ENDOMETRIOSIS, STAGES, HORMONALFACTOR, MEDICAL DISORDERS, TUBAL INFERTILITY, OVULATORY FACTOR, SPERM EJACULATE VOLUME, SPERMCONCERTRATION, COMBINED FACTOR.

These influential parameters show the list of IVF tests and information which has an impact on determining the success rate of IVF treatment on particular patients. A receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the positives (TPR = true positive rate) vsthe fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings. This ROC curve used in false positive rate of batch classifier and standard voting result vs. true positive rate of batch classifier and standard voting result. The comparison shows that the Standard voting classifier result more optimal than Batch classifier performance.

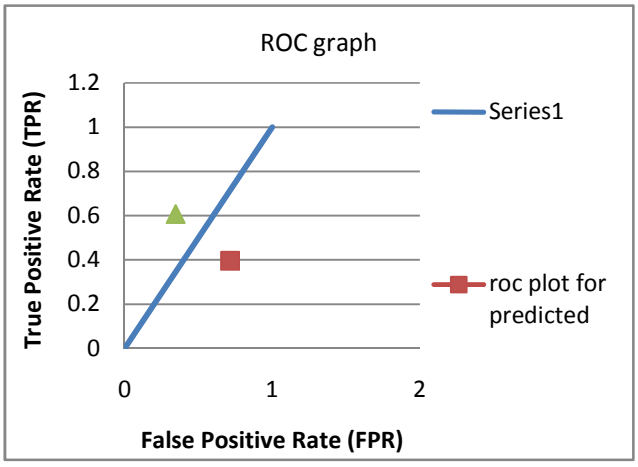


Fig. 8. ROC graph for comparing two classifiers

Comparing the two classifier results, the false standard voting classifier result is more accurate than the batch classifier performance in the same threshold value (See Fig. 8).

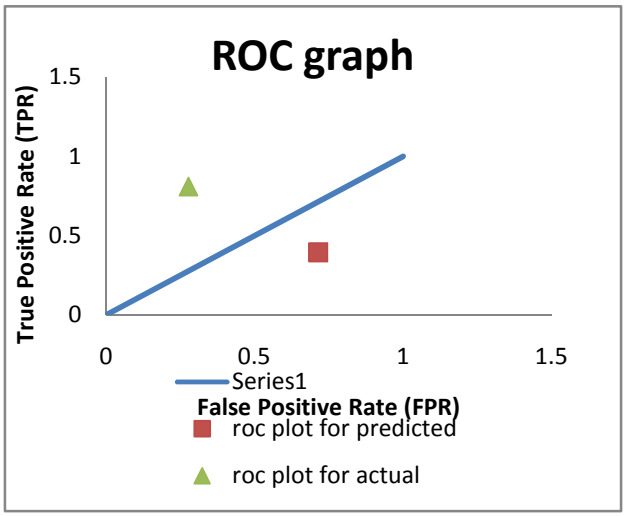


Fig. 9. Comparison between actual and predicted value

In Fig. 9, the graph shows the result of the difference between the actual and predicted values of the standard voting classifier. The classifier prediction shows the accuracy of 50%.

5.1Classification accuracy graph

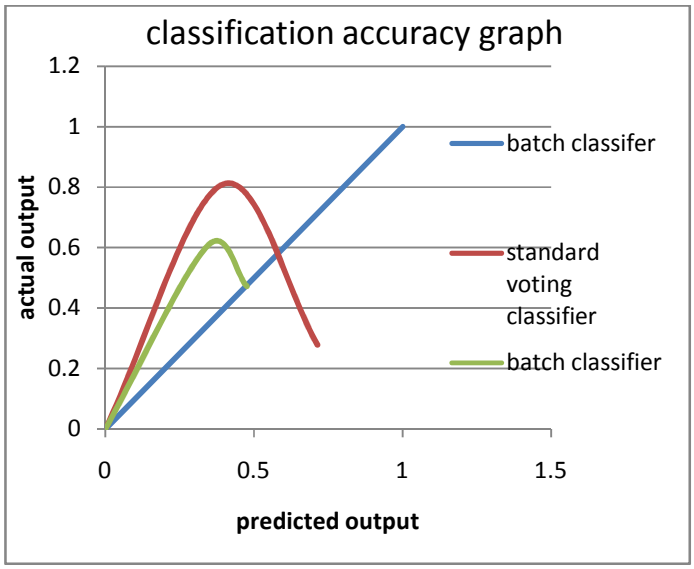


Fig. 10. Classification accuracy graph

Line curve indicates the results produces by the standard voting classifier. The dotted curve shows the results produced by batch classifier. The Line curve is close to accurate. Compare to standard voting and batch classifier results, the standard voting classifier produces the accurate result. Because the perfect classifier produces the result (1,1). The above graph results, the standard voting graph result closed to perfect classification result. Therefore the standard voting classifier is a better classifier in this application. The classification accuracy graph is illustrated in Fig. 10.

