The Role of Machine Learning in Decoding the Complexity of Bovine Pregnancy: A Review

Marilijn van Rumpt[1], M. Belen Rabaglino[1*]

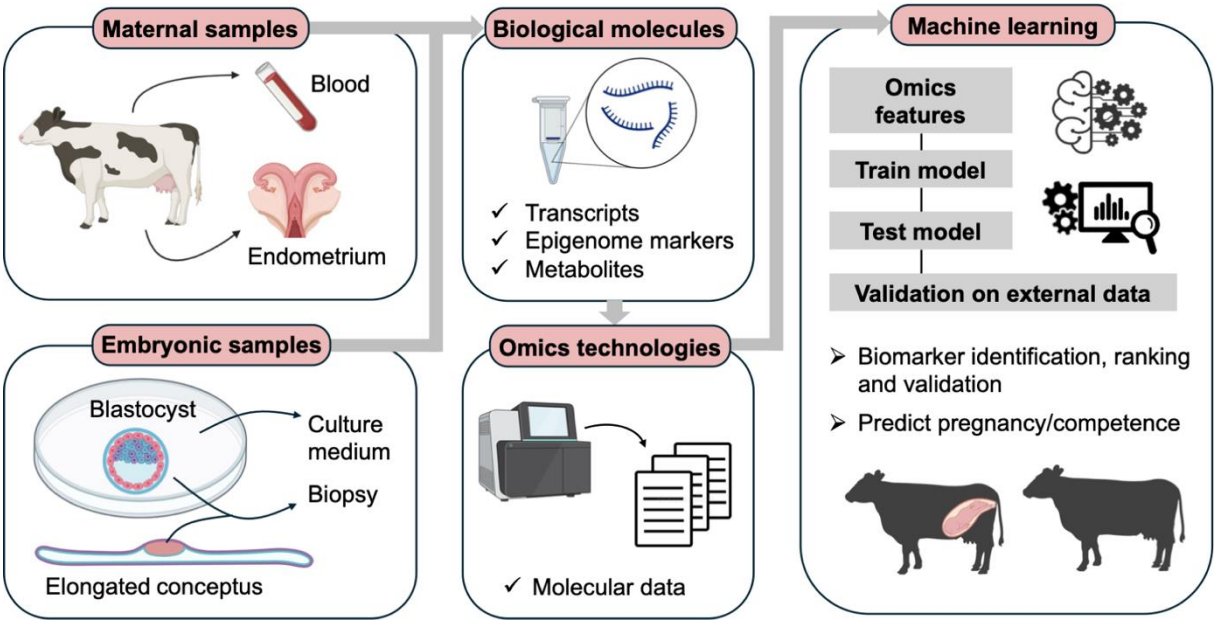
[1]Department Population Health Sciences, Faculty of Veterinary Medicine, Utrecht University

*Correspondence: m.b.rabaglino@uu.nl

Summary text for the Table of Contents

High embryo losses in cattle negatively impact the livestock industry. Modern techniques can measure the biological molecules involved in embryo-maternal communication at a large scale. These datasets are usually analysed through traditional pipelines. However, machine learning (ML) tools can effectively learn from data to make predictions, helping decipher the complexity of biological data. Here, we review the contribution of ML in understanding pregnancy establishment in cattle while also discussing the current challenges and future potential of ML in this field.

Image for the Table of Contents



The application of omics technologies allows for measuring biological molecules from samples obtained from bovine embryonic or maternal tissues at a large scale. Machine learning is a powerful approach to analyse these high-dimensional data, yielding outputs that can unravel the complexity of pregnancy establishment and progression in cattle.

Abstract

Pregnancy establishment and progression in cattle are pivotal research areas with significant implications for the industry. Despite high fertilisation rates, approximately 50% of bovine pregnancies are lost, pinpointing the need to keep studying the biological principles leading to a successful pregnancy. The increasing access and generation of omics data have aided in defining the molecular characteristics of pregnancy, i.e., embryo and foetal development and communication with the maternal environment. Large datasets generated through omics technologies are usually analysed through pipelines that could lack the power to deeply explore the complexity of biological data. Machine Learning (ML), on the other hand, a branch of artificial intelligence, offers a promising approach to address this challenge by effectively handling large-scale, heterogeneous and high-dimensional data. This review explores the role of ML in unravelling the intricacies of bovine embryo-maternal communication, including the identification of biomarkers associated with pregnancy outcome prediction and uncovering key genes and pathways involved in embryo development and survival. Through discussing recent studies, we define the contributions of ML towards advancing our understanding of bovine pregnancy, with the final goal of reducing pregnancy losses and enhancing reproductive efficiency while also addressing current limitations and future perspectives of ML in this field.

Keywords: machine learning, omics technologies, biological data, pregnancy outcome prediction, bovine pregnancy, embryo development.

1. Introduction

Understanding the complex mechanisms underlying successful embryogenesis and foetal growth is crucial for the dairy and beef industry, impacting economic efficiency, and genetic improvement programs. Embryo mortality is a significant factor affecting reproductive success. Despite achieving high fertilisation rates upwards of ~ 80%, birth rates significantly lag behind, with most embryos lost within the first month of pregnancy (Reese et al., 2020, Wiltbank et al., 2016, Berg et al., 2010). The primary reasons for this discrepancy are diverse, such as genetic factors, poor oocyte or embryo quality, suboptimal uterine receptivity, inadequate embryo-uterus interaction, failure of conceptus elongation, and abnormal foetal growth (Lonergan et al., 2016, Wiltbank et al., 2016).

Pregnancy is a complex process following tightly regulated steps. After mating or artificial insemination (AI) at oestrus, fertilisation of the oocyte occurs in the oviduct (Day 0), followed by mitotic cleavages of the embryo, resulting in a 16-cell stage developing into a morula, entering the uterus by day 4. The embryo remains transcriptionally silent during these first cleavages and relies on maternal mRNA (Schulz and Harrison, 2019). The

63 ‘maternal-to-embryonic transition’ occurs around the 8- to 16-cell stage (Kopecny et al.,
64 1989, Frei et al., 1989), involving degradation of maternal transcripts and activation of the
65 embryonic genome (Memili and First, 2000). By day 6-7, the morula develops into a
66 blastocyst, containing an inner cell mass, which will become the foetus, and outer
67 trophectoderm cells which will give rise to the placenta. At this point, the blastocyst interacts
68 with the endometrium by activating local genes in the uterus favouring an optimal intrauterine
69 environment (Passaro et al., 2018, Sponchiado et al., 2017). At day 9 to 10, the blastocyst
70 hatches from the zona pellucida. The embryo starts changing its morphology on day 12 to 14
71 from an ovoid to a tubular and finally a filamentous conceptus by day 16 to 17 (Degrelle et
72 al., 2005). The elongation and surface increase are required for elevating interferon tau
73 (IFNT) secretion by the embryo to trigger the signal for maternal pregnancy recognition and
74 thus prevention of luteolysis to maintain progesterone (P4) production by the corpus luteum.
75 Trophectoderm cells start attachment to the uterine epithelium by day 16, initiating
76 placentation around day 20 (Assis Neto et al., 2010).
77 The first seven days of gestation represent the period of most embryo loss (Reese et
78 al., 2020, Sartori et al., 2010), which are bypassed in case of embryo transfer (ET), for which
79 a 7-day old fertilised embryo is transferred to the uterus of a recipient cow. Contradictory,
80 pregnancy success is generally not higher when ET is used, compared to AI (Hansen, 2020).
81 Furthermore, there is a difference in pregnancy rates between embryos generated *in vivo* or
82 *in vitro*, with the latter having an average pregnancy rate of around 25% lower, according to
83 the outcomes of studies done between 1992 and 2014 (Ealy et al., 2019). *In vivo* produced
84 embryos are obtained by ovarian stimulation resulting in multiple ovulations, after which
85 embryos are collected and transferred to the recipient (MOET). For *in vitro* produced (IVP)
86 embryos, oocytes are either retrieved from ovaries after slaughter, or collected by
87 transvaginal aspiration of follicles in a living cow, followed by maturation and fertilisation of
88 the cultured oocyte. Another reason likely to be responsible for a significant amount of
89 embryo loss is the failure of the conceptus elongation process (Sánchez et al., 2019, Moraes
90 et al., 2018), which is not avoided by ET, as it is completely maternally driven and has not
91 been reproduced *in vitro*.
92 As both AI and ET pregnancy successes are suboptimal, research has focused on
93 uncovering mechanisms involved in the role of embryo quality and production (*in vivo* vs IVP)
94 and endometrial receptivity in pregnancy losses. With the rising availability and collection of
95 omics data, progress has been made in defining the molecular characteristics of pregnancy,
96 i.e., embryo and foetal development and communication with the maternal environment.
97 Transcriptomic analyses of the uterus revealed a different response to short versus long
98 conceptuses (Sánchez et al., 2019), and to *in vivo* versus IVP embryos (Mathew et al.,
99 2019), suggesting a sensitive uterine response to embryos with varying developmental

competence (Bauersachs et al., 2009, Mansouri-Attia et al., 2009). Similarly, the embryonic transcriptome differs between embryos able to sustain a pregnancy or not, and between MOET versus IVP embryos (Rabaglino, 2023a). Proteomic studies on early embryos have uncovered differently expressed proteins in each developmental stage (Deutsch et al., 2014, Banliat et al., 2022), and metabolomics research has shown different metabolic profiles before ET between pregnant and open cows (Gimeno et al., 2023, Gómez et al., 2020b). Even though these high throughput data have advanced reproductive research, it is hard to explore relationships hidden in these noisy, complex and high-dimensional datasets using traditional methods, such as differential analysis. In this context, the emergence of machine learning (ML) offers a promising approach to further explore the biological molecules involved in a successful pregnancy in the cow, and to complement traditional methods, considering its ability to handle large-scale, heterogenous and high-dimensional data.

