

An Automatic System for IVF Data Classification by Utilizing Multilayer Perceptron Algorithm

Gowramma G S¹, Dr.Mahesh T.R², Giridhar Gowda³,

¹PhD Scholar, Dept. of CSE, T John Institute of Technology, Bengaluru.

²Professor & Head, Dept. of CSE, T John Institute of Technology, Bengaluru.

³Associate Professor, Dept. of CSE, Don Bosco Institute of Technology, Bengaluru

Abstract—This paper depicts the data mining technique for recognize most significant factors in barrenness couples to resolve the success rate of IVF (In-vitro Fertilization) treatment. The data set used in the research includes information outcome during IVF treatment and relevant laboratory tests. It supports the medical practitioner to identify tests have high impact factors for determining the success of infertility treatment. Data mining is associated with number of techniques that are used to perform Pre-processing, normalization and data reduction. The reduced data set includes the major set of characters which are impacts the outcome that can be used to infer and predict. The initial data set are pre-processed by the supervised filter and the three different attribute selection techniques before prediction. It is significant to precisely analyze the data set and clean the superfluous data that maximize the prediction accuracy. Reduced data is then given to the Feed Forward Neural Network based classification algorithm such as Multilayer Perceptron algorithm. This algorithm performs better as compare to other algorithms

Keywords—Attribute Selection Algorithm, Multilayer Perceptron Algorithm, IVF

I. INTRODUCTION TO IVF

In vitro fertilization (IVF) is another process of fertilization that takes exterior of the body. IVF is the most efficient art. It is always used when a woman's Fallopian tubes are failed or when a man cannot produce sufficient sperms. Doctors deals with woman patient with proper drugs that causes ovaries to generate multiple eggs. The matured eggs are taken from the women. They are placed in dish with man's sperm to lead fertilization. After three to five days, vigorous embryos are embedded in the woman's uterus. Infertility, also called primary infertility, couple inability to become pregnant after more than one year. This is in contrast to secondary infertility, which refers to the inability to maintain a pregnancy until birth. Primary infertility affects to 15% of men and women of reproductive age in the world. Assisted reproductive technologies (ARTs) are a procedure to assist couple becomes pregnant, including In Vitro Fertilization (IVF), Intra Cytoplasmic Sperm Injection (ICSI), and other similar procedures. With IVF helps the reproduction, a man's sperm and the woman's egg are pooled in a laboratory dish, which leads fertilization. The consequential embryo or embryos are then placed to the woman's uterus (womb) to implant and develop naturally. Typically, two to four developing lives are put in the lady's uterus at one time. Every endeavour is

known as a cycle. Fewer than 5% of barren couples really utilize IVF. IVF is generally the treatment of decision for a woman with blocked, seriously harmed, or no Fallopian tubes. IVF is additionally used to beat infertility brought about by endometriosis or issues with the man's sperm (such as low sperm count). Couples who simply can't conceive and have tried other infertility method.

II. IVF SUCCESS FACTORS.

In vitro preparation (IVF) is a complex solution and it incorporates various strides, an assortment of medicines and systems. Every one of the factors assume an essential part in the ultimate result of the treatment, which prompts to sound live birth.. Some of the outstanding factors [1] associated with IVF are:

A. *Egg/Embryo Quality*

Reproductive age significantly impacts egg and embryo quality, live birth rates are 50% in women less than age 35 and approximately 25% in women between the ages 40 to 42 per fresh IVF cycle. Lack of success rates in IVF or any infertility treatment is due to the decreasing number of eggs within the ovaries. The drastic loss of the follicles/eggs after the age of 32-35 and the probability of fertility is significantly decreased thereafter. Several tests can identify patients who may have a suboptimal response to stimulation with IVF. These tests are day 3 FSH and estradiol levels, Clomid challenge test, inhibin B, Anti-Mullerian Hormone (AMH) levels and antral follicle count in the beginning of the cycle. At the point when there is confirmation of diminished ovarian hold by an irregular esteem in one of these tests, a more forceful way to deal with infertility treatment is for the most part taken, for example, IVF with forceful ovarian incitement.

B. *Timing of Ovulation.*

Ovulation development and advancement of the eggs ought to be planned decisively so that the eggs are not very develop or immature. In a few patients ovulation may infrequently be activated early and this may bring about immature eggs. In others, eggs may over-develop and that may not yield solid oocytes if incitement is drawn out to a point where the follicles are measuring more than 25-30 mm in size. Ideal planning of ovulation ought to be done in view of every individual case.

C. *Receptivity of the Endometrium*

The anatomical imperfections of the uterus may bring about infertility. Hydrosalpinx can be optional to earlier pelvic contaminations that harmed the tubes or from pelvic attachments or endometriosis related issues and injuries. A poor endometrial fixing most normally happens in ladies with a past filled with unexplained repetitive IVF disappointments or early intermittent unnatural birth cycles.

