

# A Review and Analysis on Artificial Intelligence Methods Performance Comparison through Seminal Quality Prediction

## I. Introduction

According to numerous previous studies, there is an obvious decrease in fertility rates during the past two decades. One possible reason behind this decrease is the decline in semen quality. In the subject literature, several factors have been considered to explain this decline, for example, the increase in the incidence of male reproductive diseases, and the effect of environmental or occupational factors to a certain lifestyle.

Meanwhile, artificial intelligence techniques has become an emerging methodology as decision support systems in medicine nowadays. Gil et al. (2012) paper evaluated and compared the performance of three different artificial intelligence techniques, decision trees, Multilayer Perceptron, and Support Vector Machines, by employing them to predict the seminal quality from the data of the environmental factors and lifestyle.

This project report will describe Gil et al. (2012) paper's contribution, discuss the issues, challenges, and limitations of this study, as well as replicate one experimental plot with R using data provided in the experiment.

## II. Issues and Challenge

## III. Methodology

## IV. Simulation Results

### 1. Data Description

The simulation data is extracted from the original data which was collected from the 100 University of Alicante students between the age of 18 and 36 years old. The researchers have done a feature selection process to the original data, and we have directly adopted the processed dataset as our simulation data. There are in total 100

records in the dataset with 9 variables and 1 column of output. The following table has listed out the variable names, descriptions and their data range.

Variable Name	Variable Description	Value (min-max)	Normalized
<b>Season</b>	Season in which the analysis was performed.	1) winter, 2) spring, 3) Summer, 4) fall	(-1, -0.33, 0.33, 1)
<b>Age</b>	Age at the time of analysis	18-36	(0, 1)
<b>ChiDiseases</b>	Childish diseases (ie , chicken pox, measles, mumps, polio)	1) yes, 2) no	(0, 1)
<b>Accident</b>	Accident or serious trauma	1) yes, 2) no	(0, 1)
<b>Surgery</b>	Surgical intervention	1) yes, 2) no	(0, 1)
<b>HighFever</b>	High fevers in the last year	1) less than three months ago, 2) more than three months ago, 3) no	(-1, 0, 1)
<b>FreAlcohol</b>	Frequency of alcohol consumption	1) several times a day, 2) every day, 3) several times a week, 4) once a week, 5) hardly ever or never	(0, 1)
<b>Smoking</b>	Smoking habit	1) never, 2) occasional 3) daily	(-1, 0, 1)
<b>HrSitting</b>	Number of hours spent sitting per day	1 - 16	(0, 1)
<b>Output</b>	Output: Diagnosis (Sperm Concentration)	normal (N), altered (O)	normal (N), altered (O)

## 2. Methodology

The aim of this simulation is to generate a Decision Tree Classification model for predicting whether the data object's sperm concentration is at a normal level.

The random forest method in caret package is used to perform the training and testing of the Decision Tree Classification model. Ten-fold cross-validation method is used to evaluate the performance of generated models without needing any other data than the training data. The ten-fold cross-validation method is embedded in caret's train function. In the paper, classification accuracy, sensitivity, specificity, positive predictive value, negative predictive value and a confusion matrix are generated in order to evaluate the three selected artificial intelligence method. Therefore, in our

simulation, these measures are also generated to be compared against the research's results.

### 3. Experiment R Script

```
fertility_training <- fertility_data

library(caret)

rf_model<-train(Output~.,data=fertility_training,method="rf",
               trControl=trainControl(method="cv",number=10),
               prox=TRUE,allowParallel=TRUE)

print(rf_model)

print(rf_model$finalModel)

plot(varImp(rf_model), top = 5)

testdata <- data.frame (Season = 1, Age = 0.65, ChiDiseases = 1, Accident = 0,
Surgery = 0, HighFever = 1, FreAlcohol = 0.6, Smoking = 0, HrSitting = 0.50)

prediction<- predict(rf_model, newdata = testdata)

print(prediction)
```

