# Cluster Analysis and Artificial Neural Network on the Superovulatory Response Prediction in Mice

Gabriela Berni Brianezi, Fernando Frei, José Celso Rocha and
Marcelo Fábio Gouveia Nogueira

Department of Biological Sciences, College of Sciences and Letters
São Paulo State University (UNESP), Campus Assis,
Av Dom Antonio 2100, Vila Tenis Clube, CEP 19806900, Assis, São Paulo, Brazil

**Abstract.** Complex biological systems require sophisticated approach for analysis, once there are variables with distinct measure levels to be analyzed at the same time in them. The mouse assisted reproduction, *e.g.* superovulation and viable embryos production, demand a multidisciplinary control of the environment, endocrinologic and physiologic status of the animals, of the stressing factors and the conditions which are favorable to their copulation and subsequently oocyte fertilization. In the past, analyses with a simplified approach of these variables were not well succeeded to predict the situations that viable embryos were obtained in mice. Thereby, we suggest a more complex approach with association of the Cluster Analysis and the Artificial Neural Network to predict embryo production in superovulated mice. A robust prediction could avoid the useless death of animals and would allow an ethic management of them in experiments requiring mouse embryo.

## 1 Superovulation and Embryonic Production

The mice superovulation (SOV) is a pharmacological treatment aiming the supernumerary induction of ovulation, by administration of eCG and hCG hormones. They have the bioactivity of FSH and LH hormones, responsible for the growth of antral follicles and the induction of ovulation. After the SOV, it could be expected a higher amount of oocytes to be delivered, captured by infundibulum and transported to the oviduct ampoule, where the fertilization will occur. The final objective of the SOV would be to produce a higher number of viable embryos compared to the physiologic pattern of the specie.

In mouse (*Mus musculus*) a mucous plug is formed in the vulva after copula, by seminal plasma components of the semen, therefore it is an indicative of the copula occurrence. Concomitantly to the SOV, aspects of the environment and manipulation of the animals may interfere with the viable embryo production. Temperature and humidity variations may lead to changes in blood circulation and in airways leading animals to a more infection susceptibility condition. The lack of cages cleaning increases ammonia concentration in the air and the probability of the animals to suffer cutaneous and pulmonary irritations. The environmental light intensity is one of the most important variables, since it stimulates (by optical nerve) the pituitary to secrete

Berni Brianezi G., Frei F., Rocha J. and Fábio Gouveia Nogueira M..
Cluster Analysis and Artificial Neural Network on the Superovulatory Response Prediction in Mice.
DOI: 10.5220/0003876600790084
In *Proceedings of the International Workshop on Veterinary Biosignals and Biodevices* (VBB-2012), pages 79-84
ISBN: 978-989-8425-94-2
Copyright © 2012 SCITEPRESS (Science and Technology Publications, Lda.)

sexual hormones. The physiological activities occur in a circadian rhythm and, therefore it is important that the light of the animal facility mimics the natural phases of light and dark, keeping the physiological pattern of the specie. The feeding of animals must be appropriate to specie and age, being the excess of energy (mainly fat) inductor of low fertility. All the variables mentioned above, resemble in a factor: they generate, directly or indirectly, stress to the animals. Furthermore, other changes in their environment, like intense odors and noise, are known stressors. Once stressed, the males tend to lose the libido and the females do not allow the copulation due to a change at their sexual receptivity. Beyond the environmental factors, the wrong administration of hormones (*e.g.* the reflux of the drug) interferes with SOV yield.

There are not reliable and accurate methods for the embryonic prediction after mice SOV. The plug is the only biological marker to copula occurrence not being enough to predict if there was fertilization and if zygotes developed to viable embryos. Currently, the predictive analysis for this purpose is subjective and unreliable. The absence of accurate methods, which could predict the viable embryos occurrence after mouse superovulation, induces to death several animals that do not produce any viable embryo. Ethically this is not accepted and the 3R principles (reduce, replace and refine) should be searched. In this way, the proposed approach could avoid or minimize unnecessary deaths.

Few articles tried a similar approach (but not making use of Cluster Analysis and ANN) to predict biological complex systems as mice SOV and embryonic production. Stevens [11] reports the use of Cluster Analysis to classify the embryonic development of Blue King crab *Paralithodes platypus*, with similar results to the image analysis. Nevertheless, Faeder [5] proposed that complex biological systems - *e.g.* interaction among molecules, cells and environment - should be analyzed by the development of multiscale and multilevel biological models.

