Predicting bovine sperm fertility: Quantity has a quality all its own

Harry Blakiston Houston, Dr Hermes Gadelha

University of Bristol, Bristol, United Kingdom hcbh96@gmail.com

Abstract. Global population and global wealth are set to continue to increase to the end of this century resulting in an increase in demand for livestock products. This is happening at a time where competition with other uses for land and water resources is intensifying and methane reduction targets are being adopted necessitating more efficient livestock production techniques. Improving our ability to predict bovine fertility and understanding the variables important for bovine fertility will play a role in meeting these challenges. Past literature shows that several sperm traits have a effect on in vitro production success rates, and some studies show a weak multi-linear correlation between sperm traits and bovine fertility. However, no study has applied non-linear machine learning techniques shown to be successful in human fertility predictions to bovine fertility studies. This forms a clear holes in current literature that this study attempts to fill. Using a database containing 316 samples of 22 features including sperm motility, morphology and concentration data, this study assesses the ability of a classification and regression tree (CART) tree and random forest (RF), to classify samples falling above or below a given blastocyst threshold. This study attempts to set an initial benchmark for future studies into bovine fertility by specifying and reporting on a metric toolset that makes it easy to select a suitable model for real world applications including area under the receiver operating curve (AUC), precision, accuracy and recall. This study then discusses the lowest number of variables required to achieve optimal classifier results. This study shows a RF (AUC 0.9 ± 0.04, precision: 0.84 ± 0.08, recall: 0.74 ± 0.21, accuracy: 0.84 ± 0.04) outperforms a CART (AUC 0.75 ± 0.06, precision: 0.75 ± 0.12, recall: 0.69 ± 0.17, accuracy: 0.78 ± 0.05). An AUC of 0.89 can be achieved with RF trained on as few as six variables including curvilinear velocity, beat cross frequency, high mitochondrial membrane potential, acrosome integrity, amplitude of lateral head, and progressive post-coital test. A RF trained on any subset of seven variables can acheive classification results similar of that of the optimal classifier (AUC = 0.87).

Keywords: Machine Learning, Fertility, Bovine, Prediction

1 Introduction

1.1 Context

The United Nations Food and Agriculture Organization (UNFAO) estimates that about 815 million people of the 7.6 billion people in the world, or 10.7%, were suffering from chronic undernourishment in 2016 [1]. Global population is expected to reach 8.5 billion in 2030, 9.7 billion in 2050 and 10.9 billion in 2100, according to the medium-variant projection [2]. On top of this long term economic growth projections released by PWC estimate the world economy to more than double in size by 2050, with growth to be driven largely by emerging markets and developing countries [3].In general the lower the income level of a country, the lower the meat consumption, the higher the income level, the higher the meat consumption [4], it follows that it is likely that the growth in global population coupled with increasing global affluence will beckon in an increase in demand for meat based protein products [4].

This is happening at a time where competition with other uses for land and water resources is intensifying and methane reduction targets are being adopted necessitating more efficient livestock production techniques [5] [6] [7]. Improving our ability to predict fertility and understanding the variables that play a crucial role in fertility will play an important role in meeting these challenges [4].

Knowledge gained from a model that successfully predicts bovine blastocyst rates from sperm traits will help in the design of In Vitro Fertilisation procedures aimed at selecting the correct sperm to increase said rates [8]. Practitioners looking to make evermore informed decisions can use machine learning techniques to aid outcome prediction and better understand the variables that have the greatest effect of bovine blastocyst rates [8].

Much research and work has already gone into applications of machine learning in fertility, specifically focusing on human fertility to answer questions such as *'What are my chances of conceiving?'*[9]. These same questions are of interest in bovine fertility which is the focus of this study. A variety of machine learning techniques have been used to answer similar questions in the areas of fertility and health prediction as discussed in section 2.

This paper asks whether it is possible to predict blastocyst rates from bovine sperm traits using machine learning techniques shown to be successful in human fertility studies. This question is important as various machine learning techniques have been effective in human fertility studies as shown in section 2 yet these techniques have never been applied to bovine fertility studies. Attempting to make make this transition is a logical step in the attempt to understand bovine fertility. This paper attempts to bridge this gap by applying two methods that have been used in human fertility studies notably a classification and regression tree (CART) and a random forest (RF) and applied them to bovine literature.

2 Literary review

As early as 1994 the race was on to identify sperm traits that would allow an accurate prediction of field fertility in cattle [10]. Several studies have shown that there exist correlations between various motility, morphology and concentration traits [10][11][12][13].

Whilst some of these tests have been shown to provide valuable information, there is yet to exist a diagnostic test can accurately predict variations in fertility among bulls that are producing apparently normal semen [12][14]. Simple correlations were discovered in 2015 between non-return rate (NRR) and mathematical models combining computer aided sperm analysis (CASA) and flow cytometry (FCM) which reported Pearson correlation coefficient (R) ranging between 0.24 and 0.40 ($p < 0.0001$) [15].

A review released in 2018 suggested that a portion of the embryo death before Day 8 is caused by the fertilising sperm, but the specific aspect of the sperm causing this effect is unclear [14], and a separate study in the same year isolated sperm traits including total motility before percoll gradient, acrosome integrity and mitochondrial membrane potential were among variables that strongly influenced embryo development ($p < 0.05$) after in vitro fertilization [16].

Several studies have noted an in vitro semen quality test, or combination of tests, which can accurately and consistently determine a bull's fertility and the optimum sperm number required for fertilisation represent the 'holy grail' in terms of semen assessment[14]. Though studies have speculated that there may exist such a test or combination of tests this has not been achieved to date [13][11][12][14].

This leaves a gap for a study to discover a sperm trait or combination thereof that is capable of predicting bovine blastocyst rates to a high degree of accuracy. This study focuses on predicting fertility rates from bovine sperm traits, and assessing which sperm traits play the most significant role in fertility estimation, this is a clear hole that can be filled in current literature.

It is notable in the above literature that only one study looked to use machine learning in the form of simple and multiple regression when assessing combinations of in vitro assessments to predict bovine fertility [15]. Furthermore the machine learning algorithm used in this analysis is unable to model non-linear relationships [15].

Several studies show there exist correlations between various sperm traits [10][11][12][13] due to the apparent interdependence between various sperm traits it is plausible that the links between these traits and the dependent variables are non-linear, and hence correlations using linear models are bound to produce weak results.

Machine learning has been an attractive methodology for human fertility studies, since 1997 it has been employed to make earlier and more accurate predictions of pregnancy in humans, at least 20 papers have reported on machine learning based prediction models in human in vitro fertilisation [17] (a break down of these paper can be seen here in figure 11). Further to this machine learning based prediction models provide a clinical decision support tool for

both clinicians and patients and lead to improvement in assisted reproductive technology success rates [17].

The studies shown in table 11 provide well balanced overviews of the different machine learning techniques that could be used in in vitro fertilisation prediction. Accuracy of above 0.90 was achieved using several measures methods including random forest (RF), support vector machine (SVM), maximum liklihood estimation (MLE), artificial neural network (ANN) [18][19] in this study for reasons outlined in section **??** a CART and RF were assessed in terms of their ability to predict in vitro fertilisation successfully as measured by accuracy, precision, recall, and area under the receiver operating curve (AUC) as defined in section 4.3.

These methods have never before been used to predict bovine fertility from bovine sperm traits. This study has taken the novel approach of applying some of the machine learning techniques shown to be successful in human fertility analysis, namely a classification and regression tree (CART) and RF, and used them to further the discussion on bovine fertility analysis. As this is the first study to applying non-linear machine learning techniques to bovine sperm analysis it would be fitting to specify a metric toolset from which to benchmark future predictive findings. This study attempts to set such a benchmark for future studies to compare against by explicitly specifying a variety of measurements commonly used to describe predictive capability of a classifier as described in section 4.3.

