# Classification of Sperm Cells According to their Chromosomic Content Using a Neural Network Trained with a Genetic Algorithm

A. F. Kuri-Morales[1] , M. R. Ortiz-Posadas[2], D. Zenteno[1] and R. Peñaloza[1]
[1]Departamento de Computación. Instituto Tecnológico Autónomo de México. México
[2]Departamento de Ingeniería Eléctrica. Universidad Autónoma Metropolitana Iztapalapa. México

*Abstract* - ***A priori*** **determination of the sex of a human individual before gestation is a desirable goal in some cases. To achieve this, it is necessary to perform the separation of sperm cells containing either X or Y chromosomes. As is well known, male sex depends on the presence of chromosome Y. Once this separation is achieved in principle, we require to determine, with a high degree of accuracy, whether the sperm cells of interest contain the desired X or Y chromosomes. If we are able to obtain certain simple measurements regarding the sperm cells under consideration we will be able to control the fertilization process reliably. In this paper we report a method which allows for non-invasive verification of the characteristics of the separated sperm. We determined a set of easily measurable characteristics. From a sample drawn from previously cropped sperm we trained a neural network with a genetic algorithm. The trained network was able to perform *a posteriori* classification with an error much smaller than 1%. This percentage of efficiency is better than the ones reported in centers of assisted fecundation.**

*Keywords* - Gender selection, neural networks, genetic algorithms, non-invasive.

## I. Introduction

One of the tasks performed by the Assisted Gestation Centers is the determination of the sex during an individual's gestation. Certain techniques, such as *in vitro* fertilization [1] allow us to achieve some limited control on the sperm which will fecundate the egg or, in the case of intrauterine insemination, on the sample of sperm used for the fertilization. Separation of X and Y sperm is a necessary step towards oriented gender selection. Once the separation is achieved, it is necessary to validate the reliability of such gender selection. In the past, some approaches looked for physical, chemical or immunological differences between the sperms with X or Y chromosomes to yield various methods for their classification. Most of the methods consist in dividing the sperm cells in two groups whose statistical reliance is defined as a percentage: 85% of X chromosomes or 75% of Y chromosomes. Any sperm separation method is based in one of the most conspicuous differences of these cells: their size. Chromosome X is larger than chromosome Y. However, to obtain this information it is necessary to dissect the sperm cell. Given this, the purpose of this work is to show that an artificial neural network (more specifically a multi-layer perceptron network) trained with a genetic algorithm (as opposed to the more conventional backpropagation learning algorithm) yields an efficient non-invasive method which permits the determination of the chromosomic contents relative to the sex in the classification of sperm cells. The method was validated with a small sample of human sperms, which was mathematically enhanced assuming a motility preclassification.

In what follows we describe the method just outlined. In part 2 we describe the proposed methodology. In part 3 we briefly discuss the type of neural networks we used in our work (multi-layer perceptron network) trained with a genetic algorithm (which, hereinafter we will refer to as GMLP). In part 4 we discuss the genetic algorithm. In part 5 we describe the detailed characteristics considered for sperm classification, the neural network and the results we obtained. Finally, in part 6 we offer our conclusions.

## II. Methodology

To achieve a generalized classification method, we trained a GMLP. It is a well known fact [2] that these networks exhibit good generalization properties when adequately trained. Therefore, the learning process consists of the following basic steps:

a) Determine the relevant characteristics which will allow the user to properly describe the selected sperm sample.

b) Obtain a sufficiently large set of training and test samples.

c) Analyze the mathematical properties of the original (before selecting the training and test subsets) sample to determine possibly large correlation ratios to insure informational relevancy.

d) Normalize the data to avoid scaling problems.

e) Supplement the original data with synthetic data drawn from a probabilistic distribution with experimentally determined population's parameters. This step is needed to train the network reliably when the original sample (as in this case) is not large enough to ensure full training.

f) Divide the original sample in a Training Sample (TS) which consists, roughly, of 85% of the total and a Test Sample (ES) which considers the remaining 15%. This step is needed to allow the cross-validation scheme resulting in good generalization properties for the trained network. That is, the GMLP will *learn* from TS but its *generalization* capabilities will depend on ES.

## III. Neural Networks

The concept of artificial neural networks is to imitate the structure and workings of the human brain by means of mathematical models.

A neuron receives signals via several input connections. These are weighted at the input to a neuron by the connection function. The weights employed here define the coupling strength (synapses) of the respective connections and are established via a learning process, in the course of which they are modified according to given patterns and a learning rule. The input function compresses these weighted inputs into a scalar value, the so-called network activity at this neuron. Simple summation is generally employed here. In such cases, the network activity, which results from the connection function and the input function, is the weighted sum of the input values. The activation function determines a new activation status on the basis of the current network activity, taking the previous status of the neuron into account. This new activation status is transmitted to the connecting structure of the network via the output function of the neuron, which is generally a linear function. By way of reference to biological neurons, the activation status at the output of a neuron is also known as the excitation of the neuron. In the case of supervised learning (as here), in addition to the input patterns, the desired corresponding output patterns are also presented to the network in the training phase. The network calculates a current output from the input pattern, and this current output is compared with the desired output. An error signal is obtained from the difference between the generated and the required output. This signal is then employed to modify the weights in accordance with the current learning rule, as a result of which the error signal is reduced.

