Araştırma Makalesi / Research Article

Seminal Quality Prediction Using Deep Learning Based on Artificial Intelligence

Hilal Benli [*1] iD **, Bülent Haznedar** [2] iD **, Adem Kalınlı** [3] iD

[1]*Erciyes University, Institute of Science and Technology, Department of Computer Engineering, 38039 Kayseri, TÜRKİYE*
[2]*Hasan Kalyoncu University, Engineering Faculty, Department of Computer Engineering, 27100 Gaziantep, TÜRKİYE*
[3]*Erciyes University, Engineering Faculty, Department of Computer Engineering, 38039 Kayseri, TÜRKİYE*

Abstract

Fertility rates have dramatically decreased in the last two decades, especially in men. It has been described that environmental factors, as well as life habits, may affect semen quality. This paper evaluates the performance of different artificial intelligence (AI) techniques for classifying fertility dataset that includes the semen sample analysed according to WHO 2010 criteria and publicly available on UCI data repository. In this context, deep neural network (DNN) which involved in many studies in recent years is proposed to classify fertility dataset successfully. For the purpose of comparing the proposed method's performance, Adaptive Neuro-Fuzzy Inference system (ANFIS) is also used for the classification problem. The results show that the performance of the DNN has the best with the average accuracy rate of 90.11%, and the results of the other ANFIS methods are also satisfactory.

Key Words
"Classification, fertility, statistical methods, artificial intelligence, deep learning, ANFIS"

*Sorumlu Yazar: hilalbenli@erciyes.edu.tr

1. INTRODUCTION

In recent years there has been a decline in fertility rates, especially in men. Fertility rate varies depending on some environmental factors and living conditions. There are several reasons that affect the amount of semen quality in men, such as environmental and seasonal factors, accident or serious trauma, childish diseases, smoking and daily activity. The analysis of the semen quality is important and a good predictor for male fertility potential.

Various studies have been carried out in this issue. Gil et al. (2012) compared three artificial intelligence methods: Decision Trees, Multilayer Perceptron (MLP) and Support Vector Machines, for prediction of the seminal quality. Girela et al. (2013) proposed a neural network approach based on MLP in order to estimate semen parameters. Wang et al. (2014) proposed a Clustering-Based Decision Forests method to tackle unbalanced class learning problem in seminal quality prediction. Bidgoli et al. (2015) proposed an optimized MLP for predicting seminal quality.

Deep learning, a subfield of machine learning, is a growing trend that is becoming increasingly common in general data analysis. In this paper, we compare the prediction performance of DNN and ANFIS trained using Hybrid (HB) algorithm, Genetic Algorithm (GA) and Simulated Annealing (SA) algorithm, by calculating the effects of environmental factors and life habits on male fertility potential. The rest of the paper is organized as follows: In the section 2, deep learning and neural network structure are presented. In the section 3, the dataset and proposed method are described, and finally in the section 4, the results and discussions are specified.

2. DEEP LEARNING: AN OVERVIEW

Deep Learning is a new area of machine learning research and has become a widely used method in recent years both academia and the industry. Deep learning is an improvement of artificial neural networks, consisting of more layers that permit higher levels of abstraction and improved predictions from data (LeCun et al., 2015).

Artificial neural networks (ANN) have been developed inspired by the working structure of the human brain and widely used in many applications. ANN is composed of a number of neurons and one of the artificial intelligence techniques trained to generate an output from a combination of input data. The elementary neuron is simulated by a model such as:
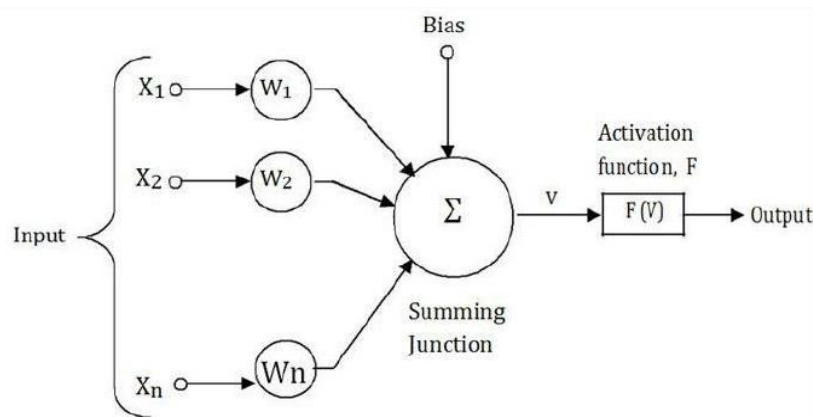