3 Exploratory Data Analysis

3.1 Understanding the data

The statistical research conducted in this project is based on a data set of bovine sperm traits collected over the course of 2 years [16]. The dataset contains over 30 variables covering sperm motility, morphology and concentration data, these traits were recorded during in vitro production as described in section 7.1.

Name	**Description**	**Type**	**Value**
Flow cytometry (FCM)			
AI	Acrosome integrity	Numeric	59-98
PI	Plasmatic membrane integrity	Numeric	17-100
ALTO	High mitochondrial membrane potential	Numeric	0-95
FRAG_CRO	Sperm DNA fragmentation	Categorical	0-4
In vitro production (IVP)			
MOT_PRE	Motility pre-percoll	Numeric	20-90
MOT_POS	Motility post-percoll	Numeric	20-90
CONC_CAMARA	Neubauer chamber sperm concentration	Numeric	10-119
VF	Final volume of sperm	Numeric	27-566
AD	Volume added for concentration adjustment	Numeric	0-476
Computer aided sperm analysis (CASA)			
VAP	Average path velocity	Numeric	56-150
VSL	Strait line velocity	Numeric	49-134
VCL	Curvilinear velocity	Numeric	85-240
ALH	Amplitude of lateral head	Numeric	3-59
BCF	Beat cross frequency	Numeric	20-47
STR	Path straightness	Numeric	65-95
LIN	Sperm linearity	Numeric	31-77
MOTILE_PCT	Post coital test	Numeric	13-88
PROGRESSIVE_PCT	Post coital test	Numeric	1-77
RAPID_PCT	Post coital test	Numeric	0-83
MEDIUM_PCT	Post coital test	Numeric	0-31
SLOW_PCT	Post coital test	Numeric	2-38
STATIC_PCT	Post coital test	Numeric	0-81
Dependent Variables			
CLIV	Day 3 cleavage rate	Numeric	49-90
BLAST_D8	Day 8 Blastocyst rate	Numeric	0-54
CELLS_COUNT	Day 8 cell count	Numeric	57-269

FCM contains four variables relating to the morphology of the sperm. IVP contains five variables, two of which MOT_PRE and MOT_POS describe the motility of the sperm, with the other three describing various concentration traits. CASA captures 13 traits, all of which assess various sperm motility measures. FCM, IVP and CASA include the independent variables.

Lower levels of motility, acrosome integrity and membrane potential have been linked to higher blastocyst rates [16]. VSL has been weakly correlated ($R = -0.12$) with a fertility measure based on non-return rate [15]. The same study suggested that using a combination of Oxidation, acrosome integrity, DNA compaction, mitochondrial activity, viability, velocity, sperm morphological abnormalities one could achieve an R of 0.40 [15]. Interestingly this optimal performance was achieved with seven variables, a point which is further analysed in figure 8. In this study acrosomal integrity relates to AI, mitochondrial activity relates to ALTO, viability relates to PI, velocity relates to VSL, the other measurements were not included in the dataset used for this study.

Finally the dependent variables can be see as potential fertility indicators. Both CLIV BLAST_D8 have been linked to non-return rate (NRR) [10] [20] [21] [22] [23].

3.2 Data Cleaning

Not all data is perfect, more frequently than not there are significant errors in the data collection process. It is often a good idea to start any machine learning project by attempting to uncover and study in more detail basic outliers and trends.

In order to decide upon a target metric for our predictions it was necessary to better understand the dependent variables, a pair plot was therefore used to uncover any abnormalities in the dependent variable space.

The first insight one can gather from figure 1 is that all variables seem to be roughly Gaussian distributed, this is to be expected and increases our confidence in the data collection process.

It should also be noted that CLIV has a moderate positive correlation with BLAST_D8 , this makes sense as CLIV can be seen as a preceding fertility measure to BLAST_D8 (both of these measurements have been correlated with NRR [10] [20] [21] [22] [23]).

While the correlation coefficient cannot capture non-linear relationships, it can capture simple relationships between variables. A heatmap of correlation coefficient data was therefore used to understand the interdependence between variables.

Figure 2 reveals some traits that can be reasonably inferred from the nature of the parameters, in that there exists strong correlations between some variables such as PROGRESSIVE_PCT with MOTILE_PCT , VAP with VSL or AD with VF, this makes it likely that one can remove some of the highly correlated "partner" variables as they provide little extra information in terms of the dependent variables. For example if one dropped MOTILE_PCT , little or no uncertainty would be lost as PROGRESSIVE_PCT captures close to all of the information provided by MOTILE_PCT . A RF's ability to generalise may increase by dropping some of the highly correlated vars [24].

The heatmap show that there exist stronger correlation within variable subgroups rather than between variable subgroups, for example any squares with deeper colour are more likely to be located within one of the coloured boxes

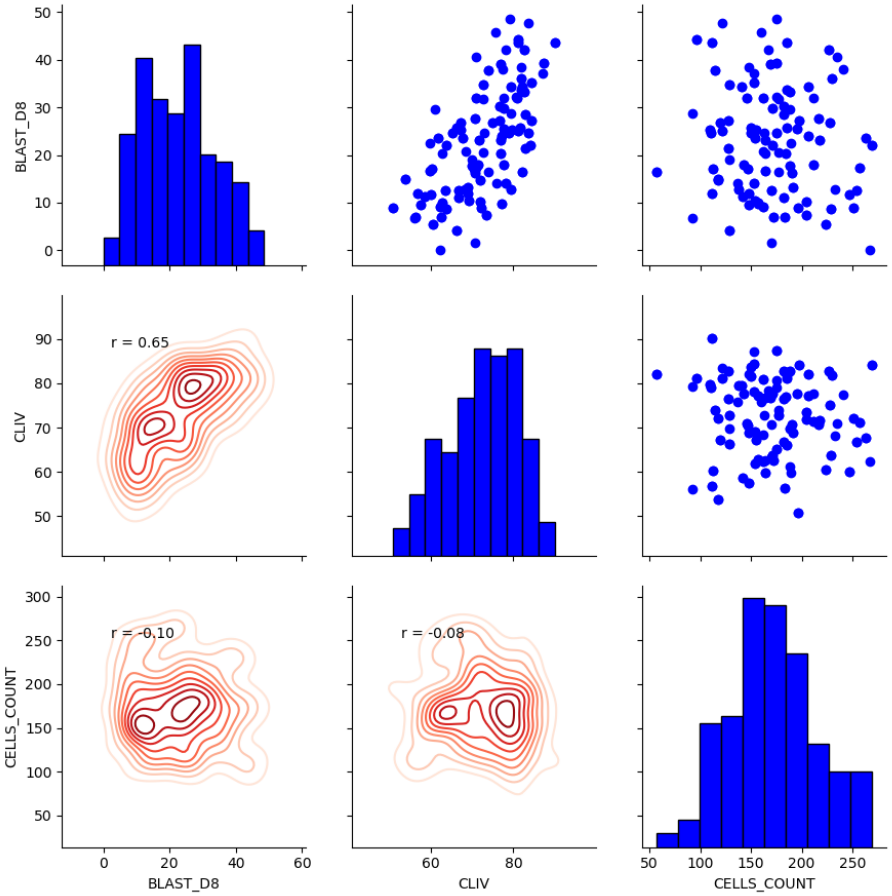

Fig. 1: Pair plot of output variables. The intersection of columns and rows describe an interaction between variables. With the upper half showing scatter plots, the diagonals portray single variable histograms and the lower half depict the density estimate alongside the Pearson correlation coefficient labeled r.

rather than outside the coloured boxes. This indicates that different subgroups may contain different information and so it may be of interest to see which subgroup can produce the best classification results. This question was analysed in figure 9.

However in general even within the individual subgroups correlations between variables are not particularly strong with most row-column intersections displaying an R value of between -0.4 and 0.4, this means that it is possible that an optimal predictor may contain one or more variables from a single subgroup. Another way of putting this is that the classification optima for the entire variable set could be the same the classification optima for a given variable subgroup.