The multilayer perceptron is a network model in which the neurons are configured in layers, whereby the neurons of a layer are generally all connected with the neurons of the following layer. As connections exist only from the input layer in the direction of the output layer, this is a feedforward network. This network is able to process analogue input patterns and learns in supervised mode. The error of the network is defined as the square distance between the required status (stipulated output pattern) and the actual status of the output pattern generated on the basis of current network weights upon definition of an input pattern [3]. This results in an error function which is dependent on the weighing factors via the activation function. Typically, these sort of NNs are trained with the backpropagation learning rule. This algorithm imposes the need for a differentiable measure of adequacy, as shown in equation (1).

$$E = \frac{1}{2}\sum_{i}(z_i - o_i)^2 \qquad (1)$$

Where Zi denotes the desired output and Oi denotes the observed output.

However, one of us [4] has shown that alternative non-differentiable measures of error yield better generalization properties on the trained networks. Since non-differentiability disallows the use of backpropagation we resorted to a genetic algorithm to train the network.

## IV. Genetic Algorithms

The optimization problem to be tackled requires an optimization algorithm that is demonstrably efficient. In the literature several genetic algorithms (GA) have been discussed and analyzed. We give a brief account of the known facts about the better known variation of a GA (the so-called *Simple* or *Canonical* GA or CGA), pointing out its advantages and shortcomings. Then we describe a non-traditional algorithm (the so-called *Vasconcelos* GA or VGA) which was designed to overcome the CGA's limitations while retaining its desirable characteristics.

Perhaps the best known GA is the one originally proposed by Holland [5]. This is the so-called *Simple* GA. The CGA is known to possess the following properties: a) It samples certain elements (the so-called *schemas*) in the genomes of its population exponentially in direct proportion to the adequacy of these elements. That is, it explores the space of solutions in a directed way such that the apparently better sections in the coded solutions under scrutiny are examined preferably. Likewise, it disregards apparently undesirable sections of the said code. 2) It explores $O(N^3)$ such elements for a population of size $N$ during every iteration. A CGA, therefore, approaches the ideal strategy of exploration/exploitation when faced with dynamic problems. These two characteristics explain why a CGA approaches a very good solution in a short number of iterations under proper conditions. The GA "assembles" an ever more complex encoding of the solution from smaller and simpler components. 3) It is also known that a CGA does not converge to the best solution even in an infinite number of steps. But by the simple expedient of retaining the best individual up to the last generation the GA (now transformed into the *eliTist* GA or TGA) does converge to the best possible solution given enough time. 4) The TGA (or the CGA, for that matter) reaches a steady state behavior regardless of the way the initial population is chosen. This explains why the number of individuals in the population is not relevant to the convergence properties of the algorithm, although it does, indeed, have bearing on its efficiency in reaching such convergence.

A CGA approaches a very good solution in a short number of iterations under proper conditions. But the CGA (or TGA) is no panacea. At least two undesirable features have been identified which impair the algorithm's performance. 1) Certain fitness functions may supply the algorithm with invalid information in terms of approaching the desired global optimum. These have been called *deceptive* functions, since they trick the CGA into "believing" it is getting closer to the result when it may not

be so. 2) When identifying the desirable "simple" elements which, hopefully, will compose the solution when properly combined, the CGA also retains some uncalled for sections of code which remain with the desired ones. This process of undesirable schema traveling along with the host has been called *spurious correlation* and is a major source of inefficiency during the CGA's execution.

In an effort to ameliorate the shortcomings of the CGA several variations of a GA have been tried. An interesting alternative (for reasons to be discussed shortly) seems to be the so-called Vasconcelos' GA or VGA [6]. In the VGA proportional selection which gives rise to *N* new individuals from *N* older ones, is replaced by what in evolutionary strategies has been called a ì + ë selection strategy, meaning that the new population comprised of ë individuals joins the older population's ì individuals of which only the *N* better individuals are retained. Furthermore, the *N* remaining individuals are crossed deterministically as follows. All individuals are sorted from best to worst. Then the *i*-th individual is crossed with the (*N-i+1*)-th with probability $p_c$ for *i=1,...,N/2*. The apparently contradictory strategy where good performers are crossed with the poor ones is explained when one considers that elitism is, here, of the strongest kind, i.e. only the best *N* individuals of every generation are retained. In this way, the VGA seeks for variety in the elements of the genome by dynamically disrupting the "good" schemas (but keeping the overall best). The mechanics of VGA (and, for that matter, almost any breed of GA) is a complex affair. As of to date, no generalized theoretical treatment which is able to model a GA with arbitrary characteristics has been developed with success. In [7] a more pragmatic approach was taken in order to establish a general methodology which would allow us to establish the relative performance of any two GAs. After applying this methodology we were able to obtain the relative performance measurements shown in table 1. We see that VGA is O(30%) more efficient than CGA as the number of generations increases when one considers an optimization process carried on a set of unbiased functions. Keeping this in mind and, in view of the known characteristics of the CGA, we have selected VGA as our training tool.

