International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 8 01 – 11

Feature Diminution by Ant Colonized Relative Reduct Algorithm for improving the Success Rate for IVF Treatment

Dr. M. Durairaj[1]
Assistant Professor
School of Comp. Sci., Engg, & Applications,
Bharathidasan University,
Tiruchirappalli – 620 023

Nandhakumar Ramasamy[2]
Research Scholar
School of Comp. Sci., Engg, & Applications,
Bharathidasan University,
Tiruchirappalli – 620 023
E-mail – Nandhakumar.897@gmail.com

Abstract: Infertility is the most common problem faced by today's generation. The factors like environment, genetic or personal characteristics are responsible for these problems. Different infertility treatments like IVF, IUI etc are used to treat those infertile people. But the cost and emotions beyond each and every cycle of IVF treatment is very high and also the success rate differs from person to person. So, there is a need to find a system which would predict the outcome of IVF to motivate the people both in psychologically and financially. Many Data Mining techniques are applied to predict the outcome of the IVF treatment. Reducing the unwanted features which affects the quality of result is one of the significant tasks in Data Mining. This paper proposes a hybrid algorithm named Ant Colonized Relative Reduct Algorithm (ACRRA) which combines the core features of Ant Colony Optimization Algorithm and Relative Reduct Theory for Feature Reduction. In this work, the proposed Algorithm is compared with the existing related algorithms. It is evident from the results that the proposed algorithm achieved its target of reducing the features to minimum numbers without compromising the core knowledge of the system to estimate the success rate.

Keywords*: IVF, Data Mining, Ant colony optimization, Relative Reduct Algorithm, Hybridization*

*****

I. INTRODUCTION

Infertility is distressing a growing number of married couples around the planet. Assisted Reproductive Technologies (ART) and In Vitro Fertilization (IVF) are the effective ways to address the problem of infertility. In IVF procedure, several eggs are collected from the woman's ovaries and fertilized with the donor's sperm to produce embryos. The best one among the embryos will be transferred to the woman's uterus, where as the biggest problem arises here since the best embryo is chosen based on recorded features, characterizing the morphology, oocytes, follicle and the sperm sample. The success rate achieved by this treatment has been increased recently up to 10%. But it still fits only to the 40% of the people [1]. It is a complicated task for an embryologist to analyze and correlate the features, since it involves a number of features which are sufficiently large. Even though there are processes customized for each case to improve the success rate in fertility and there is still some technology lagging behind to achieve it. Hence in such a case, there arises a compelling need to appeal for more advanced methods like Data Mining and Artificial Intelligence. This field is falling sort of an automated tool which can intelligently analyze the stage of IVF treatment, patient's demography and other parameters. The practitioners at IVF centres have begun to felt the absence of an expert system as a data processing tool to help them. Feature Reduction is a significant task in Data Mining since it removes the irrelevant or redundant features without loss of much information. Furthermore, a Feature Reduction technique will reduce the amount of time taken for number of tests and the predicted success rate of the treatment can be used for the patients to psychologically strengthen them knowing that their success rate is going to be positive. To achieve this objective, this paper develops a hybrid model which will reduce the maximum features to minimum number to increase the accuracy of the success rate estimation in the held at IVF clinics.

The paper is organized as follows: Section II discusses some of the existing work carried out in predicting the success rate of IVF treatment. Section III briefs the data set used for the experimentation. Section IV describes the proposed Algorithm with its framework. Section V discusses the results obtained and the paper is concluded in Section VI.

II. LITERATURE REVIEW

S.J.Kaufmann et al. [2] applied Neural Network to predict the outcome of the IVF treatment. A total of 8 different types of Neural Network is applied on the same dataset. A sensitivity of 0.55%, Specificity 0.68% and accuracy of 59% was obtained by applying Neural Network. Asli Uyar et al. [3] applied Naïve Bayes Classifier to the

dataset to classify the embryos. In order to overcome the problem arising due to the imbalanced dataset, the author analyzed the effects of oversampling, under sampling and change in threshold. The value of 0.3 is found the perfect threshold value for the correct classification of embryos. True Positive Rate 64.4% and False Positive Rate 30.6% are obtained.