D. *Embryo Transfer.*

Embryo Transfer (ET) is the last and a standout amongst the most essential strides of IVF treatment. The utilization of ultrasound direction to ideally put the incipient organisms in the uterus has turned into an essential part of the ET procedure. It has likewise been appeared by different reviews to enhance pregnancy rates. Symptoms, for example, cramping or bleeding after the techniques are pointers of a troublesome exchange.

E. *Sperm Factors.*

Sperm quality parameters and have been shown to predict poor reproductive performance. The SDI (Sperm DNA Integrity) measures DNA damage, which may be present in sperm from both fertile and infertile men. Therefore, this sperm DNA damage analysis may reveal a hidden abnormality of sperm DNA in infertile men that an abnormal SDI assay is more likely to occur in cases of abnormal semen analyses. Thus the assay is ideally suited to fertility clinics to assess male sperm DNA integrity as related to fertility potential and embryo development as well as effects of reproductive toxicant.

Table1 shows IVF influenced factors [2] and their significance in Fertility. In the above table significant influence of fertility YES indicates high predictive parameter in pregnancy. NO indicates probability of consideration in IVF is less.

III. DATA MINING PROCESS.

Data Mining is the main part of knowledge discovery in Databases. Data Mining and KDD are frequently utilized reciprocally on the grounds that data mining is fundamental part of the KDD process. Data mining is derived from computer science. It is the computational procedure of finding examples in expansive information sets including techniques for artificial intelligence, machine learning, statistics, and database systems. The primary objective of the data mining procedure is to induce data from an information set and change it into a justifiable structure for further utilize. Apart from the raw analysis step, it includes database system and data management feature, data pre-processing, model and deduction considerations. The techniques associated with data mining are Anomaly detection, Association Rules, Classification, Clustering and Regression.

IV. DATA CLEANING OR PRE-PROCESSING.

Subset selection identifies a subset of features as a group for considerations. Subset selection algorithms [4] can be classified into Wrappers, Filters and Embedded. Wrappers use a search algorithm to look through the space of conceivable components and assess every subset by running a model on the subset. Wrappers can be computationally costly and have a hazard of over fitting to the model. Filters are like Wrappers in the pursuit approach, yet as opposed to assessing against a model, an easier filter is assessed. Numerous search techniques uses greedy hill climbing, which iteratively assesses an applicant subset of components, then alters the subset and assesses if the new subset is a change over the old. Assessment of the subsets requires a scoring metric that evaluations a subset of elements. Exhaustive search is for the most part unrealistic, so at some implementer (or operator) characterized ceasing point, the subset of elements with the most noteworthy score found up to that point is chosen as the attractive component subset. The ceasing paradigm changes by calculation; conceivable criteria include: a subset score surpasses an edge, a program's greatest permitted run time has been outperformed.

Data Set

The data set contains 24 attributes of 114 patients as shown in figure1, included all influenced parameters shown in table1. After applying Attribute selection algorithm WrapperSubsetEval with BestFirst search method data set is reduced to 5 attributes. After applying Attribute selection algorithm WrapperSubsetEval with GeneticSearch search method data set is reduced to 9 attributes as shown in the figure2. After applying Attribute selection algorithm WrapperSubsetEval with RankerSearch method data set is reduced to 3 attributes.

V. MULTILAYER PERCEPTRON NEURAL NETWORK

Feed Forward Neural Networks [5] are one of the types of multilayer neural network, which allows signals to travel only in one way i.e., from input to output. First, the network is trained on a set of paired data to determine input-output mapping. The weights of the connections among neurons are fixed and the neural network is applied to identify the classifications of a new set of data. In the course of classification the signal at the input units radiates all the way throughout the net to conclude the activation values at all the output units. All input units have activation value that depicts some features exterior to the net. At that point, each input unit sends its activation value to each of the hidden units to which it is associated. Each of these hidden units computes its new actuation value and these signs are then passed on to yield units. The activation value for each receiving unit is computed by basic activation function. The function adds together all the sending units, where the commitment of a unit is recognized as the heaviness of the association between the sending and receiving units multiplied by the sending unit's activation value. This sum is normally then further changed, for instance, by modifying the initiation total to esteem somewhere around 0 and 1 or potentially by setting the activation sum to zero unless threshold level for that sum is reached. Figure3 depicts the scenario behind the perceptron neural network[6].

The Widrow-Hoff (gradient descent or Delta rule) is the most common supervised learning rule. The general function is associated with a perceptron layer which initializes the network as described with random weights and biases. Generally the random numbers are kept small and symmetrical to zero. As the input vector is assigned to the net, it delivers an output. The learning rule is then applied to the layer. A simple learning rule which is widely used is called the Widrow-Hoff rule [7].