Looking to the row labeled BLAST_D8 and CLIV, it is notable how little correlation there exists between these dependent variables and their respective

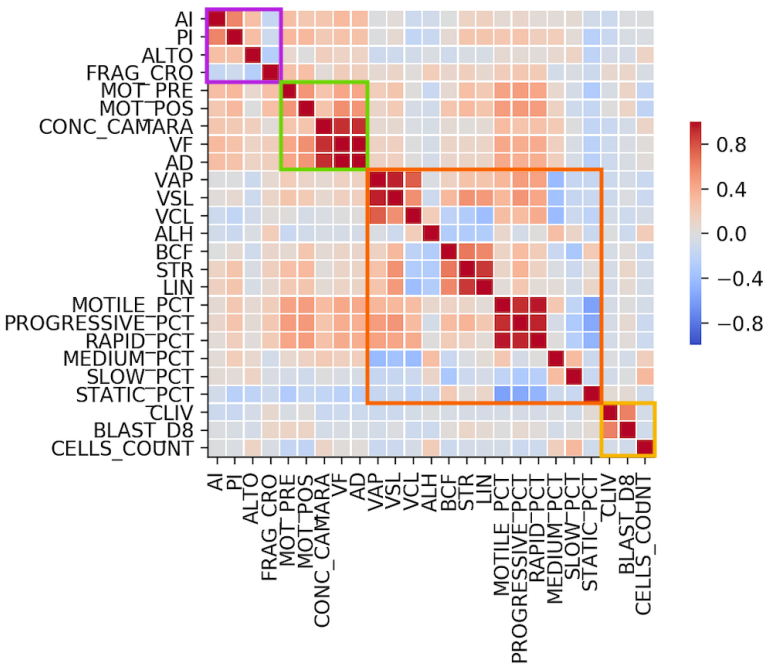

Fig. 2: Heatmap of pearson correlation coefficient data for continuous variables. The intersection of columns and rows describe the covariance between variables as described in section 7.4, with 1 representing a perfect positive correlation, 0 representing no correlation and -1 representing a perfect negative correlation. The four coloured boxes represent the 4 subgroups of variable as shown in table 3.1 with the subgroups highlighted in their respective colours: FCM, IVP, CASA, dependent variables.

independent variables. This lack of correlation across all variables shows why previous studies using linear regression, have at best shown moderate correlations R= 0.4 [15]. It is important to note that the findings in previous studies show that the R increased with more variables [15], this shows that though no single independent variable can be used to accurately predict fertility a combination thereof increases the ability of a classifier to predict a successful outcome [13] [11] [12] [14]. Interestingly optimal performance was achieved with seven variables a point which is further analysed in figure 8.

Whilst previous studies have shown some ability to predict fertility from sperm traits [15] none of the previous studies has accounted for the possibility of non-linear relationships effecting the dependent measure of fertility, this paper is the first study to applying non-linear machine learning techniques to bovine sperm analysis.

Prior research, as discussed in section 2, leads to the belief that the dataset, due to shortcomings in acrosome integrity techniques and the "bull effect" will contain variation [8]. It is therefore important that the prediction model selected

is applied in a way that looks to avoid overfitting. When predicting human in vitro fertilisation success rates MLE and SVM delivered results with the highest AUC [18]. Improperly optimised MLE, SVM and CART algorithms often suffer from overfitting [25] [**?**], whereas RF tends to surpass other methods in its ability to handle model overfitting [26], making it a particularly good candidate when setting an initial benchmark prediction for bovine fertility.

4 Models

4.1 classification and regression tree

The data structure for a classification tree is the same as that of a decision tree. The tree consists of two components: leaves and nodes, where each node represents an attribute that helps to decide the trees terminal class label, and a class label can be identified from the leaves [27]. Figure 3 is an example section of a classification tree that can be used to classify a plant as either Setosa, Versicolor or Virginica:

A more complex classification tree used in section **??**an be seen in section **??**.

There exist several variations of classification trees in literature [**?**], for the purposes of this study a CART was used. CART take the initial set and create increasingly small recursive binary partitions using the feature and threshold that yields the largest information gain at each node [27]. If no upper limit is set on the number of partitions, the tree will perfectly classify the training set but will likely under-perform on the test set, it follows that the tree can be optimised by limiting the depth of the tree allowing the classifier to generalise better [27].

4.2 random forest

RF are virtual forests containing multiple trees which have been fitted to a given dataset; the trees are then fed an input and the prediction of each tree is combined to predict the outcome [24]. rf randomly selects multiple variable subsets from the initial dataset, each subset contains approximately 63% of the variables contained in the complete set [**?**] [24]. A classification tree is then built from each subset, the classification trees are allowed to develop as defined by the hyper-parameters [24]. Each tree can then be used to give an out of bag prediction from the observation [24]. The predicted class of an observation is based on the majority vote of the out of bag predictions for that observation (with any ties split randomly) [24].

4.3 Evaluating Success

When assessing the ability of a prediction model to predict, it is paramount that one has a means of communicating the extent of that success in a relevant and comparable manor [18] [17]. The following metrics are relevant when evaluating

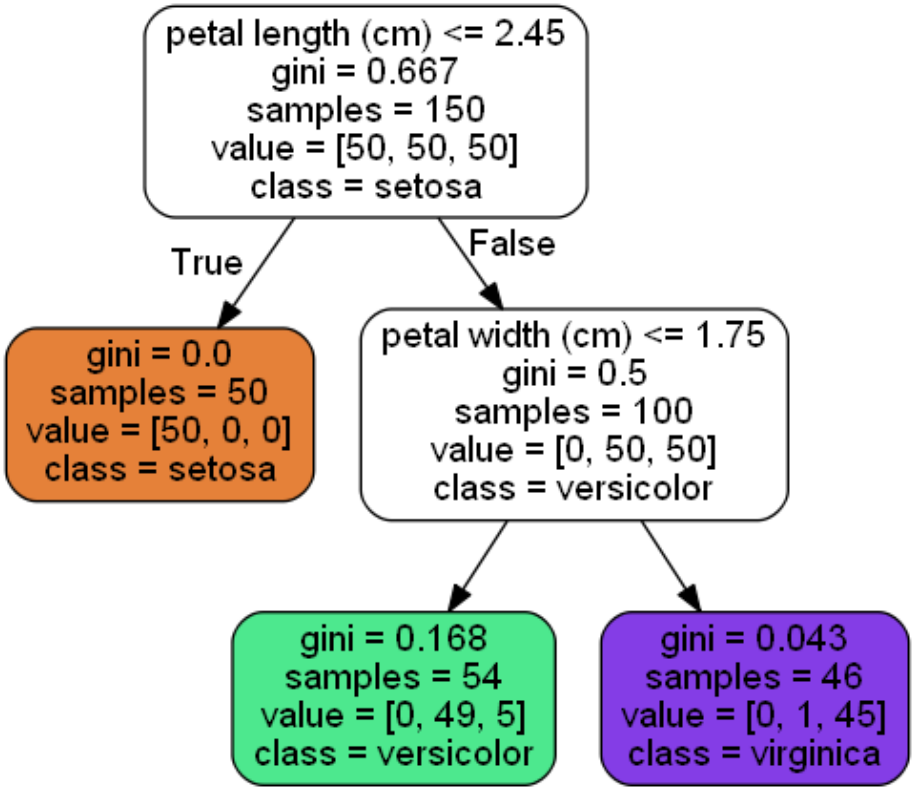

Fig. 3: Example classification tree: The tree first splits on a feature named petal length as this provides the greatest information gain i.e reduction in gini [27] [24], this completely separated of Setosa as only Setosa have a petal length under 2.45. At the second level Setosa is classified as such and so the petal leaf node is coloured orange and gini=0. The same process then occurs using petal width to separate Versicolor or Virginica.

the statistical relationship between symptoms and the presence of a disease[18], and can be applied when assessing the performance of the classifiers outlined in this paper.