Asli Uyar et al. [4] evaluated six dissimilar methods like Naïve Bayes, K- Nearest Neighbor, Decision Tree, Support Vector Machine, Multi-layer Perceptron and Radial Basis Function Network for envisaging the results for embryo implantation. Among all the six methods, Naïve Bayes and Radial Basis Function Network were observed to function better. Durairaj et al. [5] implemented a hybrid system for predicting the infertility treatment based on Rough Set and Artificial Neural Network. Rough Set Theory was used to find the reduct set. It was used as a pre processing tool to reduce the number of variables which are used as input for the Neural Network. This system worked in a better way for large and medium size of medical data. David Gil et al. [6] compared three techniques, Decision Tree, Multi Layer Perceptron and Support Vector Machine to evaluate the male partner. Clinicians obtained the data from semen analysis and compared it with the corresponding reference value. An accuracy of 86% was obtained from Multilayer Perceptron Network and Support Vector Machine. M. Durairaj et al. [7] illustrates the process of applying data mining techniques for identifying influential tests for infertility couples to determine the success rate of IVF treatment. The data set are pre processed to select only most influential parameters using attribute selection algorithm, which filters the noisy data and selected only the parameters with high impact factors. The experimental results show that the filter and classifier tool using data mining techniques employed to evaluate and produce the minimum set of data which have most influence on estimating the success rate of IVF treatment. In this paper, a data mining method of data analysis, classification is proposed for the InVitro Fertilization data analysis, and Multilayer Perceptron Network for classification or prediction.

Artificial Neural Network [8] was used for predicting the fertility success rate based on the IVF data. An accuracy of 73% was obtained. Claudio Manna et al. [9] applied Artificial Intelligence for classifying the embryo and oocytes. An integrated method based on Artificial Neural Network and Rough Set Theory [10] was adopted for analyzing the IVF data. The Rosetta tool is used for analyzing the data. An accuracy of 90% was obtained by using the integrated method [11].

III. DATA SET

The data set used for the experimentation is collected from various Fertility clinics, Hospitals and Research centres in Tamil Nadu. This data set has 42 attributes. Among all the 42 attributes, 34 attributes is taken for the experiments based on the doctor's suggestion.

Attributes used for this work				
Name	Previous Surgery	Endometriosis	Liquefaction Time	Male Factor Only
Unknown Factor	Pre-Existing Symptoms Of Depression	Tubal Infertility	Sperm Concentration	Severe Male Factor
Place	Fear And Negative Treatment Attitude	Ovulatory Factor	Sperm Motility	Female Factor Only
IVF Treatment	Psychological And Emotional Factors	Hormonal Factor	Sperm Vitality	Combined Factor
Miscarriage	Difficulty In Tolerating Negative Emotions For Extended Time	Cervical Factor	Sperm Morphology	Unknown Factor
Miscarriage Causes	Uncertainty	Unexplained Factor	No.of Oocytes Retrieved	Place
Medical Disorders	Strain Of Repeated Treatment	Semen Ejaculate Volume	No.of Embryos Transferred	IVF Treatment

Table 1: Attributes used for this work

The list of attributes given in Table 2 is taken for reduction process based on doctor's suggestion.

Age
Endometriosis
Ovulatory Factor
Hormonal Factor
Cervical Factor
Unexplained 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
Severe male factor
Female factor only
Combined factor
IVF Treatment

Table 2: List of attributes chosen for experimentation

IV. ANT COLONIZED RELATIVE REDUCT ALGORITHM (ACRRA)

At the initial stage the Pheromone and R values are initialized. The ants are created using the attributes. A solution is constructed for each Ant. A Feature Subset with Conditional Features C is selected. Then the Conditional Features are stored in R. After storing the Conditional Features, the dependency of each attribute is checked. If the dependency is equal to one, that attribute is eliminated. The remaining features are taken and stored in R. If the condition is not satisfied, then the process is repeated until the reduct set is obtained. Instead of indescernabiliity matrix in Rough Set Theory, this dependency measure is taken as a new technique. After obtaining the Reduct data set and stored in R, the best ant table is updated. The Final termination criteria are checked. If the termination criteria are reached, the optimal data is stored and the process is stopped. Else, the pheromone level is updated and the process starts from initialization.

Algorithm:

Ant Colonized Relative Reduct Algorithm ACRRA(C-> Conditional Features, D-> Decision Features)

Input: Data set

Algorithm:

Step 1: *Initialize Phermone, R*

Step 2: *Create Ants*

Step 3: *Construct a Solution for each Ant*

Step 4: *Select Feature Subset*

Step 5: $R \leftarrow C$

Step 6: $\forall_a \ \varepsilon C$

Step 7: *If* $K_{R\{a\}}(D) == 1$

$R \leftarrow R\{a\}$

Go to Step 8

Else

Go to Step 5

Step 8: *Update the best Ant Value*

Step 9: *Check for Termination Criteria*

If yes go to Step 11

Else go to step 10

Step 10: *Update the Phermone level*

Go to step 3

Step 11: *End*

Output: Optimal Reduct Data set == R

Figure 1 depicts the framework of the proposed ACRR Algorithm

 ISSN: 2321-8169
01 – 11

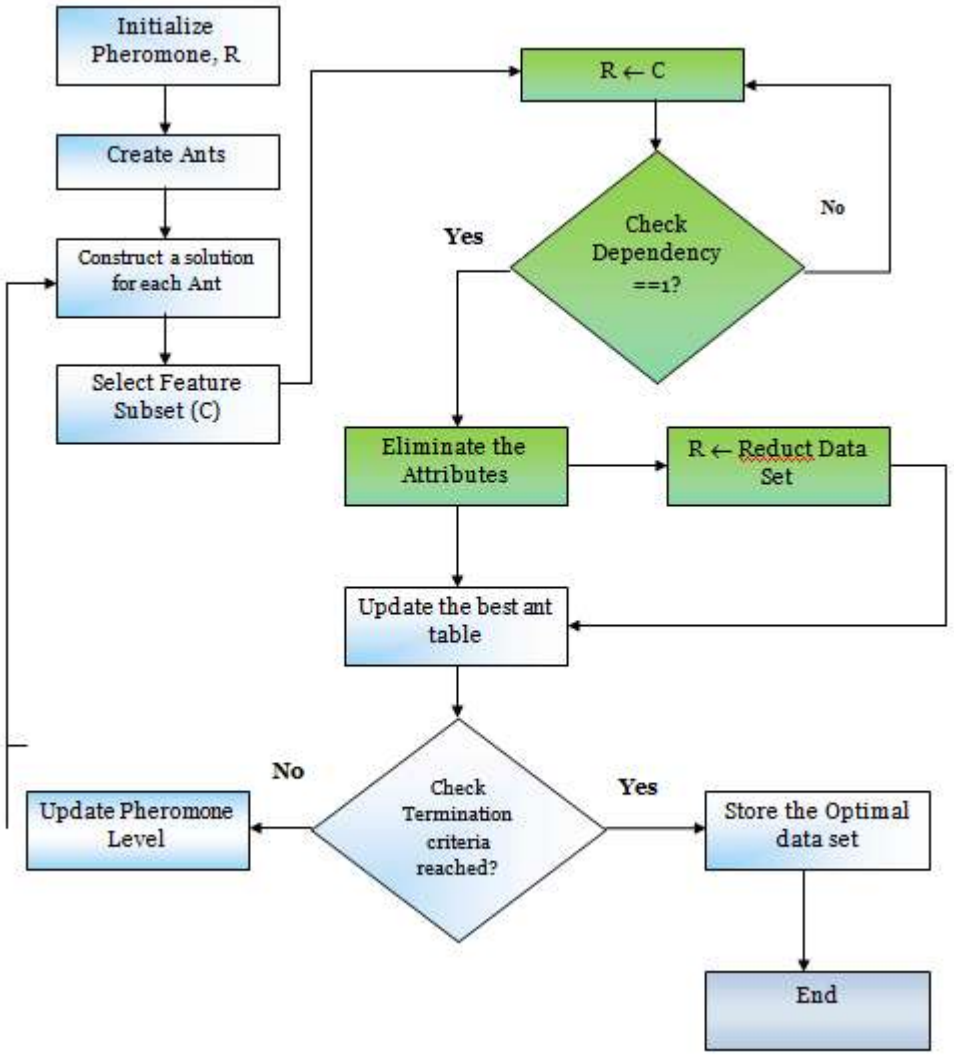

Figure 1 : The Framework of ACRRA

V. RESULTS AND DISCUSSION

The Existing Algorithms like Genetic Algorithm (GA), Ant Colony Optimization (ACO) Algorithm, Particle Swarm Optimization (PSO) Algorithm, Relative Reduct (RR) Algorithm and Quick Reduct (QR) Algorithm are taken for study with the proposed Ant Colony Relative Reduct (ACRR) Algorithm.

A total of 18 Features is selected from 41 Features after applying ACRRA. The selected features are tested with the respective algorithms. The reduced Features are listed in Table 2. .