Where $0 < \eta < 1$ positive gain function. When the output is 1 then Class will be A; when the output is 0 then Class will be B. This rule identifies a method of updating each weight. It tries to minimize the error between the target output $d(t)$ and the experimentally obtained output $y(t)$ for each neuron by calculating the error and calling it a “delta” (Δ). Each weight (W) is then adjusted by adding to delta multiplied by some attenuation

Figure1. Data set of 114 Patients

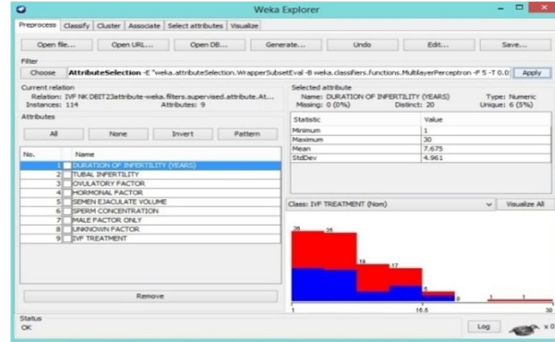


Figure2. Pre-processing with GeneticSearch using Wrapper Subset Eval

TABLE I. IVF INFLUENCED FACTORS AND THEIR SIGNIFICANCE IN FERTILITY

Sl No	Considered Parameters	Significant influence of fertility
1	Age (Female)	NO
2	Duration of Infertility	NO
3	Previous Pregnancy	NO
4	Previous Miscarriage	NO
5	Previous Surgery	NO
6	Body Mass Index	NO
7	Cervical Factor	NO
8	Hormonal Factor	NO
9	Ovulatory Factor	NO
10	Tubal Infertility	NO
11	Endometriosis	YES
12	Depression	NO
13	Medical Disorders	NO
14	Male Factor	YES
15	Sperm Morphology	NO
16	Sperm Vitality	NO
17	Sperm Motility	YES
18	Sperm Concentration	YES
19	Liquefaction Time	NO
20	Semen Ejaculate Volume	NO
21	Unexplained Factor	YES
22	No of Oocytes Retrieved	NO
23	No of Embryos Transferred	NO
24	IVF Treatment	YES

$$\Delta = d(t) - y(t) \quad (1)$$

$$W_i(t + 1) = W_i(t) + \eta \Delta X_i(t) \quad (2)$$

$$d(t) = \begin{cases} 1, & \text{if input is class A} \\ 0, & \text{if input is class B} \end{cases} \quad (3)$$

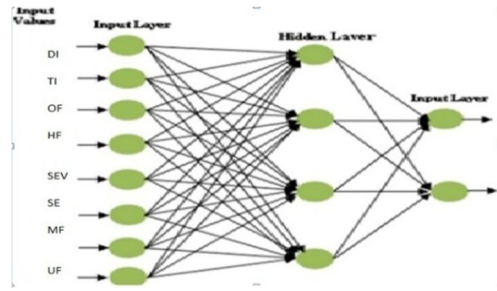


Figure3: Multilayer Perceptron Neural Network

DI: Duration of Infertility, **TI:** Tubal Infertility, **MF:** Male Factor, **OF:** Ovulatory Factor, **HF:** Hormonal Factor, **SEV:** Semen Ejaculate Volume, **SC:** Sperm Concentration, **UF:** Unexplained Factor

constant η (Learning Rate)[9]. This process is then iterated until the net error falls below some threshold value. By adding its specific error to each of the weights, we are ensured that the network is being moved towards the position of minimum error and by using attenuation constant [10], rather than the foil value of the error, we move it slowly towards this position of minimum error. When correctly trained, the perceptron shows some highly promising behaviour.

VI. RESULTS AND DISCUSSION

The multilayer perceptron neural network is constructed and generated the confusion matrix. In this process of network construction, the valuable errors, YES or NO value, true positive and false positive rate are generated, as shown in Figure 4. From figure 4 shows the classification output generated from WEKA tool using Multi Layer Perceptron Algorithm. The TP (True Positive) Rate can be calculated as

$$TP \text{ (True Positive) Rate} = \frac{TP}{(TP+FN)} \quad (4)$$

False Positive (FP)[9] is corresponds to the number of negative examples wrongly predicted as positive by the classification model. True Negative (TN)[9] is corresponds to the number of negative examples correctly predicted by the classification model. The Precision can be calculated as

$$FP \text{ (False Positive) Rate} = \frac{FP}{(TN+FP)} \quad (5)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (6)$$