- **TP**: the classifier produced a positive prediction when the actual outcome was also positive.
- **FP**: the classifier produced a negative prediction when the actual outcome was positive.
- **TN**: the classifier produced a negative prediction when the actual outcome was negative.
- **FN**: the classifier produced a positive prediction when the actual outcome was negative.

$$Total\ Samples = TP + TN + FP + FN$$

		Predicted class	
		Successful	Unsuccessful
Actual class	Successful	TP	FN
	Unsuccessful	FP	TN

Fig. 4: Example confusion matrix

A confusion matrix such as shown in figure 4 provides a visual summary of samples classified by a binary classifier. Here we define metrics which are commonly used in machine learning studies [28] [18] [17]:

- **Accuracy**: defined as the the ratio of correct classifications to total classifications.
$$Accuracy = \frac{TP + TN}{N}$$
- **Recall**: defined as the ability to correctly predict a successful outcome.
$$Recall = \frac{TP}{TP + FN}$$
- **Precision**: defined as the proportion successful cases that are predicted successful.
$$Precision = \frac{TP}{TP + FP}$$
- **Receiver operating curve (ROC)**: defined as a plot of true positive rates against false positive rates on the y and x axes respectively whilst the threshold is changed from 0 to 1.
- **Area under the receiver operating curve (AUC)**: defined as the aggregate area under the ROC with 1 representing a perfect classification and 0.5 representing a random binary classification.

It should be noted that the most common way to assess accuracy is by the area under the receiver operating curve [28], labelled AUC above. It is worth noting that no single metric conveys a complete picture of the performance of a model, AUC included. However, the AUC has an advantage of being a commonly reported metric in both clinical and recent machine learning papers [29] [17] [18].

5 Results and discussions

6 Comparing Models

As there exists little to no information from prior studies analysing the effectiveness of non-linear machine learning techniques on the prediction of bovine

blastocyst rates, the following section is aimed at providing a benchmark for future studies to compare their performance metrics with. The table is produced by running the CART and RF as classifiers described in section 4 over 30 evenly spaced thresholds between ±1 standard deviations from the mean.

Model Comparison Table		
Model	**CART**	**RF**
Mean accuracy (± STD)	0.78 (0.05)	0.84 (0.04)
Mean recall (± STD)	0.69 (0.17)	0.74 (0.21)
Mean precision (± STD)	0.75 (0.12)	0.84 (0.08)
Mean AUC (± STD)	0.75 (0.06)	0.90 (0.04)

The results in the model comparison table 6 show the ability of a relatively simple CART and RF to make accurate predictions when it comes to BLAST_D8 classification under and above a varying threshold. The model comparison table 6 shows that RF produces results which are clearly favorable in comparison to the CART across all metrics. RF standard deviation is lower in all measures bar recall, the most significant improvements are shown in Precision and AUC.

6.1 Model performance vs Number of Features

Tests were performed to assess how the models performance change as the number of features used to classify a sample was reduced. Previous literature shows optimal performance was recorded after seven variables were used in the prediction [15]. The aim was to see if there exist an optimal number of variables for predicting accuracy, recall, precision and AUC and if so what these variables were. In order to do this we ran the CART and RF models described in 4 inside a loop and after each iteration the variable that provided the lowest information gain (as specified by Breimann [24] [27]) was removed. After having run the loop and measured the average accuracy, recall, precision and AUC across 30 thresholds ±1 standard deviations from the mean, figures 5 and 6 were produced.

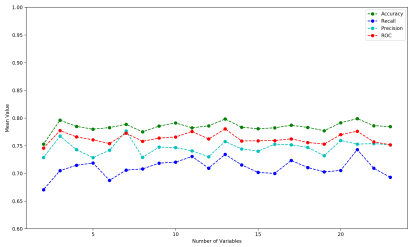

Fig. 5: CART number of features on the bottom vs mean measurement metrics for 30 uniform thresholds

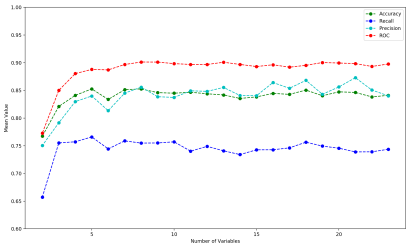

Fig. 6: RF number of features used on the bottom vs mean measurement metrics for 30 uniform thresholds

After further evaluation of the figures for the RF in figure 6 one can see that when focusing on AUC (the most commonly used classification measurement [28] [29]), the first point at which the decline in predictive capability from the full dataset is significant is when less than six variables are being used to make the prediction (p=0.05). In other words if a prediction is being made using the top six variables as defined in 6.2, then no significant predictive capacity will be gained (p=0.003) by increasing the size of the subset. The second thing that can be noted is that RF displays less variance and makes better predictions than CART across every measurement and variable set size as displayed in figures 5 and 6.

6.2 Feature Importance

Parameter Dropout Table 6.2 shows the order at which variables drop out of the dataset when running code 7.2 for both CART and RF. Given that that figure 6 displayed that most of the information is contained in the top six variables, the parameter dropout table displays this subset of six variables.

Parameter Dropout Table		
Importance	**CART**	**RF**
1.	VCL	VCL
2.	BCF	BCF
3.	ALH	ALTO
4.	ALTO	AI
5.	FRAG_CRO	ALH
6.	AI	PROGRESSIVE_PCT
No significant improvement from vars below this line		
7.	PROGRESSIVE_PCT	FRAG_CRO
8.	STATIC_PCT	VAP
9.	SLOW_PCT	VF
10.	VAP	SLOW_PCT
11.	RAPID_PCT	STR
12.	STR	PI
13.	VF	RAPID_PCT
14.	MEDIUM_PCT	CONC_CAMARA
15.	MOT_PRE	MOT_PRE
16.	CONC_CAMARA	STATIC_PCT
17.	VSL	MEDIUM_PCT
18.	LIN	VSL
19.	MOTILE_PCT	MOT_POS
20.	PI	MOTILE_PCT
21.	AD	LIN
22.	MOT_POS	AD

A question that follows from above is whether it is specifically these top six variables that provide the information required to make a successful prediction

or could any subset containing six or more variables be used as a comparable predictor. In order to do this we ran the CART and RF models described in 4 inside a loop and after each iteration the variable that provided the highest information gain (as specified by Breimann [24] [27]) was removed. In order to provide a comparison we plotted the AUC curve produced by the RF against the AUC curve shown in figure 6.

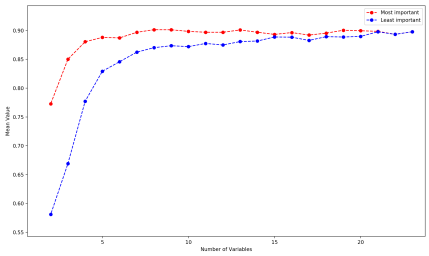

Fig. 7: Comparison of CART vs RF when reducing the number of features on the bottom vs mean AUC for 30 uniform thresholds and removing the variable that provided the highest information gain on each iteration.

After further evaluation of the figures for the RF in figure 7 one can see that when focusing on AUC (the most commonly used classification measurement [28] [29]), the first point at which the decline in mean predictive capability is significant in comparison to the initial parameter dropout is when two variables have been removed from the dataset (p=0.003). In other words the classifier only performs consistently optimally when 20 of the 22 least important variables are included in the dataset (p=0.003). The reduction in predictive capability then remains consistent and gradual between a subset of 20 variables and a subset of seven variables (p=0.003) after which the reduction in predictive capability accelerates significantly (p=0.003).

The predictive capability when using the worst seven variables is 0.87 in comparison to 0.89 for the seven best variables. To bring this number into perspective the classifier comparison table 6 shows that AUC has a mean of 0.9 across 30 thresholds with a standard deviation of 0.04 meaning the mean of 0.87 lies within one standard deviation of the optimal mean. In other words whilst on average the classifier performs worse over thresholds within ± 1 standard deviation of the mean, it can not be concluded that it perform worse over every threshold within ± 1 standard deviation from the mean in terms of AUC.