### 4. Results and Discussion

The generated confusion matrix for simulation is as follows:

```
Confusion matrix:
  N O class.error
N 84 4  0.04545455
O 10 2  0.83333333
```

The confusion matrix generated by the researchers of this study is as follows:

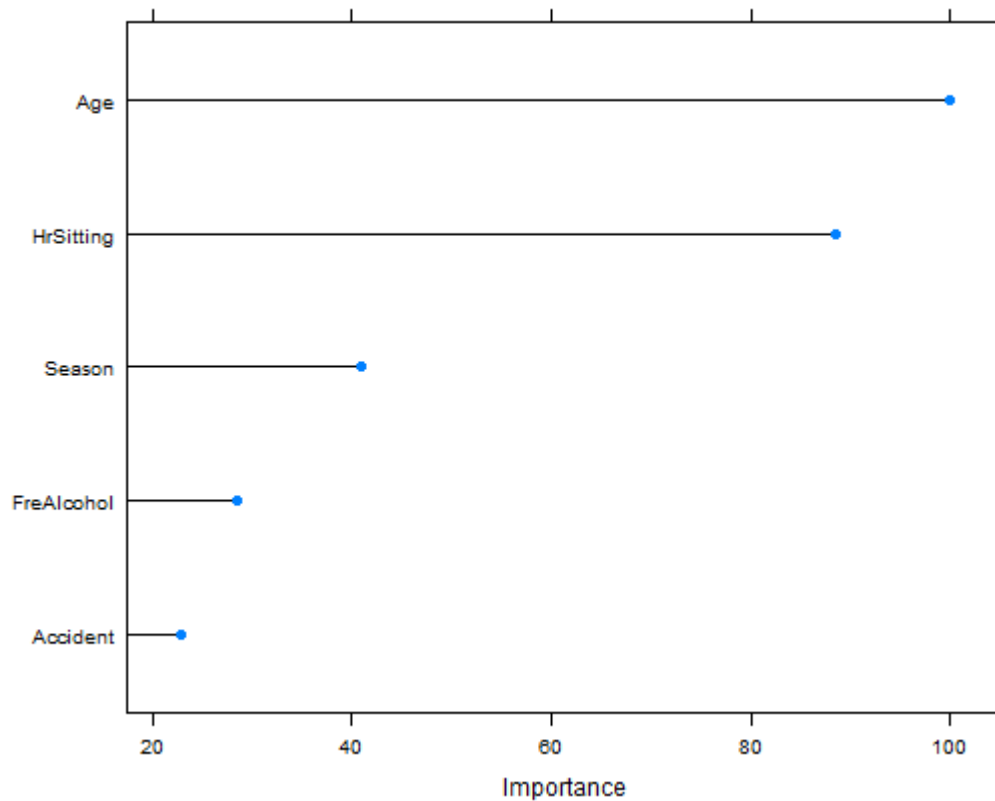
Actual	MLP Predicted		SVM Predicted		DT Predicted	
	P	N	P	N	P	N
P	80	5	83	2	82	3
N	9	6	12	3	13	2

The other measures are listed in the following table:

	Experiment Results	Simulation Results
<b>Classification accuracy (%)</b>	84	86.09091
<b>Sensitivity (%)</b>	96.5	95.45455
<b>Specificity (%)</b>	13.3	16.66667
<b>Positive predictive value (%)</b>	86.3	89.36170
<b>Negative predictive value (%)</b>	40	33.33333

From the comparison of results we can see that our results are similar to the results generated by researchers. Moreover, we have also created a new data record to test out the prediction. The result of the prediction is “Normal”, which should have an accuracy of 86.1%.

Furthermore, we have also identified the top five most important variables in determining the output, i.e. affecting the sperm concentration. These variables are Age, Number of hours spent sitting per day, Season, Frequency of alcohol consumption as well as Accident or serious trauma. The ranking of these variables are shown in the following graph:



## V. Contributions

There are mainly two contributions that this study has made in the field of development of artificial intelligence methods as well as its application in healthcare industry.