Fig. 1. An elementary neuron with n inputs

The output of the neuron is $a = f(z)$, where $z = \sum_j w_j x_j + bias$ is the weighted sum of the inputs. Feed-forward neural network architectures composed of multiple layers of neurons and each layer feeds the subsequent layer.
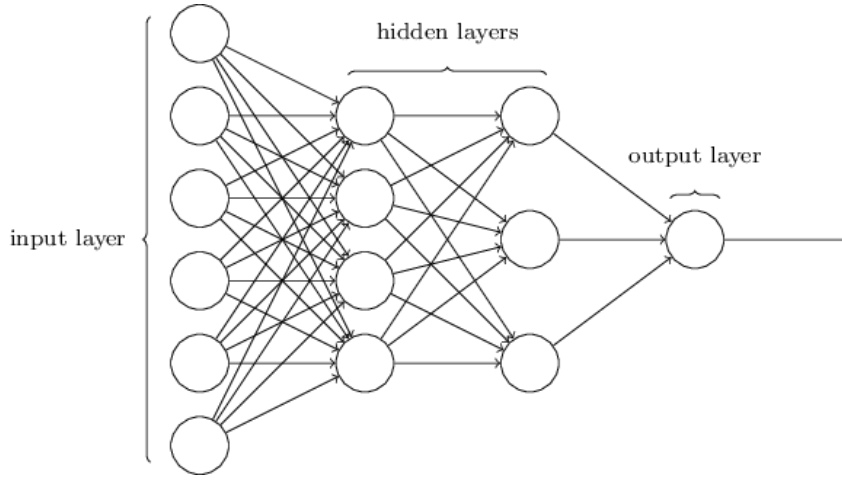
Fig. 2. An illustration of the neural network structure.

A feed-forward fully-connected multi-layer neural network is known as Deep Neural Network (DNN). DNN has become a widely used method in classification problems because of its excellent classification performance in recent years. In DNN each layer of nodes trains on a distinct set of features based on the previous layer's output and can be modelled as Eqs. (1) and (2).

$$h(n) = y(n-1)w(n) + bias \tag{1}$$

$$y(n) = f(h(n)) \tag{2}$$

$h(n)$ denotes the hidden layer from *1* to *N*. $y(n-1)$ is the output of the previous layer, $w(n)$ is the learnable weights and $bias$ is the learnable biases of the layer. f is the nonlinear activation function. The activation function used in hidden layers for feed-forward architecture is the Rectified Linear Unit (ReLU) the most used nonlinear function in recent years and modelled as Eq. (3).

$$f(x) = \max(0, x) \tag{3}$$

Sigmoid function given in Eq. (4) is used on the output layer for predicting the class.

$$f(x) = sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

Root Mean Square Propagation (RMSprop), an objective function is used in order to correctly classify the instances/classes, update the parameters of the model and reduce the error. RMSprop is a robust optimizer and applicable to mini batch learning (Tieleman&Hinton, 2012). In RMSprop, a moving average of the squared gradient is kept for each weight time step *t*:

$$r_t = (1 - \gamma) f'(\theta_t)^2 + \gamma r_{t-1} \tag{5}$$

$$v_{t+1} = \frac{\eta}{\sqrt{r_t}} f'(\theta_t) \tag{6}$$

$$\theta_{t+1} = \theta_t - v_{t+1} \tag{7}$$

Eqs. (5), (6) and (7) are the formulas of the parameter updates, where η is the learning rate, γ is decay term and θ is parameters of the network.

3. THE PROPOSED METHOD

The study is performed to classify fertility dataset (Gil et al., 2012) obtained from the UCI Machine learning repository. The fertility dataset includes a labelled dataset of 100 sample who were between 18 and 36 years old provided a semen sample for analysis as well as their socio-demographic data, environmental factors, health status and life habits. 100 volunteers provide a semen sample analysed according to the World Health Organization 2010 criteria (WHO, 1999). The dataset includes two classes

as semen quality: "normal" (N) and "altered" (O). There are 12 samples as altered and 88 as normal. This implies that the dataset is so unbalanced and classification of labelled as "altered" instances in the whole dataset is very hard. The details of the data features with their values normalized can be seen from Table 1.