6.3 Subsets

Figure 8 was produced to evaluate how any subset of seven permutations fairs in its ability to classify blastocyst rate as above or below the mean threshold value. There exist more than 170,000 unique permutations containing seven variables from the initial dataset. Due to the computational overhead the optimisation search algorithm was removed and the number of threshold values over which the classifier was evaluated was reduced from 30 to 1 (specifically the mean

threashold), this means that this result is not directly comparable with those in the Model Comparison Table 6 as it has only be tested over the mean threshold.

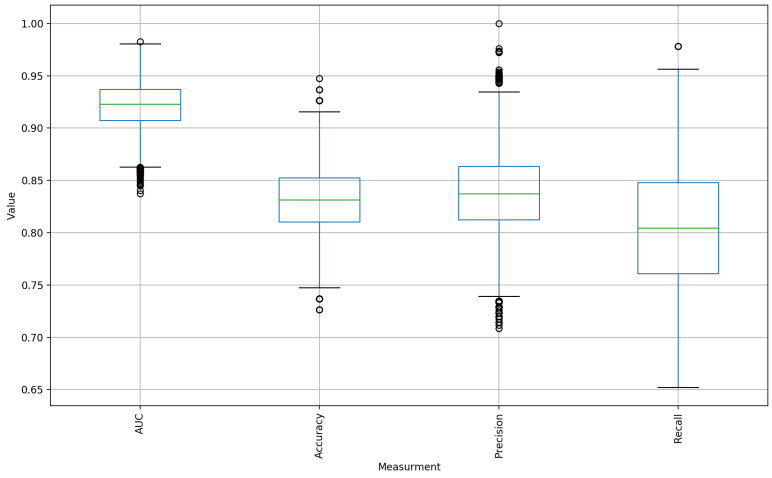

Fig. 8: Permutations box plots displaying the mean, variance, range and outliers for accuracy, precision, recall, AUC considering $> 170,000$ samples. The box shows the quartiles of the dataset while the whiskers extend to show the rest of the distribution, except for points that are determined to be "outliers" using a method that is a function of the interquartile range.

When considering AUC, analysis of figure 8 shows that almost any subset of seven variables can be used to build a model than can successfully classify a sample above or below the mean threshold with an AUC of 0.93 $\pm$0.02. This analysis corroborates that found in previous literature when seven variables including Oxidation, acrosome integrity, DNA compaction, mitochondrial activity, viability, velocity, sperm morphological abnormalities one could achieve an optimal of R of 0.40 [15].

In order to assess whether there existed any linear or non-linear relationships between the variables contained in the subset and the AUC measurement a CART was trained and tested over the space, this resulted in an AUC of 0.82 confirming that there exists some underlying predictable relationship, in otherwords there exists a relationship between the subset of seven features used to train the model and the performance of the model.

Given that CASA is the only subset of groups mentioned in section 3 to contain more than seven features, a subset containing only CASA features was assessed in terms of their AUC, accuracy, precision and recall and compared with that of non-CASA measurements, the results can be seen in figure 9.

As can be seen in figure 9 there is no significant difference between any group when considering accuracy, recall, precision or AUC. This makes sense as we displayed in figure 8 that optimal results could be obtained for all permutations

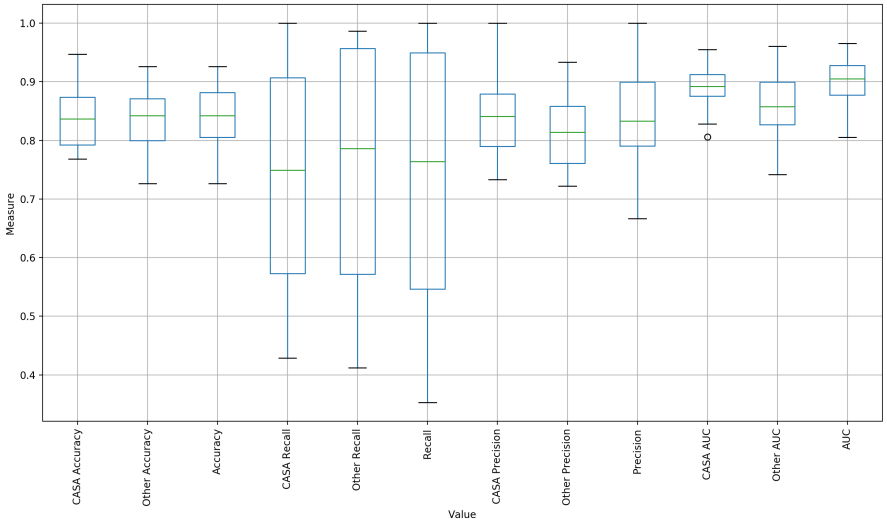

Fig. 9: Boxplot comparison of CASA, Other (non-CASA) and the initial dataset over metrics including accuracy, recall, precision and AUC. The box shows the quartiles of the dataset while the whiskers extend to show the rest of the distribution, except for points that are determined to be "outliers" using a method that is a function of the interquartile range.

that contain at least seven variables and since all groups tested contain at least seven variables none of the groups were significantly different from the others.

6.4 Conclusion

This study supports the existence of interdependence between various fertility measurements [8] as demonstrated by the variable heatmap, however results in section 3 display that there exists greater interdependence between measurements falling within the same measurement groups as displayed in the correlations heatmap 2 in section 3.

It has been shown that both RF and CART can be used to predict bovine blastocyst rate from bovine sperm traits. RF (AUC: 0.90 $\pm$0.04) performs significantly better than CART (AUC: 0.75$\pm$0.06) when used to predict bovine fertility from sperm traits over 30 thresholds spread evenly over $\pm$1 standard deviation of the mean as shown in table 6, this finding is consistent with findings in human fertility studies [18].

Although no single test can accurately predict variations in fertility recent studies suggested that a combination of laboratory tests were predictive of fertility [8], results in this paper displayed in figure 6, corroborate this finding and advanced the argument by specifying that the minimum number of variables capable of producing an optimal prediction is six and these variables as stated in the Feature Importance Table 6.2 are VCL, BCF, ALTO, AI, ALH and PROGRESSIVE_PCT. Interstingly two of these variables come from the FCM subset

and four from the CASA subset, but none from the IVP subset as displayed in table 3.1.

Previous literature found that optimal model performance was acheived when seven variables were used to train the data on a multi-linear model with a resulting R of 0.40 [15]. This paper has shown that results which are statistically indifferent to the optimal classifier can be obtained when using any subset of seven variables as shown in figure 8, this means that there are likely to be several tests that are capable of accurately predicting bovine blastocyst rates from sperm traits as displayed in figure 8. However the average classification accuracy across multiple thresholds is significantly higher if the 6 best variables are used as displayed in figure 7.

CASA as it has 13 features to train on can acheive optimal classification results, however both FCM and IVP when considered on their own will produce sub optimal results as both groups contain under seven features 9.

This study has taken note of measurment commonly used in human fertility literature [18] and specified a metric tool set and initial benchmarks for future studies into bovine fertility prediction including measurments such as accuracy, precision, recall and AUC as layed out in section 4.3.

This project has helped uncover fresh opportunities for future projects looking to predict bovine fertility. The most obvious is to continue trying and testing machine learning algorithms such as MLE or SVM, and compare their effectiveness using the benchmark values layed out in section 4.3, number of variables required to optimise the model and the variables of greatest importance.

A second line of study may be the applications of similar methods to other animal reproduction procedures such as in porcine, ovine, caprine or equine families or the application of similar methods using alternative datasets including other features normally assimilated with breeding soundness such as scrotal circumference.

A final route may be to run the same predictive analysis whilst focusing on an alternative outcome such as non-return rate, cleavage rate or cell count.