Machine learning, a branch of artificial intelligence, encompasses algorithms that enable computers to learn from data patterns and make predictions without explicit programming. ML can be divided into two broad classes: supervised learning and unsupervised learning. In supervised learning, the dataset is randomly divided into training and validation datasets with labelled data such as pregnancy outcomes. The supervised model trains on the dataset and learns to recognize patterns associated with the label and test its predictive ability to assign correct labels on new data in the validation dataset. This is mainly used for problems in classification (in case of discrete labels) or regression (in case of continues label values), on which the model ranks the sample features according to its importance in predicting the label. One practical application of supervised learning is in biomarker discovery, where e.g., models are trained to predict pregnancy outcome in cattle based on metabolites or transcripts, and select the top predictive markers (Gimeno et al., 2023, Gómez et al., 2021, Hoorn et al., 2024, Rabaglino and Kadarmideen, 2020, Rabaglino et al., 2023b). In contrast, unsupervised models operate without labelled data, identifying data patterns and searching for similarities between data samples, with the main goal to perform clustering or dimensionality reduction. Unsupervised methods such as principal component analysis (PCA) and hierarchical clustering have been broadly applied in the past decades in reproductive science. The use of supervised ML, however, has surged in recent years, with a growing appliance in evaluating and predicting embryonic viability and potential interaction with the maternal environment, with the ultimate goal of reducing pregnancy losses.

This review aims to explore the role of supervised ML applied to omics data in understanding the intricacies of bovine embryonic and foetal development, and how ML may be implemented to predict the risk of pregnancy losses. We first briefly introduce commonly used supervised ML models, followed by discussing their applications in recent studies in the

field of bovine reproductive biology. The studies are organised in those conducted before day 7 (before the first interactions between embryo and uterus) and after day 7 of gestation. Finally, we discuss the limitations, current challenges, and perspectives of ML in studying pregnancy in cattle.

2. Supervised machine learning models

This section introduces the supervised ML models applied in the articles discussed in the next section.

2.1. Decision tree (DT), Random Forest (RF) and extreme gradient boosting

A DT is a flowchart-like structure of an inverted tree, existing of a root, nodes, branches and leaves. The hierarchical nodes represent a series of feature tests connected by branches, with the root being the first test and the leaves being the final class labels (Kingsford and Salzberg, 2008). The aim of using DT analyses is creating the best model for allocating all samples into the right segment. After learning the best DT from the training dataset, new instances are passed through the tree to predict their class. For RF (Breiman, 2001), multiple hundreds to thousands DTs are ensembled, each grown on a smaller random subset of samples and features of the original dataset. Hence, each DT will differ slightly. A new sample is passed down all the trees and its class is predicted by the majority vote. RF methods work well with high dimensional data, having many predictor variables, even when the number of samples is small ('large p, small n') (Chen and Ishwaran, 2012). RF often performs with higher accuracy compared to DT, lowering the prediction variance and bias associated with a single DT. Furthermore, RF is less prone to overfitting and is efficient in estimating missing values. Among other models used for classification tasks, RF takes a top position in predictive accuracy (Fernández-Delgado et al., 2014). However, interpreting a large RF may be difficult, and they can be computationally expensive when working with large datasets. Another method based on DTs is extreme gradient boosting (XGboost), which builds multiple decision trees sequentially, with each new tree correcting the errors made by the previous ones (Chen and Guestrin, 2016). This boosting technique focuses on optimising the gradient of the loss function, leading to highly accurate and efficient models, while regularisation is also incorporated to avoid overfitting.

2.2. K-nearest neighbour (KNN)

For KNN models, the distances between samples in the (high dimensional) feature space are calculated, after which the class of a new sample is predicted by considering the class of the majority of its closest neighbour samples (Mucherino et al., 2009). KNN may be computationally expensive, mostly because of the distance calculations. The value of k

(number of evaluated neighbours) and the distance method used are important for the model's performance. If k is too small, noise and outliers can have negative effects, resulting in overfitting and high variance. However, if k is too large, overrepresented classes can overwhelm smaller classes, resulting in bias and underfitting. KNN is severely affected by 'the curse of dimensionality' (phenomena that arise when analysing data in high-dimensional spaces) and its accuracy decreases if more sample features (or dimensions) are used, as all datapoints tend to be far away, without meaningful neighbours (Elkan, 2011).

2.3. Bayesian Network (BN)

A BN is a probabilistic graphical model, specifically a 'directed acyclic graph', consisting of nodes depending on each other, connected with edges that have direction, but do not form a loop (Stephenson, 2000). Each node represents a variable and has a conditional probability attached, representing the chance of the node being in a given state, influenced by the given state of its parents. The node dependencies propagate through the network, influencing the probabilities of other nodes. However, BN assumes that, given its parents, a node is conditionally independent of its non-descendants. BNs can work well with missing data and small datasets, preferring small sets of parent variables. It is less suited for large datasets and data with feedback loop relationships affecting the outcome.

2.4. Partial least squares discriminant analysis (PLS-DA)

PLS-DA is a tool for dimensionality reduction but is also widely applied for classification. High dimensional data is projected into a lower dimensional space, preserving distances between samples. Unlike PCA, which identifies principal components that explain the highest variance in the data, PLS-DA focuses on maximising the covariance between the data and class labels, making it a 'supervised' version of PCA (Ruiz-Perez et al., 2020). This approach allows PLS-DA to effectively handle high-dimensional data, but it can be susceptible to overfitting, necessitating the use of cross-validation. In sparse PLS-DA (sPLS-DA), a variant of PLS-DA, a sparsity constraint is incorporated, selecting only the most informative variables for class prediction through using penalties (Lê Cao et al., 2011). This makes sPLS-DA particularly useful for identifying key features in complex datasets while maintaining model simplicity and interpretability

2.5. Support vector machine (SVM)

The goal of SVM is to find the optimal hyperplane that separates the datapoints into different classes while maximising the margin between classes (Pisner and Schnyer, 2020). SVM uses the training data to define this optimal hyperplane, and when separation cannot be done linearly, a kernel method is used to transform and map the data into a high-dimensional

space where a simple hyperplane separates the data. Finding the optimal hyperplane can be computationally costly. A soft margin boundary is used that allows misclassification and lowers the effect of outliers on the boundary. SVM works effectively with high-dimensional data and in cases where the number of dimensions exceeds the number of samples but is less suited for noisy data and when datapoints with different classes are overlapping.

2.6. Artificial neural networks (ANN)

ANNs are inspired by the structure and function of the neural networks in the human brain and consist of connected hierarchical layers of neurons. Each neuron receives connection weights of other neurons and a bias, which are summed up, after which an activation function (often sigmoid) is applied to determine if the neuron gets activated (Priddy and Keller, 2005). ANN consists of an input layer where the raw data enters, one or more hidden layers, and a final output layer performing the last prediction. The output of a neuron of one layer is the input to neurons in the next layer. During training of the ANN, the connection weights are updated, with the goal to make the output of the last layer as close as possible to the desired output. The updating of biases and weights is often done by backpropagation: propagating the error back through the neurons. ANNs are difficult to interpret, can be prone to overfitting, are computationally costly and often require a lot of training data. They perform better when the dataset is larger and are good at modelling complex and non-linear relationships.

3. Machine learning to study embryonic or maternal molecular features before day 7 of gestation.

As most embryonic losses occur in the first week of pregnancy (Reese et al., 2020), even before uterus-embryo interaction, much research applying the predictive abilities of ML has focused on this critical period (see Table 1). Successful pregnancy requires both a competent embryo and a receptive endometrium. Hence, both embryonic and maternal omics are deployed for understanding early pregnancy.

3.1. Machine learning in embryo transcriptomics

Embryo quality evaluation and selection is conventionally done by visual inspection of morphological features by embryologists, grading the embryos generally following the guidelines from the International Embryo Technology Society (IETS) (Bó and Mapletoft, 2018). This method is however subjective and lacks accuracy and repeatability (Hansen, 2020). Instead, although scarcely applied, combining embryo transcriptomics with ML offers a more precise approach to predicting embryonic competence and identifying molecular differences between viable and non-viable embryos.