TABLE 1
RELATIVE PERFORMANCE OF VGA AND CGA

| | Generations | | | |
|---|---|---|---|---|
| | 30 | 50 | 100 | 150 |
| æ VGA / æ CGA | 1.08 | 1.29 | 1.20 | 1.267 |

## V. TRAINING A GMLP FOR CHROMOSOME IDENTIFICATION

The six characteristics considered are related with known tests which are, at present, used for sperm cells. Notice that none of them involves invasive action. A brief description follows.

*1) Elongation.* Defined as the ratio between the width and length of the head. In normal sperm cells it is between [0.40-0.75].

*2) Length of DNA.* This characteristic is one of the most important ones in the known classification methods, since it is known, in general, that X chromosome is larger than Y and, therefore, the length of the DNA is larger in the former.

*3) Proportion of the achrosome in the head.* The achrosome is the part of the head which contains the DNA. On the average, it has been observed that for chromosome X it is between 65-66% and, for Y, between 63-65%.

*4) Length of the tail.* The length of the tail varies between [50-55]µm. It is important to point out that this characteristic, in itself, is not a determining feature for proper classification. However, when considered along with the width of the tail it assumes high relevancy.

*5) Width of the Tail.* Generally speaking, the width of the tail in the tip is one tenth of the same width in the base. Because of this, we only considered this last characteristic (the width on the base: generally on the order of 1µm).

*6) Weight.* Since the X chromosome is larger than Y, then X weights more. Therefore, the sperms which carry an X chromosome are heavier.

Our model consisted of a three layered NN where the input layer has of 6 neurons (one for each of the six characteristics considered); 4 neurons in the hidden layer and one output neuron. The activation functions were linear in layer 1 and logistic in layers 2 and 3. A VGA was run with 100 individuals in the population with crossover probability of 0.985 and mutation probability of 0.005. We used a small sample of sperm cells described in terms of the six characteristics mentioned above and the corresponding kind of chromosome: the individuals containing chromosome X were tagged with "1"; the complementary individuals were tagged with "0". Using Student's t test we were able to ascertain that the data was approximately normal. Since training the neural network requires a larger training/test set, we determined the basic parameters of the distribution (ì, ó) and generated a set of 1,000 random samples. Using the VGA the NN was trained in only 12 generations.

Once the GMLP was trained we generated a different set of 400 samples and were able to obtain a correct classification for this test sample in all but 3 of the elements of the sample. A remarkable 99.25% efficiency.

## VI. CONCLUSIONS

After performing the training and cross validation process we found that classification via GMLPs turned out to be much more efficient than other methods in use [8]. Interestingly, as mentioned before, the learning process

required of only 12 generations. We could have, as well, hardened the stopping conditions for the algorithm (in an attempt to get even better results) but we feel that the possible advantages would be minimal whereas the training process could become significantly longer.

We must stress that the data we used (a total of 1400 samples) were mostly simulated even though their validity was derived from statistical considerations. Simulations were obtained from experimental data and taking into consideration possible measurement errors. An immediate follow-up to our results would be to increase the experimental data, re-train our network, compare the reported results with the new ones and compare the relative accuracy and cost, since experimental data is relatively expensive and difficult to obtain.

At any rate, we believe that the method reported is a viable and interesting alternative in trying to ease gender oriented artificially induced fertilization techniques.

## REFERENCES


[1] E.F. Fugger, S.H. Black, K. Keyvanfar, Y.D. Schulman, "Births of Normal Daughters after MicroSort Sperm Separation and Intrauterine Insemination, in-vitro fertilization, or Intracytoplasmic Sperm Injection", *Human Reproduction*, 1998.

[2] H. Morgan , H. Bourlard, "Continuous Speech Recognition using Multi-Layer Perceptrons with Hidden Markov Models", P*roc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Albuquerque, vol. 1, pp. 413-416, 1990.

[3] D. Cohn, G. Tesauro, "How Tight are the Vapnik-Cervonenkis Bounds?", *Neural Computation*, vol. 4, pp. 249-269, 1992.

[4] A. Kuri, "Training Neural Networks using Non-standard Norms – Preliminary Results", *Lecture Notes in Computer Science*, Springer-Verlag, pp. 350-364, 2000.

[5] J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, Michigan: University of Michigan Press, 1975.

[6] A. Kuri, "A Universal Genetic Algorithm for Constrained Optimization", *Proc. 6th European Congress on Intelligent Techniques and Soft Computing*, Aachen, Germany, pp. 518-522, 1998.

[7] Kuri, A., "A Methodology for the Statistical Characterization of Genetic Algorithms", MICAI 2002: Advances in Artificial Intelligence, Lecture Notes in Artificial Intelligence, pp. 79-88, Springer-Verlag, 2002.

[8] S. P. Flaherty, C. D. Mathews, "Application of Modern Molecular Techniques to Evaluate Sperm Sex Selection Methods", *Molecular Human Reproduction*, vol. 2, pp. 937-942, 1996.