Table 1. List of the variables with their descriptions and their values range.

Feature description	Values	Normalized
Season in which the analysis was performed	1) winter, 2) spring, 3) Summer, 4) fall	(-1, -0.33, 0.33, 1)
Age at the time of analysis	18-36	(0, 1)
Childish diseases	1) yes, 2) no	(0, 1)
Accident or serious trauma	1) yes, 2) no	(0, 1)
Surgical intervention	1) yes, 2) no	(0, 1)
High fevers in the last year	1) less than three months ago, 2) more than three months ago, 3) no	(-1, 0, 1)
Frequency of alcohol consumption	1) several times a day, 2) every day, 3) several times a week,	(0, 1)
	4) once a week, 5) hardly ever or never	
Smoking habit	1) never, 2) occasional , 3) daily	(-1, 0, 1)
Number of hours spent sitting per day	0-16	(0, 1)

To evaluate the performances of the methods, the dataset is randomly divided into two subsets. The train set (70%) is used for training, the remaining 30% is used for testing. This method is used only in the first part of the work. In addition, because of the limited number of "altered" class in dataset, the cross-validation method is used to accurately evaluate the model's ability to generalize. The goal of the cross-validation method is to test the validity of the results by repeating an experiment on independent conditions. For example, in statistical classification problems, after dividing a set of data into approximately equal k clusters, use k-1 cluster to construct the classifier for each k, and test the classifier with the remaining cluster. This method is given in Fig. 3. In this study, 5-fold and 10-fold cross-validation are applied for the performance assessment of network. In this study, model assessment is done for 15 times. For this goal the whole dataset is divided into 5 and 10 subsets randomly. Every time the one part of the data is determined as test set and the remaining form training data. The accuracy rate (%) of each model at every iteration are accumulated to provide the mean accuracy rate (%).
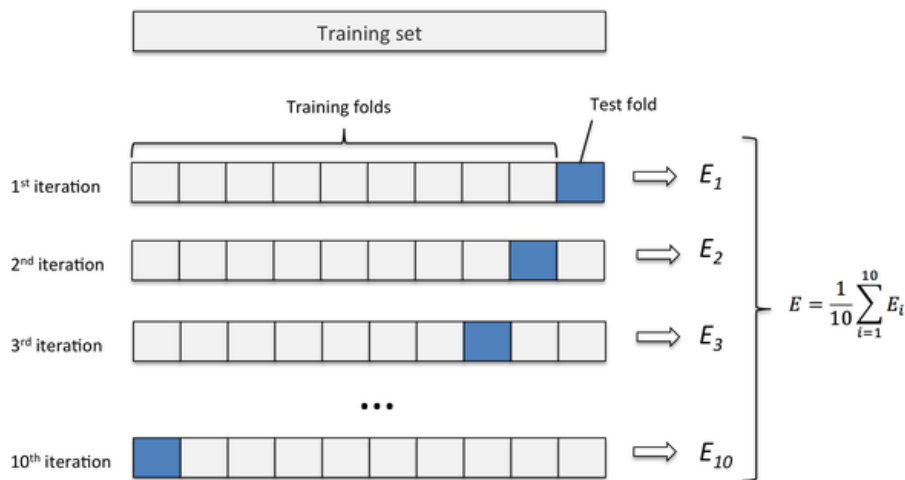

Fig. 3. 10-Fold Cross Validation Method

In this study, in order to classify fertility dataset successfully, DNN method which involved in many studies in recent years is proposed. For the simulation studies, Keras deep learning library is used to create a feed-forward neural network. Many attempts are made to decide on the various control parameter for the DNN. Afterwards, 4 hidden layers with 48, 36, 12, 6 neurons for per layer, respectively are used and the sigmoid logistic regression function is applied to the output layer. Also, RMSprop algorithm is chosen for optimisation of the proposed model. The momentum coefficient is 0.9, learning rate is 0.03 and the weight decay is 0.00005. The dropout rate is set to 0.2 in order to prevent network from overfitting.