7 Appendix

7.1 Data collection

The ovaries used in this study came from slaughterhouses. Once in the lab, oocytes were selected for in vitro maturation over the course of one day after which they were fertilised. The following process was used to evaluate the sperm on day 0. Firstly the straws containing the sperm were thawed and assessed under a microscope (MOT_PRE). The semen was then exposed to a density gradient called Percoll to select viable sperm, assimilating selection that occurs in the female reproductive tract. The sperm was then diluted so that each drop contains 1×10^6 sperm/ml. The sperms then underwent a second motility assessment post Percoll gradient. A flow cytometer was used to analyse plasma membrane integrity, acrosome integrity, mitochondrial potential, and sperm chromatin susceptibility. Computer aided sperm analysis was used to measure a range of sperm traits many of which have been shown to play significant roles in fertility [8].

7.2 Number of variables loop

```
1. read in data
2. create array of n threshold values
3. while variables in set > 1
    a. loop through n threshold values:
        i. for each threshold set values above threshold as 1 and below
            as 0
        ii. split the data train 70%, test 30%
        iii. train classifier on training set
        iv. calculate test metrics on test set
    b. remove variable that provides lowest information gain
4. plot figures
```

7.3 Information Gain

The algorithm used to calculate information gain in this paper is that provided by scikit-learn, which bases feature importance on "gini importance" or "mean decrease impurity". It is defined as th total decrease in node impurity

7.4 Pearson Correlation coefficient

Pearson correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. Given a pair of random variables (X, Y) the formula for Pearson correlation coefficient ρ is:

$$\rho = \frac{cov(X, Y)}{\sigma_x \sigma_y} \tag{1}$$

– cov is the covariance
– σ_x is the standard deviation of x
– σ_y is the standard deviation of y

7.5 Classification Tree

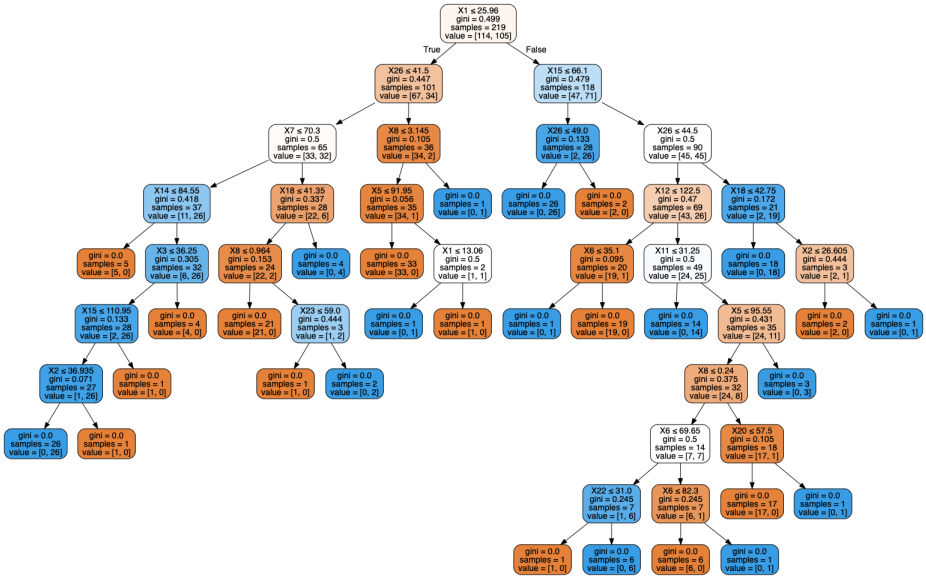

Fig. 10: Example classification tree: The tree first splits on a feature named petal length as this provides the greatest information gain i.e reduction in gini [27] [24], this completely separated of Setosa as only Setosa have a petal length under 2.45. At the second level Setosa is classified as such and so the petal leaf node is coloured orange and gini=0. The same process then occurs using petal width to separate Versicolor or Virginica.

7.6 Human fertility machine learning studies

Study	Technique(s)	ART method	Target (outcome)	External validation
Kaufmann et al. (1997)	Artificial Neural Networks (ANN)	IVF	Pregnancy	No
Jurisica et al. (1998)	Case-based reasoning (CBR)	IVF	Pregnancy	No
Kim and Jung (2003)	Bayesian network	IVF	Pregnancy	No
Passmore et al. (2003)	C5.0 Decision Tree	IVF	Pregnancy	No
Wald et al. (2005)	4-hidden node neural network	ICSI/IVF	intrauterine pregnancy	No
Morales et al. (2008)	Bayesian classification	IVF	Embryo implantation	No
Linda et al. (2009)	Bayesian network	IVF	ongoing pregnancy	No
Chen et al. (2009)	PSO, Decision Tree J48, Naïve Bayes, Bayes Net, MLP ANN	ICSI/IVF	Pregnancy	No
Nanni et al. (2010)	SVM, NN, DT	ICSI	Pregnancy	No
Guh et al. (2011)	genetic algorithm and decision tree	ICSI	Pregnancy	No
Corani et al. (2013)	Bayesian network	IVF	Pregnancy	No
Durairaj and Ramasamy (2013)	MLP ANN	IVF	pregnancy	No
Malinowski et al. (2013)	ANN	IVF/ICSI	Pregnancy	No
Uyar et al. (2014)	NB, KNN, SVM, DT, MLP, radial basis function network	IVF/ICSI	Implantation	No
Güvenir et al. (2015)	NB and RF	IVF	clinical pregnancy	No
Chen et al. (2016)	multivariable logistic regression (LR) and multivariate adaptive regression splines (MARS)	IVF/ICSI	clinical pregnancy	No
Mirroshandel et al. (2016)	NB, SVM, MLP, IBK, KStar, Bagging (KStar), RandomCommittee, J48, RF	ICSI	1) 2PN degree prediction 2) Embryo quality prediction 3) Clinical pregnancy (Beta test) prediction	No
Hafiz et al. (2017)	SVM, RPART, RF, Adaboost, 1NN	IVF/ICSI	Implantation	No
Blank et al. (2018)	RF	IVF/ ICSI	Ongoing pregnancy	No
Hassan et al. (2018)	MLP, SVM, C4.5, CART, RF	IVF	pregnancy	No

Fig. 11: The characteristics of machine learning–based prediction models on assisted reproductive technologies citations in order [30], [31], [32], [33], [34], [35], [36], [37], [38], [9], [39], [19], [40], [41], [42], [37], [43], [44], [45], [18]

Acronyms

AD volume added for concentration adjustment. 5, 6, 13
AI acrosome integrity. 1, 5, 6, 8, 13, 15, 16
ALH amplitude of lateral head. 1, 5, 13, 16
ALTO high mitochondrial membrane potential. 1, 5, 6, 13, 16
ANN artificial neural network. 4
AUC area under the receiver operating curve. 1, 4, 9, 11–17

BCF beat cross frequency. 1, 5, 13, 16
BLAST_D8 day 8 blastocyst rate. 5–7, 12

CART classification and regression tree. 1, 2, 4, 9, 12–16
CASA computer aided sperm analysis. 3, 5, 15–17
CELLS_COUNT day 8 cells count. 5
CLIV day 3 cleavage rate. 5–7
CONC_CAMARA Neubauer chamber sperm concentration. 5, 13

FAO Food and Agriculture Organization. 2
FCM flow cytometry. 3, 5, 16, 17
FN false negative. 10
FP false positive. 10, 11, 25
FRAG_CRO sperm DNA fragmentation. 5, 13

IVP in vitro production. 5, 17

LIN sperm path linearity. 5, 13

MEDIUM_PCT medium post-coital test. 5, 13
MLE maximum liklihood estimation. 4, 9, 17
MOT_POS motility post-percoll. 5, 13
MOT_PRE motility pre-percoll. 5, 13, 18
MOTILE_PCT motile post-coital test. 5, 6, 13