6. CONCLUSION

Rough set is an efficient tool for handling the large amounts of data and extracting useful information. When comparing with other traditional technique, rough set produces the optimal result for large quantities of data. In this research work, rough set reduction technique is used to computes the optimal reduct set without changing the knowledge of the original set. In the experiments, in vitro fertilization medical datasets are used in this analysis process, the reduction algorithm produces the result factors which affect the success rate of the IVF treatment. ROSETTA toolkit is implemented in this analysis process. In ROSETTA toolkit, the Johnson reduction algorithm used in this analysis process and to predict the optimal reduct set. The experimental results show that the rough set theory is an efficient tool for identifying influential parameters in determining the success rate of IVF treatment and this technique could be useful for Doctors to prescribe the treatment type to the infertility patients.

REFERENCES

[1]. Z. Pawlak. (1991). Rough Sets - Theoretical Aspect of Reasoning about Data, Kluwer Academic Publishers.

[2]. Z. Pawlak. (1982). "Rough Sets", International Journal of Computer and Information Sciences, Vol.11.

[3]. M. Durairaj and K.Meena"Application of Artificial Neural Network for Predicting Fertilization Potential of Frozen Spermatozoa of Cattle and Buffalo" "International Journal of Computer and Information Sciences" (June 2008).

[4]. M. Durairaj, K.Meena and K.R.Subramanian "IRNNS: A Hybrid Prediction System for Medical Database", National Journal of System and Information Technology (Dec 2008).

[5]. M. Durairaj, K.Meena and S.Selvaraju,"Applying a Data Mining Approach of Rough Sets on Spermatological Data Analysis as Predictors of In-Vitro Fertility of Bull Semen", International Journal of Computer, Mathematical Sciences and Application (September 2008).

[6]. M. Durairaj and K.Meena ,"A Hybrid Approach of Neural Network and Rough Set Theory for Prediction of Fertility Rate From IVF Outcomes", The Icfai University Journal of Science & Technology (2009).

[7]. AleksanderØhrn, AleksanderØhrn ,Department of Computer and Information Science, Norwegian University of Science and Technology,N-7491 Trondheim, Norway "Discernibility and Rough Sets in Medicine: Tools and Applications"

[8]. Torgeir R. Hvidsten"A tutorial-based guide to the ROSETTA system : A Rough Set Toolkit for Analysis of Data" Edition 1: May, 2006 Edition 2: April, 2010

[9]. A. Øhrn, J. Komorowski, A. Skowron, P. Synak (1998), 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, L. Polkowski and A. Skowron (eds.), Studies in Fuzziness and Soft Computing.

[10]. M. Durairaj and K.Meena"Intelligent Classification using Rough Sets and Neural Networks" "The IcfaiJournal of Information Technology" (Dec 2007).

[11]. M. Durairaj, K.Meena and K.R.Subramanian"Machine Learning Techniques to Predict Fertility Rate of Sperm from the Outcome of IVF Functional Tests" "The IcfaiJournal of Information Technology" (March 2009).

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

[13]. A.Øhrn(2000), ROSETTA Technical Reference Manual, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

AUTHOR'S PROFILES

He is currently working as an Assistant Professor, Dept. of Computer Science, Engineering &Technology, Bharathidasan University, Trichy, Tamilnadu, India. He completed his Ph.D. in Computer Science as a full time research scholar at Bharathidasan University on April, 2011. Prior to that, he received a master degree (M.C.A.) in 1997 and bachelor degree (B.Sc. in Computer Science) in 1993 from Bharathidasan University. Prior to this assignment of Assistant Professor in Computer Science at Bharathidasan University, he was working as a Computer Programmer at National Research Centre on Rapeseed-Mustard (Indian Council of Agricultural Research), Rajasthan, India, and as a Technical Officer (Computer Science) at the National Institute of Animal Nutrition and Physiology (ICAR), Bangalore for 12 years. He has published 20 research papers in both national and international journals. His areas of interest include Artificial Neural Network, Soft Computing, Rough Set Theory and Data Mining.

She received the Master Degree (M.Sc) in 2012 from Bharathidasan University, Trichy and Bachelor degree (B.Sc) in 2009 in the same University. She is currently pursuing as aM.Phil Research Scholar in the Department of Computer Science, Engineering and Technology at the Bharathidasan University. Her areas of interest are Rough Set Theory and Data Mining.