Original Attributes	GA	ACO Algorithm	PSO Algorithm	RR Algorithm	QR Algorithm	ACRR Algorithm
Age	✓	✓	✓		✓	✓
Endometriosis	✓	✓	✓	✓	✓	✓
Ovulatory Factor	✓	✓	✓	✓	✓	✓
Hormonal Factor	✓	✓	✓	✓	✓	
Crevical Factor	✓	✓	✓	✓	✓	✓

Unexplained 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	✓	✓		✓		✓
Severe male factor						
Female factor only	✓	✓		✓		
Combined factor						
IVF Treatment	✓	✓	✓	✓	✓	✓

Table 3: List of Attributes Reduced by Existing and Proposed Algorithm

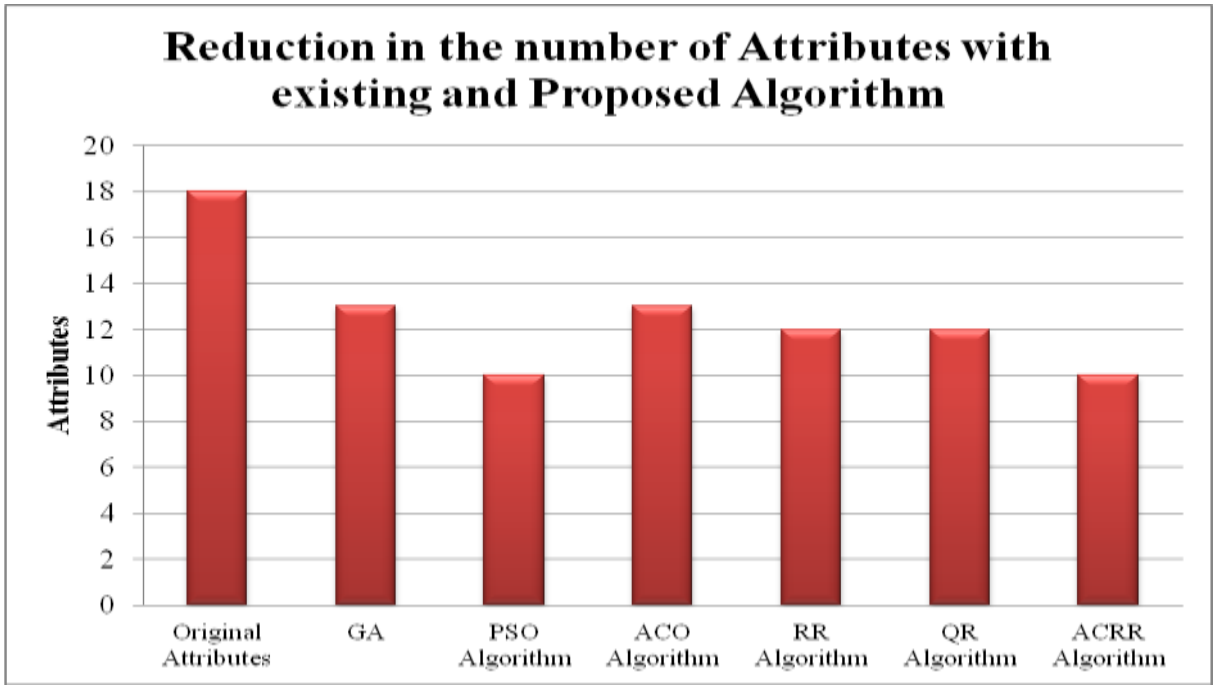

Figure 2: Reduction in the number of Attributes

Figure 2 depicts the number of features obtained by using the existing and proposed algorithms. After reducing the number of attributes, the reduced data set is classified. The existing classifiers like Naïve Bayes (NB), Multi Layer Perceptron Network (MLPN), Radial Basis Function (RBF) Network and J48 are used. The metrics like Accuracy, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Absolute Error (RRAE), True Positive Rate (TPR),

 ISSN: 2321-8169
01 – 11

False Positive Rate (FPR), Precision, Recall, F-Measure, and Receiver Operating Characteristic (ROC) Area are taken to measure the performance of each classifier.

	Original Attributes				**Attributes obtained by using GA**				**Attributes obtained by using the PSO Algorithm**			
	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48
Accuracy	72.81	95.61	77.19	73.68	67.54	91.22	71.05	79.83	66.67	83.33	68.42	69.30

Table 4 : Comparison of Accuracy of different Algorithm with Different Classifiers

	Attributes obtained by using the ACO Algorithm				**Attributes obtained by using RR Algorithm**				**Attributes obtained by using QR Algorithm**			
	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48
Accuracy	66.67	92.98	74.56	73.68	69.30	87.72	75.44	73.68	64.04	85.09	72.81	69.30

Table 5 : Comparison of Accuracy of different Algorithm with Different Classifiers

	Attributes obtained by using Proposed ACRR Algorithm			
	NB	MLPN	RBF	J48
Accuracy	75.44	90.35	78.07	73.68

Table 6: Comparison of Accuracy of different Algorithm with Different Classifiers

The accuracy obtained while classifying the attributes obtained using every classifier is listed in Table 4, 5 and Table 6 respectively

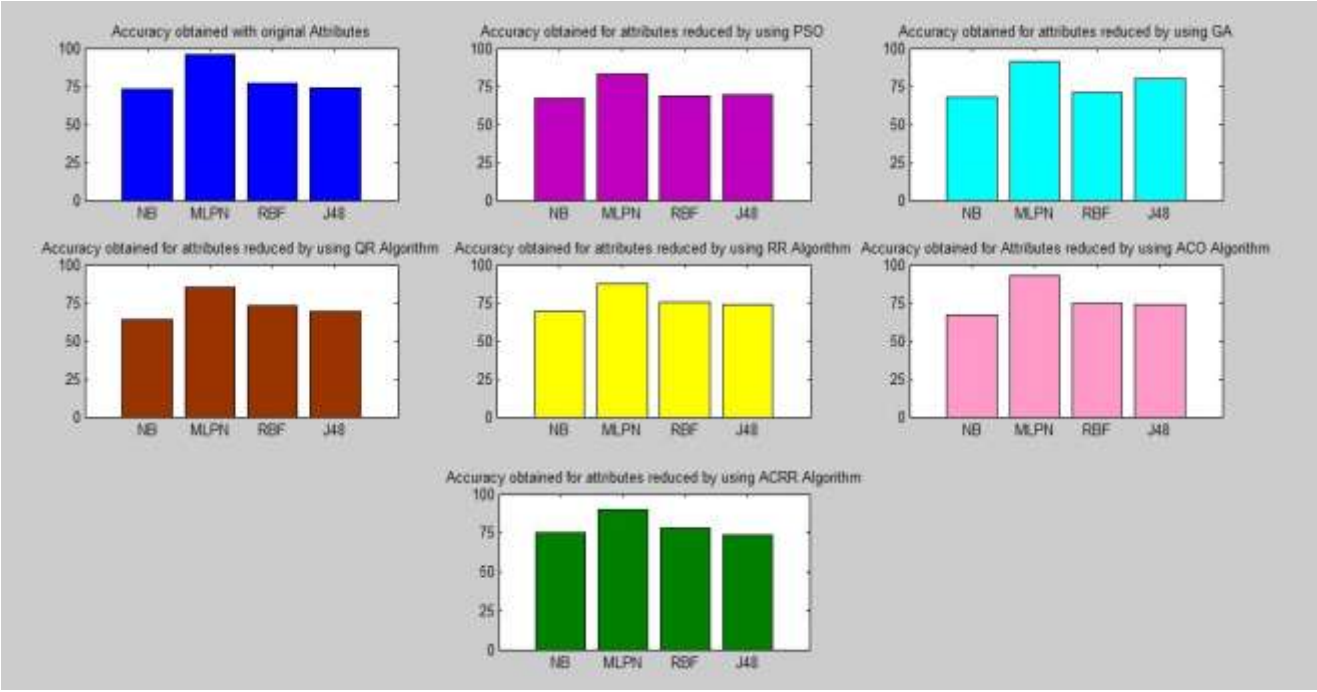

Figure 3: Comparison of Accuracy for different Classifiers by using attributes obtained by Existing and Proposed Algorithm

Figure 3 depicts the comparison of Accuracy for different classifiers by using the attributes obtained by using the Existing and Proposed Algorithm for Feature Reduction.

	Original Attributes				**Attributes obtained by using GA**				**Attributes obtained by using the PSO Algorithm**			
	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48
MAE	0.21	0.05	0.22	0.25	0.24	0.08	0.25	0.23	0.26	0.16	0.26	0.27
RMSE	0.38	0.16	0.33	0.35	0.38	0.21	0.35	0.33	0.39	0.27	0.36	0.37
RAE	62.39	16.76	66.45	73.72	70.32	25.40	74.85	67.40	76.50	47.94	77.36	81.69
RRAE	93.68	39.89	81.86	86.22	93.28	50.55	86.98	80.62	95.61	67.74	88.55	90.76

Table 7 : Comparison of Error values of different Algorithm with Different Classifiers

	Attributes obtained by using the ACO Algorithm				**Attributes obtained by using RR Algorithm**				**Attributes obtained by using QR Algorithm**			
	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48
MAE	0.23	0.077	0.23	0.25	0.22	0.12	0.21	0.25	0.25	0.14	0.26	0.27
RMSE	0.39	0.18	0.34	0.35	0.37	0.23	0.33	0.35	0.4	0.26	0.4	0.37
RAE	69.24	23.00	68.88	86.22	67.14	36.07	63.51	73.72	75.18	42.06	76.76	81.69
RRAE	96.07	43.97	83.52	86.22	92.03	57.11	80.44	86.22	98.58	62.70	88.14	90.76