NRR non-return rate. 3, 6

PI plasmatic membrane integrity. 5, 6, 13
PROGRESSIVE_PCT progressive post-coital test. 1, 5, 6, 13, 16

R Pearson correlation coefficient. 3, 6–8, 15, 17
RAPID_PCT rapid post-coital test. 5, 13
RF random forest. 1, 2, 4, 6, 9, 12–14, 16
ROC receiver operating curve. 11, 23

SLOW_PCT slow post-coital test. 5, 13
STATIC_PCT static post-coital test. 5, 13

STR sperm path straitness. 5, 13
SVM support vector machine. 4, 9, 17

TN true negative. 10
TP true positive. 10, 11, 25

UN United Nations. 2

VAP average path velocity. 5, 6, 13
VCL curvilinear velocity. 1, 5, 13, 16
VF final volume of sperm. 5, 6, 13
VSL strait line velocity. 5, 6, 13

Glossary

accuracy the the ratio of correct classifications to total classifications.. 11, 12, 15

acrosome integrity Acrosomal Integrity and Function of Spermatozoa. The acrosome reaction allows the sperm by exocytotic release of acrosomal enzymes to penetrate the zona pellucida of the egg.. 5

amplitude of lateral head Amplitude of lateral head displacement (ALH) is defined as twice the maximum displacement of a sperm head from its fitted moving axis in a track segment (unit: m). It is directly related to the level of bending in the proximal region of the tail i.e., a larger ALH value corresponds to stronger bending.. 5

area under the receiver operating curve the aggregate area under the receiver operating curve with 1 representing a perfect classification and 0.5 representing a random binary classification.. 11

average path velocity Average path velocity (µm/s). Time-averaged velocity of a sperm head along its average path.. 5

beat cross frequency Beat-cross frequency is defined as the frequency that the sperm head moves across the middle plane of the "straightened" trajectory (unit: Hz). The value of BCF is in general sub-linearly proportional to the beating frequency of the sperm tail and is roughly double the frequency of head wobbling.. 5

blastocyst rate The day 8 blastocyst rate is the total number of blastocysts among the total oocytes evaluated on day 8.. 2, 3, 5, 6, 12, 14, 16, 17

computer aided sperm analysis Computer aided sperm analysis evaluates certain characteristics of a male's semen and the sperm contained therein.. 5

curvilinear velocity Curvilinear velocity is the average velocity measured over the actual point-to-point track followed by the cell.. 5

day 3 cleavage rate Day 3 cleavage rate is defined as total number of day-3 embryos by total number of fertilized oocytes.. 5

day 8 cell count A count of fertilised embryo cells on day 8 especially in a standard volume (as a cubic millimeter).. 5

final volume of sperm Final volume of sperm refers to the volume of semen in the sample before the concentration adjustment was made.. 5

flow cytometry Flow cytometry is a technique used to detect and measure physical and chemical characteristics of a population of cells or particles. In this process, a sample containing cells or particles is suspended in a fluid and injected into the flow cytometer instrument.. 5

generalise adapt to previously unseen data. 6, 9

high mitochondrial membrane potential The mitochondrial membrane potential generated by proton pumps (Complexes I, III and IV) is an essential component in the process of energy storage during oxidative phosphorylation. Together with the proton gradient, mitochondrial membrane potential forms the transmembrane potential of hydrogen ions which is harnessed to make ATP.. 5

in vitro production Bovine in vitro embryo production is an animal biotechnology important for both agriculture and biomedical research. IVP comprises ultrasound-guided oocyte recovery from donor by follicular aspiration, in vitro maturation of oocytes, in vitro fertilization and in vitro embryo culture.. 5

motility post-percoll Motility post-percoll refers to the movement and swimming of sperm after percoll gradient. Poor sperm motility means that the sperm do not swim properly, which can lead to male infertility.. 5

motility pre-percoll Motility pre-percoll refers to the movement and swimming of sperm before percoll gradient. Poor sperm motility means that the sperm do not swim properly, which can lead to male infertility.. 5

neubauer chamber sperm concentration Neubauer Chamber Sperm Concentration is the number of sperm/mL in a semen sample. The number of sperm/mL in a sample is in this case determined by counting sperm in a Neubauer chamber.. 5

non-return rate This is the percentage of inseminated cows which are not inseminated again for a set period - usually between 30 - 60 days, often 49 days as this covers two potential heats. It is an estimate of the proportion of cows that get pregnant to each insemination.. 6

out of bag refers to any variables that are in the complete set but not present in a given subset.. 9

path straightness Path straightness is defines as the straightness of the trajectory of the average sperm path in the sample. 5

pearson correlation coefficient Pearson's correlation coefficient also referred to as Pearson's r, the Pearson product-moment correlation coefficient or the bivariate correlation, is a measure of the linear correlation between two variables X and Y. According to the Cauchy–Schwarz inequality it has a value between +1 and 1, where 1 is total positive linear correlation, 0 is no linear correlation, and 1 is total negative linear correlation. It is widely used in the sciences.. 8, 18

percoll Percoll is a tool for more efficient density separation in biochemistry that was first formulated by Pertoft and colleagues. It is used for the isolation of cells, organelles, and/or viruses by density centrifugation.. 3

plasmatic membrane integrity Cell disruption using chemical additives or through physical methods relies upon disruption of the cell membrane integrity through either direct or indirect means. As the cell membrane is

disrupted, an increase in porosity allows the release of intracellular proteins and cellular components.. 5

post coital test The post coital test (Sims-Huhner Test) is an evaluation of the sperm's interaction within cervical mucus postcoitally.. 5

precision the proportion successful cases that are predicted successful.. 11, 12, 15

recall the ability to correctly predict a successful outcome.. 11, 12, 15

receiver operating curve a plot of true positive rates against false positive rates on the y and x axes respectively whilst the threshold is changed from 0 to 1. 11

sperm DNA fragmentation Sperm DNA fragmentation is a term used to denote abnormal genetic material within the sperm, which in turn may lead to male subfertility, IVF failure and miscarriage. Men with normal sperm parameters are also found to have high sperm DNA fragmentation.. 5

sperm linearity Sperm linearity refers to the percentage of sperm swimming with a fairly strait path. 5

strait line velocity Straight line velocity is the time-average velocity of a sperm head along the straight line between its first and last detected positions.. 5

volume added for concentration adjustment A concentration adjustment was performed so that is insemination happens with a consistent sperm concentration of $1X10^6$ sperm / ml medium.. 5

References

1. R. Pérez-Escamilla, "Food security and the 2015–2030 sustainable development goals: From human to planetary HealthPerspectives and opinions," vol. 1, no. 7. Publisher: Oxford Academic.
2. United Nations, Department of Economic and Social Affairs, and Population Division, *World population prospects Highlights, 2019 revision Highlights, 2019 revision*. OCLC: 1110010089.
3. p. pwc, "the world in 2050,"
4. N. Yukiko, "The future of global meat demand – implications for the grain market - mitsui & co. global strategic studies institute," Library Catalog: www.mitsui.com.
5. H. C. J. Godfray, J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin, "Food security: The challenge of feeding 9 billion people," vol. 327, no. 5967, pp. 812–818.
6. B. J. Hayes, H. A. Lewin, and M. E. Goddard, "The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation," vol. 29, no. 4, pp. 206–214.
7. M. Herrero, S. Wirsenius, B. Henderson, C. Rigolot, P. Thornton, P. Havlík, I. de Boer, and P. J. Gerber, "Livestock and the environment: What have we learned in the past decade?," vol. 40, no. 1, pp. 177–202.
8. M. D. Utt, "Prediction of bull fertility," vol. 169, pp. 37–44.