## 2 Cluster Analysis

Cluster Analysis is a generic name assigned to several statistics methods which try to elaborate judgments to group objects. These are multivariate statistics techniques, with exploratory connotation. In this way, a given sample with several objects (animals for instance) which was measured by several variables, seeks a classification scheme to cluster the animals on similar groups. Nowadays, Cluster Analysis is applied in many areas. The set of results from those techniques could contribute to the definition of a formal classification scheme; it also may suggest a set of rules to classify new objects in new classes intending some diagnostics; it presents statistics models suggestions to describe populations, find objects which might represents groups or classes and obtain population subgroups, following certain study features. Thereby, the Cluster Analysis results may help the multidimensional space patterns discovery.

### 2.1 Similarity Measures and Group Algorithm

Similarity coefficients quantify how similar the objects are inside the analyzed meas-

ures, and their choice is very important for Cluster Analysis. For this purpose, the variable measurements should be observed, in other words, if they are qualitative, quantitative, or a mixture of these measurements, as in the proposed study. To measure the similarity among the studied objects it is used the Gower's index ($S_{ikh}$), which defines similarity coefficients in the interval from 0 to 1 between $i$ and $k$ units to each $h$ variable from a total of any $p$ variables type. The measure of these similarities results in Gower's coefficient.

$$S_{ik} = \frac{\sum_{h=1}^{p} t_{ikh} \times u_{ikh}}{\sum_{j=1}^{p} u_{ikj}} \quad (1)$$

In which $t_{ikh}$ and $u_{ikh}$ are assigned according with the $h$ variable type.
(a) If the $h$ variable is binary and

| | $t_{ikh}$ | and $u_{ikh}$ |
|---|---|---|
| present in 2 units | 1 | 1 |
| absent in 2 units | 0 | 0 |
| the units do not match | 0 | 1 |

(b) If the $h$ variable is qualitative multistate, $t_{ikh} = 1$ if the units $i$, $k$ match with $h$ variable, and $t_{ikh} = 0$ if they do not. Both cases $u_{ikh} = 1$.
(c) If the $h$ variable is quantitative

$$t_{ikh} = 1 - \frac{|v_{ih} - v_{kh}|}{(\max v_h - \min v_h)}, \; u_{ikh} = 1. \quad (2)$$

$v_{ih}$ is the result of $h$ variable, $i$ unit.

There are several algorithms and methods to perform clustering of animals after the similarity measurement. In this work we will use algorithms from Agglomerative Hierarchical Method, which has shown to be a useful tool on discovery of inherent substructures to a certain data group. The Agglomerative Hierarchical algorithms searches the objects which have the lower distance each other, inside the similarity matrix, and group them. Once grouped, those animals remain in these groups throughout the successive clustering stages. The main difficulty of Cluster Analysis is to set the quantity of groups. Furthermore, different groups come out when used different algorithms. Thereby, one of the recommendations is to use several algorithms. If the results presents similar substructures a natural partition was obtained, otherwise it is unlikely that the data represents natural distinct groups. Algorithms to be used: Complete Linkage, Average Linkage and Ward Method.

By Cluster Analysis it is expected to be found natural groups which may indicate patterns (on the animals or environment) to maximize the number of viable embryos. In this way, these patterns could be implemented in researches involving embryonic production in mice.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

## 3 Artificial Neural Networks – ANN

The Artificial Neural Network is a modern math technique which intends to mimic the biological neurons functioning through computational models, in other words it is an attempt of reproduce the human brain computationally, in order to resolve high complexity problems.

An important feature of ANN is the ability of learning from real known cases, which provides its performance improvement. The learning is made by an iterative process and happens when the ANN reaches a general solution of problems previously presented to it [1].

In this sense the ANN application to find solutions to complex problems, as the superovulation and embryonic production presented in this work, should lead the ANN to predict viable embryo production from superovulated mouse.

### 3.1 ANN Training Algorithm

To this work the ANN to be used is named multilayer and uses the Backpropagation training algorithm, which has supervisionated learning and was chosen by presenting more efficiency in this type of work. The multilayer networks have their name because they present more than one neuron interconnected layer.

We could schematize a multilayer ANN as follow in the Figure 1.