Based on the knowledge that an inherent molecular signature defines the bovine embryo ability for pregnancy establishment, we integrated seven transcriptomic datasets of both pre-ET blastocyst biopsies and elongated conceptuses of varying competence to identify biomarkers predictive of pregnancy outcome (Rabaglino et al., 2023b). Differently expressed genes (DEG) were identified between competent blastocysts (blastocysts resulting in pregnancy and long conceptuses) and incompetent blastocysts (blastocysts not resulting in pregnancy and short conceptuses), after which eight biomarker genes most discriminative between pregnant and non-pregnant samples were selected by sPLS-DA and linear discriminant analysis. The predictive ability of these biomarkers was tested on an independent dataset, consisting of competent embryos (resulting in pregnancy, cultured in normal conditions, or long conceptuses), and incompetent embryos (not resulting in pregnancy, cultured in suboptimal conditions, or short conceptuses). Using Bayesian logistic regression (BLR) and ANN, prediction accuracies from 85 to 100% were achieved, depending on the validation dataset, with ANN having the same or higher accuracies than BLR. Upregulated biomarkers in competent embryos were involved in cellular metabolism, including glycolysis/gluconeogenesis, whereas downregulated biomarkers were related to cell cycle processes (Rabaglino et al., 2023b). The glucose metabolism is known to be a critical process in embryo survival, with glucose deprivation leading to apoptosis (Riley and Moley, 2006). Hence, activation of glycolysis/gluconeogenesis pathways could have a positive effect on embryo development. The downregulation of cell cycle processes in competent embryos is in agreement with the “quiet embryo hypothesis”, which states that viable embryos have a less active metabolism (Leese, 2002). Furthermore, we have recently developed a formula to estimate an embryonic competence index based on the expression of the identified eight biomarker genes, which is available as function for the R software (Rabaglino and Hansen, 2024). Estimation of a quantitative index value can be employed in experiments to objectively identify interventions in embryo production that could increase embryo survival after transfer.

3.2. Machine learning in maternal transcriptomics

Transcriptomic data of the endometrium before embryo interaction have been used to uncover different gene expression profiles between cows that will become pregnant or not. Hoorn et al. (2024) identified biomarkers capable of predicting pregnancy status at day 30 through analysis of the transcriptome of endometrial cells of Holstein cows at day 0 before AI. For this, we applied the BORUTA algorithm (a wrapper built around the RF algorithm (Kursa and Rudnicki, 2010)) to identify biomarkers among transcripts differently expressed between cows that became pregnant or not, after which sPLS-DA was used to determine the combination of transcripts most discriminative for pregnancy status at day 30. Transcript

combinations were evaluated by applying SVM with linear kernels, which resulted in a set of 57 transcripts with an average prediction accuracy of 77%. The functional analysis of these biomarkers indicated that uterine immunological condition may be important for maternal fertility, with cows experiencing less immune activation being more likely to become pregnant (Hoorn et al., 2024).

A similar research was performed by Diniz et al. (2022), using endometrial transcriptomics data collected three days before ET from Angus-Brahman crossbred cows. Using BioDiscML (software that automises ML steps in feature and model selection in omics data (Leclercq et al., 2019)), 225 genes and five ML models based on SVM, KNN and an ordinal learning method were selected for further biomarker identification and day 30 pregnancy status prediction. Prediction accuracies on all 225 genes ranged from 80% to higher than 90%, while prediction accuracies using nine genes selected as potential biomarkers ranged from 53.85% to 61.54%, with the models based on SVM and KNN having the greatest accuracies. The lower prediction ability of the nine genes is likely due to the smaller dataset size and only nine genes not being able to reflect the complex process involved in pregnancy loss or success. Contrary to the study of Hoorn et al. (2024), immune pathways were not found to be significantly associated to analysed genes, however one of the nine selected biomarker genes, *PDCD1*, is likely to be involved in immune inhibition during pregnancy (Taglauer et al., 2008, Diniz et al., 2022). Other pathways related to the nine biomarker genes were focal adhesion, remodelling of endometrial tissue and embryonic development.

We conducted another study focused on predicting pregnancy outcome using day 6-7 endometrial transcripts from cows of four different European cattle breeds (Rabaglino and Kadarmideen, 2020). Using five ML models, of which three based on BN and two on logistic regression, 50 genes overlapping between methods were selected as potential biomarkers. An average accuracy of 96.1% was achieved, by applying SVM to predict pregnancy outcome with these biomarkers, training on all samples from all but one breed, and using the left-out breed for validation. Among pathways in which up-regulated genes related to the biomarkers were involved are embryonic development, circadian rhythm and Wnt pathways.

The three studies discussed above all identified different endometrial biomarkers for pregnancy outcome prediction (except for one biomarker in common between two studies). A potential factor influencing this lack of overlap among the biomarkers is the timing of sampling, since samples were collected on day 0, 4 and 6-7, according to the study. Transcriptomic research has shown differences in endometrial gene expression at different timepoints in the perioestrus period, likely through fluctuations in oestradiol and progesterone (Alfattah et al., 2024), which may lead to different biomarkers associated at each timepoint. Furthermore, cow breeds differed between the studies, which may affect biomarker selection.

Only the study of Rabaglino & Kadarmideen (2020) used multiple breeds for training and validation of the ML model. Thus, the identified biomarkers can be used for pregnancy outcome prediction across different breeds, avoiding the selection of breed specific biomarkers.

A different approach for pregnancy outcome prediction applied by multiple studies is the use of blood omics data, having the benefit of blood being easily available and collection is minimally invasive. Moorey et al. (2020) used peripheral white blood cell RNA at day 0 to predict pregnancy outcome. Based on 198 DEG between pregnant and non-pregnant cows, a prediction accuracy of >90% was achieved using a parallelised RF. More blood-based research has however focused on metabolomics, as discussed in the next section.

3.3. Machine learning used on maternal and embryonic metabolomics

Several studies have analysed metabolites from recipient blood and/or culture medium CM of embryos used for ET, providing insights into the metabolic status during early pregnancy and embryonic development. Blood and CM have the advantage of being easily accessible and minimally invasive, compared to endometrial samples. Furthermore, using CM for embryo assessment does not interfere with its developmental competence, unlike embryonic biopsy. However, CM composition might be impacted by external factors such as contact with plastic dishes or variations in temperature, pH and osmolarity (Sciorio and Rinaudo, 2023). Therefore, ML models based on CM metabolites might not perform similarly among different laboratories, as mentioned below.

Muñoz et al. (2014b) found recipient blood metabolites having a higher predictive accuracy for day 60 pregnancy compared to CM metabolites of MOET embryos when employing KNN for prediction. The study also revealed variability in pregnancy status prediction when CM samples were processed in two separate laboratories. Using the cumulative CM data from both laboratories, a predictive accuracy of 64.4% was achieved, while the separate accuracies were 74.6% and 74.8%. This difference might be due to divergent laboratorial procedures altering embryonic produced metabolites or their measurement, indicating that the used training dataset may alter predictions and must be selected carefully, considering the origin of the data. In contrast, predictive accuracies based on blood were highest when using cumulative data compared to separate data (74.2% versus 59.6% and 69.1%). A second study performed by Muñoz et al. (2014a) also showed higher prediction accuracies for blood metabolites compared to CM metabolites. Metabolites in CM of fresh and vitrified/warmed (VW) embryos, and recipient blood, were analysed by employing a similar approach using KNN for pregnancy outcome prediction. Birth prediction accuracies based on fresh embryo CM and blood of fresh embryo recipients reached 71.9% and 74.9% respectively. Interestingly, increased birth accuracies were obtained when only

expanded blastocyst were considered (CM: 82.8%, blood: 85.0%), so the CM of more developed embryo stages might contain a metabolite profile with higher predictive ability (Muñoz et al., 2014a). For VW embryos, only metabolites in blood plasma and not in CM were able to give a relevant prediction with 69.3% accuracy for birth. Considering the higher achieved prediction accuracies for fresh embryos compared to VW embryos, the embryo treatment likely affects the embryo's metabolome.

Further underpinning the possible effect of embryo treatment on its metabolome and pregnancy outcome prediction, Gómez et al. (2020b) revealed distinct blood biomarkers in cows transferred with differently produced embryos. RF was applied for ranking potential metabolite biomarkers based on their importance in pregnant and non-pregnant classification for Holstein cows having either fresh or VW embryos transferred. Metabolite ranking varied between different ET embryos, with ornithine and oxoglutaric acid being among top metabolites for fresh embryos, and L-glutamine and L-lysine being among top metabolites for VW embryos. The same approach in a study with Asturiana de los Valles (AV) also ranked oxoglutaric acid highly for pregnancy outcome prediction in recipient receiving fresh embryos, whereas dimethylamine and 2-hydroxybutyric acid were top metabolites for VW embryos (Gómez et al., 2020a). In both studies, only metabolites from day 7 and not day 0 had predictive ability for pregnancy outcome. Oxoglutaric acid, an intermediate in the Krebs cycle and an antioxidant, was found to have positive effects on embryo development in mice, in which oxogluratic acid treatment of embryos before ET increased blastocyst development rate and foetal growth (Zhang et al., 2019). If oxoglutaric acid blood levels reflect its uterine levels, this metabolite may positively affect embryo development and survival in the bovine uterus.