For the purpose of comparing the proposed method's performance, Adaptive Neuro-Fuzzy Inference system (ANFIS) is also used for the classification problem. The ANFIS network is trained with three algorithm. The one of them is HB which is derivative-

based algorithm. HB is used to optimize ANFIS network parameters. This algorithm is formed from the combination of "gradient descent" and "least squares" methods. The gradient descent method is used for regulating nonlinear input parameters and least squares method is used for regulating linear output parameters (Haznedar et al., 2018). The other ones are artificial intelligence optimization algorithms such as GA and SA algorithm. GA is a heuristic algorithm which is used for being able to find exact or approximate results in optimization or search problem (Haznedar et al., 2017). SA is a probability-based meta-heuristic algorithm which was first proposed by Kirkpatrick et al., (1983). The SA algorithm is based on the similarity between the physical annealing of solids and the solution of the combinatorial optimisation problems (Haznedar & Kalinli, 2018). Thus, ANFIS parameters are optimized with GA and SA too.

The performance of those algorithms for training ANFIS are compared with the proposed DNN method for the classification of fertility data. In this context, the population size is taken 25, the crossing over rate is 0.8 and the mutation rate is chosen as 0.01 for the used GA. Also, the number of temperature points, the number of iterations for each temperature point and the temperature reducing parameter are chosen 10, 8 and 0.2, respectively for the SA algorithm.

4. RESULTS AND DISCUSSIONS

In this study, the performance of the proposed DNN method is compared with 3 different classifiers including ANFIS-SA, ANFIS-GA and ANFIS-HB. Accuracy (ACC), Sensitivity (SN), Specificity (SP) and Precision (PREC) criteria are calculated by considering in the confusion matrix given in Table 2 are used to evaluate the performance of the classification methods. ACC is calculated as the number of all correct predictions divided by the total number of the dataset. The best accuracy is 1.0, whereas the worst is 0.0. It can also be presented as rate (%) which is defined by Eq. (8). The accuracy measure cannot be considered alone for evaluating a method in unbalanced dataset such fertility dataset. Therefore, SN, SP and PREC that are given in Eqs. (9), (10) and (11) are used for comparing of the proposed and the other methods over the fertility dataset. SN is calculated as the number of correct positive predictions divided by the total number of positives. It is also called recall (REC). The best SN is 1.0, whereas the worst is 0.0. SP is calculated as the number of correct negative predictions divided by the total number of negatives. It is also called true negative rate (TNR). The best SP is 1.0, whereas the worst is 0.0. PREC is calculated as the number of correct positive predictions divided by the total number of positive predictions. It is also called positive predictive value (PPV). The best PREC is 1.0, whereas the worst is 0.0. The definition of these criteria given as follows (Bidgoli et al., 2015):

True Positive (TP): Observation is positive, and prediction is positive.
False Negative (FN): Observation is positive, but prediction is negative.
False Positive (FP): Observation is negative, but prediction is positive.
True Negative (TN): Observation is negative, and prediction is negative.

$$Accuracy\,(\%) = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) * 100 \tag{8}$$

$$Sensitivity\,(\%) = \left(\frac{TP}{TP + FN} \right) * 100 \tag{9}$$

$$Specificity\,(\%) = \left(\frac{TN}{TN + FP} \right) * 100 \tag{10}$$

$$Precision\,(\%) = \left(\frac{TP}{TP + FP} \right) * 100 \tag{11}$$

Table 2. The Confusion Matrix

	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Simulation studies made using all methods presented in this study are worked by diversifying data set with 3 stage. In the first stage, the data set is divided into two random subsets as the training set (70%) and the testing set (30%). Furthermore in the second and third stage, for balancing the data set 5-fold and 10-fold cross validation method are implemented. In this context, the average performance of the proposed and the other classification methods are compared in the Tables 3, 4 and 5. From the obtained results the proposed method has higher prediction accuracy for classification of the male fertility potential.