9. R.-S. Guh, T.-C. J. Wu, and S.-P. Weng, "Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes," vol. 38, no. 4, pp. 4437–4449.
10. P. Lonergan, "The application of in vitro fertilization techniques to the prediction of bull fertility," vol. 29, no. 1, pp. 12–21.
11. G. Giritharan, N. Ramakrishnappa, A. Balendran, K. M. Cheng, and R. Rajamahendran, "Development of in vitro tests to predict fertility of bulls," vol. 85, no. 1, pp. 47–52.
12. J. P. Kastelic and J. C. Thundathil, "Breeding soundness evaluation and semen analysis for predicting bull fertility," vol. 43, pp. 368–373.
13. B. Larsson and H. Rodrıguez-Martınez, "Can we use in vitro fertilization tests to predict semen fertility?," vol. 60-61, pp. 327–336.
14. S. Fair and P. Lonergan, "Review: Understanding the causes of variation in reproductive wastage among bulls," vol. 12, pp. s53–s62.
15. E. Sellem, M. L. W. J. Broekhuijse, L. Chevrier, S. Camugli, E. Schmitt, L. Schibler, and E. P. C. Koenen, "Use of combinations of in vitro quality assessments to predict fertility of bovine semen," vol. 84, no. 9, pp. 1447–1454.e5.
16. A. F. P. Siqueira, L. S. d. Castro, P. M. d. Assis, L. d. C. Bicudo, C. M. Mendes, M. Nichi, J. A. Visintin, and M. E. O. D. Assumpção, "Sperm traits on in vitro production (IVP) of bovine embryos: Too much of anything is good for nothing," vol. 13, no. 7, p. e0200273.
17. B. Raef and R. Ferdousi, "A review of machine learning approaches in assisted reproductive technologies," vol. 27, no. 3, pp. 205–211.
18. M. Hassan, S. Al-Insaif, M. Hossain, and J. Kamruzzaman, "A machine learning approach for prediction of pregnancy outcome following IVF treatment,"
19. M. Durairaj and R. N. Kumar, "Data mining application on IVF data for the selection of influential parameters on fertility," vol. 2, no. 6, p. 6.
20. C. Schneider, J. Ellington, and R. Wright, "Relationship between bull field fertility and in vitro embryo production using sperm preparation methods with and without somatic cell co-culture," vol. 51, no. 6, pp. 1085–1098.
21. M. Shamsuddin, B. Larsson, H. Gustafsson, and H. Rodriguez-Martinez, "In vitro development up to hatching of bovine in vitro-matured and fertilized oocytes with or without support from somatic cells," vol. 39, no. 5, pp. 1067–1079.
22. F. Ward, D. Rizos, D. Corridan, K. Quinn, M. Boland, and P. Lonergan, "Paternal influence on the time of first embryonic cleavage post insemination and the implications for subsequent bovine embryo development in vitro and fertility in vivo," vol. 60, no. 1, pp. 47–55. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrd.1060.
23. B. R. Zhang, B. Larsson, N. Lundeheim, and H. Rodriguez-Martinez, "Relationship between embryo development in vitro and 56-day nonreturn rates of cows inseminated with frozen-thawed semen from dairy bulls," vol. 48, no. 2, pp. 221–231.
24. L. Breiman, "Random forests," vol. 45, no. 1, pp. 5–32.
25. G. C. Cawley and N. L. C. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," vol. 11, p. 20792107.
26. K. Matsuki, V. Kuperman, and J. A. Van Dyke, "The random forests statistical technique: An examination of its value for the study of reading," vol. 20, no. 1, pp. 20–33.
27. L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees,"

28. A. Rajkomar, A. M. Dai, M. Sun, M. Hardt, K. Chen, K. Rough, and J. Dean, "Reply: metrics to assess machine learning models," vol. 1, no. 1, pp. 1–2. Number: 1 Publisher: Nature Publishing Group.
29. Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," vol. 8, no. 1, pp. 1–12. Number: 1 Publisher: Nature Publishing Group.
30. S. J. Kaufmann, J. L. Eastaugh, S. Snowden, S. W. Smye, and V. Sharma, "The application of neural networks in predicting the outcome of in-vitro fertilization.," vol. 12, no. 7, pp. 1454–1457. Publisher: Oxford Academic.
31. I. Jurisica, J. Mylopoulos, J. Glasgow, H. Shapiro, and R. F. Casper, "Case-based reasoning in IVF: prediction and knowledge mining," vol. 12, no. 1, pp. 1–24.
32. I.-C. Kim and Y.-G. Jung, "Using bayesian networks to analyze medical data," in *Machine Learning and Data Mining in Pattern Recognition* (P. Perner and A. Rosenfeld, eds.), Lecture Notes in Computer Science, pp. 317–327, Springer.
33. L. Passmore, J. Goodside, L. Hamel, L. Gonzalez, T. Silberstein, and J. Trimarchi, "ASSESSING DECISION TREE MODELS FOR CLINICAL INVITRO FERTILIZATION DATA," p. 16.
34. M. Wald, A. E. Sparks, J. Sandlow, B. Van-Voorhis, C. Syrop, and C. S. Niederberger, "Computational models for prediction of IVF/ICSI outcomes with surgically retrieved sperm," vol. 84, p. S12. Publisher: Elsevier.
35. D. A. Morales, E. Bengoetxea, P. Larrañaga, M. García, Y. Franco, M. Fresnada, and M. Merino, "Bayesian classification for the selection of in vitro human embryos using morphological and clinical data," vol. 90, no. 2, pp. 104–116.
36. L. C. van der Gaag, S. Renooij, A. Feelders, A. de Groote, M. J. C. Eijkemans, F. J. Broekmans, and B. C. J. M. Fauser, "Aligning bayesian network classifiers with medical contexts," in *Machine Learning and Data Mining in Pattern Recognition* (P. Perner, ed.), Lecture Notes in Computer Science, pp. 787–801, Springer.
37. F. Chen, D. De Neubourg, S. Debrock, K. Peeraer, T. D'Hooghe, and C. Spiessens, "Selecting the embryo with the highest implantation potential using a data mining based prediction model," vol. 14, no. 1, p. 10.
38. L. Nanni, A. Lumini, and C. Manna, "A data mining approach for predicting the pregnancy rate in human assisted reproduction," in *Advanced Computational Intelligence Paradigms in Healthcare 5: Intelligent Decision Support Systems* (S. Brahnam and L. C. Jain, eds.), Studies in Computational Intelligence, pp. 97–111, Springer.
39. G. Corani, C. Magli, A. Giusti, L. Gianaroli, and L. M. Gambardella, "A bayesian network model for predicting pregnancy after in vitro fertilization," vol. 43, no. 11, pp. 1783–1792.
40. P. Malinowski, R. Milewski, P. Ziniewicz, A. J. Milewska, J. Czerniecki, and S. Wołczyński, "The use of data mining methods to predict the result of infertility treatment using the IVF ET method," vol. 39, no. 1, pp. 67–74.
41. A. Uyar, A. Bener, and H. N. Ciray, "Predictive modeling of implantation outcome in an in vitro fertilization setting: An application of machine learning methods," vol. 35, no. 6, pp. 714–725. Publisher: SAGE Publications Inc STM.
42. H. A. Güvenir, G. Misirli, S. Dilbaz, O. Ozdegirmenci, B. Demir, and B. Dilbaz, "Estimating the chance of success in IVF treatment using a ranking algorithm," vol. 53, pp. 911–920.
43. S. A. Mirroshandel, F. Ghasemian, and S. Monji-Azad, "Applying data mining techniques for increasing implantation rate by selecting best sperms for intracytoplasmic sperm injection treatment," vol. 137, pp. 215–229.

44. P. Hafiz, M. Nematollahi, R. Boostani, and B. N. Jahromi, "Predicting implantation outcome of in vitro fertilization and intracytoplasmic sperm injection using data mining techniques," vol. 11, no. 3, pp. 184–190.
45. C. Blank, R. R. Wildeboer, I. DeCroo, K. Tilleman, B. Weyers, P. de Sutter, M. Mischi, and B. C. Schoot, "Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective," vol. 111, no. 2, pp. 318–326.