TABLE II. COMPARISON OF OUTPUT USING WRAPPERSUBSETEVAL WITH DIFFERENT SEARCHING METHODS WITH THE MLP CLASSIFIER

SI No	Pre-processing Algorithm	Reduced No of Attributes	Accuracy
1	Without Pre-processing Algorithm	24	72.08%
2	With Attribute Selection WrapperSubsetEval with search method BestFirst	5	84.21%
3	WrapperSubsetEval with search method GeneticSearch	9	87.71%
4	With Attribute Selection WrapperSubsetEval with search method RankerSearch	3	83.33%

Precision determines the fraction of records that actually turns out to be positive in the group the classifier has declared as a positive class, the higher the precision is, the lower the number of false positive errors committed by the classifier. Recall can be calculated as

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (7)$$

Recall measures the fraction of positive examples correctly predicted by the classifier. Classifier with large recall has very few positive examples misclassified as the negative class. F-Measure can be calculated as

$$F\text{-Measure} = \frac{2TP}{(2TP+FN+FP)} \quad (8)$$

F-Measure represents a harmonic mean between recall and precision.

The comparison of output using Wrappersubseteval with different searching methods with the MLP classifier is as shown in Table2.

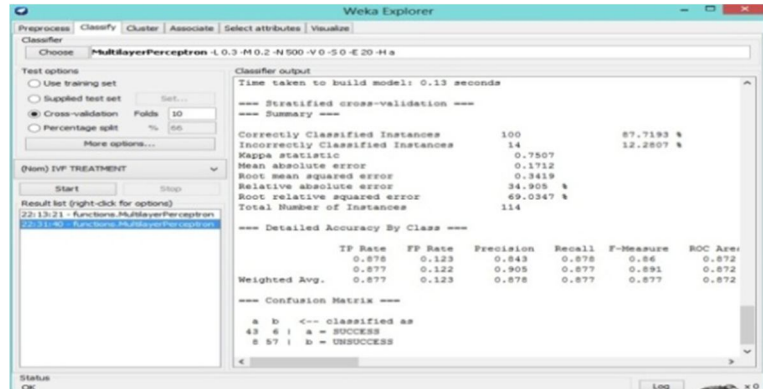


Figure 4: Multi Layer Perceptron with GeneticSearch search Output

VII. CONCLUSION

The Multilayer Perceptron algorithm of Neural Network performs well in classification as compared to linear classification algorithms such as Navie Bayes, Rule based with respect to TP rate, FP rate, Recall and Precision. This paper uses three different attribute selection algorithms in which MLP Network outputs **87.7%** with data cleaning using Attribute Selection WrapperSubsetEval with search method known as GeneticSearch which outperforms to predict most of the influenced factors for IVF that increases chances of getting successful pregnancy. Future work will be done with the same algorithm for PSO Search method.

REFERENCES.

- [1] Data Mining Application on IVF Data for the Selection of Influential Parameters on Fertility. International Journal of Engineering and Advanced Technology (IJEAT)ISSN: 2249–8958, Volume-2 Issue-6, August 2013.
- [2] 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 Applications, Serials Publications, ISSN: 0973-6786, Vol. 2(3), pp. 221-231, Dec 2008
- [3] Thangavel, K, Jaganathan, P, Pethalakshmi, A and Karnan, M.—Effective Classification with Improved Quick Reduct for Medical Database Using Rough Systeml, BIME Journal , Vol. 05, Issue (1), pp. 7-14, 2005.
- [4] Kaufmann, S.J., Eastaugh, J.L., Snowden, S., Smye, S.W. and Sharma, V.Theapplication of neural networks in predicting the outcome of in-vitro fertilization. Human Reproduction, (1997) vol.12 no. 7 pp. 1544-1457.
- [5] Larsson, B. and Rodriguez-Martinez, H. Can we use in vitro fertilization tests to predict semen fertility? Anim. Reprod. Sci. (2000) 60-61: 327-336.
- [6] A. Uyar, A. Bener, H. Ciray, and M. Bahceci, ROC based evaluation and comparison of lassifiers for IVF implantation prediction," in In Proc. Second In-ternational ICST Conference on Electronic Health-care for the 21st century, 2009
- [7] G. Venkat, R. Al-Nasser, S. Jerkovic, and I. Craft, Prediction of success in IVF treatments using neu-ral networks,"Fertility and Sterility, vol. 82, p. 215,2004
- [8] H. P. Mazurowski, M.A. and, J. Zurada, J. Lob,J. Baker, and G. Tourassi, \Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," Neural Networks , vol. 21, pp. 427-436, 2008.
- [9] Introduction to Data Mining by Pang-Ning Tan,Michael Steinbach,Vipin kumar
- [10] Applications of Artificial Neural Network for IVF Data Analysis and Prediction. Journal of Engineering, Computers & Applied Sciences (JEC&AS) ISSN No: 2319 -5606 Volume 2, No.9, September 2013.
- [11] Adriana Albu , Loredana Ungureanu "Artificial Neural Network in Medicine" International Journal of Computer science.