Table 3. The performances of the methods on fertility data for 70% training and 30% testing

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
ANFIS-HB	70.00	90.90	12.50	74.07
ANFIS-GA	86.66	92.59	33.33	92.59
ANFIS-SA	90.00	96.29	50.00	92.85
DNN	**93.33**	**100.00**	**100.00**	**93.10**

Table 4. The performances of the methods on fertility data for 5-fold

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
ANFIS-HB	83.00	93.18	14.28	88.17
ANFIS-GA	86.00	96.59	25.00	88.54
ANFIS-SA	88.00	98.86	50.00	88.77
DNN	**88.00**	**98.86**	**50.00**	**88.77**

Table 5. The performances of the methods on fertility data for 10-fold

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
ANFIS-HB	86.00	94.31	37.50	90.21
ANFIS-GA	86.00	93.18	40.00	91.11
ANFIS-SA	88.00	94.31	50.00	92.22
DNN	**89.00**	**95.45**	**55.55**	**92.30**

It can be seen that the value of Accuracy, Sensitivity, Specificity and Precision predictive of DNN is the best in compared methods. Its success criterions is approximately similar to ANFIS-SA. Generally, the specificity values of all methods are found lower. The reason for this occurrence is unbalancing data set include only 12 samples as "altered". For best generalization of the methods the cross k-fold validation is utilized. In this study, 5-fold and 10-fold cross-validation employed. Tables 3, 4 and 5 shows the performance measurements of the methods. It is demonstrated that the performance of methods enhances when the value of k increases. It is clear that when k increases the number of training samples increase, therefore the better classification accuracy is achieved.

In this study, we have applied 3 different data sets and evaluated results from these data sets. Besides, the proposed and other methods' average performance values (%) obtained by using the different type data sets are given in the Fig. 4. The Fig.4 shows that the performance of the DNN has the best with the average accuracy rate of 90.11%, and the results of the other ANFIS methods are also satisfactory.
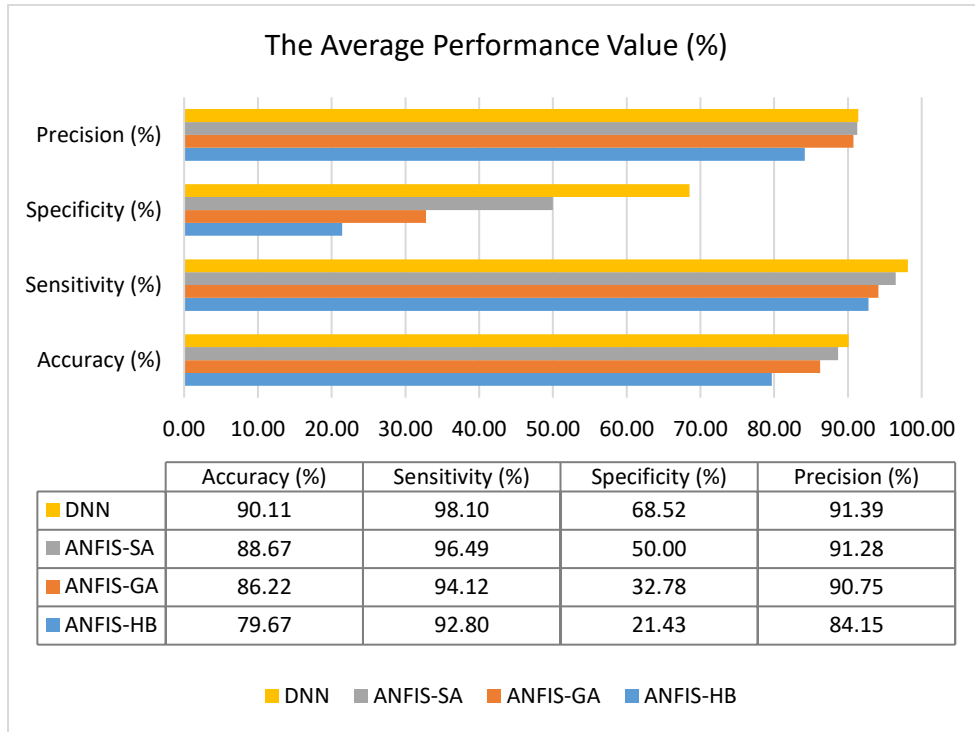
The Average Performance Value (%)

	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
DNN	90.11	98.10	68.52	91.39
ANFIS-SA	88.67	96.49	50.00	91.28
ANFIS-GA	86.22	94.12	32.78	90.75
ANFIS-HB	79.67	92.80	21.43	84.15