Table 8: Comparioson of Error Values of different Algorithm with Different Classifiers

	Attributes obtained by using Proposed ACRR Algorithm			
	NB	MLPN	RBF	J48
MAE	0.22	0.11	0.22	0.25
RMSE	0.34	0.22	0.33	0.35
RAE	65.85	32.29	66.55	73.72
RRAE	84.64	54.22	81.92	86.22

Table 9 : Comparioson of Error Values of different Algorithm with Different Classifiers

The comparison of Error Values like MAE, RMSE, RAE, RRAE for different Existing Algorithms and proposed Algorithm with different classifiers are given in Table 7, 8 and 9.

Table 10 compares the TPR, FPR, Precision, Recall, F-Measure and ROC Area obtained while implementing the existing algoithms and proposed ACRR Algorithms

 ISSN: 2321-8169
01 – 11

	Original Attributes				Attributes obtained by using GA				Attributes obtained by using the PSO Algorithm			
	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48
TPR	0.728	0.956	0.772	0.737	0.675	0.912	0.711	0.798	0.667	0.833	0.684	0.693
FPR	0.245	0.037	0.224	0.228	0.299	0.09	0.321	0.189	0.337	0.174	0.351	0.334
Precision	0.743	0.949	0.769	0.76	0.69	0.904	0.703	0.799	0.663	0.826	0.676	0.685
Recall	0.728	0.956	0.772	0.737	0.675	0.912	0.711	0.798	0.667	0.833	0.684	0.693
F-Measure	0.727	0.952	0.77	0.735	0.674	0.908	0.703	0.796	0.667	0.829	0.675	0.686

Table 10 : Comparison of Weighted Average of different Algorithm with Different Classifiers

Figure 4: Comparison of Error Values for Original Attributes and Attributes reduced by GA by using different Classifiers

Figure 5: Comparison of Error Values for Attributes obtained by using the PSO and ACO Algorithm

Figure 6: Comparison of Error Values for Attributes obtained by using the RR and QR Algorithm

Figure 7: Comparison of Error Values for Attributes reduced by using the proposed ACRR Algorithm

Figures 4, 5 and 6 shows the comparison of error values with respect to attributes obtained by using existing Algorithm. Figure 7 depicts the error value for proposed ACRR Algorithm.

Table 11 and Table 12 compares the TPR, FPR, Precision, Recall, F-Measure and ROC area obtained while implementing the existing algoithms and proposed ACRR Algorithms

	Attributes obtained by using the ACO Algorithm				**Attributes obtained by using RR Algorithm**				**Attributes obtained by using QR Algorithm**			
	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48	NB	MLPN	RBF	J48
TPR	0.667	0.93	0.746	0.737	0.693	0.877	0.754	0.737	0.64	0.851	0.728	0.693
FPR	0.311	0.072	0.28	0.228	0.297	0.137	0.263	0.228	0.351	0.156	0.287	0.334
Precision	0.679	0.922	0.738	0.76	0.696	0.87	0.746	0.76	0.644	0.843	0.72	0.685
Recall	0.667	0.93	0.746	0.737	0.693	0.877	0.754	0.737	0.64	0.851	0.728	0.693
F- Measure	0.666	0.926	0.738	0.735	0.691	0.872	0.749	0.735	0.639	0.847	0.723	0.686
ROC Area	0.769	0.988	0.83	0.758	0.775	0.964	0.849	0.758	0.727	0.947	0.783	0.735

Table 11: Comparioson of Weighted Average of different Algorithm with Different Classifiers

	Attributes obtained by using Proposed ACRR Algorithm			
	NB	MLPN	RBF	J48
TPR	0.754	0.904	0.781	0.737
FPR	0.226	0.081	0.212	0.228
Precision	0.762	0.904	0.779	0.76
Recall	0.754	0.904	0.781	0.737
F- Measure	0.753	0.9	0.778	0.735
ROC Area	0.816	0.958	0.736	0.758

Table 12: Comparioson of Weighted of different Algorithm with Different Classifiers

 ISSN: 2321-8169
01 – 11

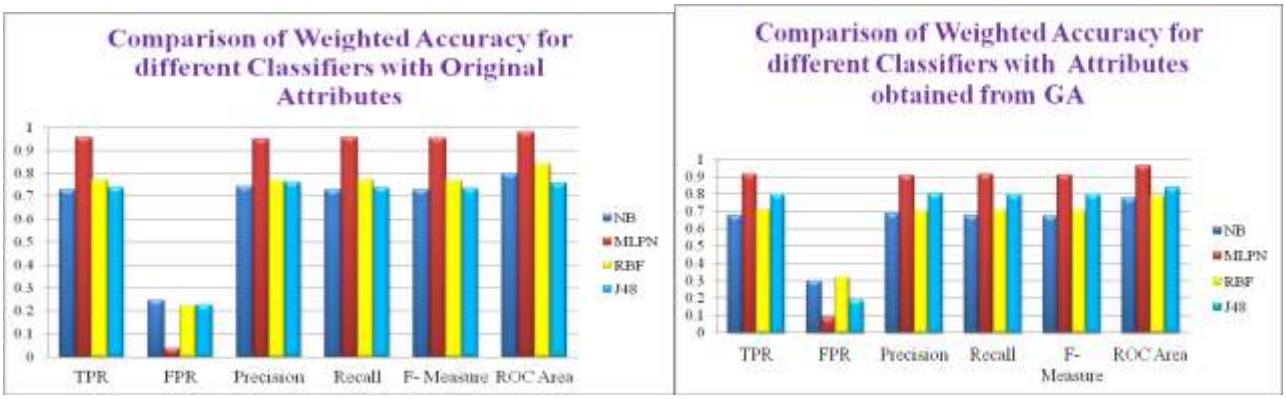

Figure 8: Comparison of the performance of different classifiers based on Weighted Accuracy by using Original Attributes and Attributes from GA

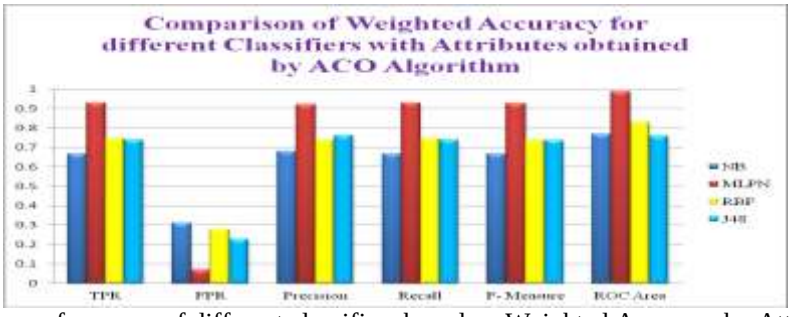

Figure 9: Comparison of the performance of different classifiers based on Weighted Accuracy by Attributes obtained from PSO and ACO Algorithm

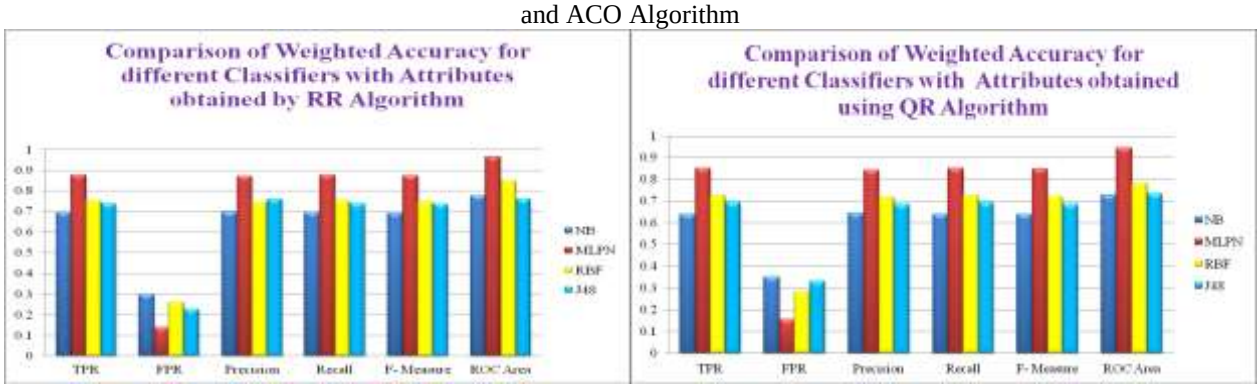

Figure 10 : Comparison of the performance of different classifiers based on Weighted Accuracy by Attributes obtained from RR and QR Algorithm