Many previous studies have tried to identify the relationship between environmental factors and semen quality using linear statistical techniques, which have shown limitations due to the complexity and diversity of information as well as heterogeneous data types. This study has adopted non-linear statistical techniques to come up with a better solution in tackling the complexity of the problem.

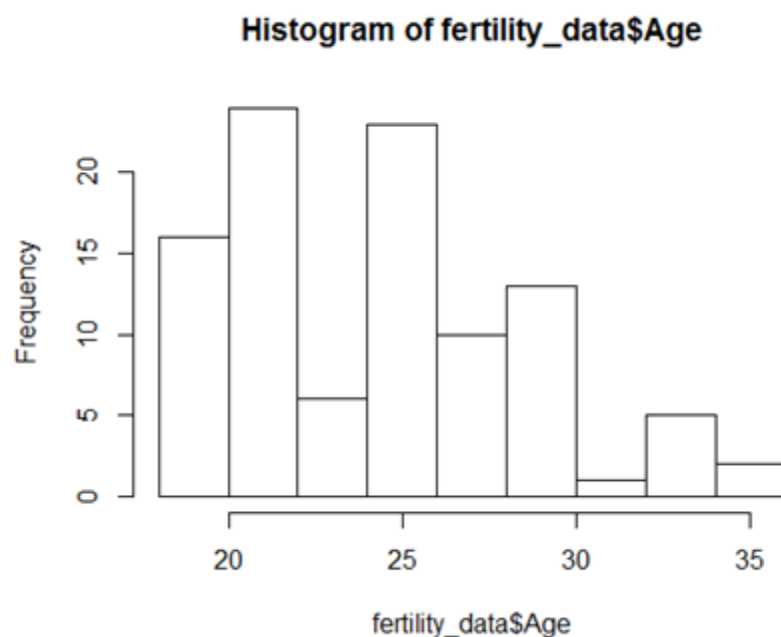
For the applications of the artificial intelligence techniques in the healthcare industry, they might be able to be served as an alternative to the more expensive laboratory tests during the initial stage of fertility tests of couples. Since the study sample mainly include healthy male university students, who represents characteristics of regular semen donor candidates, the developed models may also be used to cut costs in initial semen donor selection processes. These artificial intelligence methods could also be used in next stages

of fertility treatment to reduce and avoid painful and expensive examinations that the female partners need to go through.

## VI. Limitations

There are a few limitations observed in the experiments conducted in the study which might cast shadow on the usefulness and accuracy of the results predicted by the artificial intelligence methods developed in the study.

Firstly, the volunteers are all young healthy students from the University of Alicante in Spain between the age of 18 and 36 years old. The study did not specify to which population group the artificial intelligence methods could be applied. Moreover, the researchers used a non-probability sampling method which might not be representative enough for the target population. For example, if the targeted group is Spanish male between the age of 18 and 36 years old, as suggested by the graph below, the sample selected did not reflect proportion of the age groups according to demographic information in Spain. Moreover, in this specific situation, a sample size of 100 might also be too small to be representative of the entire target population. In future studies related to this topic, the researchers could improve sampling by clearly define the target population, refine reasonable sample size and use probability sampling method to make the study population more representative.



Secondly, for all three artificial intelligence methods used, the specificity and positive predictive values present low percentage. This might be due to the fact that the study sample population has an uneven distribution of normal and altered results. For example, 88% of the samples have normal sperm concentration while only 12% appear to have altered sperm concentration. This imbalanced distribution has affected the performance of the artificial intelligence methods, and further casted doubts on whether the predictions made are accurate. With regard to this limitation, the researchers have suggested that in the future studies they could investigate how imbalanced classes would affect classification performance when developing artificial intelligence methods such as Decision Trees, Multilayer Perceptron and Support Vector Machines. Moreover, the researchers have also suggested machine learning methods as well as data mining to be used in future studies. These methods may provide a more precise correlation between seminal data and the information collected by questionnaires (Gil, Girela, De Juan, Gomez-Torres & Johnsson, 2012).

## VII. Conclusion

Similar approaches to be adopted in identifying genetic factors