Fig. 4. The average performance values obtained by using 3 different type data set

5. CONCLUSIONS

In this paper, the performance of the DNN and ANFIS methods have been evaluated in the prediction of the male fertility potential. The relationship of life habits and environmental factors with semen parameters have been showed with the obtained results from 100 volunteers. It is seen in the Tables 3, 4 and 5. The proposed DNN method has the highest accuracy rate over the other ANFIS methods to classify the fertility dataset. In addition, the performance of the ANFIS that optimised by using derivative-based HB algorithm are poorer than the meta-heuristic GA and SA algorithms. By reason of the fact that calculating the gradient is too difficult in derivative-based algorithms. Moreover, the comparison of the results given in Tables 3, 4 and 5 demonstrate that DNN model has the highest SN, SP and PREC rates computed from confusion matrix. This rates are used to evaluate the performance of classification models used in machine learning. Thus, the evaluation of the proposed method with different measures increases the reliability and robustness of the study.

Different AI methods have been proposed for classification of fertility data set. For example, Gil et al. (2012) predicted seminal quality with artificial intelligence methods. They used MLP, SVM and DT methods in this study. They applied 10-fold cross validation for performance assessment. The results indicated that MLP and SVM had the best performance with accuracy rate of 86%, sensitivity of 94.1% and specificity of 40%, respectively. It is seen in the Table 5, DNN with 10-fold cross validation method is more successful with accuracy rate of 89%, sensitivity of 95.45% and specificity of 55.55%, respectively.

In conclusion, to our knowledge, this is the first study that the proposed DNN method has been compared for prediction of the fertility data set. From the obtained results, it is seen that DNN is quite a successful model for the classification of these data set. Furthermore in the future studies, the proposed model will be applied to different problems.

CONFLICTS OF INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this paper.

ACKNOWLEDGEMENT

REFERENCES

Bidgoli, A.A., Komleh, H.E., & Mousavirad, S.J., (2015). Seminal quality prediction using optimized artificial neural network with genetic algorithm. 9th International Conference on Electrical and Electronics Engineering (ELECO), 695-699.

Gil, D., Girela, J. L., De Juan, J., Gomez-Torres, M. J., & Johnsson, M., (2012). Predicting seminal quality with artificial intelligence methods. Expert Systems with Applications. 39, 12564-12573.

Gil, D., & Girela, J.L., (accessed on 8 November 2013). UCI Machine Learning Repository: Fertility data set. Available online: http://archive.ics.uci.edu/ml/datasets/Fertility.

Girela, J. L., Gil, D., Johnsson, M., Gomez-Torres, M. J., & De Juan, J., (2013). Semen parameters can be predicted from environmental factors and lifestyle using artificial intelligence methods. Biology of Reproduction, 88, 99.

Haznedar, B., Arslan, M.T., & Kalinli, A., (2018). Using adaptive neuro-fuzzy inference system for classification of microarray gene expression cancer profiles. Tamap Journal of Engineering. 2018 (29), 1-13.

Haznedar, B., Arslan, M.T., & Kalinli, A., (2017). Training ANFIS structure using genetic algorithm for liver cancer classification based on microarray gene expression data. Sakarya University Journal of Science. 21, 54-62.

Haznedar, B., & Kalinli A., (2018). Training ANFIS structure using simulated annealing algorithm for dynamic systems identification. NEUROCOMPUTING. 302, 66-74.

Kirkpatrick, S., Gelatt, C.D., & Vecchi, M.P., (1983). Optimization by simulated annealing. Science. 220, 671–680.

LeCun, Y., Bengio, Y., & Hinton, G., (2015). Deep learning. Nature. 521 (7553), 436–444.

Lu, J., Huang, Y., & LA, N., (2010). [WHO laboratory manual for the examination and processing of human semen: Its applicability to andrology laboratories in China]," Zhonghua nan ke xue= National journal of andrology. 16 (10), 867- 871.

Tieleman, T., & Hinton, G., (2012). Lecture 6.5-RMSprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 4.

Wang, H., Xu, Q., & Zhou, L., (2014). Seminal quality prediction using clustering-based decision forests. Algorithms. 7, 405-417.