Taking a different approach, Gómez et al. (2021) applied SVM, PLS-DA and RF on CM of VW embryos and recipient blood metabolomics for feature ranking and pregnancy outcome prediction. Among high-ranking features in CM were metabolites involved in the lipid metabolism (stearic, capric and palmitic acid), with higher levels of non-esterified saturated fatty acids (NEFA) in the CM of non-viable embryos. NEFA exposure alters the epigenetic and transcriptomic profiles of oocytes and blastocysts (Desmet et al., 2016), and negatively affects embryo viability, shown by e.g. decreased cell number, elevated apoptosis and altered metabolism (Van Hoeck et al., 2011). Top ranking amino acids in recipient blood were similar to previous findings in Holsteins and AV (Gómez et al., 2020b, Gómez et al., 2020a). Furthermore, NEFAs were among high ranked blood metabolites in AV, with lower levels among pregnant cows (Gómez et al., 2021). NEFAs are released into the blood during a negative energy balance (NEB), and its concentrations are related to the depth of NEB. Besides the negative effect of NEFAs on embryos, a negative influence also exists on

endometrial cells, causing decreased cell viability and elevated levels of pro-inflammatory cytokines (Chankeaw et al., 2018).

In the most recent study, pregnancy outcome prediction and biomarker ranking for fresh and frozen/thawed (FT) embryos was done by Gimeno et al. (2023), also applying SVM, PLS-DA and RF, using multiple iterations to reevaluate misclassified samples and improve the model, on both blood metabolites and embryo CM. More recipients than embryos were shown to be competent, which can lead to misclassifying viable recipients because they were matched with non-viable embryos. Taking viability of embryos in consideration during revaluation of recipients, improved pregnancy outcome predictions. In concordance with Gómez et al. (2020b), L-glutamine and L-glycine were top ranked recipient biomarkers. For CM biomarkers, higher prediction ability was achieved with FT embryos than fresh embryos, and top biomarkers varied between fresh and ET embryos (Gimeno et al., 2023), further supporting production procedures of embryos affecting its metabolome.

Table 1: Overview of articles applying machine learning methods to omics data obtained from embryo or maternal samples up to day 7 of gestation, listed in order of citation.

Reference	Goal of ML	ML methods	Data type	Data set	Result/highest performance
Rabaglino et al. (2023)	Biomarker discovery	sPLS-DA	Embryo transcriptomics	89 blastocysts, 48 elongated conceptuses	8 biomarkers
	Pregnancy outcome prediction	BLR, ANN			>85% accuracy
Hoorn et al. (2024)	Biomarker discovery	RF (BORUTA), sPLS-DA	Endometrium transcriptomics	193 endometrial cytobrush samples	57 biomarkers
	Pregnancy outcome prediction	SVM			77% accuracy
Diniz et al. (2022)	Biomarker discovery	Ordinal learning method and models	Endometrium transcriptomics	43 endometrial samples	9 biomarkers
	Pregnancy outcome prediction				53.85 - 61.54% (using 9 genes) and

		based on SVM, KNN			>80% accuracy (using 225 genes)
Rabaglino & Kadarmideen (2020)	Biomarker discovery	Models based on BN and logistic regression	Endometrium transcriptomics	52 endometrial samples	50 biomarkers
	Pregnancy outcome prediction	SVM			96.1% accuracy
Moorey et al. (2020)	Pregnancy outcome prediction	Parallelised RF	White blood cell transcriptomics	23 blood samples, 198 genes	>90% accuracy
Muñoz et al. (2014b)	Pregnancy outcome prediction	KNN	Embryo CM and blood metabolomics	49 MOET embryos	64.4% accuracy
				49 blood samples	74.2% accuracy
Muñoz et al. (2014a)	Pregnancy outcome prediction	KNN	Embryo CM and blood metabolomics	69 IVP fresh or VW embryos	71.9% accuracy
				69 blood samples	74.9% accuracy
Gómez et al. (2020b)	Biomarker ranking	RF	Blood metabolomics	67 blood samples (Holstein)	NA
Gómez et al. (2020a)	Biomarker ranking	RF	Blood metabolomics	74 blood samples (Asturiana de los Valles)	NA
Gómez et al. (2021)	Biomarker ranking	SVM, PLS-DA, RF	Embryo CM and blood metabolomics	36 VW embryos, 36 blood samples	NA

Gimeno et al. (2023)	Biomarker ranking	RF	Embryo CM and blood metabolomics	70 IVP fresh or FT embryos, 107 blood samples	NA

ML, machine learning; (s)PLS-DA, (sparse) partial least squares discriminant analysis; BLR, Bayesian logistic regression; ANN, artificial neural network; RF, random forest; SVM, support vector machine; KNN, k-nearest neighbour; BN, Bayesian network; IVP, in vitro produced; MOET, multiple ovulation embryo transfer; VW, vitrified/warmed; FT, frozen/thawed; NA, not applicable

4. Machine learning to study embryonic, foetal or maternal molecular features after day 7 of gestation

Endometrium-embryo cross talk starts playing an important role for embryo survival from the very first moments of interaction. The focus of research on embryo development from day 7 of gestation has not been pregnancy outcome prediction, because embryo or recipient selection needs to be done before this moment, as ET occurs at day 7. Instead, research has focused on understanding the molecular profile of both embryo and uterus and what processes are key to pregnancy maintenance during this period (see Table 2).

4.1. Machine learning in conceptus omics

More IVP embryos than MOET embryos have been transferred worldwide during the last eight years, with the last IETS report showing a significative divergence: 1,189,699 transferred embryos were IVP while 368,783 were produced *in vivo* (Viana, 2022). However, as mentioned in the introduction, IVP embryos achieve lower pregnancy rates compared to MOET embryos (Ealy et al., 2019). To explore the molecular profile induced by the lack of the oviductal and uterine environment, regardless of variables such as the maturation media, technical procedures, or parental characteristics, we performed a meta-analysis of six transcriptomic and four epigenetic datasets from day 7 blastocysts to day 13 and day 16 conceptuses to define temporally DEG and differently methylated genes (DMG) between IVP and MOET embryos (Rabaglino et al., 2021). A SVM model was trained on conceptus expression data to predict the type of embryo production, which was validated on independent conceptus datasets. This test dataset consisted of two datasets, the first one consisting of transcriptomic data of short and long conceptuses, and the second dataset consisting of conceptuses of embryos treated or not with Dickkopf-related protein (*DKK1*, Wnt signalling inhibitor). It was hypothesised that short and untreated conceptuses would be

predicted as IVP embryos and long and *DKK1*-treated embryos be predicted as MOET embryos.

Prediction accuracies of 90-100% were achieved using a cluster of 188 DEG/DMG, in which the gene expression profiles showed a clear difference at day 13 between IVP and MOET embryos, while using only DEG resulted in lower accuracies of 70%, underscoring the power of multi-omics analysis. A significantly related pathway to the DEG/DMG cluster was focal adhesion, including the 'extracellular exosome'. Adhesion is important for connection of cells with their environment and the organisation of the cytoskeleton within the embryo (Shawky and Davidson, 2015), invasion of the embryo in the endometrium, and trophectoderm-endometrium communication at implantation (Kaneko et al., 2008, Kaneko et al., 2012). Additionally, exosomes, vesicles secreted by either the endometrium or the embryo carrying bioactive molecules, play key roles in maternal pregnancy recognition, and uterus-embryo signalling in the peri-implantation period (Bridi et al., 2020). Given the different expression of key pathways in IVP and MOET embryos, and short conceptuses being predicted as IVP embryos, the genes important for conceptus elongation and focal adhesion may be de-regulated in IVP embryos, possible affecting their viability (Rabaglino et al., 2021).

4.2. Machine learning in maternal transcriptomics

From day 7, IFNT secretion by the embryo poses an important way of interaction with the endometrium, altering gene expression and changing the uterine environment, needed for maternal pregnancy recognition and prevention of luteolysis. Talukder et al. (2023) used the BORUTA algorithm to identify the most prominent interferon-stimulated genes (ISGs), using transcriptomic data from day 15 endometrial explants exposed to nothing, IFNT or a conceptus. This resulted in 54 ISGs being upregulated around the time of maternal pregnancy recognition, of which the majority are related to immunity regulation. Maternal immunomodulation is needed for controlling innate immune responses and avoiding rejection of the allogenic embryo, in which IFNT likely plays an important role, creating an uterine environment suitable for pregnancy establishment (Rocha et al., 2021).

Less well researched is the contribution of the sire in conceptus-induced changes in the uterine environment. Fertility can vary significantly among bulls, which may greatly impact pregnancy rates as one bull might fertilise thousands of cows. O'Callaghan et al. (2022) compared the transcriptomic profiles of endometrial explants exposed to conceptuses conceived with semen from high fertile and low fertile bulls. We used three methods for selection of genes differing between fertility classes: identification of DEGs, co-expression network analysis, and lastly sPLS-DA was applied to identify genes most discriminative between bull fertility classes. Subsequently, SVM was applied on selected genes to train a

model for fertility status prediction. The highest accuracy of 90% was achieved using 200 genes selected by sPLS-DA. Upregulated genes predictive of conception with a high fertile bull were mostly involved in immune regulation, further underscoring the importance of proper immune regulation for survival of the peri-implantation embryo.

4.3. Machine learning in foetal transcriptomics

A topic further explored in human research is the use of maternal blood omics and ML to predict normal pregnancy or future pregnancy complications (Rasmussen et al., 2022, Xiong et al., 2022, Camunas-Soler et al., 2022), and prediction of gestational stage (Ngo et al., 2018). With the use of ML in the bovine field being mostly applied for predictions in the early stage of pregnancy, predictive studies involving foetal development are scarce. To shed light on this area, we explored the use of maternal blood transcriptomics for predicting foetal weight, by identifying co-expressed overlapping genes between the maternal blood and day 42 foetal organs (heart, liver, gonads) positively correlated to foetal weight (Rabaglino et al., 2023c). The overlapping genes between maternal blood and each organ were used for training a regression model, applying extreme gradient boosting (XGboost), after which the model was tested using the same genes in maternal blood. The most effective training dataset consisted of 35 genes overlapping between foetal heart and blood, achieving a root-mean-square error of 0.4. Furthermore, variance in foetal heart genes explained ~93% of gene expression variance in the maternal blood. The 35 selected genes enriched ontological terms related to energy metabolism processes, including oxidative phosphorylation. These results showed a relationship between the molecular profile of the developing foetal heart and foetal weight, which can also be measured in and is associated with transcripts in the maternal blood.

Table 2: Overview of articles applying machine learning methods to omics data obtained from embryo or maternal samples after day 7 of gestation, listed in order of citation.

Reference	Goal of ML	ML methods	Data type	Data set	Result/highest performance
Rabaglino et al. (2021)	Prediction of type of embryo production (IVP vs MOET)	SVM	Embryo transcriptomics and epigenomics	34 blastocysts and 40 conceptuses, 188 DEG/DMG	90-100% accuracy

Talukder et al. (2023)	Identify ISGs	RF (BORUTA)	Endometrium transcriptomics	53 endometrial explants	54 ISGs
O'Callaghan et al. (2022)	Biomarker ranking Prediction bull fertility status	sPLS-DA SVM	Endometrium transcriptomics	32 endometrial explants	200 genes 90% accuracy
Rabaglino et al. (2023c)	Predict foetal weight	Extreme gradient boosting (XGboost)	Blood and foetal organ (heart, liver, gonads) transcriptomics	10 blood samples, 8-9 samples per foetal organ, 35 genes	0.4 root-mean-square error

ML, machine learning; IVP, in vitro produced; MOET, multiple ovulation embryo transfer; (s)PLS-DA, (sparse) partial least squares discriminant analysis; RF, random forest; SVM, support vector machine; DEG, differently expressed genes; DMG, differently methylated genes; ISG, interferon-stimulated genes

5. Discussion

Machine learning has been increasingly utilised in a wide range of medical and biological applications, demonstrating its significant value. While its application in cattle reproduction and fertility is relatively recent, ML has already shown its potential in enhancing our understanding of the biological mechanism driving bovine pregnancy establishment. ML can play a crucial role in reducing pregnancy loss through its predictive capabilities, identifying relevant biomarkers for maternal receptivity and embryonic viability.

There is no universally optimal ML model for every situation. For ML method selection, the type and amount of data needs to be considered. Often, multiple models may be suitable for a given problem, and testing several models can help determine the most effective one. This approach allows for the comparison of their performance, and, in the context of biomarker identification, which biomarkers are selected and their relevance. Among the discussed articles, SVM and KNN were the most often used models for class prediction, and only one study applied a deep learning approach through ANN. While the use of deep learning is a hot topic in multiple fields, classical ML models have shown to often outperform deep learning methods when applied to tabular data (e.g. transcriptomics, metabolomics) (Eraslan et al., 2019). Furthermore, considering the small sample sizes of discussed articles, deep learning methods may not always be a good fit as they generally

perform better with large training data sets. However, we obtained high prediction accuracies (>85%) for embryo competence using ANN (Rabaglino et al., 2023b). Additionally, ANN outperformed classical ML methods in a study predicting embryo implantation in humans, achieving 100% accuracy (Cheredath et al., 2023).

The basis of creating a good predictive model is the selection of an appropriate training dataset. Low prediction accuracies can often be traced back to noisy or mislabelled data, affecting performance. Specifically for predicting pregnancy outcome, correct labelling of data is challenging. Only the pregnancy outcome can be measured, but the exact cause of embryo mortality may not be known, as it can be due to an incompetent embryo and/or an unreceptive endometrium. Thus, when e.g. predicting pregnancy outcome based on embryo omics, an embryo may be wrongly classified as incompetent, while the cause for embryo mortality was maternal. Reevaluating misclassified samples in multiple iterations during training indeed improved the model's performance in the study by Gimeno et al. (2023). Additionally, invasive data collection such as embryo biopsy may compromise its competence (Ponsart et al., 2013), possibly resulting in a competent embryo being mislabelled as non-competent. Furthermore, the training dataset affects the generalisability of the model, and thereby its practical applicability. When one wants to be able to apply a ML model on varying external datasets, variance also needs to be included in the training dataset, lowering the risk of overfitting on a specific dataset with a specific origin, considering that prediction accuracies and/or identified top biomarkers may differ between breeds (Gómez et al., 2020b, Gómez et al., 2020a, Rabaglino and Kadarmideen, 2020), laboratories (Gómez et al., 2020b) and embryo production techniques (Gimeno et al., 2023, Gómez et al., 2020a).

It has been suggested that maternal competence has a higher variability compared to embryo viability, with differing pregnancy outcomes being more associated with varying recipient competence (McMillan, 1998, McMillan and Donnison, 1999). This would implicate that using maternal data could better predict pregnancy outcome. Supporting this, metabolomics studies showed higher pregnancy success prediction accuracies for blood metabolites compared to CM (Muñoz et al., 2014b, Muñoz et al., 2014a). Additionally, Gimeno et al. (2023) identified more competent recipients compared to embryos. Employing maternal omics may, thus, be a more effective approach in predicting pregnancy success and lowering embryo losses. A highly fertile cow will however not become pregnant if matched with an incompetent embryo, and relevant prediction accuracies have also been achieved using embryo omics (Muñoz et al., 2014b, Muñoz et al., 2014a, Rabaglino et al., 2023b), indicating that the use of both maternal and embryo omics are relevant for improving fertility rates.

Not discussed in this review is the influence of sire fertility on embryo survival. Fertility varies significantly among bulls, resulting in different field fertility rates, so the importance of sperm quality for pregnancy establishment should not be ignored. While routine sperm evaluations can detect substantially low fertile bulls, identifying sub-fertile bulls producing apparently normal sperm remains a challenge, with no accurate diagnostic test available (Kastelic and Thundathil, 2008). Hence, the use of ML to aid in fertility status prediction has also surged in this field in recent years. ML methods applied to multiple omics types and sperm variables have been utilised for predicting bull fertility (Bucher et al., 2019, Costes et al., 2024, Costes et al., 2022, Rabaglino et al., 2022) and embryo yield (Campanholi et al., 2023). Bull fertility is not only determined by the semen's ability to fertilise an oocyte but also by its impact on embryonic characteristics. Top predictive features for bull fertility were found to be related to embryonic development (Costes et al., 2024), and the endometrial transcriptome differed after being exposed to conceptuses conceived with semen from high or low fertile bulls (O'Callaghan et al., 2022). However, the true influence of sire fertility on embryo survival and identification of affected pathways requires further research.

Currently, embryo quality classification relies on visual inspection of microscopic images, following IETS grading guidelines (Bó and Mapletoft, 2018). Although higher pregnancy rates are achieved with high-quality graded embryos (Farin et al., 1999), this method suffers from subjectivity and low reproducibility (Hansen, 2020). Together with the high pregnancy losses still experienced, there is need for more reliable embryo viability prediction methods. Besides the use of omics data for embryo competence prediction as discussed in this review, ML has also been applied to microscopic embryo images, though it is more widely explored in human research than in bovine research. ANNs have been trained on bovine embryo images labelled by embryologists to predict embryo quality grades (Rocha et al., 2017). However, quality grades do not directly translate to embryo viability, so incorporating pregnancy outcomes may increase predictive importance for embryo competence. ML performs better when the correlation between the label and the data is stronger. Omics data, being more directly related to the molecular biology of the embryo, may provide a more reliable relation with embryo viability than images and might be better suited for predicting embryo survival through ML. However, further research should be performed assessing the predictive value of images and omics, alone or combined, on embryo competence.

6. Future perspectives

The future of ML in bovine embryonic and foetal development holds promising advancements, considering the use of ML has only surged in recent years. The articles discussed in this review have applied ML on relatively small datasets, which can lead to

suboptimal performance, overfitting and limited generalisability. Therefore, a key focus of future studies should be optimising ML performance and validation of the already obtained results. Nevertheless, with the increasing availability of omics data in public databases, and the ability to integrate and reuse existing datasets, ML models can be trained on larger and more diverse datasets, enhancing their robustness and accuracy. ML for biomarker discovery can be applied on integrated datasets and validated on external datasets, as we did for the study identifying biomarkers of embryo survival (Rabaglino et al., 2023b), making these markers powerful and reliable as they reached high predictive accuracies. Furthermore, the involvement of the selected biomarkers in relevant biological pathways highlighted their importance for embryo survival.

The increased understanding of embryo development through validated biomarkers and key pathways can be employed to improve embryo cultures and maternal treatments enhancing endometrial function. For example, culturing embryos in CM with amino acid concentrations similar to the uterine fluid improved embryo development and freezing viability (Li et al., 2006). Additionally, ML models can assist in embryo treatment experiments by providing a more reliable method for embryo competence estimation, allowing the treatment effect to be more accurately defined (Rabaglino and Hansen, 2024).

Nowadays, the collection of biological samples from the uterus or the embryo is not done routinely, except for some breeding companies performing embryo biopsies for genomic selection. Nevertheless, using these reproductive management practices can be justified if the application of ML will objectively select competent embryos for ET and identify receptive cows, effectively increasing pregnancy success rates. However, for these models to be suitable for clinical implementation, they must be trained in large datasets (considering the biological content of the data) and demonstrate high performance across different settings. Field validation and development of applicability standards are crucial to harnessing the power of ML and ensuring its reliability and practical utility as a supportive tool in bovine reproduction.

7. Conclusion

The application of ML in bovine reproductive research holds significant potential for reducing embryo mortality and improving pregnancy success rates. By employing the power of ML to analyse complex and high-dimensional omics data, researchers can gain deeper insights into the molecular mechanisms essential for successful embryonic and foetal development. ML models have demonstrated their ability to predict pregnancy outcomes and to identify critical biomarkers. Future efforts should focus on developing non-invasive and practical applicable assessment methods to enhance the precision and efficacy of reproductive strategies. As the cattle industry continues to evolve, ML will play a crucial role

in optimising reproductive efficiency and genetic selection, ultimately contributing to its economic sustainability.

8. Data Availability Statement: N/A

9. Conflicts of Interest: The authors declare no conflicts of interest.

10. Declaration of Funding: This research did not receive any specific funding.

References

ALFATTAH, M. A., CORREIA, C. N., BROWNE, J. A., MCGETTIGAN, P. A., PLUTA, K., CARRINGTON, S. D., MACHUGH, D. E. & IRWIN, J. A. 2024. Transcriptomics analysis of the bovine endometrium during the perioestrus period. *PLoS One,* 19**,** e0301005.

ASSIS NETO, A. C., PEREIRA, F. T., SANTOS, T. C., AMBROSIO, C. E., LEISER, R. & MIGLINO, M. A. 2010. Morpho-physical recording of bovine conceptus (Bos indicus) and placenta from days 20 to 70 of pregnancy. *Reprod Domest Anim,* 45**,** 760-72.

BANLIAT, C., MAHÉ, C., LAVIGNE, R., COM, E., PINEAU, C., LABAS, V., GUYONNET, B., MERMILLOD, P. & SAINT-DIZIER, M. 2022. Dynamic Changes in the Proteome of Early Bovine Embryos Developed In Vivo. *Front Cell Dev Biol,* 10**,** 863700.

BAUERSACHS, S., ULBRICH, S. E., ZAKHARTCHENKO, V., MINTEN, M., REICHENBACH, M., REICHENBACH, H. D., BLUM, H., SPENCER, T. E. & WOLF, E. 2009. The endometrium responds differently to cloned versus fertilized embryos. *Proc Natl Acad Sci U S A,* 106**,** 5681-6.

BERG, D. K., VAN LEEUWEN, J., BEAUMONT, S., BERG, M. & PFEFFER, P. L. 2010. Embryo loss in cattle between Days 7 and 16 of pregnancy. *Theriogenology,* 73**,** 250-60.

BÓ, G. & MAPLETOFT, R. 2018. Evaluation and classification of bovine embryos. *Anim Reprod,* 10**,** 344-348.

BREIMAN, L. 2001. Random Forests. *Machine Learning,* 45**,** 5-32.

BRIDI, A., PERECIN, F. & SILVEIRA, J. C. D. 2020. Extracellular Vesicles Mediated Early Embryo-Maternal Interactions. *Int J Mol Sci,* 21.

BUCHER, K., MALAMA, E., SIUDA, M., JANETT, F. & BOLLWEIN, H. 2019. Multicolor flow cytometric analysis of cryopreserved bovine sperm: A tool for the evaluation of bull fertility. *J Dairy Sci,* 102**,** 11652-11669.

CAMPANHOLI, S. P., GARCIA NETO, S., PINHEIRO, G. M., NOGUEIRA, M. F. G., ROCHA, J. C., LOSANO, J. D. A., SIQUEIRA, A. F. P., NICHI, M., ASSUMPÇÃO, M., BASSO, A. C., MONTEIRO, F. M. & GIMENES, L. U. 2023. Can in vitro embryo production be estimated from semen variables in Senepol breed by using artificial intelligence? *Front Vet Sci,* 10**,** 1254940.

CAMUNAS-SOLER, J., GEE, E. P. S., REDDY, M., MI, J. D., THAO, M., BRUNDAGE, T., SIDDIQUI, F., HEZELGRAVE, N. L., SHENNAN, A. H., NAMSARAEV, E., HAVERTY, C., JAIN, M., ELOVITZ, M. A., RASMUSSEN, M. & TRIBE, R. M. 2022. Predictive RNA profiles for early and very early spontaneous preterm birth. *Am J Obstet Gynecol,* 227**,** 72.e1-72.e16.

CHANKEAW, W., GUO, Y. Z., BÅGE, R., SVENSSON, A., ANDERSSON, G. & HUMBLOT, P. 2018. Elevated non-esterified fatty acids impair survival and promote lipid accumulation and pro-inflammatory cytokine production in bovine endometrial epithelial cells. *Reprod Fertil Dev,* 30**,** 1770-1784.

CHEN, T. & GUESTRIN, C. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Association for Computing Machinery.

CHEN, X. & ISHWARAN, H. 2012. Random forests for genomic data analysis. *Genomics,* 99**,** 323-9.

CHEREDATH, A., UPPANGALA, S., C, S. A., JIJO, A., R, V. L., KUMAR, P., JOSEPH, D., G, A. N., KALTHUR, G. & ADIGA, S. K. 2023. Combining Machine Learning with Metabolomic and Embryologic Data Improves Embryo Implantation Prediction. *Reprod Sci,* 30**,** 984-994.

COSTES, V., CHAULOT-TALMON, A., SELLEM, E., PERRIER, J. P., AUBERT-FRAMBOURG, A., JOUNEAU, L., PONTLEVOY, C., HOZÉ, C., FRITZ, S., BOUSSAHA, M., LE DANVIC, C., SANCHEZ, M. P., BOICHARD, D., SCHIBLER, L., JAMMES, H., JAFFRÉZIC, F. & KIEFER, H. 2022. Predicting male fertility from the sperm methylome: application to 120 bulls with hundreds of artificial insemination records. *Clin Epigenetics,* 14**,** 54.

COSTES, V., SELLEM, E., MARTHEY, S., HOZE, C., BONNET, A., SCHIBLER, L., KIEFER, H. & JAFFREZIC, F. 2024. Multi-omics data integration for the identification of biomarkers for bull fertility. *PLoS One,* 19**,** e0298623.

DEGRELLE, S. A., CAMPION, E., CABAU, C., PIUMI, F., REINAUD, P., RICHARD, C., RENARD, J. P. & HUE, I. 2005. Molecular evidence for a critical period in mural trophoblast development in bovine blastocysts. *Dev Biol,* 288**,** 448-60.

DESMET, K. L., VAN HOECK, V., GAGNE, D., FOURNIER, E., THAKUR, A., O'DOHERTY, A. M., WALSH, C. P., SIRARD, M. A., BOLS, P. E. & LEROY, J. L. 2016. Exposure of bovine oocytes and embryos to elevated non-esterified fatty acid concentrations: integration of epigenetic and transcriptomic signatures in resultant blastocysts. *BMC Genomics,* 17**,** 1004.

DEUTSCH, D. R., FRÖHLICH, T., OTTE, K. A., BECK, A., HABERMANN, F. A., WOLF, E. & ARNOLD, G. J. 2014. Stage-specific proteome signatures in early bovine embryo development. *J Proteome Res,* 13**,** 4363-76.

DINIZ, W. J. S., BANERJEE, P., RODNING, S. P. & DYCE, P. W. 2022. Machine Learning-Based Co-Expression Network Analysis Unravels Potential Fertility-Related Genes in Beef Cows. *Animals (Basel),* 12.

EALY, A. D., WOOLDRIDGE, L. K. & MCCOSKI, S. R. 2019. BOARD INVITED REVIEW: Post-transfer consequences of in vitro-produced embryos in cattle. *J Anim Sci,* 97**,** 2555-2568.

ELKAN, C. 2011. Nearest neighbor classification. 11**,** 3.

ERASLAN, G., AVSEC, Ž., GAGNEUR, J. & THEIS, F. J. 2019. Deep learning: new computational modelling techniques for genomics. *Nat Rev Genet,* 20**,** 389-403.

FARIN, P. W., SLENNING, B. D. & BRITT, J. H. 1999. Estimates of pregnancy outcomes based on selection of bovine embryos produced in vivo or in vitro. *Theriogenology,* 52**,** 659-70.

FERNÁNDEZ-DELGADO, M., CERNADAS, E., BARRO, S. & AMORIM, D. 2014. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.,* 15**,** 3133–3181.

FREI, R. E., SCHULTZ, G. A. & CHURCH, R. B. 1989. Qualitative and quantitative changes in protein synthesis occur at the 8-16-cell stage of embryogenesis in the cow. *J Reprod Fertil,* 86**,** 637-41.

GIMENO, I., SALVETTI, P., CARROCERA, S., GATIEN, J., GARCÍA-MANRIQUE, P., LÓPEZ-HIDALGO, C., VALLEDOR, L. & GÓMEZ, E. 2023. Biomarker metabolite mating of viable frozen-thawed in vitro-produced bovine embryos with pregnancy-competent recipients leads to improved birth rates. *J Dairy Sci,* 106**,** 6515-6538.

GÓMEZ, E., CANELA, N., HERRERO, P., CERETO, A., GIMENO, I., CARROCERA, S., MARTIN-GONZALEZ, D., MURILLO, A. & MUÑOZ, M. 2021. Metabolites Secreted by Bovine Embryos In Vitro Predict Pregnancies That the Recipient Plasma Metabolome Cannot, and Vice Versa. *Metabolites,* 11.

GÓMEZ, E., MUÑOZ, M., GATIEN, J., CARROCERA, S., MARTÍN-GONZÁLEZ, D. & SALVETTI, P. 2020a. Metabolomic identification of pregnancy-specific biomarkers in blood plasma of BOS TAURUS beef cattle after transfer of in vitro produced embryos. *J Proteomics,* 225**,** 103883.

GÓMEZ, E., SALVETTI, P., GATIEN, J., CARROCERA, S., MARTÍN-GONZÁLEZ, D. & MUÑOZ, M. 2020b. Blood Plasma Metabolomics Predicts Pregnancy in Holstein Cattle Transferred with Fresh and Vitrified/Warmed Embryos Produced in Vitro. *J Proteome Res,* 19**,** 1169-1182.

GÓMEZ, E., SALVETTI, P., GATIEN, J., MUÑOZ, M., MARTÍN-GONZÁLEZ, D., CARROCERA, S. & GOYACHE, F. 2020c. Metabolomic Profiling of Bos taurus Beef, Dairy, and Crossbred Cattle: A Between-Breeds Meta-Analysis. *J Agric Food Chem,* 68**,** 8732-8743.
HANSEN, P. J. 2020. The incompletely fulfilled promise of embryo transfer in cattle-why aren't pregnancy rates greater and what can we do about it? *J Anim Sci,* 98**,** 1-20.
HOORN, Q. A., RABAGLINO, M. B., AMARAL, T. F., MAIA, T. S., YU, F., COLE, J. B. & HANSEN, P. J. 2024. Machine learning to identify endometrial biomarkers predictive of pregnancy success following artificial insemination in dairy cows. *Biol Reprod*.
KANEKO, Y., LECCE, L., DAY, M. L. & MURPHY, C. R. 2012. Focal adhesion kinase localizes to sites of cell-to-cell contact in vivo and increases apically in rat uterine luminal epithelium and the blastocyst at the time of implantation. *J Morphol,* 273**,** 639-50.
KANEKO, Y., LINDSAY, L. A. & MURPHY, C. R. 2008. Focal adhesions disassemble during early pregnancy in rat uterine epithelial cells. *Reprod Fertil Dev,* 20**,** 892-9.
KASTELIC, J. P. & THUNDATHIL, J. C. 2008. Breeding soundness evaluation and semen analysis for predicting bull fertility. *Reprod Domest Anim,* 43 Suppl 2**,** 368-73.
KINGSFORD, C. & SALZBERG, S. L. 2008. What are decision trees? *Nat Biotechnol,* 26**,** 1011-3.
KOPECNY, V., FLECHON, J. E., CAMOUS, S. & FULKA, J., JR. 1989. Nucleologenesis and the onset of transcription in the eight-cell bovine embryo: fine-structural autoradiographic study. *Mol Reprod Dev,* 1**,** 79-90.
KURSA, M. B. & RUDNICKI, W. R. 2010. Feature Selection with the Boruta Package. *Journal of Statistical Software,* 36**,** 1-13.
LÊ CAO, K.-A., BOITARD, S. & BESSE, P. 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics,* 12**,** 253.
LECLERCQ, M., VITTRANT, B., MARTIN-MAGNIETTE, M., SCOTT BOYER, M., PERIN, O., BERGERON, A., FRADET, Y. & DROIT, A. 2019. Large-Scale Automatic Feature Selection for Biomarker Discovery in High-Dimensional OMICs Data. *Front Genet,* 10**,** 452.
LEESE, H. J. 2002. Quiet please, do not disturb: a hypothesis of embryo metabolism and viability. *Bioessays,* 24**,** 845-9.
LI, R., WEN, L., WANG, S. & BOU, S. 2006. Development, freezability and amino acid consumption of bovine embryos cultured in synthetic oviductal fluid (SOF) medium containing amino acids at oviductal or uterine-fluid concentrations. *Theriogenology,* 66**,** 404-14.
LONERGAN, P., FAIR, T., FORDE, N. & RIZOS, D. 2016. Embryo development in dairy cattle. *Theriogenology,* 86**,** 270-277.
MANSOURI-ATTIA, N., SANDRA, O., AUBERT, J., DEGRELLE, S., EVERTS, R. E., GIRAUD-DELVILLE, C., HEYMAN, Y., GALIO, L., HUE, I., YANG, X., TIAN, X. C., LEWIN, H. A. & RENARD, J. P. 2009. Endometrium as an early sensor of in vitro embryo manipulation technologies. *Proc Natl Acad Sci U S A,* 106**,** 5687-92.
MATHEW, D. J., SÁNCHEZ, J. M., PASSARO, C., CHARPIGNY, G., BEHURA, S. K., SPENCER, T. E. & LONERGAN, P. 2019. Interferon tau-dependent and independent effects of the bovine conceptus on the endometrial transcriptome. *Biol Reprod,* 100**,** 365-380.
MCMILLAN, W. H. 1998. Statistical models predicting embryo survival to term in cattle after embryo transfer. *Theriogenology,* 50**,** 1053-70.
MCMILLAN, W. H. & DONNISON, M. J. 1999. Understanding maternal contributions to fertility in recipient cattle: development of herds with contrasting pregnancy rates. *Anim Reprod Sci,* 57**,** 127-40.
MEMILI, E. & FIRST, N. L. 2000. Zygotic and embryonic gene expression in cow: a review of timing and mechanisms of early gene expression as compared with other species. *Zygote,* 8**,** 87-96.
MOOREY, S. E., WALKER, B. N., ELMORE, M. F., ELMORE, J. B., RODNING, S. P. & BIASE, F. H. 2020. Rewiring of gene expression in circulating white blood cells is associated with pregnancy outcome in heifers (Bos taurus). *Sci Rep,* 10**,** 16786.

MORAES, J. G. N., BEHURA, S. K., GEARY, T. W., HANSEN, P. J., NEIBERGS, H. L. & SPENCER, T. E. 2018. Uterine influences on conceptus development in fertility-classified animals. *Proc Natl Acad Sci U S A,* 115**,** E1749-E1758.

MUCHERINO, A., PAPAJORGJI, P. J. & PARDALOS, P. M. 2009. K-nearest neighbor classification. *Data Mining in Agriculture***,** 83-106.

MUÑOZ, M., UYAR, A., CORREIA, E., DÍEZ, C., FERNANDEZ-GONZALEZ, A., CAAMAÑO, J. N., MARTÍNEZ-BELLO, D., TRIGAL, B., HUMBLOT, P., PONSART, C., GUYADER-JOLY, C., CARROCERA, S., MARTIN, D., MARQUANT LE GUIENNE, B., SELI, E. & GOMEZ, E. 2014a. Prediction of pregnancy viability in bovine in vitro-produced embryos and recipient plasma with Fourier transform infrared spectroscopy. *J Dairy Sci,* 97**,** 5497-507.

MUÑOZ, M., UYAR, A., CORREIA, E., PONSART, C., GUYADER-JOLY, C., MARTÍNEZ-BELLO, D., MARQUANT-LE GUIENNE, B., FERNANDEZ-GONZALEZ, A., DÍEZ, C., CAAMAÑO, J. N., TRIGAL, B., HUMBLOT, P., CARROCERA, S., MARTIN, D., SELI, E. & GOMEZ, E. 2014b. Metabolomic prediction of pregnancy viability in superovulated cattle embryos and recipients with fourier transform infrared spectroscopy. *Biomed Res Int,* 2014**,** 608579.

NGO, T. T. M., MOUFARREJ, M. N., RASMUSSEN, M. H., CAMUNAS-SOLER, J., PAN, W., OKAMOTO, J., NEFF, N. F., LIU, K., WONG, R. J., DOWNES, K., TIBSHIRANI, R., SHAW, G. M., SKOTTE, L., STEVENSON, D. K., BIGGIO, J. R., ELOVITZ, M. A., MELBYE, M. & QUAKE, S. R. 2018. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science,* 360**,** 1133-1136.

O'CALLAGHAN, E., SANCHEZ, J. M., RABAGLINO, M. B., MCDONALD, M., LIU, H., SPENCER, T. E., FAIR, S., KENNY, D. A. & LONERGAN, P. 2022. Influence of sire fertility status on conceptus-induced transcriptomic response of the bovine endometrium. *Front Cell Dev Biol,* 10**,** 950443.

PASSARO, C., TUTT, D., MATHEW, D. J., SANCHEZ, J. M., BROWNE, J. A., BOE-HANSEN, G. B., FAIR, T. & LONERGAN, P. 2018. Blastocyst-induced changes in the bovine endometrial transcriptome. *Reproduction,* 156**,** 219-229.

PISNER, D. A. & SCHNYER, D. M. 2020. Chapter 6 - Support vector machine. *Machine Learning.* Academic Press.

PONSART, C., LE BOURHIS, D., KNIJN, H., FRITZ, S., GUYADER-JOLY, C., OTTER, T., LACAZE, S., CHARREAUX, F., SCHIBLER, L., DUPASSIEUX, D. & MULLAART, E. 2013. Reproductive technologies and genomic selection in dairy cattle. *Reprod Fertil Dev,* 26**,** 12-21.

PRIDDY, K. L. & KELLER, P. E. 2005. *Artificial neural networks: an introduction*, SPIE press.

RABAGLINO, M. B. 2023a. Review: Overview of the transcriptomic landscape in bovine blastocysts and elongated conceptuses driving developmental competence. *Animal,* 17 Suppl 1**,** 100733.

RABAGLINO, M. B. & HANSEN, P. J. 2024. Development of a formula for scoring competence of bovine embryos to sustain pregnancy. *Biochemistry and Biophysics Reports,* 39**,** 101772.

RABAGLINO, M. B. & KADARMIDEEN, H. N. 2020. Machine learning approach to integrated endometrial transcriptomic datasets reveals biomarkers predicting uterine receptivity in cattle at seven days after estrous. *Sci Rep,* 10**,** 16981.

RABAGLINO, M. B., LE DANVIC, C., SCHIBLER, L., KUPISIEWICZ, K., PERRIER, J. P., O'MEARA, C. M., KENNY, D. A., FAIR, S. & LONERGAN, P. 2022. Identification of sperm proteins as biomarkers of field fertility in Holstein-Friesian bulls used for artificial insemination. *J Dairy Sci,* 105**,** 10033-10046.

RABAGLINO, M. B., O'DOHERTY, A., BOJSEN-MØLLER SECHER, J., LONERGAN, P., HYTTEL, P., FAIR, T. & KADARMIDEEN, H. N. 2021. Application of multi-omics data integration and machine learning approaches to identify epigenetic and transcriptomic differences between in vitro and in vivo produced bovine embryos. *PLoS One,* 16**,** e0252096.

RABAGLINO, M. B., SALILEW-WONDIM, D., ZOLINI, A., TESFAYE, D., HOELKER, M., LONERGAN, P. & HANSEN, P. J. 2023b. Machine-learning methods applied to integrated transcriptomic data from bovine blastocysts and elongating conceptuses to identify genes predictive of embryonic competence. *Faseb j,* 37**,** e22809.

RABAGLINO, M. B., SÁNCHEZ, J. M., MCDONALD, M., O'CALLAGHAN, E. & LONERGAN, P. 2023c. Maternal blood transcriptome as a sensor of fetal organ maturation at the end of organogenesis in cattle. *Biol Reprod,* 109**,** 749-758.

RASMUSSEN, M., REDDY, M., NOLAN, R., CAMUNAS-SOLER, J., KHODURSKY, A., SCHELLER, N. M., CANTONWINE, D. E., ENGELBRECHTSEN, L., MI, J. D., DUTTA, A., BRUNDAGE, T., SIDDIQUI, F., THAO, M., GEE, E. P. S., LA, J., BARUCH-GRAVETT, C., SANTILLAN, M. K., DEB, S., AME, S. M., ALI, S. M., ADKINS, M., DEPRISTO, M. A., LEE, M., NAMSARAEV, E., GYBEL-BRASK, D. J., SKIBSTED, L., LITCH, J. A., SANTILLAN, D. A., SAZAWAL, S., TRIBE, R. M., ROBERTS, J. M., JAIN, M., HOGDALL, E., HOLZMAN, C., QUAKE, S. R., ELOVITZ, M. A. & MCELRATH, T. F. 2022. RNA profiles reveal signatures of future health and disease in pregnancy. *Nature,* 601**,** 422-427.

REESE, S. T., FRANCO, G. A., POOLE, R. K., HOOD, R., FERNADEZ MONTERO, L., OLIVEIRA FILHO, R. V., COOKE, R. F. & POHLER, K. G. 2020. Pregnancy loss in beef cattle: A meta-analysis. *Anim Reprod Sci,* 212**,** 106251.

RILEY, J. K. & MOLEY, K. H. 2006. Glucose utilization and the PI3-K pathway: mechanisms for cell survival in preimplantation embryos. *Reproduction,* 131**,** 823-35.

ROCHA, C. C., DA SILVEIRA, J. C., FORDE, N., BINELLI, M. & PUGLIESI, G. 2021. Conceptus-modulated innate immune function during early pregnancy in ruminants: a review. *Anim Reprod,* 18**,** e20200048.

ROCHA, J. C., PASSALIA, F. J., MATOS, F. D., TAKAHASHI, M. B., CINICIATO, D. S., MASERATI, M. P., ALVES, M. F., DE ALMEIDA, T. G., CARDOSO, B. L., BASSO, A. C. & NOGUEIRA, M. F. G. 2017. A Method Based on Artificial Intelligence To Fully Automatize The Evaluation of Bovine Blastocyst Images. *Sci Rep,* 7**,** 7659.

RUIZ-PEREZ, D., GUAN, H., MADHIVANAN, P., MATHEE, K. & NARASIMHAN, G. 2020. So you think you can PLS-DA? *BMC Bioinformatics,* 21**,** 2.

SÁNCHEZ, J. M., MATHEW, D. J., BEHURA, S. K., PASSARO, C., CHARPIGNY, G., BUTLER, S. T., SPENCER, T. E. & LONERGAN, P. 2019. Bovine endometrium responds differentially to age-matched short and long conceptuses. *Biol Reprod,* 101**,** 26-39.

SARTORI, R., BASTOS, M. R. & WILTBANK, M. C. 2010. Factors affecting fertilisation and early embryo quality in single- and superovulated dairy cattle. *Reprod Fertil Dev,* 22**,** 151-8.

SCHULZ, K. N. & HARRISON, M. M. 2019. Mechanisms regulating zygotic genome activation. *Nat Rev Genet,* 20**,** 221-234.

SCIORIO, R. & RINAUDO, P. 2023. Culture conditions in the IVF laboratory: state of the ART and possible new directions. *J Assist Reprod Genet,* 40**,** 2591-2607.

SHAWKY, J. H. & DAVIDSON, L. A. 2015. Tissue mechanics and adhesion during embryo development. *Dev Biol,* 401**,** 152-64.

SPONCHIADO, M., GOMES, N. S., FONTES, P. K., MARTINS, T., DEL COLLADO, M., PASTORE, A. A., PUGLIESI, G., NOGUEIRA, M. F. G. & BINELLI, M. 2017. Pre-hatching embryo-dependent and -independent programming of endometrial function in cattle. *PLoS One,* 12**,** e0175954.

STEPHENSON, T. A. 2000. An introduction to Bayesian network theory and usage.

TAGLAUER, E. S., TRIKHACHEVA, A. S., SLUSSER, J. G. & PETROFF, M. G. 2008. Expression and function of PDCD1 at the human maternal-fetal interface. *Biol Reprod,* 79**,** 562-9.

TALUKDER, A. K., RABAGLINO, M. B., BROWNE, J. A., CHARPIGNY, G. & LONERGAN, P. 2023. Dose- and time-dependent effects of interferon tau on bovine endometrial gene expression. *Theriogenology,* 211**,** 1-10.

VAN HOECK, V., STURMEY, R. G., BERMEJO-ALVAREZ, P., RIZOS, D., GUTIERREZ-ADAN, A., LEESE, H. J., BOLS, P. E. & LEROY, J. L. 2011. Elevated non-esterified fatty acid concentrations during bovine oocyte maturation compromise early embryo physiology. *PLoS One,* 6**,** e23183.

VIANA, J. 2022. 2022 statistics of embryo production and transfer in domestic farm animals: the main trends for the world embryo industry still stand. *Embryo Transfer Newsletter,* 41**,** 20-38.

WILTBANK, M. C., BAEZ, G. M., GARCIA-GUERRA, A., TOLEDO, M. Z., MONTEIRO, P. L., MELO, L. F., OCHOA, J. C., SANTOS, J. E. & SARTORI, R. 2016. Pivotal periods for pregnancy loss during the first trimester of gestation in lactating dairy cows. *Theriogenology,* 86**,** 239-53.

XIONG, Y., LIN, L., CHEN, Y., SALERNO, S., LI, Y., ZENG, X. & LI, H. 2022. Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques. *J Matern Fetal Neonatal Med,* 35**,** 2457-2463.

ZHANG, Z., HE, C., ZHANG, L., ZHU, T., LV, D., LI, G., SONG, Y., WANG, J., WU, H., JI, P. & LIU, G. 2019. Alpha-ketoglutarate affects murine embryo development through metabolic and epigenetic modulations. *Reproduction,* 158**,** 123-133.