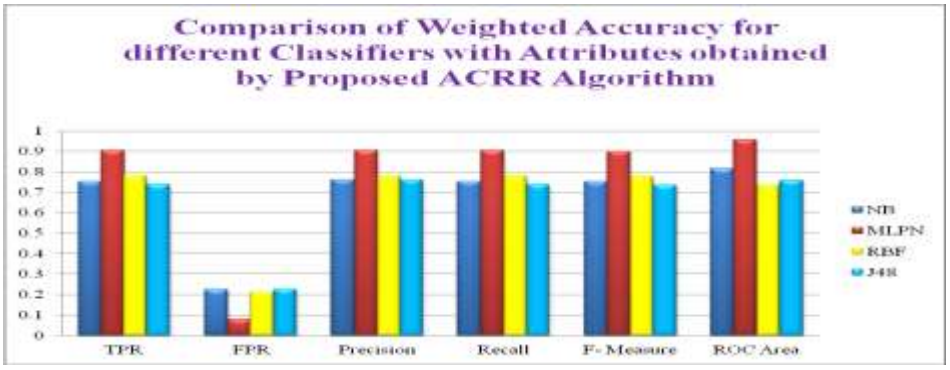

Figure 11: Comparison of the performance of different classifiers based on Weighted Accuracy by using Attributes obtained from Proposed ACRR Algorithm

Figure 8,9,10 and 11 depicts the TPR, FPR, Precision, Recall, F-Measure and ROC Area values obtained while comparing the algorithms.

VI. CONCLUSION

Feature Reduction is an important task in Data Mining. Reducing the unwanted features will improve the classification accuracy and also minimizes the time taken for classification too. The existing algorithms GA, PSO, ACO, QR and RR are taken for study and a new Algorithm ACRRA is proposed. ACO Algorithm performed better than the PSO and GA. The performance of RR Algorithm gives better performance than the QR Algorithm. The proposed ACRRA algorithm is developed by combining the core features of ACO Algorithm and RR Algorithm, which performs better than the existing Algorithms. The proposed ACRR Algorithm improves the accuracy and also performs better with other metrics, which is evident from the results obtained.

REFERENCES

[1] Milewski, R., Milewska, A. J., Czerniecki, J., Leśniewska, M., & Wołczyński, S. (2013). Analysis of the demographic profile of patients treated for infertility using assisted reproductive techniques in 2005–2010. Ginekologia Polska. 84(7), 609–614.

[2] S.J Kaufmann, J.L.Eastaugh, S. Snowden, S.W.Smye, V.Sharma, "*The application of neural network in predicting the outcome of in-vitrofertilization*", HumanReproduction vol.12 no.7 pp. 1454-1457, 1997.

[3] Asli Uyar, Ayse Bener, H.Nadir Ciray, Mustafa Bahceci, "*Handling the Imbalance Problem of IVF Implantation Prediction*", IAENG International Journal of Computer Science, May 2010.

[4] Asli Uyar, Ayse Bener, H.Nadir Ciray, Mustafa Bahceci, "*ROC Based Evaluation and Comparison of Classifiers for IVF Implantation Prediction*", Institute of Computer Sciences, Social-Informatics andTelecommunication Engineering, LNICST 27, pp. 108-111, 2010.

[5] M. Durairaj, K. Meena, "*A Hybrid Prediction System Using Rough Setsand Artificial Neural Networks*", International Journal of Innovative Technology and Creative Engineering, vol. 1, no. 7, July 2011.

[6] David Gil, Jose Luis Girela, Joaquin De Juan, M. Jose Gomez-Torres, Magnus Johnsson, "*Predicting seminal quality with artificial intelligence methods*", Elsevier, Expert Systems withApplications,2012.

[7] M. Durairaj, R. Nandhakumar, "*Data Mining Application on IVF Data For The Selection of Influential Parameters on Fertility*", IJEAT,Volume 2, Issue 6, Aug 2013.

[8] M. Durairaj, P. Thamilselvan, "*Applications of Artificial NeuralNetwork for IVF Data Analysis and Prediction*", Journal of Engineering,Computers and Applied Sciences, Volume 2, No 9, September 2013.

[9] Claudio Manna, Loris Nanni, Alessandra Lumini, Sebastiana Pappalardo, "*ArtificialIntelligence Techniques for Embryo and OocyteClassification*", Elsevier, Reproductive BioMedicine 2013.

[10] M. Durairaj, R. Nandhakumar, "*An Integrated Methodology of Artificial Neural Network and Rough Set Theory for Analyzing IVF Data*", IEEE International Conference on Intelligent Computing Application, 2014

[11] M. Durairaj, Nandhakumar Ramasamy, "*Intelligent Prediction Methods and Techniques Using Disease Diagnosis in Medical Database: A Review*", International Journal of Control theory and Applications, Volume 8, issue 5, 2015.

[12] M. Durairaj, Nandhakumar Ramasamy, "*A Comparison of the Perceptive Approaches for Preprocessing the Data Set for Predicting Fertility Success Rate*", International Journal of Control theory and Applications, Volume 9, issue 27, 2016.