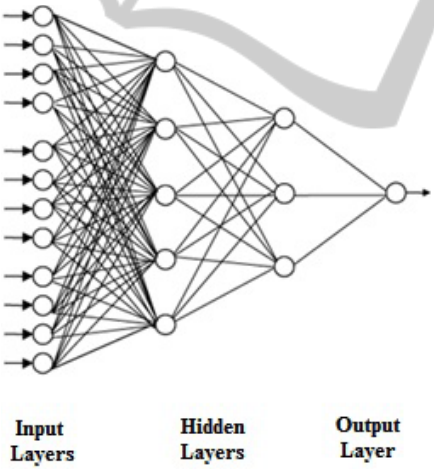


**Fig. 1.** An ANN multilayer scheme.

The input layer corresponds to the chosen variables as parameters for the ANN training. The hidden layers are intermediary layers and may lead the network to an easier understanding of the process of learning. The output layer is the one which gives the ANN answer and it is compared to the results we intend to obtain from the ANN. The Backpropagation algorithm is characterized by two stages. At the first one, named feedforward, the variables are processed and the ANN response is compared with the experimentally results obtained. If the difference between the responses is

greater than the pre-established determined error, the network starts the feedbackward stage, which recalculates the internal parameters. The stages are repeated until the final result to be inside previously established error [9]. After training, the ANN will be able to predict the occurrence of viable embryos from superovulated mouse.

## 4 Hypothesis and Variables to Be Analyzed

Due the complexity of the variables that influence the superovulatory response which involves the animal, the environment and its manipulation, an objective analysis would be potentially more effective than the current subjective one. The work hypothesis of our group is to test the Cluster Analysis or Artificial Neural Network for this prediction and to compare the obtained results to the current ones (subjectively inferred). The following variables will be use to the Cluster Analysis and the Artificial Neural Network, and produced data – prediction of variables which might be indicatives of viable embryo production – will be compared to the data of the embryo yielded by animals. The variables to be analyzed are:

1) female weight (g) at the embryo recovery day;
2) male weight (g) at the embryo recovery day;
3) copulatory plug presence or absence;
4) presence or absence of reflux from hormones administration;
5) female age (days) at the embryo recovery day;
6) male age (days) at the embryo recovery day;
7) animal lineage (Swiss Webster or C57BL/6);
8) animal facility location (A or B);
9) origin of the female (in-house or obtained out of university Campus);
10) origin of male (in-house or obtained out of university Campus);
11) amount of viable embryos yielded in the recovery day (0, 1, 2, 3, *etc*);
12) season at the embryo recovery day;
13) light intensity (lux) into copulation cage;
14) temperature (ºC) into copulation cage.

## References


1. Braga, A. P.; Ludemir, T. B. *Redes Neurais Artificiais – Teoria e Aplicações* (2nd ed.). Rio de Janeiro: LTC (2007).
2. Bryson, A. E.; Ho, Y-C. *Applied optimal control: optimization, estimation and control.* Blaisdell Publishing Company (1969).
3. Damy, S. B. *et al*. Aspectos fundamentais da experimentação animal – aplicações em cirurgia experimental. *Rev Assoc Med Bras*, (2010) 56(1): 103-11.
4. Everitt, B. S. (1993). *Cluster Analysis* (3rd. ed.). New York: John Wiley & Son.
5. Faeder, R. J. Toward a comprehensive language for biological systems. *BMC Biology*, (2011) 9:68. doi:10.1186/1741-7007-9-68.
6. Frei, F. *Introdução à Análise de Agrupamentos: Teoria e Prática*. Sao Paulo State: Editora fundação UNESP (2006).
7. Haykin, S. *Neural Networks*. New York: MacMillan College Publishing Company (1994).


SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

8. Kaufman, L.; Rousseeuw P. J. Finding *groups in data: An introduction to cluster analysis*. New York: John Wiley & Sons (1990).
9. Kovács, Z. L. *Redes neurais artificiais: Fundamentos e Aplicações* (3rd ed.). São Paulo: Livraria da Física Editora (2002).
10 Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev*, (1958) 65: 386–408.
11 Stevens, B. G. Embryo Development and Morphometry in the The Blue King Crab *Paralithodes Platypus* Studied by Using Image and Cluster Analysis. *Journal of Shellfish Research*, (2006) 25(2), 569-576.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS