

Universidade de Brasília – UnB
Faculdade de Ciências e Tecnologias em Engenharia – FCTE
Engenharia de Software

**Aplicação de Aprendizado de Máquina na
Predição Não-Invasiva de Euploidia em
Embriões Humanos com Base em Dados
Morfocinéticos**

Autor: Maria Eduarda Dos Santos Abritta Ferreira e Sabrina
Caldas Berno

Orientador: Prof Dr. George Marsicano Corrêa

Brasília, DF

2025



Maria Eduarda Dos Santos Abritta Ferreira e Sabrina Caldas Berno

**Aplicação de Aprendizado de Máquina na Predição
Não-Invasiva de Euploidia em Embriões Humanos com
Base em Dados Morfocinéticos**

Monografia submetida ao curso de graduação
em Engenharia de Software da Universidade
de Brasília, como requisito parcial para ob-
tenção do Título de Bacharel em Engenharia
de Software.

Universidade de Brasília – UnB

Faculdade de Ciências e Tecnologias em Engenharia – FCTE

Orientador: Prof Dr. George Marsicano Corrêa

Brasília, DF

2025

*Este trabalho é dedicado a todas as mulheres que
sonham em poder gerar a sua família*

Resumo

A fertilização in vitro é amplamente utilizada por casais com dificuldades reprodutivas ou que desejam postergar a gravidez, mas seu sucesso depende de vários fatores, sendo a seleção de embriões viáveis um dos mais críticos. Escolher o embrião correto aumenta as chances de implantação, reduz o risco de abortos e melhora as taxas de nascimentos saudáveis. O Teste Genético Pré-Implantacional para Aneuploidia é o método mais usado, mas, além do custo elevado, apresenta limitações de precisão diagnóstica, podendo levar à transferência de embriões inviáveis ou exclusão de viáveis. Com o avanço de tecnologias como o Time-Lapse System, que captura imagens contínuas do desenvolvimento embrionário, surge a chance de prever a euploidia usando dados morfocinéticos de forma menos invasiva e mais eficaz. A pesquisa adota uma abordagem quantitativa e experimental, focando na análise de dados numéricos de padrões morfocinéticos de embriões para o desenvolvimento da IA preditiva. Este trabalho desenvolve uma abordagem baseada em Machine Learning para identificar padrões nos dados morfocinéticos e prever a euploidia, oferecendo uma alternativa ao PGT-A sem intervenções invasivas. Após revisão de literatura e análise de correlação, foram identificadas as variáveis mais relevantes, sendo a idade e tb-t2b as com maior influência negativa. Os modelos de Machine Learning foram treinados e validados utilizando métricas como acurácia, sensibilidade e especificidade, de forma a avaliar seu desempenho na predição de euploidia a partir de dados morfocinéticos. O modelo desenvolvido, baseado em uma Rede Neural Artificial do tipo Perceptron Multicamadas (MLP), alcançou uma acurácia de 88,2% e uma Área sob a Curva ROC (AUC) de 0,944, indicando elevada capacidade discriminativa entre embriões euploides e aneuploides. A avaliação por meio da matriz de confusão evidenciou uma excelente performance na identificação de embriões aneuploides, com recall perfeito (1,00) para essa classe e robusto (0,75) para embriões euploides, sem ocorrência de falsos positivos. A integração do método LIME possibilitou tornar as predições do modelo interpretáveis, identificando quais variáveis morfocinéticas mais influenciaram cada predição individual, uma característica fundamental para garantir transparência e confiabilidade no contexto biomédico. A solução desenvolvida oferece uma ferramenta promissora para complementar o PGT-A, reduzindo custos e evitando intervenções invasivas, com potencial de aplicação real na rotina clínica da reprodução assistida.

Palavras-chave: Fertilização in vitro. Euploidia. Time-Lapse System. PGT-A. Inteligência Artificial. Aprendizado de máquina. Análise de dados.

Abstract

In vitro fertilization is widely used by couples facing reproductive difficulties or those wishing to postpone pregnancy, but its success depends on several factors, with the selection of viable embryos being one of the most critical. Choosing the correct embryo significantly increases the chances of successful implantation, reduces the risk of miscarriage, and improves healthy birth rates. Preimplantation Genetic Testing for Aneuploidy is the most commonly used method, but in addition to its high cost, it has diagnostic precision limitations, which can lead to the transfer of non-viable embryos or exclusion of viable ones. With the advancement of technologies such as the Time-Lapse System, which captures continuous images of embryonic development, the possibility arises of predicting euploidy using morphokinetic data in a less invasive and more effective manner. The research adopts a quantitative and experimental approach, focusing on the analysis of numerical data related to the morphokinetic patterns of embryos for the development of predictive AI. This work develops a Machine Learning-based approach to identify patterns in morphokinetic data and predict euploidy, offering an alternative to PGT-A without invasive interventions. After a literature review and correlation analysis, the most relevant variables were identified, with age and tb-t2b showing the highest negative influence. The Machine Learning models were trained and validated using metrics such as accuracy, sensitivity, and specificity to assess their performance in predicting embryo euploidy based on morphokinetic data. The developed model, based on a Multilayer Perceptron (MLP) Artificial Neural Network, achieved an accuracy of 88.2% and an Area Under the ROC Curve (AUC) of 0.944, indicating a high discriminative ability between euploid and aneuploid embryos. The evaluation through the confusion matrix demonstrated excellent performance in identifying aneuploid embryos, with perfect recall (1.00) for this class and robust recall (0.75) for euploid embryos, without any false positives. The integration of the LIME method made the model's predictions interpretable by identifying which morphokinetic variables most influenced each individual prediction, an essential feature to ensure transparency and reliability in the biomedical context. The developed solution offers a promising tool to complement PGT-A, reducing costs and avoiding invasive interventions, with potential for real-world application in the clinical routine of assisted reproduction.

Key-words: In vitro fertilization. Euploidy. Time-Lapse System. PGT-A. Artificial Intelligence. Machine Learning. Data analysis.

Lista de ilustrações

Figura 1 – Resultados do PGT-A e suas respectivas probabilidades de sucesso gestacional	19
Figura 2 – Discrepância potencial entre uma biópsia de células utilizando PGT-A e a composição celular em regiões adjacentes do trofotoderma	21
Figura 3 – Aprendizado Supervisionado: Identificando Euploidia e Aneuploidia . .	26
Figura 4 – Estrutura de um MLP com uma camada de entrada, três camadas ocultas e uma camada de saída.	29
Figura 5 – Dispersão entre Idade e t4 - Coeficiente de Spearman: -0.15 e entre Idade e t5 - Coeficiente de Spearman: 0.11	59
Figura 6 – Dispersão entre Idade e tSB - Coeficiente de Spearman: -0.10 Dispersão entre Idade e cc2 (t3-t2) - Coeficiente de Spearman: -0.15 Dispersão entre Idade e s2 (t4-t3) - Coeficiente de Spearman: -0.24 Dispersão entre Idade e s3 (t8-t5) - Coeficiente de Spearman: -0.28 . .	60
Figura 7 – Dispersão entre t2 e t4 - Coeficiente de Spearman: 0.89	62
Figura 8 – Dispersão entre t3 e t2 - Coeficiente de Spearman: 0.78	62
Figura 9 – Dispersão entre t4 e t5 - Coeficiente de Spearman: 0.56	62
Figura 10 – Dispersão entre t5 e t8 - Coeficiente de Spearman: 0.52	62
Figura 11 – Dispersão entre tSC e tSB - Coeficiente de Spearman: 0.75	63
Figura 12 – Dispersão entre tSC e tB - Coeficiente de Spearman: 0.74	63
Figura 13 – Dispersão entre tSB e tB - Coeficiente de Spearman: 0.93	64
Figura 14 – Dispersão entre cc2 (t3-t2) e t3 - Coeficiente de Spearman: 0.80	66
Figura 15 – Dispersão entre s3 (t8-t5) e t8 - Coeficiente de Spearman: 0.48	68
Figura 16 – Matriz de confusão para avaliação do desempenho do modelo	75
Figura 17 – Curva ROC com AUC de 0,94	75
Figura 18 – Relação entre sensibilidade e especificidade em diferentes limiares de decisão	77

Lista de tabelas

Tabela 1	– Fase 1: Análise e Preparação de Dados	41
Tabela 2	– Fase 2: Desenvolvimento e Avaliação do Modelo	45
Tabela 3	– Planilha Normalizada	57
Tabela 4	– Interpretação do coeficiente de correlação de Spearman	111

Lista de abreviaturas e siglas

IA	Inteligência Artificial
ICSI	Injeção Intracitoplasmática de Espermatozoides
FIV	Fertilização In Vitro
ML	Machine Learning
PGT-A	Teste Genético Pré-Implantacional para Aneuploidia
TRA	Transferência de Embriões
TLS	Time-Lapse System

Sumário

1	INTRODUÇÃO	12
1.1	Contexto	12
1.2	Motivação	13
1.3	Problema	14
1.4	Objetivos	15
1.4.1	Objetivos Gerais	15
1.4.2	Objetivos Específicos	15
1.5	Metodologia	15
1.6	Composição e estrutura do trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Fertilização In Vitro	17
2.2	Métodos de Avaliação Genética em Reprodução Assistida	18
2.3	Time-Lapse System	21
2.3.1	Idade	22
2.3.2	t2, t3, t4, t5, t8, s2, cc2 (t3-t2), tSC, tSB, tB, cc3 (t5-t3), s3 (t8-t5), t5-t2, tSC-t8 e tB-tSB	22
2.3.3	Estágio e Morfo	24
2.3.4	KIDScore TM	24
2.3.5	Ploidia	25
2.4	Aprendizado de Máquina	25
2.4.1	Redes Neurais Artificiais (RNA)	27
2.4.1.1	Multilayer Perceptron (MLP)	28
2.4.1.2	Características das Camadas Densas	29
2.4.1.3	Inteligência Artificial Explicável (XAI) e o Algoritmo LIME	30
2.5	Passos para o Desenvolvimento de um Algoritmo de Aprendizado de Máquina	31
2.5.1	Definição do problema e análise do panorama geral	31
2.5.2	Obtenção de Dados	32
2.5.3	Exploração de Dados	32
2.5.4	Preparação dos dados para os Algoritmos de Aprendizado de Máquina	33
2.5.5	Seleção e treinamento do modelo	34
2.5.6	Ajuste do modelo	34
2.5.7	Lançamento da Solução	35

2.6	Identificação de Padrões Morfocinéticos e Predição de Euploidia com IA e Trabalhos Correlatos	36
3	METODOLOGIA	39
3.1	Classificação da Pesquisa	39
3.1.1	Natureza	39
3.1.2	Método ou Abordagem Metodológica	39
3.1.3	Objetivos	39
3.1.4	Procedimentos De Pesquisa	40
3.2	Design da Pesquisa	40
3.3	Fases de Trabalho	40
3.3.1	Objetivo Específico 1 - Identificação de Parâmetros em Embriões	41
3.3.1.1	Atividade 1 (A1): Análise, Revisão, Seleção e Limpeza de Variáveis para Predição de Euploidia	41
3.3.1.1.1	Atividade 2 (A2): Normalização dos Dados para Otimização	42
3.3.1.1.2	Atividade 3 (A3): Identificação da Correlação e Atribuição de Pesos aos Parâmetros na Previsão da Ploidia do Embrião	43
3.3.1.1.3	Atividade 4 (A4): Separar o conjunto de dados em conjuntos de treinamento e teste, fazendo uma distribuição dos dados e aplicar técnica de aumento de dados	44
3.3.1.2	Objetivo Específico 2 - Treinamento e Ajuste de Modelo de Machine Learning para Predição de Euploidia	46
3.3.1.2.1	Atividade 5 (A5): Desenvolvimento e Treinamento do Modelo de Machine Learning para Otimização da Predição de Euploidia	46
3.3.1.3	Objetivo Específico 3 - Avaliação do modelo	47
3.3.1.3.1	Atividade 6 (A6): Utilizar métricas de avaliação mais adequadas para medir o desempenho do modelo de acordo com a natureza do problema de classificação	47
3.3.1.3.2	Atividade 7 (A7): Avaliar a precisão e eficácia do modelo em prever corretamente casos de euploidia e aneuploidia por meio da Matriz de Confusão e Curva ROC	48
3.3.1.4	Objetivo Específico 4 - Desenvolver uma Interface	49
3.3.1.4.1	Atividade 8 (A8): Prototipar uma interface básica para exibir as previsões de euploidia para o usuário final (médicos)	49
3.3.1.4.2	Atividade 9 (A9): Desenvolver a aplicação web integrando o protótipo à inteligência artificial existente	50
4	EXECUÇÃO DA PESQUISA E ANÁLISE DOS RESULTADOS . . .	53
4.1	Fase 1: Análise e Preparação de Dados	54
4.1.1	OE1 - Expandir, Processar e Analisar os Dados para Predição de Ploidia . .	54
4.1.1.1	Atividade 1 (A1): Análise, Revisão, Seleção e Limpeza de Variáveis para Predição de Euploidia	54
4.1.1.2	Limpeza dos Dados	55

4.1.1.3	Atividade 2 (A2): Normalização dos Dados para Otimização	55
4.1.1.4	Atividade 3 (A3): Identificação da Correlação entre os Parâmetros na Previsão da Ploidia do Embrião	58
4.1.1.5	Atividade 4 (A4): Separar o conjunto de dados em conjuntos de treinamento, validação e teste, fazendo uma distribuição dos dados e aplicar técnica de aumento de dados	70
4.2	Fase 2: Desenvolvimento e Avaliação do Modelo	72
4.2.1	OE2 - Treinar e Ajustar do Modelo de Machine Learning para Predição de Ploidia	72
4.2.1.1	Atividade 5 (A5): Desenvolvimento e Treinamento do Modelo de Machine Learning para Otimização da Predição de Euploidia	72
4.2.2	OE3 - Realizar Avaliação do Modelo	74
4.2.2.1	Atividade 6 (A6): Utilizar métricas adequadas para medir o desempenho do modelo	74
4.2.2.2	Atividade 7 (A7): Avaliação do Desempenho do Modelo na Predição por Meio da Matriz de Confusão e Curva ROC	76
4.2.3	OE4 - Construir Protótipo	77
4.2.3.1	Atividade 8 (A8): Prototipar uma interface básica para exibir as predições de euploidia para o usuário final (médicos)	77
4.2.3.2	Atividade 9 (A9): Desenvolver a aplicação web integrando o protótipo à inteligência artificial existente	79
5	DISCUSSÃO DOS RESULTADOS	82
5.1	Cumprimento dos Objetivos da Pesquisa	82
5.1.1	OE1: Expandir, Processar e Analisar os Dados para Predição de Ploidia	82
5.1.2	OE2: Treinar e Ajustar de Modelo de Machine Learning para Predição de Ploidia	82
5.1.3	OE3: Realizar Avaliação do Modelo	83
5.1.4	OE4: Construir Protótipo	83
5.2	Análise Crítica das Correlações Identificadas	83
5.2.1	Correlação entre Idade Materna e Euploidia	83
5.2.2	Correlação entre tB-tSB e Euploidia	83
5.2.3	Correlação entre t5 e Euploidia	84
5.2.4	Outras Correlações Significativas	84
5.2.5	Correlações fracas ou inexistentes	84
5.3	Desempenho do Modelo e Comparação com o Estado da Arte	85
5.3.1	Comparação com o Estado da Arte	85
5.3.2	Originalidade e Valor Clínico	86
5.3.3	Importância do LIME para Interpretabilidade	86
5.3.4	Exemplo real de uso do LIME na saúde	87
5.4	Implicações Práticas e Sociais	88

5.5	Limitações e Ameaças à Validade do Estudo	88
6	CONCLUSÕES E TRABALHOS FUTUROS	90
6.1	Conclusões Finais	90
6.2	Trabalhos Futuros	91
	REFERÊNCIAS	93
	GLOSSÁRIO	100
	APÊNDICES	102
	APÊNDICE A – ACESSO AO PROJETO	103
	APÊNDICE B – VARIÁVEIS UTILIZADAS NA ANÁLISE DA PLOI- DIA EMBRIONÁRIA	104
	APÊNDICE C – Z-SCORE NORMALIZATION (STANDARDIZA- TION)	107
	APÊNDICE D – MONTE CARLO	109
	APÊNDICE E – COEFICIENTE DE SPEARMAN	110
	ANEXOS	112
	ANEXO I – PARECER DO COMITÊ DE ÉTICA EM PESQUISA	113
	ANEXO II – TERMO DE CONSENTIMENTO PARA UTILIZA- ÇÃO DE DADOS DE ENTREVISTAS, GRAVAÇÃO DE REUNIÕES E USO DE GRAVAÇÃO	121
	ANEXO III – CONTRATO DE AUTORIZAÇÃO PARA UTILIZA- ÇÃO DE DADOS EM PESQUISA	126
	ANEXO IV – CORRELAÇÕES ENTRE OS PARÂMETROS	131

1 Introdução

1.1 Contexto

A fertilização in vitro (FIV) é uma das técnicas mais importantes de reprodução assistida e tem ganhado crescente relevância no campo da medicina reprodutiva, oferecendo novas possibilidades para pessoas com dificuldades para engravidar. Maria Chaves Jardim destaca que o Brasil lidera a América Latina em número de procedimentos de fertilização in vitro, evidenciando a relevância e o avanço dessa tecnologia no país ([JARDIM, 2022](#)).

O sucesso da FIV está diretamente relacionado à saúde genética dos embriões utilizados. De acordo com [Ping et al. \(2023\)](#), a qualidade genética dos embriões é um fator crucial para o êxito da fertilização, pois alterações cromossômicas estão frequentemente associadas a desfechos desfavoráveis, como abortos espontâneos ou doenças genéticas nos fetos. Cerca de 50% dos casos de abortos espontâneos no primeiro trimestre estão relacionados a alterações cromossômicas ([SILVA et al., 2023](#)). Este cenário destaca a importância de métodos de avaliação genética, que se tornam ferramentas essenciais para aumentar as taxas de sucesso da FIV e melhorar a segurança dos tratamentos.

Para avaliar a qualidade dos embriões, são comumente utilizados testes genéticos de triagem embrionária, sendo o Teste Genético Pré-Implantação (PGT) um dos mais conhecidos e amplamente aplicados, especialmente o Teste Genético Pré-Implantação de Aneuploidia (PGT-A) ([YANG et al., 2024](#)). No entanto, a realização do PGT-A apresenta alguns desafios e limitações, como o tempo necessário para obtenção dos resultados, o custo elevado e a complexidade e o risco do procedimento, que envolve a biópsia embrionária (extração de células do trofotoderma) e, por isso, pode causar dano ao embrião ([YANG et al., 2024](#)). Assim, surge a oportunidade de se desenvolver um método de avaliação embrionária mais rápido, menos custoso e menos invasivo.

Um exemplo de tecnologia que pode unir Tecnologia de Reprodução Assistida (TRA) e Inteligência Artificial (IA) é a fotografia time-lapse incorporada às incubadoras de última geração, que geram informações sobre a morfologia e a cinética do desenvolvimento embrionário, facilitando a observação de eventos dinâmicos, seus tempos e padrões, definindo-os, em conjunto, como variáveis morfocinéticas ([MESEGUER et al., 2011](#)). Postula-se que, com o maior conhecimento da dinâmica embrionária em cultivo, a identificação de marcadores do potencial de implantação poderá fornecer, em futuro próximo, informações cruciais para o processo de escolha do embrião a ser transferido para o útero materno ([LUONG; LE, 2023](#)). Entretanto, até aqui, pouco se evoluiu no

que diz respeito à aplicação dos dados ou à análise de variáveis múltiplas sobre resultados clínicos objetivos.

A proposta deste projeto é desenvolver uma abordagem de baixo custo, utilizando IA para a detecção de padrões genéticos em embriões. Mais especificamente, o objetivo é identificar a probabilidade de um embrião ser euploide, isto é, aquele que possui a quantidade correta de cromossomos (23 pares), ou seja, os cromossomos estão organizados em pares completos (ZEGERS-HOCHSCHILD et al., 2017). Ao realizar a análise dos dados gerados pelo TLS, a IA poderá oferecer uma abordagem menos invasiva, eliminando a necessidade de testes genéticos caros e invasivos, como o PGT-A. Assim, ao melhorar a seleção dos embriões com maior potencial de euploidia, a IA poderá tornar o tratamento mais acessível para um maior número de pacientes e, no futuro, aumentar as taxas de sucesso da FIV.

Para as clínicas de fertilização, a implementação dessa tecnologia pode resultar em um fluxo de trabalho mais ágil e eficiente, otimizando recursos e melhorando a satisfação dos pacientes. A integração da IA na avaliação embrionária tem o potencial de transformar os tratamentos de fertilidade, proporcionando resultados mais positivos e uma experiência aprimorada para todos os envolvidos. Assim, este estudo busca ser uma solução menos invasiva dentre os métodos de seleção embrionária em tratamentos de reprodução assistida, integrando as inovações da IA à medicina reprodutiva.

1.2 Motivação

A motivação para a elaboração deste Trabalho de Conclusão de Curso surge da necessidade de contruibir para a criação de métodos mais acessíveis na medicina reprodutiva, especialmente no que diz respeito à avaliação da qualidade genética dos embriões em tratamentos de fertilidade. Essa área, que lida com questões pessoais e delicadas, fez avanços notáveis na última década (PANDIT; SHARMA, 2022), mas ainda enfrenta desafios significativos que exigem soluções eficazes. O avanço das tecnologias de reprodução assistida possibilitou que muitas pessoas, antes incapazes de conceber, realizassem o sonho da gravidez. No entanto, essa evolução trouxe consigo um novo desafio: a falha na implantação e a gravidez não viável. A repetição de ciclos sem sucesso pode gerar profunda frustração e desespero, fazendo com que os casais busquem incessantemente respostas e soluções (MONTAGNINI et al., 2010).

A jornada para a maternidade através da FIV é marcada por uma intensa carga emocional, com expectativas e incertezas. A perda gestacional, além do sofrimento emocional, pode gerar um impacto significativo na saúde física e mental da mulher (MONTAGNINI et al., 2010). De acordo com o estudo de Montagnini et al. (2010), as mulheres apresentam, em comparação aos homens, níveis mais elevados de ansiedade e depressão,

além de uma autoestima mais baixa, com sentimentos de culpa e vergonha relacionados à infertilidade. A ansiedade, em particular, é um dos principais desafios enfrentados pelos casais, frequentemente ultrapassando níveis considerados normais (MONTAGNINI et al., 2010).

A escolha cuidadosa dos embriões para a transferência é um passo crucial nesse processo de tentativa de adicionar um membro à família, pois a qualidade e a saúde dos embriões influenciam diretamente a taxa de sucesso da gravidez (YANG et al., 2024). O impacto da seleção de embriões com defeitos genéticos pode ser devastador para o emocional dos casais, agravando o sofrimento causado por abortos espontâneos e falhas de implantação. Ao eliminar a necessidade de intervenções, como o exame PGT-A, que podem gerar ansiedade e desconforto, podemos proporcionar um ambiente mais acolhedor e propício à realização do sonho de ter filhos.

Portanto, a relevância deste projeto no contexto atual da medicina reprodutiva é indiscutível. Ao abordar as limitações dos métodos tradicionais e explorar as potencialidades da IA, este estudo permite que os médicos tomem decisões mais informadas, com base em análises detalhadas, resultando em melhores resultados clínicos. Além disso, este trabalho destaca a importância da Engenharia de Software como uma ferramenta essencial para impulsionar o desenvolvimento de outras áreas do conhecimento. Ao aplicar técnicas avançadas de análise de dados e algoritmos de Machine Learning, ele contribui diretamente para o crescimento da medicina reprodutiva, promovendo inovações que tornam os tratamentos mais precisos, acessíveis e eficientes. A integração de abordagens interdisciplinares e a personalização do tratamento têm o potencial de gerar avanços significativos em médio e longo prazo, além de reduzir o tempo necessário para se obter um nascimento saudável. Assim, este projeto não apenas beneficia a área médica, mas também reforça o papel da Engenharia de Software como transformadora em diferentes campos científicos.

1.3 Problema

A seleção de embriões euploides em procedimentos de fertilização in vitro (FIV) é determinante para o sucesso do tratamento. No entanto, esse processo ainda depende de técnicas invasivas, como o Teste Genético Pré-Implantacional de Aneuploidia, um procedimento relativamente complexo, pois requer uma biópsia do embrião, durante a qual deve ser garantido o mínimo de danos ao embrião (YANG et al., 2024).

Dessa forma, o objetivo deste trabalho é responder à seguinte pergunta: "Como utilizar a inteligência artificial para identificar padrões em dados morfocinéticos de embriões, obtidos por meio do Time-Lapse System, para prever a porcentagem de euploidia, oferecendo uma solução mais eficaz e menos invasiva do que o PGT-A?"

1.4 Objetivos

1.4.1 Objetivos Gerais

Desenvolver uma abordagem baseada em inteligência artificial para identificar padrões em dados morfocinéticos de embriões, obtidos por meio do Time-Lapse System, capaz de prever a porcentagem de euploidia, proporcionando uma solução mais eficaz e menos invasiva em comparação ao PGT-A.

1.4.2 Objetivos Específicos

- **OE1:** Expandir, Processar e Analisar os Dados para Predição de Ploidia
- **OE2:** Treinar e Ajustar o Modelo de Machine Learning para Predição de Ploidia.
- **OE3:** Realizar Avaliação do Modelo
- **OE4:** Construir Protótipo

1.5 Metodologia

A metodologia do projeto está dividida em 2 fases, onde cada uma visa resolver um Objetivo Específico (OE) do projeto, que contém suas respectivas atividades para serem alcançados:

- **Fase 1: Análise e Preparação de Dados**
 - **OE1 - Expandir, Processar e Analisar os Dados para Predição de Ploidia:**
 - * **Atividade 1 (A1):** Análise, Revisão, Seleção e Limpeza de Variáveis para Predição de Euploidia
 - * **Atividade 2 (A2):** Normalização dos Dados para Otimização.
 - * **Atividade 3 (A3):** Identificação da Correlação e Atribuição de Pesos aos Parâmetros na Previsão da Ploidia do Embrião
 - * **Atividade 4 (A4):** Divisão dos Dados e aplicação de Data Augmentation
- **Fase 2: Desenvolvimento e Avaliação do Modelo**
 - **OE2 - Treinar e Ajustar o Modelo de Machine Learning para Predição de Ploidia:**
 - * **Atividade 5 (A5):** Desenvolvimento e Treinamento do Modelo de Machine Learning para Otimização da Predição de Ploidia

- **OE3 - Realizar Avaliação do Modelo:**
 - * **Atividade 6 (A6):** Utilizar métricas adequadas para medir o desempenho do modelo
 - * **Atividade 7 (A7):** Avaliação do Desempenho do Modelo na Predição por meio da Matriz de Confusão e Curva ROC
- **OE4 - Construir Protótipo:**
 - * **Atividade 8 (A8):** Prototipar uma interface básica para exibir as predições de euploidia para o usuário final (médicos)
 - * **Atividade 9 (A9):** Desenvolver a aplicação web integrando o protótipo à inteligência artificial existente

1.6 Composição e estrutura do trabalho

Este trabalho foi organizado da seguinte maneira:

Capítulo 2, intitulado "Fundamentação Teórica", apresenta os principais conceitos que fundamentam a contextualização deste estudo.

Capítulo 3, intitulado "Metodologia", descreve os procedimentos e métodos utilizados na pesquisa, incluindo o planejamento de trabalho, as atividades realizadas e os resultados esperados. A seção 1.5 será detalhada ao longo deste capítulo.

Capítulo 4, intitulado "Execução da Pesquisa e Análise dos Resultados", apresenta a previsão de término de cada atividade e fase propostas ao longo do trabalho.

Capítulo 5, intitulado "Conclusões e Trabalhos Futuros", expõe os avanços e resultados obtidos durante o período de desenvolvimento deste TCC.

Capítulo 6, intitulado "Planejamento", apresenta todas as atividades realizadas e as atividades futuras a serem realizadas.

2 Fundamentação Teórica

Para dar início a fundamentação teórica de forma concisa e alinhada ao que foi apresentado no Capítulo 1, é fundamental abordar o contexto e o papel da Fertilização In Vitro, a qualidade dos embriões e os principais desafios envolvidos nesse processo. Além disso, é essencial evidenciar como técnicas de Aprendizado de Máquina, com destaque para as Redes Neurais Artificiais, podem contribuir para o aprimoramento das taxas de sucesso da gravidez, a mitigação de riscos relacionados a abortos e alterações cromossômicas, bem como a redução dos impactos emocionais e físicos enfrentados por pacientes submetidas a esse tipo de tratamento.

2.1 Fertilização In Vitro

A fertilidade tem sido um tema de grande relevância ao longo da história humana, visto como uma bênção divina em diversas culturas. Civilizações antigas, como a grega e a egípcia, realizavam rituais, usavam amuletos e talismãs, ou buscavam ajuda de divindades para garantir a continuidade de suas linhagens e prosperidade (MOURA; SOUZA; SCHEFFER, 2020). Esses métodos, embora enraizados em crenças religiosas e espirituais, refletem o desejo universal de superar desafios relacionados à reprodução.

Com os avanços científicos e médicos, a compreensão da fertilidade passou por uma profunda transformação. O primeiro marco da inseminação artificial em humanos ocorreu em 1978, com o nascimento de Louise Brown, o primeiro "bebê de proveta" (MOURA; SOUZA; SCHEFFER, 2020). Este feito foi possível graças ao desenvolvimento da técnica de Fertilização In Vitro (FIV), criada pelo embriologista Robert Edwards e pelo ginecologista Patrick Steptoe. A técnica permitiu a fertilização de embriões fora do corpo humano e, por sua contribuição revolucionária, Edwards recebeu o Prêmio Nobel de Fisiologia ou Medicina em 2010 (CORLETA, 2010).

As Técnicas de Reprodução Assistida (TRA) compreendem um conjunto de métodos médicos especializados que buscam ajudar indivíduos com dificuldades reprodutivas a alcançarem a concepção (SOUZA, 2024). Dentre essas técnicas, a FIV se destaca como a mais avançada e amplamente utilizada (MOURA; SOUZA; SCHEFFER, 2020). O processo envolve várias etapas, como a estimulação ovariana controlada (uso de medicamentos para estimular a produção de óvulos), coleta de óvulos por punção transvaginal, fertilização em laboratório e posterior transferência dos embriões formados para o útero (MOURA; SOUZA; SCHEFFER, 2020).

Além de estabelecer as bases da medicina reprodutiva moderna, o nascimento

de Louise Brown abriu caminho para avanços na medicina reprodutiva. Desde então, a combinação de avanços médicos e tecnológicos permitiu não apenas a realização da fertilização in vitro, mas também a análise genética detalhada dos embriões (MOURA; SOUZA; SCHEFFER, 2020). Esses exames identificam anomalias cromossômicas e genéticas, proporcionando uma maior chance de sucesso na implantação e no desenvolvimento de gestações saudáveis.

Os esforços históricos e as inovações científicas ilustram a busca contínua da humanidade por soluções eficazes contra a infertilidade. No próximo tópico, discutiremos detalhadamente a análise de embriões, um procedimento crucial para aumentar as chances de concepção saudável e seu papel na e seu papel na seleção de embriões para a FIV.

2.2 Métodos de Avaliação Genética em Reprodução Assistida

No contexto das Tecnologias de Reprodução Assistida, os métodos de avaliação genética desempenham um papel essencial na identificação de anomalias cromossômicas, como a aneuploidia, que é a alteração no número normal de cromossomos da espécie humana, representa a principal causa de falhas de implantação quando sua origem é embrionária. Dessa forma, a presença de aneuploidia nas células do embrião pode impactar diretamente a taxa de sucesso das técnicas de reprodução humana assistida. Além de dificultar a implantação, essa alteração cromossômica pode levar a abortos espontâneos ou até mesmo a malformações em bebês nascidos vivos (SOUZA, 2022b). Após décadas de avanços científicos, os testes genéticos tornaram-se cada vez mais precisos, com o desenvolvimento de técnicas sofisticadas e integradas à tecnologia, como a Testagem Genética Pré-implantacional.

Pelo final do século XX, o método do PGT começou a ser usado para realizar o rastreamento de doenças genéticas que possuíam uma alta taxa de incidência nas populações de amostra (YANG et al., 2024). Com ele, foi propiciada a triagem de embriões antes da implantação, permitindo a seleção de embriões que possuíam menos riscos (YANG et al., 2024). Tais métodos incluem a Testagem Genética Pré-implantacional para Doenças Monogênicas (PGT-M), como distrofia miotônica e fibrose cística; para rearranjos estruturais cromossômicos (PGT-SR); para aneuploidias (PGT-A), como Síndrome de Down e Síndrome de Turner; e mais recentemente, o PGT-P para doenças poligênicas (YANG et al., 2024). No caso do nosso objeto de estudo, PGT-A, se destaca por sua capacidade de aumentar as chances de implantação embrionária bem-sucedida, reduzir a probabilidade de perdas gestacionais espontâneas e garantir maior probabilidade de nascimento de crianças com o número de cromossomos normais (YANG et al., 2024). Além disso, o embrião considerado “normal” (euploide) apresenta 23 pares de cromossomos (46 cromossomos no total), sendo metade proveniente do espermatozoide e metade do óvulo. Já

o embrião “anormal” (aneuploide) possui uma contagem incorreta de cromossomos em uma célula, sendo a maioria dos embriões com aneuploidias não compatíveis com a vida (ZEGERS-HOCHSCHILD et al., 2017).

Os testes genéticos realizados durante a fase de blastocisto, formados cerca de 5–6 dias após a inseminação (ZEGERS-HOCHSCHILD et al., 2017), são geralmente preferidos por especialistas em reprodução assistida porque essa abordagem envolve a coleta de células do trofotoderma, a camada externa do embrião que futuramente dará origem à placenta e não diretamente do interior do embrião (LEAVER; WELLS, 2019). Essa técnica, o PGT-A, é considerada menos invasiva em comparação com as biópsias realizadas na fase de clivagem, quando o embrião possui apenas algumas células e está em um estágio muito inicial de desenvolvimento (LEAVER; WELLS, 2019). Na fase de clivagem, a retirada de células do interior do embrião (blastômeros), pode causar danos mais significativos ao embrião, afetando sua capacidade de se desenvolver adequadamente e reduzindo as chances de implantação bem-sucedida no útero (LEAVER; WELLS, 2019).

No entanto, é fundamental mencionar que a técnica de micromanipulação necessária para realizar esses procedimentos ainda não foi completamente padronizada. Isso significa que diferentes laboratórios podem adotar práticas distintas para a biópsia embrionária, o que pode resultar em variações na segurança e eficácia dos testes. Além disso, os impactos potenciais dessa intervenção, tanto nos desfechos reprodutivos (como taxas de gravidez e nascimento) quanto na saúde a longo prazo dos bebês nascidos a partir desses embriões, ainda não estão completamente compreendidos.

O PGT-A classifica os embriões em três categorias: euploide (normal, com 46 cromossomos), mosaico (mistura de células normais e anormais) e aneuploide (todas as células com número anormal de cromossomos).

Resultados Possíveis do PGT-A	Euplóide	Mosaico	Aneuplóide
Número de cromossomos	Normal	Misto: Alguns normais e outros anormais	Anormal
Probabilidade de produzir uma gravidez bem-sucedida	Alta	Intermediária	Baixa

Figura 1 – Resultados do PGT-A e suas respectivas probabilidades de sucesso gestacional

Apesar de suas vantagens, o PGT-A apresenta limitações significativas, como a variabilidade nos resultados das biópsias do trofotoderma e o risco de diagnósticos falso-positivos ([GLEICHER; PATRIZIO; BRIVANLOU, 2021](#)). A The Preimplantation Genetic Diagnosis International Society (PGDIS) e a European Society of Human Reproduction and Embryology (ESHRE) apontam questões críticas, incluindo:

- As divergências no conteúdo de DNA aneuploide entre diferentes regiões da trofoectoderma e a massa celular interna demonstram que a biópsia de cinco células pode apresentar resultados variados;
- O número exato de células na biópsia nunca é conhecido, o que impossibilita determinar com precisão a porcentagem de DNA aneuploide;
- A biópsia da trofoectoderma danifica células individuais, causando vazamento de DNA e contaminação das células vizinhas, dificultando a medição precisa da aneuploidia;
- O limiar de 20% entre euploidia e mosaicismismo é baseado apenas na sensibilidade atual do sequenciamento de nova geração (NGS), que não detecta níveis inferiores a 20% de DNA aneuploide. Consequentemente, qualquer mosaicismismo abaixo de 20% é considerado euploide normal;
- Dentro da faixa de mosaicismismo (20% a 80%), os desfechos de implantação e nascimento são semelhantes, indicando que o uso de limiares rígidos para prever resultados de FIV é incorreto.

Estudos indicam que embriões descartados como aneuploides pelo PGT-A resultaram em nascimentos normais ([GLEICHER; PATRIZIO; BRIVANLOU, 2021](#)), evidenciando a necessidade de revisar as diretrizes para evitar o desperdício de embriões viáveis. A figura abaixo ilustra a seleção de um pedaço do embrião e como isso pode influenciar a definição da ploidia do mesmo. A biópsia mencionada nas etapas do PGT-A, envolvendo a remoção de uma célula do embrião para análise genética. [Phillips et al. \(2024\)](#), levantam a preocupação de que a remoção dessas células em crescimento possa comprometer o desenvolvimento do embrião, afetando os resultados neonatais, visto que as técnicas de micromanipulação utilizadas na biópsia não são totalmente isentas de riscos, como observado por [Leaver e Wells \(2019\)](#).

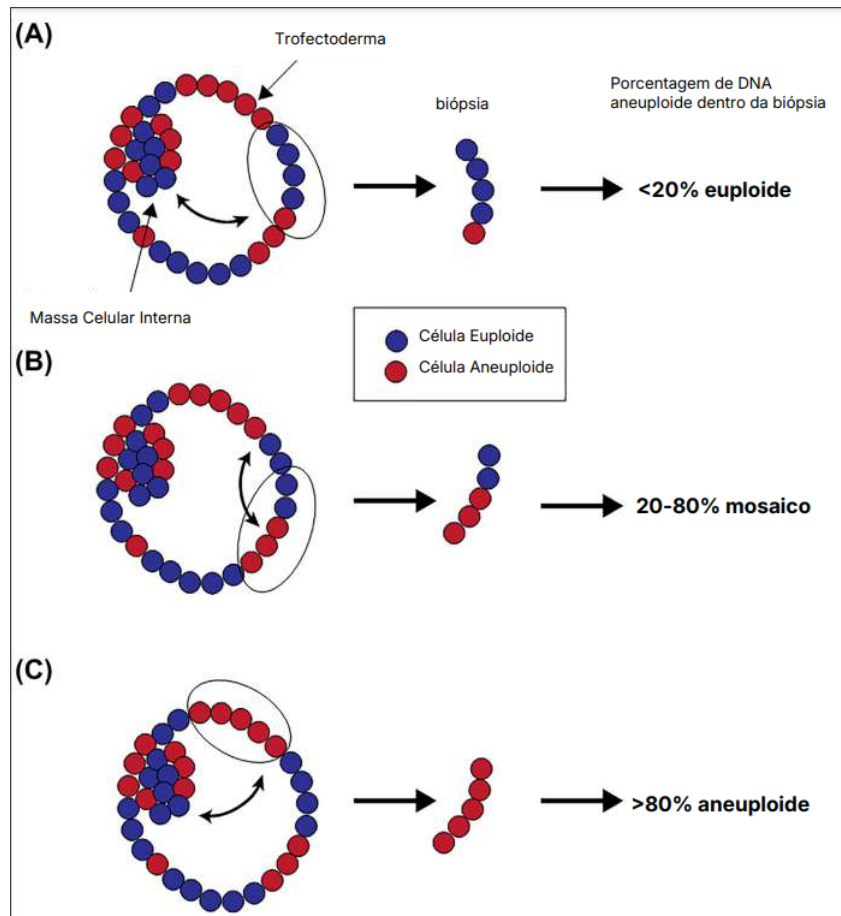


Figura 2 – Discrepância potencial entre uma biópsia de células utilizando PGT-A e a composição celular em regiões adjacentes do trofotoderma

Embora o PGT-A permita detectar aneuploidias e aumente as chances de uma gravidez bem-sucedida, ele pode impactar negativamente o potencial de implantação do embrião (GLEICHER; PATRIZIO; BRIVANLOU, 2021). Por esses motivos, métodos não invasivos estão sendo estudados como alternativas eficientes e seguras, tornando-se cada vez mais relevantes. Um exemplo de técnica não invasiva é a análise morfocinética a partir de imagens obtidas de incubadoras de última geração equipadas com a tecnologia Time-Lapse System.

2.3 Time-Lapse System

O TLS é um sistema que captura imagens contínuas dos embriões em desenvolvimento, em intervalos regulares, sem alterar o ambiente de cultivo (MOUSTAKLI et al., 2024). Essa análise morfocinética, gerada pelas imagens adquiridas pelo TLS, permite o monitoramento quase contínuo do desenvolvimento do embrião, possibilitando a observação de eventos dinâmicos e frequentemente transitórios que não seriam visíveis em observações estáticas (BOUCRET et al., 2021). O uso do TLS não interrompe as condi-

ções de cultura, mantendo a viabilidade do embrião durante o processo de monitoramento (MOUSTAKLI et al., 2024).

As variáveis morfocinéticas incluem aspectos como a forma e a estrutura do embrião (morfológicas) e o movimento e o desenvolvimento do embrião ao longo do tempo (cinéticas), os quais são essenciais para uma análise detalhada de seu progresso (GLEICHER; PATRIZIO; BRIVANLOU, 2021). Com esse monitoramento contínuo, é possível observar a regularidade das divisões celulares e identificar momentos críticos do crescimento, o que pode auxiliar na diferenciação de embriões euplóides e aneuplóides com base no seu padrão de desenvolvimento (BOUCRET et al., 2021). De acordo com Moustakli et al. (2024), o TLS oferece insights valiosos sobre a saúde e o potencial de desenvolvimento dos embriões, utilizando uma abordagem não invasiva, em contraste com a biópsia de embriões. Alguns estudos indicam que, ao ser combinado com pontuações morfocinéticas, o TLS pode aumentar as taxas de implantação e gravidez clínica em comparação aos métodos tradicionais (BOUCRET et al., 2021).

As variáveis da planilha de dados, dados esses extraídos pelo TLS, que estamos utilizando para esse trabalho incluem informações sobre a qualidade morfológica e cinética dos embriões, como a taxa de divisão celular e a regularidade do desenvolvimento. Essas variáveis são essenciais para a análise morfocinética e a classificação dos embriões em diferentes estágios de desenvolvimento (BOUCRET et al., 2021). A seguir, discutiremos as variáveis cientificamente comprovadas que influenciam a ploidia.

2.3.1 Idade

As informações trazidas pelo Fertility and Ageing (Baird et al. (2005)) corroboram a relevância de incluir variáveis relacionadas à **idade materna**, pois estudos apontam que o aumento da aneuploidia em embriões está diretamente associado ao envelhecimento materno. O estudo de Yuan et al. (2023) aponta que a taxa de euploidia dos embriões está correlacionada com a idade feminina. À medida que a idade avança, há um declínio no número total e na qualidade dos ovócitos, um fator crítico para a fecundidade reduzida observada após os 35 anos (YUAN et al., 2023).

2.3.2 t_2 , t_3 , t_4 , t_5 , t_8 , s_2 , cc_2 (t_3-t_2), t_{SC} , t_{SB} , t_B , cc_3 (t_5-t_3), s_3 (t_8-t_5), t_5-t_2 , $t_{SC}-t_8$ e t_B-t_{SB}

Os **sistemas de Time-Lapse** ajudam a identificar **marcadores morfocinéticos**, que mostram como as células se dividem durante o desenvolvimento do embrião. Esses marcadores, junto com características físicas tradicionais, são fundamentais para selecionar o embrião mais adequado para a transferência (SOUZA, 2022a). O desenvolvimento embrionário é um processo dinâmico, com mudanças perceptíveis em um curto

período (CRUZ et al., 2012). Estudos detalhados sobre o ritmo das divisões celulares, assim como características como tamanho e organização das células, demonstram que o tempo necessário para atingir certos estágios de desenvolvimento está diretamente relacionado ao potencial de implantação do embrião (SOUZA, 2022a).

Embriões que se dividem muito rapidamente apresentam menor chance de implantação quando comparados aos que seguem um ciclo celular dentro do intervalo considerado normal (CRUZ et al., 2012). Isso acontece porque mudanças no tempo que o embrião leva para se desenvolver nas primeiras fases estão ligadas a uma maior chance de problemas genéticos no número de cromossomos (CRUZ et al., 2012). O projeto *Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality* de Cruz et al. (2012) utilizou o sistema Time-Lapse para identificar marcadores morfocinéticos e monitorar com precisão os tempos das divisões celulares durante o desenvolvimento embrionário. Os tempos considerados ótimos para previsões de desenvolvimento embrionário foram: **t2 (24,3–27,9 horas)**, **t3 (35,4–40,3 horas)**, **t5 (48,8–56,6 horas)**, **s2 (<0,76 horas)** e **cc2 (<11,9 horas)**. A explicação detalhada dessas variáveis está disponível no **Apêndice B**.

Especificamente, no nível morfológico, **t5** destacou-se como o indicador mais relevante do potencial de implantação (CRUZ et al., 2012). Observa-se que a capacidade de diferenciar embriões viáveis daqueles não viáveis melhora significativamente quando os critérios se baseiam em eventos de divisão celular mais tardios (CRUZ et al., 2012). Embriões com **t5** entre 48,8 e 56,6 horas demonstram não apenas um maior potencial de implantação, mas também uma maior propensão a se desenvolverem em blastocistos de morfologia superior (CRUZ et al., 2012).

Ao criar um modelo de IA para a previsão de euploidia, é importante que considere as variáveis **t4**, **t8**, **tSC**, **tSB**, **tB**, **cc3 (t5 - t3)** e **s3 (t8 - t5)**. Essas métricas são fundamentais para o desenvolvimento embrionário, conforme evidenciado no estudo que comprovou a efetividade da IA em detectar embriões viáveis com uma precisão de 70% (RIENZI et al., 2020). A incorporação dessas variáveis em um modelo de IA é crucial, pois possibilita a captura das sutilezas do desenvolvimento embrionário, que, de acordo com a pesquisa, são melhoradas através da análise automatizada fundamentada em Inteligência Artificial. Isso se torna especialmente relevante pois os sistemas de time-lapse disponibilizam informações detalhadas e contínuas, que podem ser combinadas para detectar padrões relacionados à euploidia (RIENZI et al., 2020).

A avaliação de fatores morfocinéticos, como o tempo necessário para clivagem e a extensão das fases subsequentes, tem sido alvo de pesquisa para antecipar a probabilidade de implantação embrionária. Por exemplo, um estudo publicado pelo Instituto Sapientiae por Desai et al. (2019) destacou a importância desses parâmetros na predição do potencial de implantação embrionária. Apesar dessa pesquisa não tratar especificamente os

intervalos **t5-t2**, **tSC-t8** e **tB-tSB**, ela sugere que a avaliação de intervalos de tempo entre eventos específicos no desenvolvimento embrionário pode oferecer percepções valiosas sobre a qualidade e a capacidade de desenvolvimento dos embriões (DESAI et al., 2019).

2.3.3 Estágio e Morfo

Modelos de IA têm sido empregados na avaliação de embriões produzidos até o quinto dia ou mais, com o objetivo de aprimorar a escolha com base em informações objetivas e de alta exatidão (LASSEN et al., 2022). Esses modelos priorizam a análise dos estágios "**Dia 5+**", ou seja, aqueles que atingem o estágio de blastocisto no quinto dia ou mais tarde, devido à maior disponibilidade de dados morfológicos e dinâmicos do desenvolvimento embrionário (LASSEN et al., 2022).

De acordo com os critérios definidos por Gardner (1999), com base na **morfologia do embrião** se tem a categorização dos blastocistos, um fator determinante para o potencial de implantação e a qualidade embrionária (CAPALBO et al., 2014). Os blastocistos são agrupados em quatro categorias principais, considerando tanto a massa celular interna (ICM, Inner Cell Mass) quanto o TE (trophectoderma):

- **Grupo 1 (Excelente):** Blastocistos com classificação $\geq 3AA$. Blastocistos altamente desenvolvidos com massa celular interna densa e trophectoderma bem organizado (CAPALBO et al., 2014).
- **Grupo 2 (Bom):** Blastocistos com classificação **3, 4, 5 ou 6** e com notas **AB ou BA**. Apresentam características boas, mas menos consistentes em relação ao grupo excelente (CAPALBO et al., 2014).
- **Grupo 3 (Médio):** Blastocistos com classificação **3, 4, 5 ou 6** e notas **BB, AC ou CA**. Qualidade moderada com irregularidades tanto na ICM quanto no TE (CAPALBO et al., 2014).
- **Grupo 4 (Ruim):** Blastocistos com classificação $\leq 3BB$. Blastocistos de menor qualidade, com poucas células organizadas na ICM e TE menos coeso (CAPALBO et al., 2014).

O estudo de Capalbo et al. (2014) enfatizou a relação entre a morfologia padrão dos blastocistos, a euploidia e as taxas de implantação. Blastocistos de excelente morfologia, particularmente os biopsiados no dia 5, mostraram uma maior probabilidade de serem euploides e apresentaram taxas de implantação superiores.

2.3.4 KIDScoreTM

O KIDScoreTM, um algoritmo baseado em IA aplicado à análise de imagens em sistemas Time-Lapse, tem se mostrado uma ferramenta importante na avaliação de embriões durante os tratamentos de reprodução assistida (KATO et al., 2021). O algoritmo combina variáveis morfocinéticas e parâmetros de desenvolvimento embrionário para fornecer uma pontuação que auxilia na seleção de embriões com maior potencial de implantação e viabilidade genética (GAZZO et al., 2020). A pontuação vai de **0 a 10**. Pontuações baixas, entre 0 e 3, indicam embriões de qualidade inferior, com baixo potencial de implantação. Pontuações médias, de 4 a 6, correspondem a embriões de qualidade moderada, com um potencial razoável de implantação. Já pontuações altas, de 7 a 10, representam embriões de alta qualidade, com grande potencial de implantação (GAZZO et al., 2020). Kato et al. (2021) cita que o modelo apresentou uma alta precisão na previsão de resultados de gravidez, sendo especialmente útil tanto em pacientes com idade materna avançada quanto em pacientes mais jovens. Por fim, Gazzo et al. (2020) informaram que o uso do algoritmo no processo de seleção embrionária levou a um aumento expressivo nas taxas de implantação após a transferência de embriões congelados (FET). Então o modelo quando combinada com informações sobre os tempos de divisão celular, sendo **t2**, **t3**, **t5**, **s2** e **cc2** descritos por Cruz et al. (2012), possibilita um exame mais completo do crescimento embrionário, melhorando a acurácia na seleção de embriões com maior probabilidade de êxito em tratamentos de FIV.

2.3.5 Ploidia

Para a elaboração do nosso modelo de previsão de euploidia, optamos por empregar a coluna de **Ploidia**, que proporciona uma categorização minuciosa dos embriões em diversas categorias de euploidia. As categorizações contidas nesta coluna são: Aneuploide complexo, Aneuploide/Triploide XXX, Caótico, Haploide, Mosaico de alto grau, Mosaico de baixo grau e Normal/Euploide. De acordo com o Bastida et al. (2019), considera-se os embriões com Caótico, Haplóide e Mosaico de alto grau como Aneuploides, enquanto os com mosaico de baixo grau como Euploides. Diante disso, optamos por reorganizar os valores na tabela de dados, agrupando as seguintes categorias sob o termo **Aneuploide**: Aneuploide complexo, Aneuploide/Triploide XXX, Caótico, Haploide e Mosaico de alta complexidade. Em contrapartida, os embriões categorizados como Mosaico de baixo grau e Normal/Euploide serão reunidos na categoria **Euploide**.

2.4 Aprendizado de Máquina

As dificuldades em definir IA não são, portanto, o resultado de alguma deficiência ou descuido, mas surgem do fato de que fomos incapazes de determinar precisamente qual

inteligência desejávamos replicar artificialmente (SHEIKH; PRINS, 2023). Dessa forma, definimos a Inteligência Artificial como sistemas que exibem comportamento inteligente ao analisar seu ambiente e tomar ações, com algum grau de autonomia, para atingir objetivos específicos (SHEIKH; PRINS, 2023).

Na medicina reprodutiva, a IA tem se mostrado promissora na melhoria de processos como a fertilização in vitro (FIV). Essa busca por imitar a inteligência humana e entender seus processos cognitivos levou ao desenvolvimento de diversas abordagens, entre elas o Aprendizado de Máquina (Machine Learning, ML), que envolve a capacidade de computadores de interpretar grandes volumes de dados, construir modelos baseados nesses dados e, assim, gerar hipóteses ou previsões sobre o mundo ao seu redor (RUSSELL; NORVIG, 2016). Na medicina, algoritmos podem ser treinados para reconhecer padrões genéticos em embriões, classificar o melhor embrião para implantação e prever características genéticas de novos embriões.

Os métodos de ML são geralmente classificados em três tipos principais: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. Neste estudo, opta-se pelo aprendizado supervisionado como abordagem principal, dada sua eficácia na análise de dados rotulados, permitindo decisões mais precisas e embasadas para otimizar os tratamentos de FIV.

O aprendizado supervisionado consiste no treinamento de algoritmos com base em conjuntos de dados rotulados, nos quais as variáveis de entrada (inputs) e os resultados esperados (outputs) já são conhecidos. O algoritmo aprende a correlacioná-los de forma eficiente (RUSSELL; NORVIG, 2016), ajustando seus parâmetros com base nas diferenças entre previsões e resultados reais (TRASK, 2019). Uma técnica amplamente usada dentro do aprendizado supervisionado é a classificação, cujo objetivo é atribuir rótulos ou classes pré-definidas aos dados. Por exemplo, no contexto da medicina reprodutiva, um modelo pode ser treinado para diferenciar embriões euploides e aneuploides, aprendendo a reconhecer padrões associados a cada grupo (IZBICKI; SANTOS, 2020).

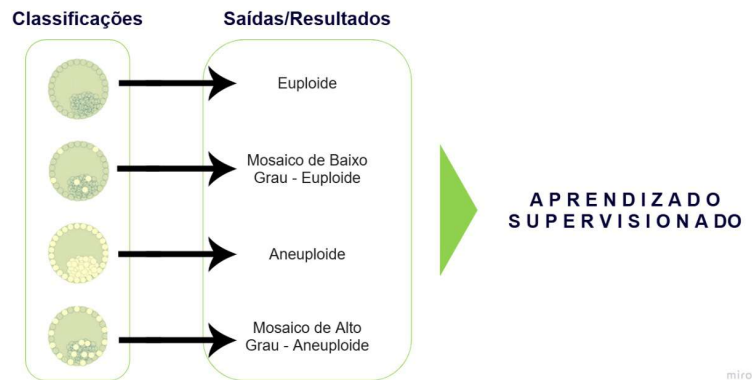


Figura 3 – Aprendizado Supervisionado: Identificando Euploidia e Aneuploidia

De maneira geral, um algoritmo de aprendizado supervisionado separa o banco de dados em três subconjuntos: treinamento, validação e teste (IZBICKI; SANTOS, 2020). Na fase de treinamento, o algoritmo identifica padrões nos dados de entrada e os associa às classes desejadas (IZBICKI; SANTOS, 2020). Na validação, um subconjunto de dados não utilizado no treinamento avalia o desempenho do modelo (IZBICKI; SANTOS, 2020), permitindo ajustes nos hiperparâmetros. Após resultados satisfatórios, o conjunto de testes mensura métricas como acurácia, recall e precisão, garantindo o desempenho esperado (IZBICKI; SANTOS, 2020).

A seleção aleatória das amostras para treinamento, validação e teste é uma boa prática (IZBICKI; SANTOS, 2020), evitando problemas decorrentes de ordenações previamente estabelecidas nos bancos de dados. Isso assegura uma visão representativa e imparcial dos dados, fundamental para a construção de modelos robustos e confiáveis (IZBICKI; SANTOS, 2020).

Para a exploração inicial, utilizaremos as Redes Neurais Artificiais (RNAs) como alternativa metodológica. De acordo com Haykin (2009), as RNAs são estruturas computacionais compostas por unidades de processamento simples, interconectadas, capazes de armazenar conhecimento adquirido por meio da experiência e aplicá-lo em novas situações. Além disso, o autor destaca que as redes neurais se assemelham ao cérebro humano em dois aspectos fundamentais: o conhecimento é adquirido a partir da interação com o ambiente por meio de um processo de aprendizado e esse conhecimento é armazenado nas conexões entre as unidades de processamento, chamadas de pesos sinápticos.

Outro ponto de destaque é a capacidade de generalização das RNAs, ou seja, a habilidade do modelo de fornecer respostas coerentes para dados que não foram apresen-

tados durante o treinamento. Segundo Haykin (2009), essa propriedade permite que as redes neurais sejam aplicadas em problemas de alta complexidade, onde as relações entre as variáveis não são facilmente modeladas por técnicas convencionais, que é o nosso caso. Considerando a complexidade dos dados deste projeto, que incluem informações clínicas, morfológicas e morfocinéticas de embriões humanos, a aplicação de Redes Neurais Artificiais se mostra adequada, oferecendo maior potencial de precisão na predição da saúde genética dos embriões em tratamentos de FIV.

2.4.1 Redes Neurais Artificiais (RNA)

As Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no funcionamento do cérebro humano. Elas são compostas por unidades chamadas neurônios artificiais, que se interconectam para formar estruturas capazes de aprender a partir de dados, identificando relações complexas e não lineares (HAYKIN, 2009).

Segundo Haykin (2009), uma Rede Neural Artificial (RNA) pode ser compreendida como um processador distribuído em larga escala, formado por unidades simples de processamento. Sua principal característica é a capacidade de adquirir conhecimento por meio da experiência e armazená-lo nos pesos sinápticos (valores que conectam os neurônios). Essa estrutura se assemelha ao cérebro humano principalmente por dois aspectos: (1) o aprendizado ocorre com base em experiências e (2) o conhecimento é armazenado nos pesos dessas conexões.

O objetivo central de uma RNA é aproximar uma função desconhecida, ou seja, mapear um conjunto de entradas $x = (x_1, x_2, \dots, x_n)$ a uma saída y , por meio da combinação ponderada dos sinais de entrada processados em múltiplas camadas (HAYKIN, 2009). Por essa razão, as redes neurais são amplamente utilizadas em tarefas como classificação, regressão, reconhecimento de padrões e outros problemas com alto grau de complexidade.

As RNAs geralmente são organizadas em camadas: uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada é composta por neurônios artificiais interligados e os sinais de entrada percorrem essas camadas, sendo processados até gerar o resultado final. A forma como os neurônios estão interconectados e como os pesos são ajustados determina o desempenho do modelo (HAYKIN, 2009). Entre as arquiteturas mais conhecidas, destaca-se o Perceptron de Múltiplas Camadas. Segundo Haykin (2009), essa estrutura foi essencial para o avanço das redes neurais, pois permitiu resolver problemas complexos, como o XOR, que não podem ser resolvidos por modelos lineares simples.

2.4.1.1 Multilayer Perceptron (MLP)

O Multilayer Perceptron (MLP) é uma das arquiteturas mais clássicas de Redes Neurais Artificiais (RNA). Trata-se de uma rede feedforward totalmente conectada, ou seja, o fluxo de informações ocorre sempre em uma única direção - da camada de entrada, passando por camadas ocultas, até a camada de saída, sem ciclos ou conexões de retorno (HAYKIN, 2009).

Segundo Haykin (2009), o MLP é composto por camadas sequenciais:

1. uma **camada de entrada**, formada por n neurônios, onde cada neurônio representa uma variável do problema;
2. uma ou mais **camadas ocultas**, densamente conectadas, com m e p neurônios, que extraem representações cada vez mais abstratas dos dados;
3. uma **camada de saída**, geralmente com um único neurônio, responsável por gerar a predição — neste caso, indicando a probabilidade de o embrião ser euploide ou aneuploide.

Cada neurônio em uma camada recebe entradas da camada anterior, aplica um somatório ponderado seguido de uma função de ativação não linear e transmite o resultado para a próxima camada. O desempenho do modelo depende da quantidade de camadas ocultas, da escolha das funções de ativação e da forma como os pesos são ajustados durante o treinamento (HAYKIN, 2009). A figura abaixo ilustra um MLP típico com uma camada de entrada, duas camadas ocultas e uma camada de saída. As setas representam as conexões entre os neurônios, demonstrando o fluxo dos dados.

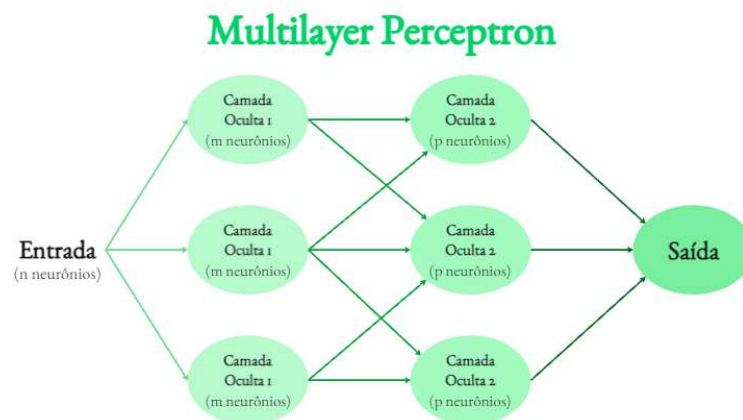


Figura 4 – Estrutura de um MLP com uma camada de entrada, três camadas ocultas e uma camada de saída.

As camadas ocultas desempenham um papel fundamental ao permitir que a rede aprenda funções complexas e não lineares, transformando os dados de forma progressiva e hierárquica (HAYKIN, 2009). No modelo adotado neste trabalho, o MLP será configurado com duas camadas ocultas, contendo 64 e 32 neurônios, respectivamente. Essa configuração busca um equilíbrio entre capacidade de representação e eficiência computacional, permitindo um bom desempenho mesmo com um conjunto de dados reduzido.

2.4.1.2 Características das Camadas Densas

As camadas do MLP são chamadas densas (*fully connected*) porque cada neurônio em uma camada está conectado a todos os neurônios da camada seguinte. Essa estrutura permite que a rede aprenda combinações complexas dos atributos de entrada (HAYKIN, 2009).

O cálculo realizado por um neurônio em uma camada densa pode ser expresso como:

$$z = \sum_{i=1}^n w_i x_i + b \quad (2.1)$$

Onde w_i representam os pesos ajustáveis, x_i as entradas e b o termo de viés bias. Em seguida, aplica-se uma função de ativação z , resultando na saída do neurônio:

$$a = \phi(z) \quad (2.2)$$

2.4.1.3 Inteligência Artificial Explicável (XAI) e o Algoritmo LIME

As Redes Neurais Artificiais (RNAs), especialmente aquelas com múltiplas camadas ocultas, são frequentemente classificadas como modelos de caixa-preta (black-box models). Isso porque, apesar de sua alta precisão e capacidade de generalização, os processos internos que levam às suas decisões são difíceis de interpretar por seres humanos (ADADI; BERRADA, 2018). Isso ocorre devido à sua estrutura altamente complexa, composta por milhares, ou até milhões, de parâmetros ajustáveis, como pesos sinápticos, vieses e funções de ativação, distribuídos ao longo de diversas camadas ocultas. Assim, mesmo que uma RNA indique que um embrião tem alta probabilidade de ser euploide, não é possível saber claramente quais variáveis influenciaram essa decisão (RIBEIRO; SINGH; GUESTRIN, 2016).

Essa falta de transparência é especialmente crítica em áreas sensíveis, como a medicina reprodutiva, onde decisões automatizadas precisam ser não apenas precisas, mas também compreensíveis. Surge, nesse contexto, a Inteligência Artificial Explicável (XAI – Explainable Artificial Intelligence), um campo que busca criar métodos para

explicar como os modelos de aprendizado de máquina tomam suas decisões (GILPIN et al., 2018). A XAI tem como objetivo reduzir o conflito entre desempenho e interpretabilidade, tornando viável o uso seguro e confiável de modelos complexos em contextos que exigem responsabilidade, rastreabilidade e validação ética e técnica (GILPIN et al., 2018).

No caso das RNAs, essa necessidade é ainda maior, pois sua arquitetura não linear impede a extração direta de regras ou justificativas lógicas. Isso levanta uma pergunta central: podemos confiar em modelos cujas decisões não são auditáveis? Na área da saúde, essa falta de explicações pode dificultar a aceitação por parte de médicos, reguladores e pacientes. A confiança em sistemas de IA vai além da acurácia — ela depende de entender os critérios que fundamentam cada predição.

Para lidar com esse desafio, surgiram técnicas de XAI, entre as quais se destaca o LIME (Local Interpretable Model-Agnostic Explanations), desenvolvido por Ribeiro, Singh e Guestrin (2016). O LIME é considerado agnóstico ao modelo, ou seja, pode ser aplicado a qualquer tipo de preditor (como árvores de decisão, ensembles ou redes neurais), sem precisar conhecer sua estrutura interna (RIBEIRO; SINGH; GUESTIN, 2016). O LIME gera explicações locais, focadas em uma predição específica, ao invés de tentar entender todo o comportamento global do modelo. Para isso, ele gera pequenas variações (perturbações) nos dados de entrada e observa como o modelo responde. A partir dessas variações, o LIME ajusta um modelo linear simples (como uma regressão linear com regularização) que imita o comportamento do modelo original apenas na vizinhança daquela entrada específica (RIBEIRO; SINGH; GUESTIN, 2016). Os coeficientes desse modelo indicam a importância de cada variável para aquela decisão.

Neste trabalho, o LIME será usado para explicar as predições do modelo MLP treinado para estimar a probabilidade de um embrião ser euploide. Com isso, será possível identificar a contribuição de cada variável para cada embrião individualmente, tornando o processo mais transparente e confiável. O uso de ferramentas de XAI, como o LIME, aumenta o valor clínico do modelo, permitindo uma compreensão interpretável dos fatores morfocinéticos envolvidos na predição da euploidia.

2.5 Passos para o Desenvolvimento de um Algoritmo de Aprendizado de Máquina

O desenvolvimento de sistemas baseados em aprendizado de máquina é um processo que envolve uma série de etapas bem definidas. Essas etapas vão desde a definição inicial do problema até a implementação e manutenção do sistema em um ambiente de produção. Segundo Géron (2017), em *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, seguir uma abordagem estruturada permite que os desenvolvedores abordem os desafios de forma sistemática, garantindo não apenas a eficácia técnica

do modelo, mas também sua aplicabilidade prática. O processo começa com a definição do problema e o entendimento do contexto geral. Depois, os dados são coletados, explorados para descobrir padrões e preparados para melhorar os resultados dos algoritmos. Em seguida, diferentes modelos são testados, e os melhores são ajustados para formar uma solução final. Essa solução é apresentada, implementada em produção e monitorada continuamente para garantir que funcione bem ao longo do tempo (GÉRON, 2017).

2.5.1 Definição do problema e análise do panorama geral

A definição clara do problema e a análise do panorama geral são etapas essenciais para o sucesso de projetos de aprendizado de máquina, pois orienta todas as decisões subsequentes, desde a coleta de dados até a implementação final. Conforme abordado por Géron (2017) e Müller (2017), a etapa inicial estabelece as bases para decisões estratégicas durante todo o processo.

É necessário, em primeiro lugar, compreender o objetivo principal do projeto, tanto do ponto de vista técnico quanto do ponto de vista de negócio ou pesquisa. Isso inclui identificar como o modelo será utilizado, qual é o tipo de problema (por exemplo, classificação ou regressão) e quais métricas serão mais apropriadas para mensurar seu desempenho (GÉRON, 2017).

Além disso, é importante avaliar as soluções existentes ou alternativas em uso, que podem servir como ponto de comparação para medir o impacto do modelo para servir como referencial de qualidade e impacto. Também é fundamental validar as hipóteses iniciais sobre os dados e a viabilidade do problema, garantindo que ele esteja bem formulado e que eventuais limitações sejam reconhecidas desde o início.

Segundo Müller (2017), uma compreensão profunda dos dados e suas características é essencial para a escolha dos algoritmos e para o sucesso do projeto. Perguntas como “Quantos dados possuo?”, “Há dados faltantes?” e “Esses dados são suficientes para responder às perguntas do projeto?” guiam essa análise preliminar (MÜLLER, 2017).

Por fim, essa etapa também exige atenção ao alinhamento entre o problema técnico e os resultados esperados em termos de negócio ou impacto social. Isso garante que o desenvolvimento não seja apenas tecnicamente sólido, mas também relevante e eficaz em seu contexto de aplicação.

2.5.2 Obtenção de Dados

A etapa de obtenção de dados é um dos pilares fundamentais para o sucesso de um projeto de aprendizado de máquina, pois a qualidade e a relevância das informações coletadas impactam diretamente a eficácia do modelo (GÉRON, 2017). Esse processo tem início com a identificação dos dados necessários, considerando sua quantidade, for-

mato, tipo e origem. Também é importante documentar as fontes utilizadas, garantir que os dados estejam acessíveis de forma eficiente e assegurar o armazenamento adequado (GÉRON, 2017).

Além dos aspectos técnicos, essa etapa requer atenção especial às obrigações legais e éticas. No contexto brasileiro, destaca-se a necessidade de conformidade com a Lei Geral de Proteção de Dados (LGPD), que exige o consentimento para uso de informações pessoais e a adoção de práticas de segurança, como a anonimização de dados sensíveis (MÜLLER, 2017). Segundo Müller (2017), a proteção da privacidade e a integridade das informações são fundamentais, sobretudo quando se trata de dados pessoais ou clínicos.

Para garantir que os dados coletados sejam adequados ao treinamento do modelo, é importante adotar práticas de pré-processamento, como limpeza, normalização e engenharia de atributos. Essas técnicas ajudam a mitigar ruídos, inconsistências e redundâncias nos dados (GÉRON, 2017). Além disso, recomenda-se a separação criteriosa entre os conjuntos de treinamento e teste, de forma a validar corretamente a capacidade de generalização do modelo em dados nunca vistos.

Adotar uma abordagem cuidadosa e ética na obtenção de dados contribui para o desenvolvimento de soluções mais robustas, seguras e confiáveis, promovendo tanto a qualidade técnica quanto a responsabilidade social dos sistemas de aprendizado de máquina (GÉRON, 2017).

2.5.3 Exploração de Dados

A exploração de dados é uma etapa fundamental em projetos de aprendizado de máquina, pois estabelece a base sobre a qual os modelos serão construídos (GÉRON, 2017). Antes de aplicar qualquer algoritmo, é necessário compreender as características do conjunto de dados, iniciando-se com uma cópia dos dados originais e, se necessário, reduzindo sua escala para facilitar a análise preliminar (GÉRON, 2017).

A exploração dos dados deve seguir uma abordagem sistemática que envolve:

1. **Estudo das características dos atributos:** Identificar o nome, o tipo (categórico, numérico, texto, etc.), o percentual de valores ausentes, o nível de ruído (outliers, erros de arredondamento), a utilidade potencial para a tarefa e o tipo de distribuição (gaussiana, uniforme, logarítmica) dos dados disponíveis.
2. **Identificação de atributos-alvo:** No caso de aprendizado supervisionado, determinar qual atributo será o alvo da predição.
3. **Visualização de dados:** Criar gráficos de dispersão, histogramas ou outros métodos visuais para identificar padrões, correlações e tendências. Géron (2017) sugere,

por exemplo, experimentar combinações de atributos, como comparar o número de quartos por domicílio, em vez de analisar apenas o número total de quartos.

4. **Correlação entre atributos:** Analisar as relações entre as variáveis para identificar combinações promissoras que possam melhorar a precisão do modelo.

Müller (2017) destaca que a inspeção visual é crucial para entender a estrutura dos dados e detectar inconsistências, como unidades divergentes ou valores atípicos, comuns em bases reais. Gráficos de dispersão e matrizes de pares são úteis especialmente quando há poucas variáveis, permitindo identificar padrões relevantes. Essa análise também ajuda a verificar se o problema é bem definido e se os dados disponíveis são suficientes para resolvê-lo.

Além disso, Géron (2017) enfatiza a criação de atributos derivados mais informativos, como "população por domicílio", e a importância de documentar todas as descobertas. Caso necessário, essa etapa também pode indicar a necessidade de obter dados adicionais. Assim, a exploração de dados contribui diretamente para o desenvolvimento de modelos mais robustos e eficazes.

2.5.4 Preparação dos dados para os Algoritmos de Aprendizado de Máquina

A preparação dos dados é uma etapa essencial no pipeline de aprendizado de máquina, composta por transformações que tornam os dados adequados para uso pelos algoritmos (GÉRON, 2017). É recomendável automatizar essas transformações por meio de funções reutilizáveis, facilitando sua aplicação em diferentes conjuntos e projetos futuros.

Uma das primeiras ações é a limpeza dos dados, especialmente no tratamento de valores ausentes. De acordo com Géron (2017), isso pode ser feito de três formas: excluir os registros incompletos, remover o atributo com dados ausentes ou substituí-los por valores como a média ou a mediana da coluna.

Outro ponto fundamental é a padronização das escalas das variáveis numéricas. Algoritmos como redes neurais e máquinas de vetores de suporte são sensíveis a diferenças de escala entre atributos (GÉRON, 2017). As principais técnicas são a **normalização** (reduz os dados a um intervalo entre 0 e 1) e a **padronização** (Transforma os dados para que tenham média zero e desvio padrão igual a um). A escolha entre essas técnicas depende das características do algoritmo e da distribuição dos dados.

Por fim, a preparação adequada dos dados garante uma base sólida para a próxima etapa: a seleção e o treinamento de modelos. Nessa fase, técnicas como validação cruzada e análise de erro contribuem para identificar o algoritmo mais eficaz para o problema em questão (GÉRON, 2017).

2.5.5 Seleção e treinamento do modelo

A seleção e o treinamento de modelos em aprendizado de máquina são etapas essenciais para desenvolver soluções eficazes (GÉRON, 2017). Após a exploração e preparação dos dados, é necessário escolher um modelo apropriado para o problema e realizar seu treinamento com os dados disponíveis. (GÉRON, 2017).

Inicialmente, é comum treinar modelos simples com parâmetros padrão, a fim de obter uma avaliação preliminar de desempenho (GÉRON, 2017). No entanto, para uma estimativa mais robusta, recomenda-se o uso de técnicas como a validação cruzada. Essa técnica divide o conjunto de treinamento em K subconjuntos (folds), treinando e avaliando o modelo K vezes, cada vez com um subconjunto diferente reservado para validação. Esse processo reduz a variabilidade da avaliação e fornece uma medida mais confiável da performance do modelo (GÉRON, 2017).

Para problemas mais complexos, podem ser utilizadas abordagens baseadas em comitês de modelos, como o Gradient Boosting, que combina múltiplas árvores de decisão de forma sequencial. Cada árvore é treinada para corrigir os erros das anteriores, aumentando a precisão e reduzindo o risco de sobreajuste (overfitting) (GÉRON, 2017).

Além da escolha do modelo, a engenharia de atributos e a seleção de variáveis relevantes são estratégias essenciais para melhorar o desempenho. Essas técnicas ajudam os modelos a generalizarem melhor, ajustando-se para cometer menos erros ou erros menos impactantes. Conforme destacado por Géron (2017), esse processo é iterativo: os modelos devem ser ajustados e reavaliados continuamente com base nos resultados de desempenho e nas análises de erro, promovendo uma evolução progressiva da solução.

2.5.6 Ajuste do modelo

O processo de ajuste fino de modelos (fine-tuning) é uma etapa crucial no desenvolvimento de modelos de aprendizado de máquina, especialmente após a seleção de modelos promissores (GÉRON, 2017). Uma das abordagens mais comuns para realizar esse ajuste é o uso do *Scikit-Learn*, que automatiza a busca pelos melhores hiperparâmetros do modelo. Em vez de testar manualmente combinações de valores para os hiperparâmetros, o *Scikit-Learn* avalia todas as possibilidades de uma lista de valores, utilizando validação cruzada para escolher a melhor configuração (GÉRON, 2017). Isso facilita o processo, economizando tempo e aumentando a precisão na escolha dos parâmetros ideais para o modelo.

Além do ajuste de hiperparâmetros, outra técnica importante para aprimorar o modelo é o uso de *Métodos de Ensemble*. Esses métodos combinam os melhores modelos individuais, muitas vezes resultando em um desempenho superior ao de qualquer modelo isolado (GÉRON, 2017). A combinação de modelos com erros diferentes pode reduzir

a variabilidade e melhorar a precisão geral. Um exemplo clássico é o *Random Forest*, que utiliza múltiplas árvores de decisão para obter melhores resultados do que uma única árvore. A estratégia de ensemble pode ser fundamental para melhorar o desempenho do modelo, especialmente em tarefas complexas (GÉRON, 2017).

Após o ajuste, é importante analisar os melhores modelos e seus erros para entender melhor o desempenho do sistema. Inspeccionar os atributos mais importantes para a previsão pode revelar insights valiosos sobre o problema (MÜLLER, 2017). Além disso, entender os tipos de erros cometidos pelo modelo e as razões por trás deles pode ajudar a ajustar o modelo, adicionando ou removendo features, tratando outliers ou refinando a transformação dos dados (MÜLLER, 2017). Após realizar essas melhorias, o modelo deve ser avaliado no conjunto de teste para estimar sua capacidade de generalização. A avaliação no conjunto de teste fornece uma medida objetiva da performance do modelo em dados não vistos, sendo fundamental para garantir que o modelo não esteja superajustado aos dados de treinamento (MÜLLER, 2017).

2.5.7 Lançamento da Solução

Após a aprovação do lançamento de um projeto, é crucial preparar a solução para produção, conectando as fontes de dados de entrada e escrevendo os testes necessários para garantir que o sistema funcione conforme esperado. A integridade e a qualidade do sistema, bem como sua adaptação ao ambiente de produção, são essenciais para que ele se mantenha eficiente.

A qualidade dos dados de entrada também deve ser constantemente monitorada, já que problemas como sensores defeituosos ou dados desatualizados podem impactar significativamente a acuracidade do modelo, prejudicando sua performance e a confiança nos resultados gerados (MÜLLER, 2017).

A manutenção do sistema é outro aspecto crítico após o lançamento, exigindo que o modelo seja re-treinado regularmente com dados frescos (GÉRON, 2017). Isso é importante para evitar que a performance do sistema flutue de forma inesperada, o que pode acontecer se o modelo for atualizado de maneira esporádica (GÉRON, 2017). A automação desse processo de re-treinamento é essencial para garantir que o modelo seja atualizado sempre que necessário, sem depender de intervenções manuais.

Portanto, o lançamento de um sistema de aprendizado de máquina não se limita à integração inicial dos dados e à validação do modelo. O sucesso contínuo depende de um monitoramento eficiente e de uma manutenção constante, garantindo que o modelo se adapte às mudanças nos dados ao longo do tempo. Implementando uma infraestrutura de monitoramento e re-treinamento robusta, as equipes podem assegurar que o sistema continue atendendo aos objetivos de negócios de forma eficaz e precisa, minimizando riscos

e maximizando a performance ao longo de sua vida útil.

2.6 Identificação de Padrões Morfocinéticos e Predição de Euploidia com IA e Trabalhos Correlatos

Os dados que são coletados pela tecnologia do TLS, são chamados de “dados morfocinéticos”, que são definidos como dados do desenvolvimento dos embriões (MOUSTAKLI et al., 2024). Essa informação reunida proporciona noções detalhadas sobre o padrão do desenvolvimento e divisão celular embrionário. Atualmente, após recorrentes estudos sobre esses dados, sabe-se que as características morfocinéticas dos embriões têm sido associadas à avaliação de sua potência de desenvolvimento, ou seja, se um embrião analisado pelo TLS tenha um melhor desenvolvimento, ele terá mais probabilidade de ser euplóide, pois um bom desempenho de um embrião é capaz de prever a implantação (YUAN et al., 2023).

Os modelos de TLS, de acordo com Yuan et al. (2023), tem uma avaliação contínua na etapa do desenvolvimento embrionário por meio de suas imagens e, por observações estáticas, monitora as características do embrião, como tempo e padrões de divisão celular, fornecendo uma base para prever a euploidia. O TLS por si só, não opera com a IA, mas é frequentemente mesclado com essa tecnologia para maiores análises. Um exemplo é o estudo do Yuan et al. (2023), o artigo “Development of an artificial intelligence based model for predicting the euploidy of blastocysts in PGT-A treatments” o qual teve como objetivo utilizar o TLS e desenvolver um modelo de IA usando uma técnica de regressão logística, para predizer a euploidia de blastocistos—fase do desenvolvimento embrionário que ocorre após a clivagem do óvulo fertilizado—em tratamentos de PGT-A, ajudando a identificar embriões com maiores possibilidades de serem geneticamente normais antes da etapa de transferência. O modelo foi avaliado com uma boa precisão, indicando que ele consegue distinguir entre embriões euploides e aneuploides.

Outro estudo é o de Souza (2022a), “Análise da ploidia de embriões humanos por meio da inteligência artificial com o uso de variáveis de morfologia, morfocinética e variáveis relacionadas com a paciente”, que também utiliza IA, mas combinando dados morfológicos, morfocinéticos e clínicos para prever a ploidia dos embriões, utilizando uma rede neural artificial para classificá-los como euploides ou aneuploides. Em contraste, o modelo proposto neste projeto busca prever a porcentagem de aneuploidia, oferecendo uma análise quantitativa detalhada, em vez de uma classificação binária. Essa abordagem visa proporcionar uma compreensão mais profunda da saúde genética dos embriões, permitindo decisões mais precisas durante a seleção.

Divergente do trabalho de Yuan et al. (2023) e de Souza (2022a), que focam na predição binária de euploidia (ou seja, identificar se um embrião é euploide ou aneuploide),

o modelo proposto neste projeto visa prever a porcentagem de aneuploidia dos embriões. Essa abordagem oferece um indicador quantitativo em vez de uma simples classificação binária, permitindo uma análise mais detalhada e informativa sobre a saúde genética dos embriões. Com isso, embriologistas poderão avaliar não apenas se um embrião é geneticamente normal, mas também entender o grau de aneuploidia presente, possibilitando decisões mais precisas durante o processo de seleção.

Além disso, nosso modelo busca ser uma alternativa menos invasiva e mais acessível. Embora técnicas atuais, como o PGT-A (Testagem Genética Pré-implantacional), forneçam informações precisas sobre a euploidia, elas dependem de métodos invasivos, como a biópsia embrionária, e de infraestrutura avançada, o que pode limitar o acesso a essas tecnologias em algumas clínicas. Com o uso de IA e dados de morfocinética obtidos via TLS (Time-Lapse System), o objetivo é desenvolver uma solução que permita uma avaliação robusta sem a necessidade de procedimentos invasivos, ampliando o alcance e a aplicação da tecnologia.

Um exemplo de modelo amplamente utilizado na prática clínica é o *KIDScoreTM Day 5*, que classifica embriões com base no seu potencial de implantação, sendo frequentemente usado em dispositivos como o *TM*(REIGNIER et al., 2019). Esse modelo utiliza grandes bancos de dados multicêntricos para atribuir uma pontuação ao embrião, ajudando a ranquear os embriões do mesmo ciclo, facilitando a escolha do embrião com maior potencial para transferência (REIGNIER et al., 2019).

Apesar dos avanços e do sucesso clínico do *KIDScoreTM*, algumas limitações importantes foram destacadas na literatura. Segundo Reignier et al. (2019), o desempenho desses modelos pode ser influenciado por variáveis como idade materna, número de oócitos, origem dos oócitos e características específicas de cada centro de FIV, como as condições de embrionária e uso de oxigênio reduzido. Isso destaca a necessidade de criar modelos robustos fundamentados em grandes conjuntos de dados multicêntricos e amplamente representativos, a fim de assegurar maior generalização e confiabilidade nos resultados.

Dentro desse cenário, o sistema *CHLOETM* (Cultivating Human Life through Optimal Embryos), desenvolvido pela *FairtilityTM*, se destaca como uma ferramenta inovadora no campo da FIV por usar IA e visão computacional para prever com precisão a implantação e o desenvolvimento dos blastocistos. (FAIRILITYTM, 2020). Enquanto nosso modelo é mais quantitativo e menos invasivo, o CHLE oferece uma solução mais integrada e transparente, visando otimizar a seleção de embriões de formas complementares, com foco na acurácia e confiança no processo de FIV. Essa tecnologia utiliza análises baseadas em imagens, o que indica um método não invasivo, além de oferecer uma análise transparente das decisões tomadas pela IA (FAIRILITYTM, 2020).

De forma semelhante ao *CHLOETM*, nosso modelo também combina IA e dados

morfocinéticos para oferecer uma avaliação robusta e detalhada dos embriões. No entanto, enquanto o CHLOETM se destaca por sua integração e foco na confiança do processo, nosso modelo propõe uma abordagem que enfatiza a estimativa quantitativa da porcentagem de aneuploidia, permitindo uma análise ainda mais detalhada da qualidade genética dos embriões.

Com essas similaridades e diferenças em mente, nosso objetivo é desenvolver um modelo preditivo que combine precisão e acessibilidade, oferecendo uma análise quantitativa detalhada da saúde genética dos embriões. Essa abordagem tem o potencial de melhorar as taxas de sucesso nos tratamentos de fertilização in vitro, contribuindo para decisões mais informadas e eficazes no processo de seleção embrionária.

3 Metodologia

3.1 Classificação da Pesquisa

3.1.1 Natureza

Em relação à natureza desta pesquisa, trata-se de uma pesquisa aplicada. Temos como objetivo principal gerar um conhecimento que possa ter um impacto direto e uma utilidade prática, ambos em contextos reais ([NASCIMENTO, 2016](#)).

Com esta pesquisa, o desenvolvimento do modelo de IA tem a potencialidade de melhorar os sucessos dos tratamentos de FIV, fazendo o processo de seleção embrionária mais eficaz, menos invasivo e mais acessível a um maior número de pessoas.

3.1.2 Método ou Abordagem Metodológica

A metodologia ou abordagem metodológica dessa pesquisa é quantitativa ([NASCIMENTO, 2016](#)). Nosso foco é a análise dos dados numéricos referentes aos padrões morfocinéticos de embriões. Esses dados serão utilizados para o desenvolvimento da IA, que será capaz de prever a porcentagem de euploidia, auxiliando na seleção de embriões com maior possibilidade de saúde genética.

Escolher a abordagem quantitativa nos ajudará a atingir os objetivos desta pesquisa, permitindo explorar e validar os dados com precisão, oferecendo resultados objetivos.

3.1.3 Objetivos

Quanto aos objetivos, o objetivo desta pesquisa é exploratório ([NASCIMENTO, 2016](#)). Este trabalho procura identificar e investigar padrões em dados morfocinéticos de embriões, usando o TLS, explorando a possibilidade de realizar essa predição juntamente com as tecnologias de IA.

Ao focar na concepção de um modelo que terá a capacidade de identificar padrões nos dados, exploraremos a relação entre esses dados e a importância de cada padrão para o resultado desejado, compreendendo os fatores que influenciam a saúde genética dos embriões, mas sem um conhecimento prévio estabelecido que explique completamente essas relações.

3.1.4 Procedimentos De Pesquisa

O procedimento adotado neste trabalho é experimental. Definimos este procedimento por causa do objetivo de investigar as relações entre as variáveis, o que é uma característica da pesquisa experimental (NASCIMENTO, 2016). Buscamos estabelecer uma relação de causa e efeito entre as características morfocinéticas dos embriões, mais especificamente a porcentagem de euploidia.

3.2 Design da Pesquisa

Esse estudo adotará o uso de IA para realizar a análise dos dados morfogenéticos dos embriões, desenvolvendo um modelo de predição baseado em machine learning. O modelo será treinado para identificar os padrões nos dados coletados pelo TLS, com foco em prever a porcentagem de euploidia, o que indica a saúde genética dos embriões.

Para desenvolver e testar o modelo, utilizaremos a linguagem de programação Python, aproveitando as bibliotecas disponíveis para a construção da IA. Quanto à validação do modelo, será elaborada uma fase experimental, na qual o modelo será testado com dados reais de embriões já classificados, a fim de compararmos e testarmos seu desempenho, refinando-o quando necessário.

Nesta pesquisa, buscaremos identificar e mapear os padrões em um campo que ainda está em desenvolvimento. A prática será testada em um ambiente controlado com dados obtidos pelo TLS, avaliando a efetividade do modelo com base na sua capacidade de prever, em porcentagem, a ploidia do embrião.

3.3 Fases de Trabalho

As fases do nosso trabalho se dividem em duas etapas: **Fase 1: Análise e Preparação de Dados** (Tabela 1), com o objetivo de compreender e organizar os dados para realizar a análise da influência dos parâmetros na porcentagem de euploidia, e a **Fase 2: Desenvolvimento e Avaliação do Modelo** detalhada na Tabela (Tabela 2), que foca no desenvolvimento, ajuste e avaliação de um modelo de ML para efetuar a predição de euploidia, finalizando com a entrega de um protótipo de uma interface a ser evoluída em trabalhos futuros, realizando a criação e junção dos dois.

Nas seções a seguir, demonstraremos os objetivos de cada fase, mostrando suas atividades, nas quais estão descritos resumidamente o que será feito, qual método será utilizado e o resultado esperado. Em seguida, detalharemos cada parte.

Tabela 1. Fase 1: Análise e Preparação de Dados

Fase 1: Análise e Preparação de Dados				
Objetivos Específicos	Atividades	Método de Pesquisa	Resultados Esperados	
OE1 Expandir, Processar e Analisar os Dados para Predição de Ploidia	Atividade 1 (A1) Análise, Revisão e Seleção de Variáveis para Predição de Euploidia	- Pesquisa bibliográfica - Python - Biblioteca Pandas	- Analisar, Revisar e Selecionar as variáveis - Limpeza dos dados	
	Atividade 2 (A2) Normalização dos Dados para Otimização	- Z-Score	- Normalização das variáveis numéricas	
	Atividade 3 (A3) Identificação da Correlação e Atribuição de Pesos aos Parâmetros na Previsão da Ploidia do Embrião	- Coeficiente de correlação de Spearman	- Análise das correlações entre variáveis pelo Gráfico de dispersão	
	Atividade 4 (A4) Divisão de Dados e Aplicação de Data Augmentation	- Divisão do conjunto de dados - Data augmentation com o Algoritmo de Monte Carlo	- Dados para treinamento e teste - Aumento do conjunto de dados para treinamento	

3.3.1 Objetivo Específico 1 - Identificação de Parâmetros em Embriões

3.3.1.1 Atividade 1 (A1): Análise, Revisão, Seleção e Limpeza de Variáveis para Predição de Euploidia

Começaremos com a análise das variáveis existentes na planilha de dados dos embriões, avaliando sua pertinência e verificando a necessidade de introdução de novas variáveis que possam aprimorar a exatidão da análise ou, se necessário, eliminar aquelas consideradas irrelevantes ou redundantes. Esse processo visa aprimorar a correlação entre os parâmetros estudados e a porcentagem de euploidia. Em seguida, será realizada a limpeza dos dados, com a detecção e correção de inconsistências, como valores em branco ou discrepantes, substituindo valores nulos por outros mais apropriados, especialmente em colunas que dependem de cálculos entre variáveis. Esse conjunto de ações garantirá que os dados estejam organizados, consistentes e prontos para análises futuras, prevenindo distorções nos resultados.

Para realizar a verificação da pertinência das variáveis já existentes e avaliar a viabilidade de introduzir novas variáveis, utilizaremos a pesquisa bibliográfica, nos norteados pelos estudos apresentados no capítulo 2: o de [Yuan et al. \(2023\)](#), o artigo 'Development

of an artificial intelligence-based model for predicting the euploidy of blastocysts in PGT-A treatments’ e o de Souza (2022a), ‘Análise da ploidia de embriões humanos por meio da inteligência artificial com o uso de variáveis de morfologia, morfocinética e variáveis relacionadas com a paciente’. Ambos os artigos descrevem o uso de IA para fazer a predição da ploidia de embriões, o que se assemelha com o que queremos propor, com a diferença de que temos o objetivo final de prever a porcentagem de aneuploidia.

Dessa forma, analisaremos os estudos feitos por ambos pesquisadores e utilizaremos para entender o poder que cada variável tem para o objetivo final e se é necessário adicionar outras variáveis que não existem na nossa planilha, ou até mesmo excluí-las. Além disso, conduziremos entrevistas com o especialista encarregado de fornecer os dados, para uma análise prática da pertinência das variáveis disponíveis e para debater possíveis variáveis extras que possam ser relevantes para a análise. As entrevistas possibilitarão alinhar a seleção das variáveis ao conhecimento clínico e experiência prática do profissional, garantindo que as variáveis selecionadas sejam aplicáveis no cenário real de previsão de euploidia. Em relação a limpeza dos dados, utilizaremos a linguagem *Python* que possui a biblioteca *Pandas*, que permite o carregamento de planilhas do Excel, do formato .xlsx, como a que possuímos. De acordo com Chen (2018): “O Pandas é uma biblioteca Python de código aberto para análise de dados. Ele dá a Python a capacidade de trabalhar com dados do tipo planilha, permitindo carregar, manipular, alinhar e combinar dados rapidamente, entre outras funções.”. Usaremos as funções *isnull()* e *info()*, da biblioteca *Pandas*.

Essa atividade corresponde ao Passo 2 do desenvolvimento de algoritmos de aprendizado de máquina, que trata da obtenção dos dados necessários para o projeto.

3.3.1.1.1 **Atividade 2 (A2):** Normalização dos Dados para Otimização

Antes de iniciarmos a normalização, algumas variáveis serão transformadas para o formato numérico, pois estão originalmente representadas como strings. A primeira variável a ser alterada será a ploidia, que será convertida para valores binários: 0 para euploide e 1 para aneuploide. Em seguida, a variável estágio será modificada removendo-se a letra “D” e mantendo apenas o número que representa o dia de desenvolvimento embrionário, como 5, 6 ou 4. Por fim, a variável classificação Morfo será substituída pelo seu grupo numérico equivalente, conforme descrito na fundamentação teórica, com valores de 1 a 4.

As variáveis numéricas serão normalizadas por meio de uma técnica de normalização, o **Z-Score**, detalhado no APÊNDICE C. A normalização será essencial, pois, de acordo com Milewski et al. (2016), estudos anteriores demonstraram que a incorporação de dados morfológicos normalizados na avaliação da qualidade embrionária aumenta consideravelmente o poder preditivo dos modelos construídos. A normalização é uma forma de

dimensionar recursos, transformando o intervalo deles em uma escala padrão (JAISWAL, 2024). Essa técnica transformará os dados para uma escala padrão (JAISWAL, 2024), facilitando a interpretação e a comparação entre diferentes variáveis.

Essa atividade corresponde ao Passo 3 do desenvolvimento de algoritmos de aprendizado de máquina, que envolve a preparação e transformação dos dados para garantir que estejam em um formato adequado para o treinamento dos modelos.

3.3.1.1.2 **Atividade 3 (A3):** Identificação da Correlação e Atribuição de Pesos aos Parâmetros na Previsão da Ploidia do Embrião

O objetivo desta atividade é identificar as relações entre os diferentes parâmetros presentes na planilha de dados dos embriões, avaliando a intensidade e o sentido dessas relações, com foco em sua influência na porcentagem de euploidia. Após a pesquisa bibliográfica, utilizaremos o coeficiente de correlação de Spearman, explicado com mais detalhes no APÊNDICE E. Esse coeficiente mede a relação monótona entre duas variáveis, considerando as ordens atribuídas às observações em vez dos valores originais (SOUSA, 2019). A correlação será calculada para todas as combinações possíveis de variáveis, possibilitando uma análise mais detalhada de suas interações (SOUSA, 2019). Um gráfico de dispersão será gerado para complementar a análise visual, facilitando a identificação de padrões. Também com o conhecimento adquirido pela A1, determinaremos a relevância relativa de cada parâmetro na previsão da ploidia do embrião, atribuindo pesos que reflitam sua influência, identificando a relevância de cada parâmetro.

A fórmula do coeficiente de correlação de Spearman será aplicada, utilizando as ordens atribuídas, assegurando que o método se adapte a diferentes formatos de relação, como curvas monótonas crescentes ou decrescentes. Além disso, será realizada uma análise complementar com gráficos de dispersão, que ajudarão a identificar a inclinação dos dados e o sentido da correlação, sendo positiva (próximas ao valor 1) quando as variáveis variam no mesmo sentido e negativa (próximas ao valor -1) quando variam em sentidos opostos (SOUSA, 2019). O resultado numérico do coeficiente será avaliado em relação à sua magnitude, indicando se a correlação é forte, moderada ou fraca, e seu sinal indicará o tipo de relação (positiva ou negativa).

Usaremos a biblioteca Pandas para manipulação dos dados e a SciPy para calcular o coeficiente de Spearman. A metodologia para a definição e atribuição de pesos específicos aos parâmetros relevantes para a ploidia do embrião combina análise teórica e prática. Inicialmente, será realizada uma pesquisa bibliográfica em publicações científicas e revisões sistemáticas que explorem a influência dos parâmetros na ploidia embrionária.

Essa atividade corresponde ao Passo 3 do desenvolvimento de algoritmos de aprendizado de máquina, que envolve a preparação e transformação dos dados para garantir

que estejam em um formato adequado para o treinamento dos modelos.

3.3.1.1.3 **Atividade 4 (A4):** Separar o conjunto de dados em conjuntos de treinamento e teste, fazendo uma distribuição dos dados e aplicar técnica de aumento de dados

A aplicação de técnicas de Aprendizado de Máquina em conjuntos de dados pequenos impõe desafios metodológicos significativos. Nesses cenários, o risco de overfitting (sobreajuste) é amplificado. Esse fenômeno ocorre quando o modelo aprende padrões específicos e ruídos do conjunto de treinamento, em vez de generalizações úteis para novos dados, resultando em uma performance artificialmente elevada durante o treinamento, mas fraca diante de dados independentes (BASHIR et al., 2020).

Como o tamanho do conjunto de dados inviabiliza uma divisão tradicional em três subconjuntos (treinamento, validação e teste), optamos por uma abordagem mais viável: a divisão em apenas dois conjuntos — 80% para treinamento e 20% para teste. Essa decisão visa preservar a maior quantidade possível de dados para o treinamento, garantindo ao mesmo tempo uma avaliação minimamente robusta do desempenho do modelo. Ng (2018) sugere que conjuntos de dados menores que 10.000, a divisão deve ser cuidadosamente ponderada, pois qualquer porção reservada a validação pode comprometer a capacidade de aprendizado e a confiabilidade da avaliação.

Além da divisão, será feito um **balanceamento das classes** durante a divisão dos dados, a fim de garantir que a escolha aleatória não acabe segregando mais amostras de uma classe em um conjunto do que em outro. Esse cuidado é essencial para evitar viés nos dados e assegurar que o modelo tenha a oportunidade de aprender e ser avaliado com amostras representativas de todas as classes, o que pode impactar diretamente na precisão do modelo.

Dado o volume reduzido de dados, aplicamos a técnica de aumento de dados (data augmentation), utilizando o algoritmo de Monte Carlo (ver APÊNDICE D), com o objetivo de gerar amostras sintéticas de alta qualidade a partir dos dados existentes. Essa técnica foi aplicada exclusivamente ao conjunto de treinamento, de forma a evitar interferências na avaliação final e permitir que o modelo aprenda a partir de padrões legítimos.

Essa atividade corresponde ao Passo 4 do desenvolvimento de algoritmos de aprendizado de máquina, que trata da preparação final dos dados, garantindo um aprendizado mais robusto e eficaz do modelo.

Tabela 2. Fase 2: Desenvolvimento e Avaliação do Modelo

Fase 2: Desenvolvimento e Avaliação do Modelo				
Objetivos Específicos	Atividades	Método de Pesquisa	Resultados Esperados	
OE2 Treinar e Ajustar o Modelo de Machine Learning para Predição de Ploidia	Atividade 5 (A5) Desenvolvimento e Treinamento do Modelo de Machine Learning para Otimização da Predição de Ploidia	- Python (Bibliotecas pandas, joblib, scikit-learn, matplotlib, seaborn e a lime)	- Treinamento do modelo bem-sucedido com no mínimo 70% de precisão, usando Rede Neural MLP com LIME.	
	Atividade 6 (A6) Utilizar métricas adequadas para medir o desempenho do modelo	- Python (Biblioteca scikit-learn e pandas)	- Acurácia - ROC-AUC - Recall	
OE3 Realizar Avaliação do Modelo	Atividade 7 (A7) Avaliação do Desempenho do Modelo na Predição por Meio da Matriz de Confusão e Curva ROC	- Matriz de confusão (Random Forest - scikit-learn)	- Verdadeiros positivos (TP) - Verdadeiros negativos (TN) - Falsos positivos (FP) - Falsos negativos (FN) - Gráfico exibindo a relação entre a sensibilidade e a especificidade para diferentes valores de limiar.	
	Atividade 8 (A8) Prototipar uma interface básica para exibir as predições de euploidia para o usuário final (médicos)	- Figma	- Protótipo funcional	
OE4 Construir Protótipo	Atividade 9 (A9) Desenvolver a aplicação web integrando o protótipo à inteligência artificial existente	- Next.js, TypeScript e TailwindCSS - Vercel	- Aplicação web com leitura de planilhas .xlsx integrada à IA	

3.3.1.2 **Objetivo Específico 2** - Treinamento e Ajuste de Modelo de Machine Learning para Predição de Euploidia

3.3.1.2.1 **Atividade 5 (A5):** Desenvolvimento e Treinamento do Modelo de Machine Learning para Otimização da Predição de Euploidia

Nesta etapa, será realizado o desenvolvimento e o treinamento de modelos de aprendizado de máquina supervisionado com o objetivo de prever a euploidia de embriões, utilizando dados previamente tratados e normalizados (conforme a Atividade 3). O foco principal é construir um classificador binário que seja capaz de distinguir entre embriões euploides e aneuploides, alcançando, no mínimo, 70% de acurácia nos dados de teste.

Utilizaremos Rede Neural do *Scikit-learn*, importando as bibliotecas essenciais para manipulação de dados, avaliação de desempenho, visualização gráfica e para o método de explicação LIME. A base de dados com as variáveis preditoras e o alvo seria carregada a partir de um arquivo Excel, junto com o modelo MLP previamente treinado e o scaler utilizado para normalização dos dados. Para garantir a compatibilidade entre os dados atuais e o modelo, ajustaria as colunas do conjunto para que correspondessem exatamente às esperadas pelo scaler, adicionando colunas faltantes preenchidas com zero quando necessário, de forma a evitar erros na transformação.

Na etapa de avaliação, calcularia métricas fundamentais como acurácia, área sob a curva ROC, matriz de confusão e recalls para cada classe, a fim de entender a performance do modelo de forma abrangente. Para complementar essa análise quantitativa, criaria visualizações gráficas da matriz de confusão e da curva ROC, facilitando a interpretação dos resultados. Por fim, exportaria o conjunto de dados enriquecido com as previsões e probabilidades para um novo arquivo, encerrando o processo com a garantia de que o modelo não apenas oferece resultados precisos, mas também explicações interpretáveis que podem ser utilizadas para suporte à decisão clínica.

Ao final, espera-se obter um modelo capaz de classificar embriões com pelo menos 70% de precisão, fornecendo uma base confiável para apoio a decisões clínicas e pesquisas científicas.

Essa atividade corresponde ao Passo 5 do desenvolvimento de algoritmos de aprendizado de máquina, que envolve a seleção, desenvolvimento e treinamento do modelo para realizar as previsões desejadas com base nos dados preparados nas etapas anteriores.

3.3.1.3 Objetivo Específico 3 - Avaliação do modelo

3.3.1.3.1 Atividade 6 (A6): Utilizar métricas de avaliação mais adequadas para medir o desempenho do modelo de acordo com a natureza do problema de classificação

O objetivo desta atividade é aplicar e avaliar métricas adequadas para medir o desempenho e a confiabilidade de um modelo de IA considerando as particularidades do problema abordado. Serão utilizadas as métricas Acurácia, *ROC-AUC* e *Recall* (Sensibilidade). A análise será realizada com base em um conjunto de dados previamente definido, utilizando *Python* e bibliotecas como *scikit-learn* e *pandas* para implementação, análise e visualização dos resultados.

Cada métrica será detalhada a seguir:

- **Acurácia:** Mede a proporção de previsões corretas em relação ao total de previsões realizadas. Essa métrica é útil em problemas com classes balanceadas e quando não há maior preocupação com erros específicos, como falsos positivos ou falsos negativos (JUNIOR et al., 2022). Será calculada com a função *accuracy_score* do *scikit-learn*.
- **ROC-AUC:** Permite avaliar a capacidade geral do modelo de distinguir entre classes, mesmo quando estas estão desbalanceadas (JUNIOR et al., 2022). Essa métrica complementa a análise das demais, oferecendo uma visão ampla do desempenho do classificador.
- **Recall:** Mede a capacidade do modelo de identificar corretamente os exemplos pertencentes à classe positiva. É especialmente importante em situações onde falsos negativos têm maior impacto, como na detecção de aneuploidia de embriões (JUNIOR et al., 2022). Será calculado com a função *recall_score*.

Espera-se que a análise dessas métricas proporcione uma avaliação abrangente do desempenho do modelo, destacando seus pontos fortes e limitações, e possibilitando ajustes para melhorar sua confiabilidade e alinhamento com os objetivos do problema. Além disso, a escolha criteriosa das métricas será essencial para orientar decisões estratégicas, especialmente em aplicações sensíveis a erros de classificação.

Essa atividade corresponde ao Passo 6 do desenvolvimento de algoritmos de aprendizado de máquina, que envolve a seleção, desenvolvimento e treinamento do modelo, que trata da avaliação do desempenho do modelo utilizando métricas apropriadas e técnicas de análise para compreender sua eficácia, pontos fortes e limitações antes da implementação final.

3.3.1.3.2 **Atividade 7 (A7):** Avaliar a precisão e eficácia do modelo em prever corretamente casos de euploidia e aneuploidia por meio da Matriz de Confusão e Curva ROC

O objetivo desta atividade é avaliar o desempenho do modelo de classificação desenvolvido para prever corretamente os casos de euploidia e aneuploidia, utilizando a Matriz de Confusão e a Curva ROC (Receiver Operating Characteristic). A Matriz de Confusão será construída após o treinamento e teste do modelo, permitindo a identificação das taxas de verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN). Essa análise contribuirá para a compreensão do desempenho geral do modelo e identificação de possíveis pontos de melhoria. Adicionalmente, será gerada a Curva ROC, que avalia a performance do modelo na separação entre as classes euploide (classe positiva) e aneuploide (classe negativa), com base em diferentes valores de limiar (*threshold*) (JUNIOR et al., 2022). A partir dela, será possível observar a sensibilidade e a especificidade do modelo em diversos pontos de corte. A métrica AUC (Área sob a Curva) será utilizada como indicador global da capacidade do modelo em distinguir entre as classes, sendo especialmente útil em cenários de classificação desbalanceada (JUNIOR et al., 2022).

A avaliação será realizada em duas etapas principais: a construção da Matriz de Confusão e a geração da Curva ROC. A Matriz de Confusão é uma ferramenta essencial para compreender os tipos de acertos e erros do modelo. Segundo Sathyanarayanan e Tantri (2024), se trata de uma tabela de dimensão, onde N representa o número de classes. Para este estudo binário, a matriz permitirá verificar diretamente os valores de TP, TN, FP e FN, a partir dos quais serão calculadas métricas como acurácia e recall. Essas métricas são importantes para indicar não apenas o desempenho global, mas também a sensibilidade do modelo para cada classe.

A Curva ROC será gerada com base nas probabilidades de pertencimento à classe euploide, fornecidas pela função `predict_proba` da biblioteca `scikit-learn`. Essa curva será traçada com o auxílio da função `roc_curve`, que calcula os valores de taxa de falsos positivos (1 - especificidade) e de verdadeiros positivos (sensibilidade) para diferentes valores de limiar (JUNIOR et al., 2022). O gráfico resultante, com a especificidade no eixo *x* e a sensibilidade no eixo *y*, permitirá uma análise detalhada do comportamento do modelo à medida que o limiar de decisão varia. Essa análise é particularmente útil para identificar o ponto ótimo de corte, equilibrando os *trade-offs* entre falsos positivos e falsos negativos.

Por fim, será calculada a AUC com a função `roc_auc_score` da biblioteca `scikit-learn`. A AUC varia de 0 a 1 e resume a capacidade do modelo de distinguir corretamente entre as classes. Valores mais próximos de 1 indicam excelente desempenho, enquanto valores próximos de 0,5 indicam que o modelo se comporta de maneira semelhante a um classificador aleatório (JUNIOR et al., 2022). A partir dessas análises, espera-se obter uma base

sólida para avaliar a eficácia do modelo e sua viabilidade de aplicação prática na predição de euploidia, com suporte tanto quantitativo quanto interpretativo.

Essa atividade corresponde ao Passo 6 do desenvolvimento de algoritmos de aprendizado de máquina, que envolve a seleção, desenvolvimento e treinamento do modelo, que trata da avaliação do desempenho do modelo utilizando métricas apropriadas e técnicas de análise para compreender sua eficácia, pontos fortes e limitações antes da implementação final.

3.3.1.4 **Objetivo Específico 4** - Desenvolver uma Interface

3.3.1.4.1 **Atividade 8 (A8):** Prototipar uma interface básica para exibir as predições de euploidia para o usuário final (médicos)

O objetivo desta etapa é desenvolver um protótipo interativo da interface que será utilizada pelos médicos para visualizar as predições de euploidia de embriões com base nos dados morfocinéticos fornecidos. Este protótipo deve contemplar os princípios fundamentais de design de interfaces, experiência do usuário (UX) e acessibilidade, garantindo que a navegação seja intuitiva e que os resultados sejam apresentados de forma clara e eficiente.

Para isso, será utilizada a ferramenta online **Figma**, amplamente reconhecida por sua versatilidade no desenvolvimento colaborativo de protótipos interativos de alta fidelidade (Figma, 2025). O Figma permitirá o trabalho simultâneo da equipe, com recursos para edição vetorial, prototipagem rápida e compartilhamento de fluxos interativos.

O protótipo incluirá os seguintes elementos principais:

- **Campos para entrada de dados morfocinéticos:** Área destinada ao *upload* de arquivos que conterão as informações necessárias para a predição;
- **Botões de interação:** Elementos visuais que possibilitarão o envio e processamento dos dados inseridos;
- **Área de exibição de resultados:** Seção responsável por apresentar, de forma clara e acessível, os resultados da predição de ploidia;
- **Layout responsivo:** Interface adaptável a diferentes dispositivos e tamanhos de tela, garantindo acessibilidade e boa usabilidade.

O protótipo será construído com base no **Design Centrado no Usuário** (DCU), uma abordagem que colocará as necessidades, capacidades e limitações dos usuários no centro do processo de design (NORMAN, 2013). Considerando o ambiente médico, onde

decisões rápidas e seguras são fundamentais, será imprescindível garantir uma interface intuitiva, eficiente e de fácil interpretação.

Para isso, serão aplicadas como referência as **Heurísticas de Usabilidade de Nielsen** (NIELSEN, 1994), um conjunto amplamente reconhecido de diretrizes que auxiliará no desenvolvimento de sistemas mais acessíveis, eficazes e agradáveis. As heurísticas de Jakob Nielsen, publicadas em 1994, foram formuladas com base em décadas de estudos sobre interação humano-computador e têm sido, desde então, utilizadas como padrão para avaliação e desenvolvimento de sistemas interativos (MOLICH; NIELSEN, 1990).

Essas heurísticas representam princípios universais de usabilidade e visam assegurar que os usuários possam interagir com o sistema de forma segura, confortável e produtiva. Duas dessas heurísticas, em especial, serão incorporadas ao protótipo:

1. **1. Visibilidade do status do sistema (Visibility of system status:)** É essencial que o sistema mantenha os usuários informados sobre o que está acontecendo, por meio de feedbacks adequados e em tempo razoável. No protótipo, isso será implementado através de mensagens claras de carregamento, feedback visual após o envio dos dados e indicações sobre o status da previsão. Dessa forma, os médicos terão maior segurança sobre o funcionamento do sistema.
2. **2. Flexibilidade e eficiência de uso (Flexibility and efficiency of use:)** A interface deve ser eficiente tanto para usuários experientes em tecnologia quanto aqueles com menor familiaridade digital. Por isso, o protótipo será planejado para realizar o envio de dados de forma ágil, com suporte a preenchimentos automáticos ou reaproveitamento de planilhas padronizadas, reduzindo o tempo de interação e o esforço necessário para acessar as predições

A importância da aplicação dessas diretrizes será ainda maior considerando o contexto de uso médico, onde uma interface mal projetada pode impactar diretamente na tomada de decisão clínica e, consequentemente, na segurança e no bem-estar dos pacientes. Conforme destacado por Rangraz Jeddi et al. (2020), a avaliação da usabilidade em sistemas de informação em saúde é fundamental não apenas para melhorar a experiência do usuário, mas também para reduzir erros humanos, aumentar a confiabilidade dos processos e garantir a segurança do paciente.

A Atividade 8 (A8) corresponderá ao Passo 7 do ciclo de desenvolvimento do algoritmo de aprendizado de máquina, quando a solução será disponibilizada por meio de uma interface amigável e acessível, conectando a inteligência artificial à prática clínica.

3.3.1.4.2 **Atividade 9 (A9):** Desenvolver a aplicação web integrando o protótipo à inteligência artificial existente

Após criação do protótipo interativo na Atividade 8, esta etapa corresponderá ao desenvolvimento da aplicação web, implementando de forma prática a interface projetada com tecnologias modernas e integrando-a ao modelo de inteligência artificial desenvolvido nas etapas anteriores. O objetivo será transformar a solução teórica em uma aplicação funcional e acessível, permitindo o uso por médicos especialistas em reprodução assistida. A implementação será realizada com:

- **Next.js com App Router:** Framework React com arquitetura moderna para organização modularizada entre páginas, componentes e bibliotecas, facilitando a escalabilidade e manutenção do código.
- **TypeScript:** Linguagem de programação com tipagem estática para maior robustez, facilitando a detecção precoce de erros, especialmente relevante no contexto médico.
- **TailwindCSS:** Framework utilitário para estilização rápida, promovendo uma interface responsiva e consistente com o esquema de cores definido previamente no protótipo.
- **Leitura de arquivos XLSX:** Implementação de funcionalidade para upload e leitura de planilhas contendo dados morfofocinéticos, que serão processados e enviados ao modelo de machine learning.
- **FastAPI:** Framework backend em Python escolhido para transformar o modelo de IA em uma API moderna e de alta performance, com documentação automática via Swagger, validação de dados e integração simplificada com o front-end.
- **Hospedagem da API na Render:** A FastAPI será implantada na plataforma Render, garantindo disponibilidade, escalabilidade e facilidade de manutenção.
- **Hospedagem da aplicação web na Vercel:** A aplicação Next.js será hospedada na Vercel, permitindo deploy contínuo, agilidade nas entregas e fácil acesso para testes e validações ao longo do desenvolvimento.

A interface será cuidadosamente construída para refletir o fluxo de uso planejado, garantindo uma transição fluida entre o protótipo e o ambiente real de utilização. Durante o desenvolvimento, será priorizado o **design centrado no usuário**, buscando promover clareza, legibilidade e acessibilidade, aspectos essenciais no contexto médico, onde qualquer ambiguidade pode comprometer a tomada de decisão.

Outro aspecto fundamental nesta etapa será a aplicação sistemática de práticas de **Clean Code**, conforme os princípios descritos por [Martin \(2009\)](#). A adoção dessas boas práticas contribuirá para a legibilidade, manutenibilidade e robustez do código, assegurando a qualidade da solução mesmo em contextos sensíveis como o da saúde. Os seguintes princípios serão aplicados:

- **Nomenclatura clara e intuitiva:** Os nomes de variáveis, funções e componentes serão descritivos e precisos, refletindo suas responsabilidades no sistema.
- **Consistência nos padrões:** Será mantida uniformidade nos padrões de nomenclatura, indentação e formatação, promovendo um código organizado e previsível.
- **Funções pequenas e focadas:** As funções serão escritas para desempenhar uma única tarefa, mantendo-se concisas e especializadas.
- **Comentários relevantes e sucintos:** Os comentários serão utilizados com moderação, apenas quando realmente necessários para esclarecer intenções que não estejam evidentes no código.
- **Evitar repetição de código (DRY):** O projeto buscará reutilização inteligente de componentes e funções para evitar duplicação desnecessária.
- **Tratamento adequado de erros:** Serão implementadas estratégias robustas para lidar com falhas e fornecer mensagens de erro claras e informativas ao usuário.
- **Princípio KISS (Keep It Simple, Stupid):** O código será mantido o mais simples e direto possível, evitando complexidades desnecessárias que possam dificultar a manutenção e o entendimento futuro.

Ao final desta etapa, espera-se concluir o desenvolvimento do protótipo da interface, representando a finalização prática do trabalho. Essa interface terá papel fundamental ao permitir que médicos utilizem o sistema de forma clara, acessível e eficiente, apoiando a análise de dados morfofocinéticos e auxiliando na tomada de decisão em casos clínicos. Dessa forma, a proposta visa não apenas validar a aplicação da inteligência artificial no contexto médico, mas também promover uma ferramenta com potencial de impacto real na rotina de profissionais da saúde reprodutiva.

A Atividade 9 (A9) corresponde ao Passo 7 do desenvolvimento do algoritmo de aprendizado de máquina, que envolve o lançamento da solução.

4 Execução da Pesquisa e Análise dos Resultados

Antes de começar a "Execução da Pesquisa e Análise dos Resultados", é crucial enfatizar que todas as fases deste estudo foram realizadas de acordo com as normas éticas e legais pertinentes à investigação científica. Os dados utilizados neste estudo foram obtidos por meio do especialista em Reprodução Humana Assistida, Dr. Bruno Ramalho, e de suas pacientes, que autorizaram o uso dessas informações para fins de pesquisa. As informações foram coletadas na clínica Bruno Ramalho Reprodução Humana, bem como na clínica Genesis, onde o Dr. Bruno também atua. A documentação que comprova isso está devidamente incluída nos anexos deste estudo, incluindo os seguintes documentos essenciais:

- **Parecer da Plataforma Brasil:** Inclui a permissão ética que valida a utilização dos dados dos embriões para a execução deste estudo, garantindo a conformidade com os padrões éticos definidos para pesquisas que envolvem seres humanos (disponível no Anexo I)
- **Contrato de Autorização para Utilização de Dados em Pesquisa:** Assinado pelo Dr. Bruno Ramalho, que autoriza o uso dos dados clínicos de suas pacientes, essenciais para a modelagem e análise realizadas durante esta pesquisa (disponível no Anexo III)
- **Termo de Consentimento para Utilização de Dados de Entrevistas, Gravação de Reuniões e Uso de Gravação:** Contrato que permite o uso de discursos, dados e gravações provenientes das reuniões e entrevistas conduzidas com o Dr. Bruno Ramalho, garantindo que as informações e contribuições fornecidas por ele fossem utilizadas com total autorização (disponível no Anexo II).

Estes documentos evidenciam o comprometimento ético e a transparência desta pesquisa, enfatizando que todas as informações empregadas foram obtidas de maneira responsável e autorizada. Os códigos mencionados nas atividades estão disponíveis no repositório do projeto no GitHub, podendo ser acessado aqui: [GitHub](#), onde são detalhadamente explicados, incluindo a lógica utilizada para sua implementação e os resultados obtidos.

4.1 Fase 1: Análise e Preparação de Dados

4.1.1 OE1 - Expandir, Processar e Analisar os Dados para Predição de Ploidia

4.1.1.1 Atividade 1 (A1): Análise, Revisão, Seleção e Limpeza de Variáveis para Predição de Euploidia

O desenvolvimento desta atividade teve início com a verificação da relevância das variáveis com base nos estudos da fundamentação teórica. Esse processo envolveu a análise das variáveis presentes na tabela de dados que serão utilizadas nesta pesquisa, garantindo sua adequação ao objetivo do estudo.

A partir da revisão bibliográfica, conduzimos um processo de seleção das variáveis mais relevantes para o nosso modelo de Inteligência Artificial, fundamentado em evidências científicas. Dessa forma, foi possível identificar um conjunto de variáveis da base de dados que apresentam respaldo teórico quanto à sua influência na evolução do modelo. As variáveis selecionadas para o estudo são: **Idade, t2, t3, t4, t5, t8, s2, cc2 (t3-t2), tSC, tSB, tB, cc3 (t5-t3), s3 (t8-t5), t5-t2, tSC-t8, tB-tSB, Estágio, Morfo e KIDScore.**

A coluna de **Plodia** também foi selecionada, pois nos possibilita agrupar os embriões em duas categorias claras, distinguindo entre aqueles com euploidia normal e aqueles com alterações cromossômicas, o que é crucial para a elaboração de um modelo sólido.

Não identificamos estudos que estabelecem uma ligação direta entre o parâmetro **st2** e a previsão de euploidia. Apesar do movimento citoplasmático antes da citocinese ser um marco significativo no desenvolvimento embrionário, a ausência de provas científicas que liguem esse movimento à qualidade do embrião e à euploidia nos levou a remover o **st2** da lista de variáveis para o modelo. Igualmente, não foram identificados estudos que analisassem especificamente o intervalo entre **t2** (o instante em que o embrião alcança a fase de duas células) e **st2** (movimento citoplasmático pré-citocinese) para prever a euploidia. Como **st2** foi eliminada, também removemos o parâmetro **t2-st2** do nosso grupo de variáveis.

A partir disso, conseguimos determinar quais variáveis são fundamentais para a elaboração do nosso modelo de previsão de euploidia. As colunas **Id, Data da biópsia e Embrião n.** também foram eliminadas, uma vez que não contribuem para o valor do modelo. Portanto, a planilha revisada inclui somente as variáveis que possuem uma ligação comprovada com a previsão de euploidia, fundamentada nas evidências científicas revisadas.

4.1.1.2 Limpeza dos Dados

Ao lidar com os dados ausentes, utilizamos o método de Análise de Casos Completos (ACC), que envolve a remoção de observações que possuem pelo menos um valor ausente (CAMARGOS et al., 2011). Este procedimento é frequentemente empregado quando o número de dados ausentes é reduzido, assegurando que a remoção de algumas observações não interfira de forma significativa na análise estatística e preserva a consistência do modelo (CAMARGOS et al., 2011). Em nosso cenário, por se tratar de uma base de dados pequena, optamos por excluir manualmente as linhas com valores ausentes. Essa abordagem foi escolhida por ser mais prática e eficiente, considerando o tamanho reduzido da amostra. Após essa etapa, a planilha foi ajustada com apenas os dados completos.

4.1.1.3 Atividade 2 (A2): Normalização dos Dados para Otimização

A etapa de normalização dos dados é um passo fundamental na criação de modelos de aprendizado de máquina, particularmente em situações onde as variáveis têm diferentes escalas e distribuições. Esta tarefa foi executada com o uso do método Z-Score. Esta metodologia modifica os dados de forma que cada variável possua uma média de 0 e um desvio padrão de 1, explicado na APÊNDICE C.

Antes da aplicação da normalização por Z-Score, foi necessário realizar o pré-processamento dos dados para garantir que todas as variáveis relevantes estivessem em formato numérico. Essa etapa é fundamental em pipelines de aprendizado de máquina, uma vez que a maioria dos algoritmos não consegue operar diretamente com dados categóricos ou textuais.

1. **Conversão da variável alvo “Ploidia”:** A coluna *Ploidia*, originalmente categórica com os valores “Euplóide” e “Aneuplóide”, foi convertida para valores binários (0 e 1, respectivamente). Essa codificação binária é necessária para que o modelo classificador interprete corretamente a variável-alvo.
2. **Transformação da coluna “Estágio”:** A coluna *Estágio*, que indicava o dia de desenvolvimento embrionário com um prefixo textual (“D3”, “D5”, etc.), teve a letra “D” removida com a função `str.replace()`, deixando apenas o valor numérico correspondente ao dia. Em seguida, os valores foram convertidos para o tipo `int` com a função `pd.to_numeric()`, assegurando a correta tipagem para análise quantitativa.
3. **Conversão da classificação morfológica (“Morfo”):** Essa coluna originalmente armazenava uma combinação alfanumérica representando o estágio morfológico do blastocisto (ex: “4AA”, “3BB”). Para torná-la compatível com os modelos, foi

criada uma função chamada `classificar_morfo()` que categoriza os embriões em quatro grupos (1 a 4), com base na qualidade morfológica indicada no Capítulo 2.

4. **Extração das colunas numéricas:** Após as transformações, foi extraído um subconjunto contendo apenas colunas numéricas por meio da função `select_dtypes()`. Isso assegura que o arquivo de saída contenha apenas dados prontos para normalização e posterior uso em modelos de aprendizado de máquina.
5. **Exportação do conjunto processado:** Os dados finalizados foram exportados para uma nova planilha Excel, chamada “`PlanilhaNumerica.xlsx`”, por meio do método `to_excel()`, permitindo seu uso direto nas etapas posteriores do pipeline de modelagem.

Seguindo para a normalização, ela foi conduzida com Python e a biblioteca Pandas. O procedimento foi automatizado para simplificar a análise. A normalização foi conduzida utilizando a linguagem Python e a biblioteca **Pandas**, com foco na automação do processo para aplicação em larga escala aos dados morfocinéticos dos embriões. A transformação foi aplicada especificamente às variáveis numéricas previamente selecionadas na fase de escolha de atributos, de forma a garantir consistência na entrada do modelo.

A sequência de passos executados pelo código para a normalização dos dados foi:

1. **Análise dos Dados:** Os dados foram inicialmente importados do arquivo Excel que possui os dados dos embriões, usando a função `pd.read_excel()`.
2. **Identificação das Colunas Numéricas:** Foi aplicada uma função para identificar automaticamente as colunas do tipo numérico. Em seguida, a variável-alvo “`Ploidia`” foi removida da lista, já que não deve passar pelo processo de normalização.
3. **Cálculo da Média e Desvio Padrão:** Para cada variável numérica, a média e o desvio padrão foram calculados.
4. **Utilização do Z-Score:** Para cada valor de uma variável, a média é subtraída e o desvio padrão é dividido, de acordo com a fórmula. Isso modifica os dados para que todas as variáveis apresentem média zero e desvio padrão igual a 1.
5. **Armazenamento dos Dados Normalizados:** Depois de aplicar o Z-Score, os dados normalizados foram guardados em uma nova planilha Excel, denominada “`Planilha_normalizada.xlsx`”.

A escolha do método Z-Score justifica-se por sua capacidade de padronizar variáveis com distribuições distintas, facilitando a convergência de algoritmos de aprendizado

de máquina e aumentando a comparabilidade entre atributos (JAISWAL, 2024). Além disso, essa padronização é especialmente recomendada em modelos que são sensíveis à escala dos dados, como Redes Neurais Artificiais e Máquinas de Vetores de Suporte (SVM). Ao final do processo, o conjunto de dados normalizado encontra-se adequado para ser utilizado nas etapas seguintes do pipeline de modelagem.

Análise da Planilha Normalizada

Depois de realizar a normalização, a Tabela 3 gerada apresenta os dados convertidos para que cada variável possua uma média de 0 e um desvio padrão de 1.

Tabela 3. Planilha Normalizada

Kidscore	t2	t3	t4
-0,1651468441	-1,04331772	-0,2061446738	-0,1365241993
-0,3621219891	0,0828611265	0,1102377935	-0,04951654895
1,509141888	-0,7805426557	-0,4547308981	-0,8035828516
-0,1651468441	0,233018306	0,08769304581	-0,1365241993
1,75536082	-0,4426890018	-0,4095334028	-0,8909505013
1,410654316	-0,9306998352	-0,6129221317	-1,151613453
0,8689726671	-1,231014194	-0,5677246364	-1,064605802
-0,3128782029	-0,8180819506	-0,4547308981	-0,7745803015
0,9182164534	-1,11839631	-0,9067058513	-1,586651704
-0,06665927163	0,833647024	0,607410242	0,7625458359
-0,5095833479	0,833647024	1,217576429	1,197593105
-1,888679363	0,3831754855	0,0650429815	-0,07851909905

Fonte: Autoras (2024)

Isso reduz o efeito das variações de escala entre as variáveis e simplifica a compreensão dos dados. A tabela normalizada foi verificada e todos os valores foram ajustados de acordo com a fórmula do Z-Score, com a média e o desvio padrão calculados para cada linha. Por exemplo, a variável "t2" tinha valores que oscilavam entre 19 e 38 horas antes da normalização, indicando uma ampla dispersão na faixa de dados. Depois de normalizados, os dados da variável foram convertidos em uma escala com média zero e desvio padrão de 1. Isso permitiu que todos os valores da variável estivessem comparáveis, mesmo com a grande variação original.

Efeito da Normalização na Comparabilidade das Variáveis

O principal benefício de normalizar as variáveis é que agora é possível compará-las de forma mais justa. Antes da normalização, variáveis como o "t8"(que apresentava uma

amplitude maior) poderiam impactar de forma desproporcional o modelo de aprendizado de máquina, enquanto variáveis como o "t2", com uma amplitude menor, poderiam ser desconsideradas. Com a normalização, todas as variáveis estão na mesma escala, possibilitando uma avaliação justa de cada uma durante o treinamento do modelo.

Este efeito é especialmente relevante em algoritmos que lidam com escalas de variáveis, como as redes neurais. Com os dados normalizados, as conexões entre as variáveis podem ser examinadas de forma mais nítida, simplificando a elaboração de modelos preditivos mais confiáveis.

4.1.1.4 Atividade 3 (A3): Identificação da Correlação entre os Parâmetros na Previsão da Ploidia do Embrião

Após os dados estarem normalizados, iniciamos a utilização do coeficiente de **correlação de Spearman**, explicado no APÊNDICE E, que foi escolhido para esta atividade devido à necessidade de reconhecer relações entre os parâmetros documentados nos dados dos embriões e a porcentagem de euploidia. A metodologia leva em conta a classificação ordinal das observações, minimizando os efeitos de valores atípicos e permitindo a análise de variáveis com distribuições não normais (SOUSA, 2019).

O programa em Python criado para essa análise foi desenvolvido para ser modular, eficaz e gerar resultados claros tanto em formato visual quanto textual. Ele emprega as bibliotecas *Pandas*, *SciPy*, *Matplotlib* e *python-docx* para manipular dados, determinar correlações, produzir gráficos e elaborar relatórios em Word. A seguir, detalha-se a lógica do código:

- **Carregamento e Preparação dos Dados:** As informações dos embriões foram processadas a partir da Planilha de Dados Refinada, utilizando todas as colunas disponíveis. Essa etapa garante que todas as variáveis relevantes sejam consideradas.
- **Cálculo da Correlação de Spearman:** O coeficiente de Spearman foi calculado para todas as combinações possíveis de variáveis, utilizando a função `spearmanr` da biblioteca SciPy. O método "omit" foi empregado para lidar com valores ausentes, garantindo maior robustez mesmo que não existam dados faltantes nesta planilha.
- **Criação de Gráficos de Dispersão:** Para cada par de variáveis, foram criados gráficos de dispersão que auxiliam na análise numérica e permitem a identificação visual de padrões. As cores foram padronizadas, com a variável 1 (`var1`) em verde claro e marcador "o", e a variável 2 (`var2`) em azul escuro e marcador "x".
- **Elaboração de Relatório Automatizado:** O relatório gerado contém descrições textuais dos coeficientes e gráficos correspondentes. Este documento automatizado facilita a comunicação visual e escrita dos resultados.

O código analisa a correlação entre as variáveis numéricas em uma sequência de dados, empregando o coeficiente de Spearman. Ele inicia carregando a planilha Excel com as informações dos embriões e seleciona todas as colunas relevantes para a análise. O coeficiente de Spearman é calculado para cada par de variáveis, avaliando a intensidade e a direção da relação monótona entre elas. Os resultados são armazenados em um **DataFrame**, que é uma estrutura de dados bidimensional do Pandas, semelhante a uma tabela, permitindo fácil manipulação e análise. Esta tabela é então exportada para um arquivo Excel chamado **correlation_results.xlsx**.

Após isso, o programa gera gráficos de dispersão para observar as correlações, ressaltando os padrões das variáveis associadas. Esses gráficos, juntamente com os coeficientes de correlação, são automaticamente incorporados a um documento Word. No final, temos um relatório completo com as análises e visualizações, salvo como um arquivo chamado **relatorio_correlacoes.docx**, que contém todos os gráficos de dispersão para as combinações examinadas e valores de correlação relacionados.

Em seguida, apresentaremos a avaliação dos resultados mais relevantes obtidos a partir dos dois documentos gerados pelo código, destacando as variáveis com correlações mais significativas, bem como aquelas com pouca ou nenhuma interação. Para uma visão mais abrangente dos dados, a matriz de correlação entre todas as variáveis está disponível no Anexo IV.

Idade

A correlação com **t4 (-0,15)** **sugere que a idade exerce uma influência negativa leve sobre eventos específicos do desenvolvimento embrionário**. Mulheres mais velhas podem apresentar embriões com ligeiro atraso no tempo necessário para atingir o estágio *t4*. Em relação ao tempo para 5 células, *t5* (0.11), obteve uma correlação positiva. **Isso sugere uma tendência muito sutil de que embriões em fases mais avançadas estejam ligados a mães de mais idade.**

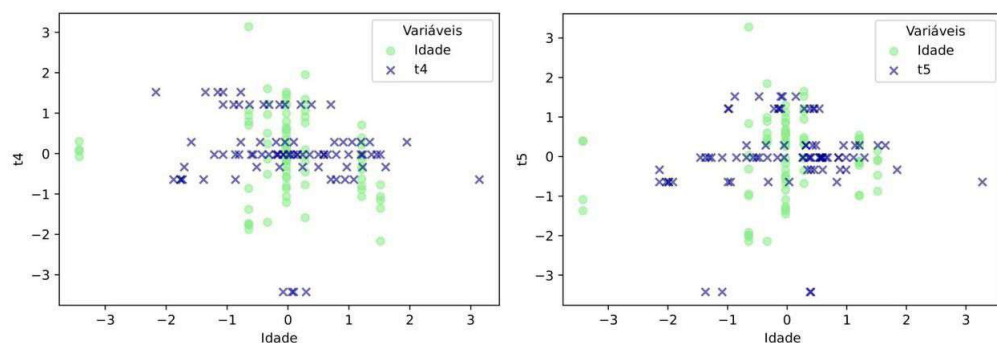


Figura 5. Dispersão entre Idade e t4 - Coeficiente de Spearman: -0.15 e entre Idade e t5 - Coeficiente de Spearman: 0.11

Ao reparar em alguns índices negativos, observamos os índices **tSB (-0.10)**, **cc2 (t3-t2) (-0.17)**, **s2 (t4-t3) (-0.24)**, **s3 (t8-t5) (-0.28)** e a **Ploidia(-0.50)**. Nota-se que os coeficientes mais próximos de 0 (como -0,10 a -0,28) indicam uma correlação negativa fraca. Isso traduz que há uma tendência muito sutil de que, quando uma variável aumenta, a idade, a outra diminui. Em *tSB* Figura 6, o coeficiente negativo insinua que o tempo de formação inicial da blastocisto tende a ser menor em embriões provenientes de mães mais velhas. Na variável *cc2*, Figura 6, a correlação desfavorável sugere uma maior irregularidade no intervalo entre a segunda e a terceira divisão celular (t2 para t3) em embriões de mães mais velhas. Em *s2* (Figura 6) e *s3* (Figura 6), mostra que a idade materna também está ligada a uma diminuição na eficácia do intervalo entre as divisões celulares de t3 para t4. Isso indica um efeito na fase intermediária do ciclo celular. Nos embriões de mães mais velhas, o período entre a fase de 8 células e a formação do blastocisto final é estendido, sinalizando obstáculos no progresso dessas fases. Todos esses atrasos podem ser cruciais, já que fases iniciais bem sincronizadas são fundamentais para um desenvolvimento embrionário adequado, mostrando como uma idade avançada pode afetar o desenvolvimento embrionário, afirmando a bibliografia estudada, fato já citado por [Junior \(2022\)](#), que reitera que **a idade materna exerce maior influência sobre a qualidade embrionária**.

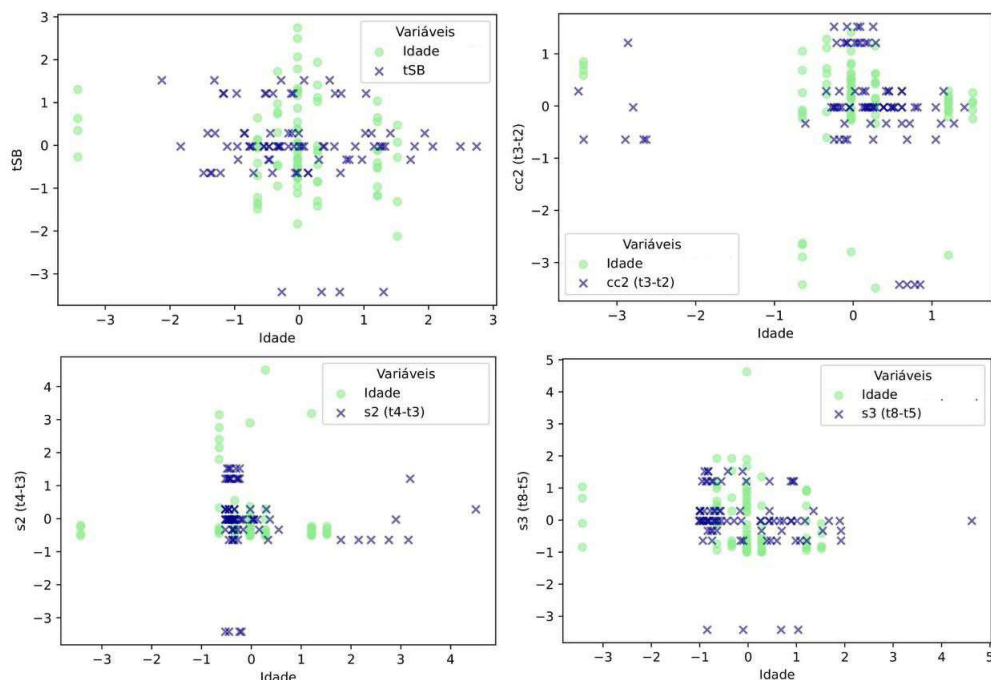


Figura 6. Dispersão entre Idade e tSB - Coeficiente de Spearman: -0.10 | Dispersão entre Idade e cc2 (t3-t2) - Coeficiente de Spearman: -0.15 | Dispersão entre Idade e s2 (t4-t3) - Coeficiente de Spearman: -0.24 | Dispersão entre Idade e s3 (t8-t5) - Coeficiente de Spearman: -0.28

E por fim temos a *Ploidia (-0,50)*, a correlação negativa mais relevante entre to-

das as outras variáveis demonstrando que **a alta idade materna está ligada a uma diminuição na proporção de embriões euploides**. Esta informação indica que **embriões de mães mais velhas contêm uma proporção reduzida de euploidia, o que pode estar ligado a uma queda na qualidade genética dos embriões**. Portanto, a idade materna é um dos elementos chave na diminuição da euploidia embrionária, alinhando com as informações de *Fertility and Ageing* (BAIRD et al., 2005), que também reitera que o aumento da aneuploidia em embriões está diretamente associado ao envelhecimento materno.

Estágio

Os coeficientes de correlação de Spearman entre os *estágios* e os tempos de transição celular (t_2 , t_3 , t_4 , t_5 , t_8) revelam relações positivas moderadas. Esses dados indicam que, **conforme o embrião progride para estágios mais avançados, os tempos associados às divisões celulares iniciais tendem a aumentar de forma sutil**. Esses resultados sugerem que o progresso do estágio embrionário está associado a um padrão de desenvolvimento, possivelmente, mais metódico nos estágios iniciais e intermediários do ciclo celular.

As correlações positivas observadas entre o *estágio* embrionário e os tempos tSC ($0,50$), tSB ($0,53$) e tB ($0,59$) indicam uma conexão significativa entre o progresso do estágio embrionário e o desenvolvimento contínuo das estruturas celulares. O incremento nos índices de correlação sugere que, conforme **o embrião avança para fases mais desenvolvidas, há uma maior regularidade no cumprimento dos marcos temporais dessas transições**. Isso implica que o estágio embrionário atua como um indicador abrangente da qualidade e da evolução do desenvolvimento embrionário.

A relação inversa entre o *estágio* de desenvolvimento e a *ploidia* ($-0,24$) indica que, **embora essa correlação seja moderada, sugere que o estágio de desenvolvimento pode ter um impacto negativo na ploidia**.

Morfo

A avaliação da variável *Morfo*, que categoriza a expansão morfológica dos embriões em fases iniciais, revelou correlações moderadamente negativas com variáveis temporais como t_2 ($-0,38$), t_3 ($-0,43$) e t_5 ($-0,45$). Esses achados sugerem que **embriões com características morfológicas menos expansivas costumam apresentar atrasos nos primeiros tempos de divisão celular**. Dessa forma, os dados estão alinhados com a literatura, que cita que a morfologia do embrião é um fator determinante para o potencial de implantação e a qualidade embrionária Capalbo et al. (2014).

Ao examinarmos os intervalos $cc2$ (t_3-t_2) ($-0,38$) e $cc3$ (t_5-t_3) ($-0,31$), os gráficos

corroboram essa tendência, destacando que embriões que apresentam alterações morfológicas menos favoráveis quando enfrentam flutuações no ritmo de desenvolvimento celular, a medida que uma variável aumenta, a outra vem a diminuir. Esses resultados enfatizam a relevância de uma divisão celular sincronizada para preservar características morfológicas ideais, evidenciando como o ritmo do ciclo celular influencia diretamente a qualidade estrutural do embrião.

A correlação entre *Morfo* e *Ploidia* exibe um coeficiente de correlação de 0.05 , sugerindo uma ligação positiva, porém extremamente fraca. Isso sugere que, neste conjunto de dados, as características morfológicas dos embriões não parecem estar diretamente ligadas à euploidia.

t2, t3, t4, t5 e t8

Existe uma forte interdependência entre os tempos de transição celular, como evidenciado pela correlação de 0.89 entre $t2$ e $t4$, indicando que esses eventos de desenvolvimento estão fortemente associados, à medida que uma variável aumenta, a outra também aumenta de forma consistente. A correlação entre $t3$ e $t2$ (0.78), $t4$ e $t5$ (0.56) e entre $t5$ e $t8$ (0.52) também demonstra alinhamento nas fases iniciais e intermediárias do desenvolvimento embrionário.

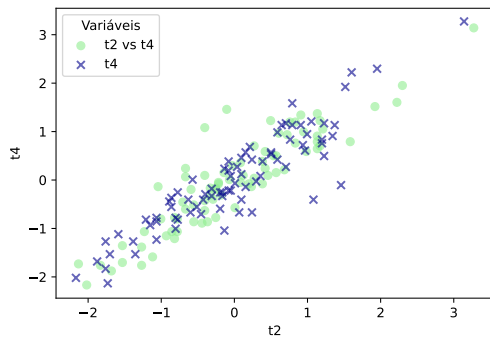


Figura 7. Dispersão entre t2 e t4 - Coeficiente de Spearman: 0.89

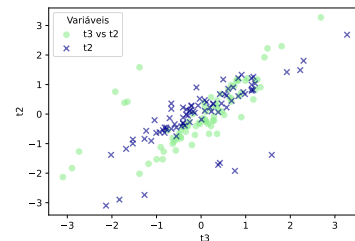


Figura 8. Dispersão entre t3 e t2 - Coeficiente de Spearman: 0.78

Fonte: Autoras (2025)

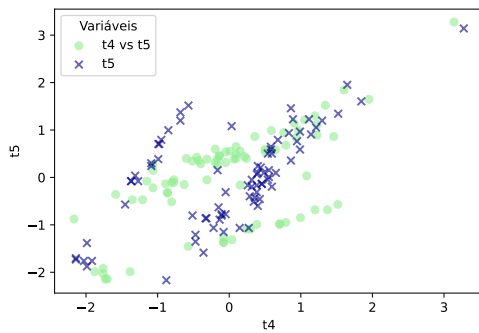


Figura 9. Dispersão entre t_4 e t_5 - Coeficiente de Spearman: 0.56

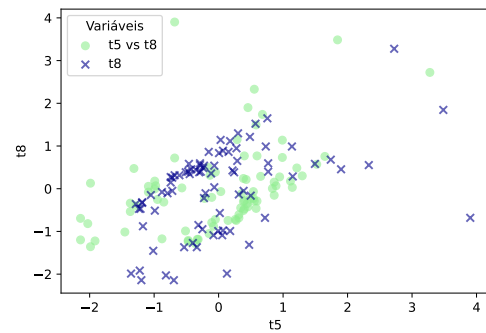


Figura 10. Dispersão entre t_5 e t_8 - Coeficiente de Spearman: 0.52

A correlação quase nula entre t_2 (-0.08), t_3 (-0.06), t_4 (-0.02), t_8 (-0.07) e a *ploidia*, demonstram uma falta de correlação entre essas variáveis para a determinação de um embrião euploidia ou aneuplóide. Já **t_5 e *ploidia* (-0.24)** indica que atrasos nesse estágio de desenvolvimento, estão ligados a uma qualidade genética inferior (menor *ploidia*), alinhando com a literatura, que destaca t_5 como o indicador mais relevante do potencial de implantação [Cruz et al. \(2012\)](#).

tSC

Ao relacionar a variável de Tempo de formação do estágio de clivagem sincronizada (Time to Synchronized Compaction) com as variáveis de tempos de transição celular, t_2 (0.40), t_3 (0.42), t_4 (0.43), t_5 (0.35) e t_8 (0.35), indicam uma relação de intensidade moderada e positiva durante o desenvolvimento embrionário, sendo mais acentuada nos estágios iniciais e intermediários. Esses padrões indicam que o **tSC tem um papel mais significativo nas etapas iniciais e intermediárias do desenvolvimento, diminuindo seu impacto progressivamente conforme o tempo passa.**

As correlações de *tSC* com *tSB* (0.75) e *tB* (0.74) demonstram uma forte ligação entre essas variáveis, indicando que estão fortemente conectadas, em que a medida que uma variável aumenta, a outra também aumenta de forma consistente. Isso indica uma relação quase linear, em que mudanças no *tSB* impactam diretamente no *tSC*, evidenciando uma ligação funcional direta entre os dois acontecimentos.

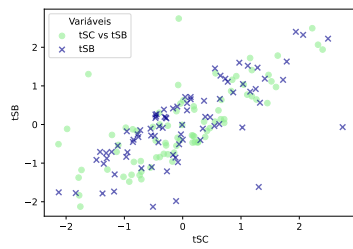


Figura 11. Dispersão entre tSC e tSB -
Coeficiente de Spearman: 0.75

Fonte: Autoras (2025)

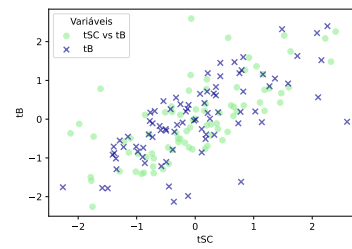


Figura 12. Dispersão entre tSC e tB -
Coeficiente de Spearman: 0.74

Fonte: Autoras (2025)

Os coeficientes de correlação de tSC com os intervalos de tempo *cc2* ($t3-t2$) (0.41), *cc3* ($t5-t3$) (0.24) e *t5-t2* (0.34) apresentam relações positivas que variam de moderadas a fracas, sinalizando variados níveis de concordância com esses tempos acumulados. Estes achados indicam que a influência do tSC se **reduz** conforme os períodos de tempo se estendem.

Por fim, a correlação com a *Ploidia* (-0.04), é fraca e negativa, sugerindo que não existe uma conexão relevante entre essas variáveis. Este resultado indica que *Ploidia* pode estar ligada a processos diferentes que não afetam diretamente *tSC*.

tSB

Ao analisar sua realção com *t2* (0.43), *t3* (0.56), *t4* (0.57), *t5* (0.39) e *t8* (0.40) mostram relações positivas de intensidade moderada, conforme uma variável aumenta, a outra segue um aumento consistente, aumentando nos estágios iniciais e intermediários e apresentando um ligeiro declínio nos estágios posteriores. A relação com *t2* **indica que os eventos iniciais exercem um efeito significativo no comportamento de tSB**. A forte ligação com *t3* e *t4* indica que **esses estágios intermediários têm uma influência mais significativa sobre tSB**. Em contrapartida, as correlações moderadas com *t5* e *t8* sugerem uma redução na conexão nos estágios subsequentes, evidenciando um efeito menos significativo de *tSB* conforme os processos progridem no desenvolvimento.

As correlações entre *tSB* e *tSC* (0.75), *tB* (0.93) e *cc2* ($t3-t2$) (0.63) evidenciam ligações fortes e relevantes, indicando uma interdependência entre essas variáveis. A forte ligação com **tSC** sugere que alterações em uma variável têm um **impacto direto** na outra. **A correlação com tB indica uma dependência quase total entre as duas variáveis**, indicando que são praticamente comparáveis em termos de sua interação. Ademais, a correlação robusta com *cc2* ($t3-t2$) indica que **o período entre t3 e t2 tem um impacto considerável sobre tSB**, possivelmente por causa do efeito direto dos eventos temporais iniciais no comportamento dessa variável. O gráfico ilustra a forma que *tSB* e *tB* possuem uma relação de crescimento mútuo entre si, quase

se espelhando. Dessa forma, alterações em tB podem ser significativas para mudanças em tSB.

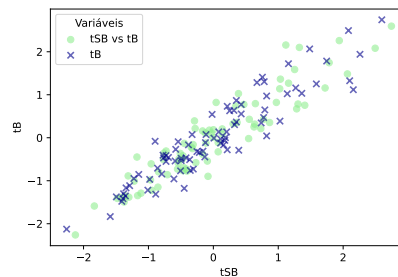


Figura 13. Dispersão entre tSB e tB - Coeficiente de Spearman: 0.93

O coeficiente de correlação de *tSB* com *Ploidia* (-0.11) é fraco e negativo, sugerindo que não existe uma conexão evidente ou relevante entre essas variáveis. Este resultado indica que *Ploidia* é afetada por elementos diferentes, que não estão diretamente ligados aos processos ligados à *tSB*.

tB

As correlações entre *tB* e os tempos de desenvolvimento *t2* (0.40), *t3* (0.48), *t4* (0.50), *t5* (0.35) e *t8* (0.37) exibem relações moderadamente positivas, sugerindo que *tB* está ligado a eventos temporais durante o ciclo embrionário, sendo mais intensa nos estágios intermediários.

As correlações de *tB* com os intervalos de tempo *cc2* ($t3-t2$) (0.54) e *tSC-t8* (0.47) ressaltam a influência de processos particulares no comportamento dessa variável. A correlação moderada-alta com *cc2* ($t3-t2$) indica que **os períodos de transição entre t3 e t2 têm uma influência relevante no comportamento de tB**, sugerindo que as dinâmicas nos estágios iniciais influenciam diretamente o desenvolvimento subsequente. Em contrapartida, a correlação moderada entre *tSC* e *t8* indica que a diferença temporal entre **tSC e t8** está positivamente ligada a *tB*.

A fraca e negativa correlação entre *tB* e *Ploidia* (-0.17) indica uma conexão fraca entre essas variáveis. A direção negativa sugere que aumentos em *tB* podem estar ligados a pequenas diminuições em *Ploidia*. Esta associação tênue enfatiza que *tB* não desempenha um papel crucial no comportamento da *Ploidia*, porém o afeta fracamente de forma negativa.

cc2 (t3-t2)

A relação entre o intervalo *cc2* e as diversas fases de crescimento do embrião, simbolizadas pelos tempos *t2*, *t3*, *t4*, *t5* e *t8*, mostra variações de intensidade. A correlação

com t_2 (0.39) é moderada e positiva. Com t_3 (0.80), se nota uma correlação positiva significativa, demonstrando um **alinhamento quase imediato entre cc2** e o momento de desenvolvimento correspondente a t_3 , sugerindo uma **conexão robusta entre ambos**. Por outro lado, a correlação com t_4 (0.59) é moderadamente alta, indicando que **cc2 é também impactado por eventos subsequentes representados por t_4** . A conexão com t_5 (0.64), também moderada-alta, **evidencia a persistência da influência de eventos temporais**, evidenciando que t_5 desempenha um papel significativo na progressão de cc2. Por fim, com t_8 (0.46), a correlação é moderada, porém o efeito de t_8 sobre cc2 é menos significativo do que nos estágios anteriores, **sugerindo uma dependência que diminui conforme o embrião progride para fases mais avançadas de desenvolvimento**. O gráfico abaixo mostra a maior relação que cc2 possui durante as fases de crescimento, que é na fase **t_3 (0.80)**, em que a variável cc2 (t_3-t_2), cresce de forma consistente à medida que a outra também cresce.

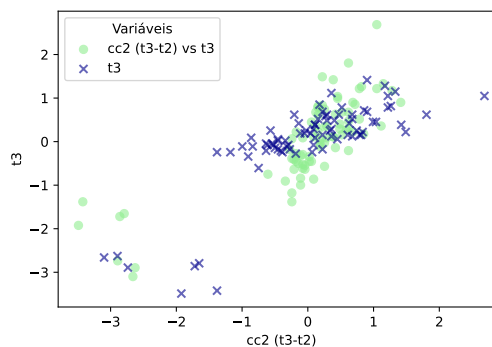


Figura 14. Dispersão entre cc2 (t_3-t_2) e t_3 - Coeficiente de Spearman: 0.80

A correlação entre *cc2* e *Ploidia* (-0.03) é muito fraca e negativa, sugerindo que não existe uma conexão relevante entre esses dois componentes. Isso indica que **o período temporal entre t_3 e t_2 não tem um impacto significativo na Ploidia**, o que sugere que o desenvolvimento temporal do embrião, simbolizado por cc2, não está diretamente ligado à ploidia do embrião.

cc3 (t_5-t_3)

A correlação com t_2 (0.15) é fraca e positiva. Com t_3 (0.25), a correlação é moderada e positiva, mostrando um ligeiro crescimento em cc3 à medida que t_3 se eleva. A correlação com t_4 (0.26) é igualmente moderada e positiva. A correlação com t_5 (0.81) é extremamente forte e positiva, sugerindo que **conforme t_5 cresce, cc3 também cresce quase de maneira linear**. Isso é coerente, já que cc3 é determinado como a diferença entre t_5 e t_3 , e qualquer alteração em t_5 afeta diretamente cc3. Por fim, a correlação com

o **t8 (0.40)** é moderada e positiva, indicando que, conforme o t8 se eleva, o cc3 tende a crescer.

A correlação entre o intervalo *t5-t3* (*representado por cc3*) e o intervalo *t5-t2* (*0.90*) é extremamente forte e positiva, **sugerindo que o comportamento de cc3 está fortemente ligado ao intervalo t5-t2**. Portanto, o intervalo t5-t2 pode ser um indicador relevante para prever o comportamento de cc3, indicando que **a evolução do embrião entre os estágios t5 e t2 desempenha um papel relevante no comportamento de cc3 (t5-t3)**. Isso indica que alterações nesse intervalo temporal afetam significativamente a diferença de tempo entre os estágios t5 e t3.

A correlação de cc3 com a *Ploidia* (*-0.281*) é bastante elevada quando comparada a outras variáveis, como a correlação de Ploidia com t2 (*-0.075*) ou t5 (*-0.237*). Isso sugere que **cc3 exerce uma influência moderada e negativa sobre a Ploidia, indicando que alterações na cc3 podem ter um impacto relevante sobre a ploidia**. Este efeito adverso indica que alterações em cc3 estão ligadas a uma diminuição no valor da ploidia.

t5-t2

A correlação com *Morfo* (*-0.38*) é moderadamente negativa, sugerindo que, conforme o valor de Morfo cresce, o intervalo t5-t2 tende a se estreitar. Isso indica que **as características morfológicas do embrião podem influenciar o comportamento temporal entre os estágios t5 e t2**, possivelmente indicando um desenvolvimento mais devagar ou desordenado conforme as características morfológicas se modificam.

Com *t3* (*0.51*), a correlação é moderada e positiva, sugerindo uma correlação mais intensa entre os intervalos temporais dos estágios t5 e t3. A correlação com *t4* (*0.39*) também é moderada e positiva. Com *t5* (*0.92*), a correlação é extremamente intensa e positiva, sinalizando que **o período entre t5 e t2 está fortemente ligado ao tempo de t5**. Isso é esperado, pois t5 representa o término desse intervalo. A correlação robusta indica que **alterações em t5 afetam consideravelmente o intervalo entre t5 e t2**. A correlação com *t8* (*0.45*) é positiva e moderada, indicando uma conexão entre t8 e o intervalo t5-t2.

Com *tSB* (*0.40*) e *tB* (*0.37*), as correlações são moderadas e positivas, indicando que **a mudança nos tempos tSB e tB também afeta o intervalo entre t5 e t2**. No caso de *s3* (*indicado por t8-t5, -0.41*), a correlação é moderada e negativa, indicando que, **conforme o intervalo t8-t5 cresce, o intervalo t5-t2 tende a se reduzir, indicando um possível comportamento compensatório entre os dois períodos de tempo**.

O coeficiente de correlação de **Ploidia (-0.23)** com t5-t2 é moderado e negativo,

sugerindo **uma relação inversa entre as alterações na ploidia e o intervalo entre t5 e t2**. Isso indica que, **conforme t5-t2 evolui, a ploidia tende a se reduzir, indicando uma possível restrição da ploidia no comportamento temporal entre esses estágios**.

s2 (t4-t3)

A correlação com *t3* (-0.22) é moderada e negativa, sugerindo que conforme o valor de *t3* cresce, a distância s2 (t4-t3) tende a se afilar. Isso indica que uma alteração no tempo *t3* pode diminuir a distância entre *t4* e *t3*, sugerindo que as alterações no tempo *t3* estão um pouco inversamente ligadas a esse intervalo.

A correlação com *cc2* (*t3-t2*) (-0.21) é moderada e negativa. Isso pode sugerir uma relação inversa entre o comportamento temporal dos intervalos t4-t3 e t3-t2, com alterações em *cc2* impactando negativamente o intervalo s2, possivelmente por causa de um comportamento mais rápido no começo do desenvolvimento do embrião.

Com uma correlação moderada e positiva de *s3* (*t8-t5*) (0.25), indica que um crescimento no intervalo s3 está ligado a um crescimento no intervalo s2. Isso pode sugerir que **o aumento do intervalo entre t8 e t5 está ligado ao aumento do intervalo entre t4 e t3**, o que pode representar uma compensação nos tempos entre as diversas fases do embrião.

A correlação com *Ploidia* (0.11) é leve e positiva, sugerindo que alterações em s2 exercem um efeito levemente positivo na ploidia. Apesar do efeito ser mínimo, ele propõe que **uma alteração em s2 pode levar a uma pequena alteração na ploidia**.

s3 (t8-t5)

A correlação com *t5* (-0.39) é moderada e negativa. Isso indica que **o período entre t8 e t5 diminui à medida que o estágio t5 progride, sugerindo que uma elevação em t5 pode levar a uma aceleração do progresso**, reduzindo o intervalo subsequente até t8.

Com *t8* (0.48), a correlação é positiva e moderada mais alta que *t5*, indicando que, conforme *t8* se eleva, o intervalo s3 também tende a se expandir. Isso sugere que um aumento no intervalo de tempo entre *t8* e *t5* está ligado a um incremento no valor de *t8*, sinalizando um processo de evolução mais extenso ou mais lento em comparação com outros intervalos. **Este resultado indica que, conforme o embrião se aproxima do estágio t8, o período de tempo entre os estágios t8 e t5 aumenta**, assim demonstrado no gráfico abaixo, em que s3 está em ver claro.

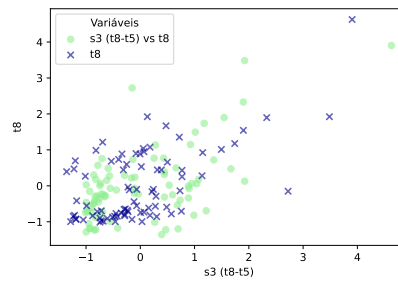


Figura 15. Dispersão entre $s3$ ($t8-t5$) e $t8$ - Coeficiente de Spearman: 0.48

A correlação com $cc3$ ($t5-t3$) (-0.36) é moderada e negativa. Este efeito indica que uma extensão prolongada entre $t5$ e $t3$ pode estar ligada a uma diminuição no tempo entre $t8$ e $t5$.

Finalmente, a correlação com a *Ploidia* (0.25) é moderada e positiva, indicando que, conforme a ploidia cresce, o intervalo $s3$ ($t8-t5$) também tende a se expandir. Este efeito pode sugerir que a **aneuploidia está ligado a uma diminuição no tempo de desenvolvimento, estendendo o intervalo entre $t8$ e $t5$.**

tSC-t8

A relação com $t8$ (-0.29) é moderadamente negativa sugere que, conforme o intervalo de tempo entre tSC e $t8$ cresce, o tempo de $t8$ também tende a crescer. Este resultado indica que, **ao aumentar o intervalo entre tSC e $t8$, pode ocorrer uma diminuição na velocidade do desenvolvimento embrionário entre tSC e $t8$, estendendo o período até $t8$.** Isso pode indicar uma etapa de crescimento mais lenta ou prolongada, influenciando o ritmo geral de avanço do embrião.

A correlação com tSC (0.70) é forte e positiva, que indica que o intervalo $tSC-t8$ tem uma ligação direta com o intervalo tSC . **Conforme o intervalo de tempo entre tSC e $t8$ se amplia, o intervalo tSC também tende a se expandir. Este efeito indica uma relação temporal na qual o estágio tSC desempenha um papel significativo no progresso do embrião no período subsequente até $t8$.**

A correlação de $tB-tSB$ (0.20) é positiva e indica que um crescimento no intervalo entre tB e tSB pode estar estreitamente ligado a um crescimento no intervalo $tSC-t8$.

A correlação muito tênue (0.04) e positiva sugere que *Ploidia* exerce uma influência quase nula no intervalo $tSC-t8$.

tB-tSB

A correlação com a *Idade* (0.20) sendo positiva e fraca indica que, conforme a idade avança, o intervalo tB-tSB também tende a crescer, mesmo que de maneira sutil. Isso pode sugerir que a idade da paciente pode ter um pequeno impacto no desenvolvimento embrionário, possivelmente indicando uma variação no ritmo de desenvolvimento.

A correlação moderadamente positiva do *Estágio* (0.38) indica que, conforme o estágio de desenvolvimento progride, o intervalo tB-tSB tende a se estender.

A correlação moderadamente positiva de *tB* (0.38) indica que, conforme o intervalo tB se amplia, o intervalo tB-tSB também tende a se expandir. Isso é coerente, já que **o período tB é um estágio inicial do desenvolvimento embrionário, e sua duração pode impactar diretamente o tempo até o estágio seguinte, o tSB**, impactando a evolução do desenvolvimento embrionário. Uma maior extensão do tB geralmente indica um período mais extenso até o tSB.

A correlação com *Ploidia* (-0.284) exerce um efeito significativo, visto que **é a segunda correlação de maior magnitude entre as variáveis examinadas**. Desempenha um papel significativo, uma vez que indica que a qualidade genética do embrião, avaliada por tB e tSB, pode impactar diretamente a ploidia.

Ploidia

A avaliação das variáveis relacionadas à ploidia indica que **a idade materna tem o maior efeito na qualidade genética do embrião**. A idade materna, **com uma correlação de -0,50**, apresenta uma correlação inversa significativa com a ploidia. Isso significa que, **conforme a mulher envelhece, a quantidade de embriões euploides diminui, o que indica um crescimento na taxa de aneuploidia**. Isso está alinhado com o que a literatura já aponta sobre o impacto do envelhecimento materno na qualidade genética dos embriões, onde embriões de mulheres mais idosas têm maior probabilidade de conter erros cromossômicos.

Adicionalmente, as variáveis **tB-tSB e cc3 (t5-t3)** possuem uma correlação moderada de -0,28 com a ploidia, sugerindo que o período entre as fases de desenvolvimento **t5, t3, tB e tSB podem afetar diretamente a qualidade genética do embrião**. Este intervalo de tempo está ligado ao ritmo de divisão celular e ao comportamento dos cromossomos. Mudanças nesse processo podem comprometer a criação de embriões euploides, levando a um aumento na probabilidade de aneuploidia.

Outros fatores, como o **estágio de desenvolvimento (-0,24)** e o **t5 (-0,24)**, também mostraram uma correlação negativa com a ploidia. Isso indica que **o avanço para fases mais avançadas de desenvolvimento e o atraso no t5 podem estar ligados a uma queda na qualidade genética dos embriões**.

- 4.1.1.5 Atividade 4 (A4): Separar o conjunto de dados em conjuntos de treinamento, validação e teste, fazendo uma distribuição dos dados e aplicar técnica de aumento de dados

Divisão do Conjunto de Dados em Treinamento e Validação

Foi realizada a divisão do conjunto de dados normalizado em dois subconjuntos principais: **conjunto de treinamento** (80%) e **conjunto de validação** (20%). Essa segmentação foi feita com estratificação em relação à variável-alvo *Ploidia*, assegurando que a proporção entre embriões euploides e aneuploides fosse preservada em ambos os conjuntos.

A divisão foi implementada de forma aleatória, com o intuito de evitar vieses associados à ordenação dos dados. Essa aleatoriedade contribui para a geração de subconjuntos mais representativos e imparciais, como indicado por Ng (2018), e é particularmente relevante quando se trabalha com conjuntos de dados de pequeno porte.

Balanceamento das Classes no Conjunto de Treinamento

Para evitar que o modelo aprenda de forma enviesada devido a desbalanceamentos entre as classes, foi implementada uma etapa de **balanceamento do conjunto de treinamento**. O procedimento consistiu na aplicação de uma técnica inspirada em *under-sampling*: a classe majoritária foi reduzida até igualar-se à classe minoritária em número de amostras. No entanto, ao invés de descartar os dados excedentes, eles foram realocados para o conjunto de validação. Isso permitiu manter o conjunto de treinamento equilibrado sem perder as informações contidas nas instâncias adicionais.

Essa abordagem mitigou os riscos de sobreajuste (*overfitting*) para a classe dominante e fortaleceu a avaliação do modelo com um conjunto de validação mais representativo. Conforme discutido por Moriconi et al. (2024), esse tipo de técnica é especialmente útil em contextos biomédicos, onde o volume de dados pode ser restrito e a preservação da representatividade é essencial para garantir confiabilidade.

Aumento de Dados (*Data Augmentation*) utilizando o Método de Monte Carlo

Para aumentar a capacidade de generalização do modelo e reduzir o risco de *overfitting*, foi aplicada uma técnica de **aumento de dados** (*data augmentation*) baseada em amostragem estocástica — uma variação do método de Monte Carlo.

Este processo gerou novas amostras sintéticas com base nas propriedades estatísticas das variáveis numéricas do conjunto original. A geração seguiu os seguintes critérios:

- Para **variáveis numéricas**, os novos valores foram gerados a partir de uma dis-

tribuição normal, usando como parâmetros a média e o desvio padrão da variável original.

- Para **variáveis categóricas**, os valores foram replicados aleatoriamente com reposição, de forma proporcional à sua distribuição original.

A aplicação da média e do desvio padrão assegura que os novos dados numéricos correspondam às características do conjunto original, ao mesmo tempo que as categorias mantêm suas proporções, mantendo os padrões originais. A produção de dados por meio de distribuições estatísticas amplia a diversidade do conjunto sem adicionar ruído excessivo ou padrões inverossímeis, assegurando uma diversificação controlada. Este procedimento garante que o conjunto de treinamento ampliado seja robusto o suficiente para treinar o modelo, prevenindo sobrecargas e aprimorando a generalização, levando a um modelo mais eficiente e seguro.

Depois de executar o código, a base de dados de treinamento foi **ampliada em 20 vezes**, conforme o fator de aumento definido. As amostras sintéticas geradas foram concatenadas ao conjunto original, totalizando um volume significativamente maior de dados. Essa abordagem é especialmente indicada em contextos onde o conjunto de dados é escasso, como em estudos biomédicos, e contribui para a construção de modelos mais estáveis, com maior capacidade de reconhecer padrões complexos e realizar previsões confiáveis em novos dados.

4.2 Fase 2: Desenvolvimento e Avaliação do Modelo

4.2.1 OE2 - Treinar e Ajustar do Modelo de Machine Learning para Predição de Ploidia

4.2.1.1 Atividade 5 (A5): Desenvolvimento e Treinamento do Modelo de Machine Learning para Otimização da Predição de Euploidia

Nesta etapa do projeto, foi desenvolvido e treinado um modelo de Aprendizado de Máquina (Machine Learning) com o objetivo de otimizar a predição de ploidia em embriões, utilizando dados morfocinéticos extraídos do Time-Lapse System (TLS). O modelo escolhido foi uma Rede Neural Artificial do tipo Perceptron Multicamadas (MLP, do inglês Multilayer Perceptron), implementada com a biblioteca Scikit-Learn em linguagem Python.

A **Scikit-Learn** foi empregada para *diversas tarefas essenciais do pipeline de modelagem*, incluindo: a divisão dos dados em conjuntos de treino e teste com estratificação, a padronização dos atributos com **StandardScaler**, o treinamento da rede neural com **MLPClassifier** e a avaliação do modelo por meio de métricas como acurácia, curva ROC

e matriz de confusão. Além disso, a Scikit-Learn também foi responsável pela serialização do modelo e do scaler com a ferramenta `joblib`, permitindo seu reaproveitamento em outras etapas do sistema. Essa integração modular e coesa entre diferentes etapas do processo de modelagem é um dos principais pontos fortes da Scikit-Learn, se tornando particularmente adequada para projetos como esse, que exigem reprodutibilidade, clareza na implementação e robustez computacional (GÉRON, 2017).

A arquitetura de Rede Neural Artificial utilizada foi o **Perceptron Multicamadas** (MLP, do inglês *Multilayer Perceptron*), que é uma rede supervisionada composta por três tipos de camadas: uma de entrada, uma ou mais ocultas e uma de saída (HAYKIN, 2009). O modelo MLP foi construído utilizando a classe `MLPClassifier`, da biblioteca Scikit-Learn, que é um algoritmo de aprendizado supervisionado baseado em Redes Neurais Artificiais que utiliza o algoritmo de retropropagação do erro para realizar a otimização dos pesos durante o treinamento (GÉRON, 2017). Essa classe faz parte do módulo `sklearn.neural_network` e oferece uma interface padronizada para treinamento e predição, integrando facilmente ao restante do ecossistema da Scikit-Learn, como ferramentas de validação cruzada, seleção de atributos e ajuste de hiperparâmetros (GÉRON, 2017).

A configuração do modelo incluiu os seguintes hiperparâmetros principais:

- **activation='relu'**: Utiliza a função de ativação ReLU (Rectified Linear Unit), para evitar o problema do desvanecimento do gradiente (GÉRON, 2017).
- **solver='adam'**: Aplica o otimizador *Adam*, um método de descida do gradiente adaptativo eficiente para volumes de dados e parâmetros.
- **max_iter=1000**: Define o número máximo de iterações para o treinamento, garantindo convergência adequada.
- **random_state=attempt**: Usado para reprodutibilidade dos resultados.

Após o treinamento, os objetos do modelo e do *scaler* foram serializados em arquivos `.pkl` utilizando a ferramenta `joblib`. O **scaler** é um objeto responsável pela normalização dos dados, neste caso, utilizando a padronização por Z-Score (GÉRON, 2017), garantindo que todas as variáveis contribuam de forma equilibrada para o aprendizado. É fundamental aplicar exatamente a mesma transformação aos dados futuros que serão analisados pelo modelo, razão pela qual o objeto do *scaler* é salvo junto ao modelo treinado. Já o arquivo com extensão `.pkl` (do inglês *Pickle*) é utilizado para armazenar os objetos Python de forma persistente no disco. A serialização dos objetos, modelo treinado e scaler, permite que eles sejam carregados posteriormente sem a necessidade de repetir todo o processo de treinamento, otimizando tempo, desempenho e consistência da aplicação.

Foi incorporado ao pipeline do projeto o método *LIME* (Local Interpretable Model-Agnostic Explanations). O LIME é uma técnica de Inteligência Artificial Explicável (XAI — *Explainable Artificial Intelligence*) proposta por [Ribeiro, Singh e Guestrin \(2016\)](#), que visa interpretar modelos complexos de forma local e intuitiva. Sua principal função é explicar individualmente cada predição feita por modelos considerados “caixas-pretas”, como redes neurais. O LIME faz isso criando um modelo interpretable treinado em uma vizinhança gerada artificialmente em torno da instância original de entrada. Assim, ele estima quais atributos (características) mais influenciaram a decisão do modelo naquela predição específica ([RIBEIRO; SINGH; GUESTIN, 2016](#)).

Para tornar as predições do modelo mais compreensíveis para embriologistas e profissionais da área clínica, foi utilizada a ferramenta LIME (Local Interpretable Model-Agnostic Explanations), uma técnica de Inteligência Artificial Explicável (XAI) proposta por [Ribeiro, Singh e Guestrin \(2016\)](#). Essa técnica permite interpretar individualmente cada predição feita pela rede neural, identificando quais atributos mais contribuíram para a classificação de cada embrião como euploide ou aneuploide. No presente projeto, a porcentagem de euploidia de cada embrião foi estimada com base na probabilidade de predição da classe "euploide"(classe 1) fornecida pelo modelo MLP, por meio do método `predict_proba()`. Esse valor probabilístico foi convertido em percentual na planilha final gerada pelo sistema. Esse tipo de abordagem interpretável é especialmente importante em contextos biomédicos, onde decisões automatizadas exigem justificativas claras e auditáveis ([ARRIETA, 2020; RIBEIRO; SINGH; GUESTIN, 2016](#)).

Ao aplicar o modelo preditivo em novos dados, a entrada inicial, uma planilha contendo os dados morfofocinéticas dos embriões, passa por um processo de pré-processamento automatizado. Esse pipeline transforma variáveis originalmente categóricas, como classificações textuais da morfologia embrionária, em valores numéricos padronizados. Um exemplo dessa transformação é a classificação morfológica, que é convertida automaticamente em categorias numéricas que representam níveis de qualidade embrionária, conforme descrito na metodologia. Após essa etapa, os dados numéricos são submetidos a um escalonamento por meio de um *scaler*, que normaliza as variáveis para uma mesma escala, evitando que diferenças de magnitude influenciem indevidamente a predição. Em seguida, o modelo de Rede Neural MLP realiza a predição da classe para cada embrião, classificando-os como euploides ou aneuploides. Assim, o modelo fornece a probabilidade associada à predição da classe euploide, expressa em termos percentuais. Finalmente, os resultados (que incluem as informações originais da planilha, a classificação prevista e a probabilidade de euploidia) são exportados para um novo arquivo. Essa saída organizada facilita a interpretação dos dados e suporta a tomada de decisão clínica, tornando o processo mais transparente e acessível.

4.2.2 OE3 - Realizar Avaliação do Modelo

4.2.2.1 Atividade 6 (A6): Utilizar métricas adequadas para medir o desempenho do modelo

Nesta etapa, foram aplicadas métricas quantitativas para avaliar o desempenho do modelo de aprendizado de máquina desenvolvido para a predição de euploidia em embriões. A **acurácia** de 88,2% indica que o modelo classificou corretamente a grande maioria dos embriões, tanto euploides quanto aneuploides, mostrando uma performance geral satisfatória. A **Área sob a Curva ROC (AUC)** de 0,944 evidencia a elevada capacidade do modelo em distinguir entre as classes, mesmo ao variar o limiar de decisão.

O **recall**, que mensura a taxa de verdadeiros positivos, apresentou resultados diferenciados para as classes: foi perfeito para embriões aneuploides (1,00), garantindo a identificação correta de todos os casos com anomalias genéticas, e robusto para embriões euploides (0,75), indicando uma boa sensibilidade, ainda que com espaço para melhorias. Esses indicadores superaram os critérios mínimos estabelecidos para aceitação do modelo.

A matriz de confusão gerada (Figura 16) revela que, dentre os 17 embriões avaliados na base de teste, 9 embriões aneuploides foram corretamente classificados (verdadeiros negativos) e 6 euploides foram corretamente identificados (verdadeiros positivos). Apenas dois casos de euploidia foram incorretamente classificados como aneuploides (falsos negativos) e nenhum aneuploide foi erroneamente identificado como euploide (falso positivo).

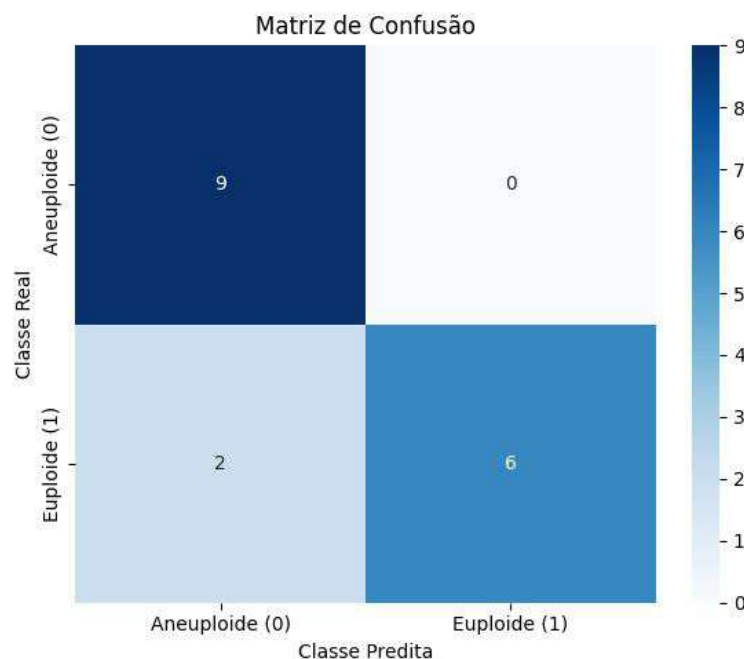


Figura 16. Matriz de confusão para avaliação do desempenho do modelo

A curva ROC (Receiver Operating Characteristic), ilustrada na Figura 17, expressa graficamente a performance do classificador sob diferentes limiares de decisão. A área sob

a curva ROC (AUC) foi de 0,94, o que indica excelente capacidade do modelo em distinguir entre as duas classes.

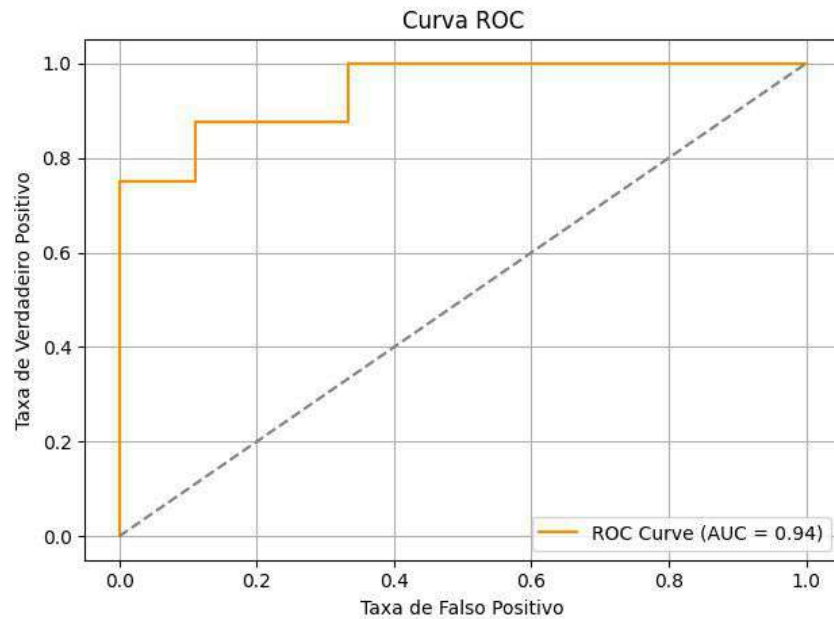


Figura 17. Curva ROC com AUC de 0,94

Com base nas métricas obtidas e nas representações gráficas, se conclui que o modelo treinado apresentou desempenho consistente. A elevada acurácia e AUC, aliadas ao recall para embriões aneuploides, reforçam a confiabilidade do modelo como uma ferramenta auxiliar na triagem embrionária. Esses resultados evidenciam a viabilidade de aplicar técnicas de aprendizado de máquina, como redes neurais, no suporte à tomada de decisão em contextos de reprodução assistida.

4.2.2.2 Atividade 7 (A7): Avaliação do Desempenho do Modelo na Predição por Meio da Matriz de Confusão e Curva ROC

A **matriz de confusão** explicada na Atividade 6 (A6) é uma representação tabular que resume o desempenho do classificador ao comparar os rótulos reais com as previsões do modelo. Ela permite a identificação de quatro métricas essenciais:

- **Verdadeiros Positivos (TP):** Casos em que o modelo previu corretamente a euploidia (classe 1). Foram identificados 6 TP.
- **Verdadeiros Negativos (TN):** Casos em que o modelo previu corretamente a aneuploidia (classe 0). Foram registrados 9 TN.
- **Falsos Positivos (FP):** Casos em que o modelo previu incorretamente um embrião aneuploide como euploide. Neste caso, $FP = 0$.

- **Falsos Negativos (FN):** Casos em que o modelo classificou incorretamente um embrião euploide como aneuploide. Foram observados 2 FN.

Essas métricas permitem calcular indicadores importantes, como **precisão**, **recall**, **especificidade** e **acurácia**. A matriz resultante confirma o excelente desempenho do modelo na detecção de aneuploidias, com zero falsos positivos e todos os embriões com anomalias genéticas corretamente identificados.

O gráfico da Figura 18 apresenta a variação das métricas de **sensibilidade** e **especificidade** do modelo conforme diferentes valores de limiar de decisão, indo de 0 a 1. A **sensibilidade** (também chamada de recall para a classe euploide) representa a proporção de embriões euploides corretamente identificados como tais. Observa-se que ela decresce à medida que o limiar aumenta, indicando que o modelo se torna mais conservador na predição de euploidia, o que reduz a quantidade de verdadeiros positivos (euploides corretamente classificados). Por outro lado, a **especificidade** (ou recall para a classe aneuploide) tende a aumentar com o aumento do limiar, o que significa que o modelo passa a cometer menos falsos positivos, classificando embriões aneuploides de forma mais precisa como negativos. Esse gráfico revela o clássico *trade-off* entre sensibilidade e especificidade. Ao escolher limiares mais baixos, há maior sensibilidade (maior detecção de euploidia), mas menor especificidade (mais risco de falsos positivos). Já limiares mais altos priorizam a especificidade, reduzindo a chance de classificar incorretamente um aneuploide como euploide, mas podem deixar de identificar euploidias verdadeiras. Essa análise é essencial para aplicações clínicas, nas quais a escolha do limiar pode ser ajustada com base em prioridades médicas, como evitar falsos negativos (não detectar um embrião viável) ou falsos positivos (implantar um embrião inviável).

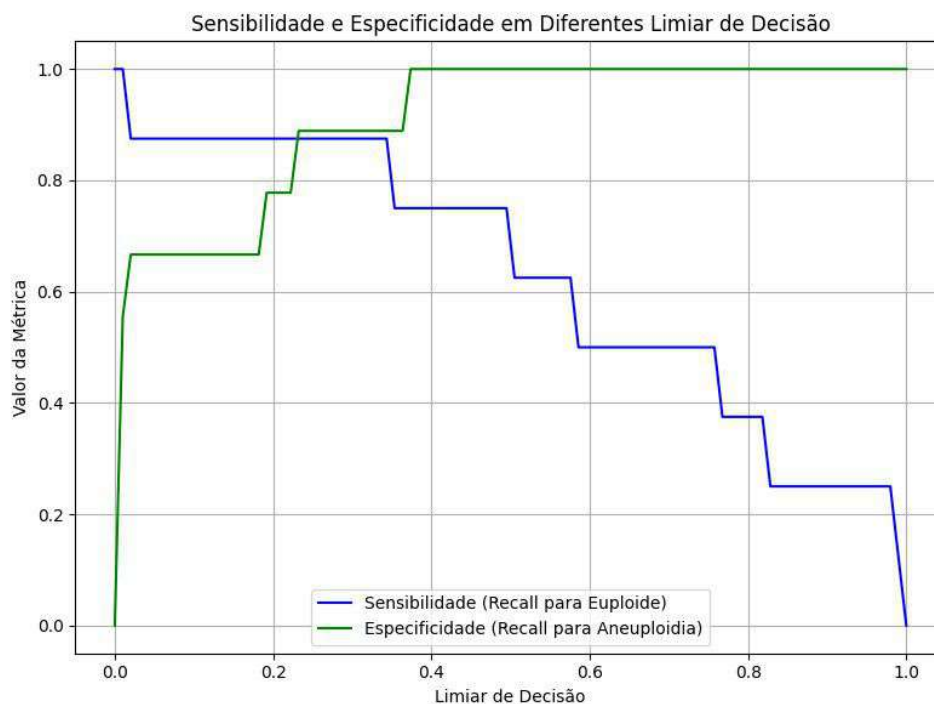


Figura 18. Relação entre sensibilidade e especificidade em diferentes limiares de decisão

Dessa forma, a análise da matriz de confusão e da curva de sensibilidade versus especificidade confirma que o modelo apresenta excelente desempenho, especialmente na identificação de embriões aneuploides. A ausência de falsos positivos e a possibilidade de ajustar o limiar de decisão conforme a prioridade clínica tornam o modelo uma ferramenta promissora.

4.2.3 OE4 - Construir Protótipo

4.2.3.1 Atividade 8 (A8): Prototipar uma interface básica para exibir as predições de euploidia para o usuário final (médicos)

A execução da Atividade 08 teve início com a análise das necessidades essenciais para a construção de uma interface funcional, clara e objetiva voltada aos profissionais da área médica. O principal objetivo desta etapa foi desenvolver um protótipo viável de interface, um *Mínimo Produto Viável* (MVP), que possibilitasse a visualização das predições de euploidia de maneira intuitiva, prática e segura. Como a intenção é evoluir progressivamente esse sistema, priorizou-se uma entrega enxuta, mas funcional, capaz de ser testada e aprimorada com base no feedback dos usuários finais.

A prototipação foi realizada utilizando a ferramenta **Figma**, que permitiu o desenvolvimento visual colaborativo e iterativo da solução. O MVP da interface consiste em uma estrutura simples com foco total na funcionalidade central: permitir que o usuário realize o *upload* de uma planilha com dados morfocinéticos dos embriões e receba como

resultado a predição da ploidia de cada um deles, incluindo a probabilidade estimada de euploidia.

O processo de construção da interface considerou os seguintes elementos:

- **Header e Identidade do Projeto:** A interface possui um cabeçalho simples, com a identidade visual do projeto e um título claro que indica a finalidade da página.
- **Apresentação do Projeto:** Logo abaixo do cabeçalho, é exibido um resumo introdutório do projeto, explicando de forma objetiva a proposta da solução e sua relevância no contexto da medicina reprodutiva.
- **Equipe Envolvida:** A seção seguinte apresenta os participantes do projeto, incluindo o professor orientador, as alunas desenvolvedoras e o médico especialista em medicina reprodutiva
- **Restrições e Regras para o Upload:** Uma seção dedicada às instruções detalhadas orienta o usuário sobre os requisitos da planilha a ser submetida. São explicadas as colunas numéricas obrigatórias, seus nomes exatos e o formato esperado dos dados, com o intuito de evitar erros de entrada que comprometam a interpretação dos dados pelo modelo de inteligência artificial.
- **Upload e Validação de Arquivo:** A funcionalidade principal da interface é o envio da planilha. Foram integradas validações para evitar o envio incorreto de arquivos — etapa crítica para garantir a integridade dos dados analisados e a segurança da predição.
- **Página de Resultados:** Após o envio, o usuário é redirecionado a uma segunda página, que apresenta de forma clara os resultados da análise. São exibidos o total de embriões analisados, o número de embriões classificados como euploides e como aneuploides e a probabilidade de euploidia associada a cada embrião.

A simplicidade do design foi uma escolha estratégica. Ao concentrar esforços no essencial, garantiu-se que o protótipo fosse funcional e acessível, respeitando as boas práticas de UX e as heurísticas de Nielsen já mencionadas. A proposta é que esse MVP sirva como base para iterações futuras, nas quais novas funcionalidades, como visualizações gráficas, histórico de análises ou exportação de relatórios, possam ser integradas com base nas necessidades dos médicos e especialistas que utilizarão o sistema.

A figura interativa do protótipo foi desenvolvida na plataforma Figma e representa a interface básica criada para visualização das predições de euploidia. O link para acesso do [Figma](#) está disponível no Apêndice A, juntamente com o repositório do projeto no [GitHub](#). No repositório do projeto é possível acessar as páginas do protótipo feita, garantindo acesso futuro mesmo que o link original se torne indisponível.

4.2.3.2 Atividade 9 (A9): Desenvolver a aplicação web integrando o protótipo à inteligência artificial existente

A implementação da solução web foi realizada com base nas tecnologias modernas descritas na metodologia: **Next.js com App Router, TypeScript, TailwindCSS**, e bibliotecas auxiliares para leitura de arquivos e comunicação com a API de predição. A execução teve início com a estruturação da aplicação através do comando *create-next-app*, que gerou a base do projeto com suporte ao roteamento moderno e a organização de pastas voltada para escalabilidade.

Com a arquitetura inicial definida, a implementação do protótipo funcional foi iniciada com base no layout desenvolvido no Figma. A interface foi cuidadosamente construída para refletir o fluxo de uso planejado, garantindo uma transição fluida entre prototipagem e desenvolvimento. A cada componente implementado, houve o cuidado de seguir os princípios de **design centrado no usuário**, considerando clareza, legibilidade e acessibilidade, especialmente em um contexto sensível como o médico.

Durante o desenvolvimento, foram feitas diversas **validações e verificações** para assegurar a integridade dos dados. A aplicação possui mecanismos de checagem da planilha enviada pelo usuário, validando a presença e a estrutura correta das colunas esperadas, prevenindo erros durante o processamento e evitando que predições equivocadas sejam geradas por causa de entradas malformadas. Esse cuidado garante maior robustez à solução e confiança no uso clínico.

Para a **integração com o modelo de inteligência artificial**, foi utilizada a **FastAPI** como framework para a construção da API backend. A escolha da FastAPI se deu por sua performance elevada, suporte a tipagem com Pydantic, documentação automática via Swagger e facilidade de integração com aplicações front-end. Essa estrutura garantiu que a comunicação entre a interface e o modelo de IA ocorresse de forma fluida, eficiente e segura. Os dados enviados pela planilha são processados e retornam as predições com porcentagens de euploidia associadas a cada embrião, exibidas em tempo real para o usuário.

A aplicação foi **hospedada na plataforma Vercel**, que possibilitou um processo de **deploy contínuo** desde as primeiras fases de desenvolvimento. Isso permitiu testar cada nova funcionalidade em tempo real, facilitando ajustes iterativos e garantindo que a versão de produção estivesse sempre atualizada e funcional.

As práticas de **Clean Code** foram aplicadas ao longo de todo o desenvolvimento da aplicação, com base nos princípios descritos por [Martin \(2009\)](#). Tais práticas visam garantir legibilidade, manutenibilidade e robustez do código, aspectos essenciais em soluções que envolvem dados sensíveis e contextos clínicos. Dentre os princípios adotados, destacam-se:

- **Nomenclatura clara e intuitiva:** todos os nomes de variáveis, funções e componentes foram definidos de forma descritiva, refletindo com precisão sua responsabilidade no sistema;
- **Consistência nos padrões:** foi mantida uniformidade nos padrões de nomenclatura, indentação e formatação ao longo de todo o projeto, promovendo organização e previsibilidade;
- **Funções pequenas e focadas:** cada função foi estruturada com uma única responsabilidade, mantendo-se concisa e especializada em uma única tarefa;
- **Comentários relevantes:** os comentários foram empregados com parcimônia, utilizados apenas para indicar intenções específicas do código, evitando explicações redundantes de lógica já evidente;
- **Evitar repetição de código (DRY):** foram aplicadas práticas de reutilização por meio de funções reutilizáveis e componentes modulares, reduzindo duplicações e facilitando a manutenção;
- **Tratamento de erros adequado:** implementações robustas de tratamento de exceções e mensagens de erro foram incluídas para garantir confiabilidade na comunicação com o usuário;
- **Princípio KISS (Keep It Simple, Stupid):** o código foi mantido o mais simples possível, evitando estruturas desnecessariamente complexas e favorecendo soluções diretas;

A adoção desses princípios contribuiu significativamente para a qualidade técnica do projeto, promovendo um código mais claro, confiável e sustentável ao longo do tempo.

Como resultado, a aplicação desenvolvida entrega uma experiência simples, funcional e robusta, integrando diretamente o modelo de IA em uma interface amigável. A página permite o **upload da planilha com os dados morfocinéticos**, realiza a análise via IA, e apresenta os resultados com a **classificação binária da ploidia** e a **probabilidade percentual associada à predição euploide**. Tudo isso foi projetado e construído com foco em entregar o **Mínimo Produto Viável (MVP)**, viabilizando futuras evoluções com base no feedback dos usuários médicos.

O deploy da aplicação web pode ser acessado por meio do link da [Ferramenta de Predição de Ploidia de Embriões](#), disponível no Apêndice A, juntamente com o repositório do projeto no [GitHub](#). No repositório é possível acessar também os códigos-fonte e informações complementares, garantindo acesso futuro mesmo que o link original da aplicação se torne indisponível.

5 Discussão dos Resultados

Este trabalho teve como objetivo geral desenvolver uma abordagem baseada em inteligência artificial para identificar padrões em dados morfocinéticos de embriões, obtidos por meio do Time-Lapse System, capaz de prever a classificação e a porcentagem de ploidia, proporcionando uma solução mais eficaz e menos invasiva em comparação ao PGT-A. A seguir, revisitam-se os objetivos específicos estabelecidos no Capítulo 1, discutindo como cada um foi cumprido ao longo da pesquisa.

5.1 Cumprimento dos Objetivos da Pesquisa

5.1.1 OE1: Expandir, Processar e Analisar os Dados para Predição de Ploidia

Esse objetivo foi alcançado por meio da execução das Atividades 1 a 4, que envolveram a análise, revisão e seleção criteriosa das variáveis mais relevantes para a predição de euploidia, incluindo idade materna, tempos de divisão celular (como $t2$, $t3$, $t4$, $t5$, $t8$), intervalos derivados (como $cc2$, $s2$, $cc3$, $s3$) e indicadores clínicos como estágio e morfologia do embrião. A normalização dos dados, utilizando técnicas como *Z-Score*, permitiu a padronização das variáveis para o treinamento do modelo, enquanto a identificação de correlações contribuiu para fundamentar a seleção de pesos, evidenciando variáveis com maior influência na previsão da euploidia. Além disso, a divisão dos dados em conjuntos de treinamento, validação e teste, juntamente com a aplicação de técnicas de aumento de dados (*Data Augmentation*), possibilitou uma melhor generalização do modelo.

5.1.2 OE2: Treinar e Ajustar de Modelo de Machine Learning para Predição de Ploidia

Este objetivo foi alcançado com a construção e treinamento de modelos de *Machine Learning* supervisionados, incluindo o *Multilayer Perceptron* (*MLP*), utilizado para prever a euploidia com base nos parâmetros morfocinéticos e clínicos selecionados. Os ajustes de hiperparâmetros foram realizados para otimizar o desempenho, considerando métricas de acurácia, sensibilidade e especificidade. Os resultados demonstraram alta performance preditiva, com destaque para o modelo que obteve *AUC* superior a 0,94, indicando excelente capacidade discriminatória entre embriões euploides e aneuploides.

5.1.3 OE3: Realizar Avaliação do Modelo

A avaliação do modelo foi realizada por meio da análise das métricas de desempenho, incluindo matriz de confusão, curva *ROC* e análise da relação entre sensibilidade e especificidade. Os resultados confirmaram a robustez do modelo desenvolvido, validando sua aplicabilidade prática como ferramenta não invasivo de predição da saúde genética embrionária. A alta acurácia alcançada sugere a viabilidade do uso de dados morfocinéticos como preditores confiáveis de euploidia, complementando métodos invasivos como o *PGT-A*.

5.1.4 OE4: Construir Protótipo

Por fim, foi desenvolvido um protótipo de interface para a aplicação web, integrando o modelo de inteligência artificial de forma a apresentar os resultados das predições de maneira clara e acessível aos usuários finais, como médicos e embriologistas. Após isso, a interface foi projetada considerando usabilidade e experiência do usuário, com foco em otimizar o processo de análise e seleção embrionária nas clínicas de fertilização *in vitro*.

5.2 Análise Crítica das Correlações Identificadas

A análise dos coeficientes de correlação de Spearman realizada neste estudo permitiu identificar relações estatísticas relevantes entre variáveis morfocinéticas e a ploidia dos embriões. A seguir, se apresenta as interpretações críticas dessas correlações, acompanhadas de comparações com os achados da literatura científica revisada no Capítulo 2.

5.2.1 Correlação entre Idade Materna e Euploidia

Os resultados indicaram correlação negativa entre idade materna e euploidia, confirmando que o aumento da idade está associado à redução da probabilidade de embriões euploides. Essa evidência está alinhada com o estudo de [Yuan et al. \(2023\)](#), que demonstrou que mulheres acima de 35 anos apresentam declínio significativo na taxa de euploidia devido à menor qualidade e quantidade de ovócitos disponíveis. O coeficiente de correlação de *Spearman* identificado neste trabalho reforça a necessidade de considerar a idade como uma variável crítica em modelos preditivos de saúde embrionária.

5.2.2 Correlação entre *tB-tSB* e Euploidia

Foi observada correlação negativa entre o intervalo *tB-tSB* e a euploidia, sugerindo que embriões que apresentam maior diferença temporal entre o início da expansão do blastocisto (*tSB*) e o blastocisto totalmente expandido (*tB*) tendem a ter menor probabilidade

de serem euploides. Essa interpretação corrobora os achados de Cruz et al. (2012), que indicam que embriões com tempos prolongados de desenvolvimento apresentam maior risco de aneuploidias ou falhas no potencial de implantação.

5.2.3 Correlação entre $t5$ e Euploidia

O tempo $t5$, que representa o momento em que o embrião atinge o estágio de 5 células, mostrou correlação negativa com a euploidia quando excede o intervalo ótimo (48,8–56,6 horas). Esse achado é consistente com Cruz et al. (2012), que identificaram $t5$ como um dos preditores mais significativos do potencial de implantação, sendo que tempos fora do intervalo ótimo podem estar relacionados a alterações cromossômicas ou aneuploidias. Embriões que atingiram $t5$ dentro do intervalo descrito apresentaram maior probabilidade de serem euploides, demonstrando que o modelo de *IA* treinado neste estudo capta essa relação de forma adequada.

5.2.4 Outras Correlações Significativas

Além das variáveis destacadas, foram observadas correlações significativas entre:

- $t2$ e $t4$ ($r = 0,89$) e $t3$ e $t2$ ($r = 0,78$): correlações positivas fortes, indicando alta dependência entre os tempos de divisão sequenciais, como descrito por Cruz et al. (2012), e reforçando que a cadência adequada das divisões celulares está diretamente associada ao desenvolvimento embrionário saudável.
- tSC e tSB ($r = 0,75$) e tSB e tB ($r = 0,93$): correlações positivas fortes, demonstrando que os tempos relacionados à expansão e eclosão do blastocisto são interdependentes. Estudos como o de Capalbo et al. (2014) reforçam que blastocistos que atingem estágios avançados de forma sincronizada tendem a apresentar maior euploidia.

5.2.5 Correlações fracas ou inexistentes

Apesar de algumas variáveis apresentarem uma correlação direta fraca ou inexistente com a ploidia — como é o caso do *Morfo*, tSC , tSB , $cc2$ e o intervalo tSC - $t8$, que demonstra um efeito praticamente nulo — é importante destacar que essas variáveis mantêm entre si relações estatísticas relevantes, capazes de influenciar, de forma indireta, o comportamento da ploidia.

As inter-relações entre os parâmetros temporais do desenvolvimento embrionário constroem um cenário dinâmico e multifatorial, no qual o impacto de uma variável sobre a ploidia pode não ser imediato ou isolado, mas sim mediado por sua influência em outras etapas do desenvolvimento. Por exemplo, mesmo que o $cc2$ ($t3$ - $t2$) não apresente uma

relação significativa direta com a ploidia, ele se conecta a outros parâmetros, como o $t3$ e o $cc3$, sendo que este último já demonstra uma correlação moderada e negativa com a ploidia (-0.281).

Da mesma forma, variáveis como o tSC e o intervalo $tSC-t8$, apesar de mostrarem uma correlação direta praticamente nula com a ploidia, estão integradas a outros intervalos temporais que, por sua vez, têm potencial de afetar o desfecho genético do embrião. Isso evidencia que o sistema de desenvolvimento embrionário funciona de maneira interdependente, onde alterações em um ponto específico podem reverberar ao longo do processo, impactando de forma indireta variáveis de interesse clínico, como a ploidia.

Portanto, ao considerar a exclusão de variáveis do modelo ou da análise, é fundamental adotar uma abordagem sistêmica, reconhecendo que a ausência de uma relação direta com a ploidia não implica na irrelevância da variável dentro do contexto geral do desenvolvimento embrionário. Desconsiderar essas variáveis pode levar a interpretações incompletas ou equivocadas do comportamento biológico em estudo.

5.3 Desempenho do Modelo e Comparação com o Estado da Arte

O modelo de *Machine Learning* desenvolvido neste trabalho apresentou acurácia superior a 88%, AUC (Área sob a Curva ROC) de 0,94 e *recall* elevado, indicando alta sensibilidade para identificar embriões euploides. Esses resultados evidenciam a capacidade discriminatória do modelo em diferenciar embriões saudáveis daqueles com aneuploidias, garantindo menor risco de falsos negativos, fator crítico na prática clínica.

A matriz de confusão revelou alta taxa de verdadeiros positivos e negativos, demonstrando que o modelo é equilibrado em sua classificação, sem tendência significativa a erros de um único tipo. Já a Curva ROC , com AUC de 0,94, comprova excelente performance geral, confirmando a robustez do modelo em diferentes limiares de decisão. Na prática, quanto mais próxima de 1,0 for a AUC , maior a capacidade do modelo de distinguir corretamente entre as classes (euploide/aneuploide).

5.3.1 Comparação com o Estado da Arte

Os resultados obtidos foram comparados aos principais trabalhos correlatos revisados no Capítulo 2:

- [Yuan et al. \(2023\)](#) identificaram a forte correlação entre idade materna e euploidia, mas não desenvolveram um modelo preditivo automatizado baseado em *Machine Learning*. Portanto, o presente estudo avança ao operacionalizar essa relação em um modelo prático de apoio clínico.

- [Souza \(2022a\)](#) destacou a importância de variáveis morfocinéticas como $t2$, $t3$, $t5$ e $cc2$, confirmadas neste trabalho como relevantes para o modelo preditivo. No entanto, Souza não apresentou um algoritmo de classificação e predição, diferentemente deste estudo.
- *KIDScoreTM* é um sistema comercial que utiliza variáveis morfocinéticas e fornece uma pontuação de 0 a 10 baseada em *IA* para estimar o potencial de implantação do embrião ([GAZZO et al., 2020](#)). Embora apresente boa acurácia preditiva para implantação, não realiza classificação direta de euploidia ou aneuploidia ou porcentagem de ploidia, limitando seu uso para avaliação genética embrionária.
- *CHLOETM* é outro algoritmo de *IA* citado pela literatura que, similarmente ao *KIDScore*, utiliza análise de imagens *Time-Lapse* para estimar viabilidade embrionária, mas não realiza predição quantitativa de euploidia.

5.3.2 Originalidade e Valor Clínico

O modelo desenvolvido apresenta como grande diferencial a predição quantitativa da porcentagem de ploidia (0–100%), em contraste com os sistemas existentes que se restringem a classificações binárias (euploide/aneuploide) ou categóricas de implantação. Essa abordagem original oferece aos embriologistas e médicos informações adicionais, permitindo decisões mais seguras sobre a transferência embrionária, principalmente em casos de mosaico ou quando há dúvidas sobre o potencial genético do embrião. A predição da porcentagem permite, ainda, planejar estratégias personalizadas de acordo com o risco detectado.

5.3.3 Importância do LIME para Interpretabilidade

Outro aspecto inovador do modelo é a incorporação do algoritmo *LIME* (*Local Interpretable Model-agnostic Explanations*) para interpretabilidade de predições individuais. No contexto médico, a interpretabilidade é fundamental para garantir:

- [Ribeiro, Singh e Guestrin \(2016\)](#) destacam que o *LIME* permite a geração de explicações locais aproximadas, evidenciando as variáveis que mais influenciam cada predição. Segundo os autores, essa ferramenta “explica as previsões de qualquer classificador de maneira interpretável e fiel” ([RIBEIRO; SINGH; GUESTRIN, 2016](#)).
- Utilizar o *LIME* ajuda a diminuir o chamado efeito “caixa-preta”, que ocorre quando um modelo de *inteligência artificial* faz previsões ou decisões, mas as pessoas não conseguem entender como ele chegou àquelas conclusões. Em áreas como a saúde,

onde a confiança e a compreensão dos resultados são fundamentais, tornar o funcionamento do modelo mais transparente aumenta as chances de aceitação e uso por profissionais clínicos (SAMEK; WIEGAND; MÜLLER, 2017).

5.3.4 Exemplo real de uso do LIME na saúde

Naz et al. (2023) desenvolveram uma *Convolutional Neural Network* (CNN) para a classificação de múltiplas doenças pulmonares, incluindo edema, tuberculose e pneumonia associada à COVID-19, a partir de radiografias de tórax. Para tornar o processo de decisão do modelo mais transparente, os autores utilizaram o método *LIME* (*Local Interpretable Model-agnostic Explanations*), que permite explicar, de forma visual e interpretável, o resultado de cada predição, destacando automaticamente as regiões da imagem consideradas mais relevantes pelo modelo. A análise demonstrou que as áreas identificadas pelo *LIME* coincidiram com as regiões apontadas por radiologistas como críticas para o diagnóstico, evidenciando o potencial clínico da ferramenta para apoiar decisões médicas de forma confiável e compreensível.

Aldughayfiq et al. (2023) aplicaram técnicas de *deep learning* em imagens fundoscópicas com o objetivo de diagnosticar retinoblastoma, um câncer ocular infantil. Após o treinamento de um classificador, os autores utilizaram as ferramentas *LIME* (*Local Interpretable Model-agnostic Explanations*) e *SHAP* (*SHapley Additive exPlanations*) para gerar mapas de salientização local, que destacam as regiões da imagem mais relevantes para a decisão do modelo. Os resultados demonstraram que o *LIME* foi capaz de identificar corretamente as áreas com presença tumoral, além de evidenciar as características visuais que mais contribuíram para a predição. Essa abordagem interpretável fornece informações adicionais aos médicos, permitindo compreender quais aspectos do tumor, como tamanho, forma e localização, influenciaram o diagnóstico, favorecendo maior confiança clínica no uso do modelo.

Alabi et al. (2023) se concentraram no prognóstico de pacientes com câncer de nasofaringe, utilizando dados clínicos e registros epidemiológicos. Eles treinaram um robusto modelo de *machine learning* do tipo ensemble e empregaram as técnicas *LIME* e *SHAP* para explicar as predições individuais da probabilidade de sobrevida. O *LIME* evidenciou o nível de confiabilidade de cada predição, mostrando como variáveis como idade, estágio tumoral e etnia contribuíram para o resultado previsto. Nas palavras dos autores, o *LIME* revelou “o grau de confiabilidade da predição feita pelo modelo”. Dessa forma, para cada paciente, o *LIME* destacou as características que fundamentaram o prognóstico estimado, auxiliando oncologistas na interpretação e uso clínico do modelo.

5.4 Implicações Práticas e Sociais

Os resultados obtidos neste estudo apresentam implicações práticas significativas para a área de reprodução assistida. O modelo desenvolvido possui potencial para aumentar substancialmente a agilidade e a confiabilidade na tomada de decisão clínica, fornecendo aos embriologistas e médicos informações preditivas objetivas sobre a euploidia embrionária de forma rápida e não invasiva.

Na prática, o modelo pode ser utilizado como ferramenta de triagem inicial ou como suporte complementar à decisão médica, permitindo priorizar embriões com maior probabilidade de euploidia para transferência.

Além dos benefícios clínicos, o estudo apresenta importantes implicações sociais e emocionais. O principal avanço está relacionado à possibilidade de redução do uso de biópsias embrionárias invasivas (como o *PGT-A*) para análise cromossômica. Ao propor uma abordagem preditiva não invasiva, o modelo contribui para reduzir os custos globais do tratamento de fertilização in vitro, o tornando mais acessível.

Outro impacto social relevante está na redução do estresse emocional das pacientes. O processo de *FIV* já é, por si só, psicologicamente desgastante e a expectativa quanto à qualidade genética dos embriões é um fator adicional de ansiedade. Uma ferramenta de predição confiável, não invasiva e de fácil aplicação pode proporcionar maior segurança emocional aos casais, que se sentem mais confiantes nas decisões clínicas tomadas pela equipe multidisciplinar.

Por fim, ao contribuir para a seleção de embriões mais saudáveis de forma segura, rápida e acessível, o modelo desenvolvido fortalece a medicina reprodutiva baseada em dados, promove a democratização do acesso a tecnologias de ponta e reforça o compromisso ético com a melhoria da qualidade de vida de pacientes e famílias.

5.5 Limitações e Ameaças à Validade do Estudo

Este estudo apresenta algumas limitações e ameaças à validade que devem ser consideradas para a interpretação adequada dos resultados e para direcionar pesquisas futuras.

Primeiramente, o tamanho reduzido da base de dados utilizada, embora ampliado por técnicas de *data augmentation*, pode comprometer a generalização do modelo para populações maiores ou diferentes, especialmente em dados provenientes de outras clínicas com protocolos e características distintas. Essa limitação é comum em estudos iniciais e reforça a necessidade de validações futuras com *datasets* mais amplos, diversos e representativos.

Outro ponto importante é que, apesar do bom desempenho demonstrado, o modelo desenvolvido deve ser compreendido como uma ferramenta auxiliar à decisão clínica, e não como substituto do julgamento médico. A interpretação e aplicação dos resultados devem sempre considerar o contexto clínico e a experiência dos profissionais envolvidos.

Em termos de interpretabilidade, embora o *SHAP* seja considerado teoricamente mais consistente por garantir propriedades como aditividade e justiça na atribuição das importâncias das variáveis, o objetivo inicial deste estudo foi gerar explicações locais, rápidas e clinicamente interpretáveis, justificando a escolha do *LIME* como primeira abordagem. Como aprimoramento futuro, propõe-se a implementação conjunta do *LIME* e do *SHAP*, especialmente em cenários com bases de dados maiores, pois o *SHAP* oferece explicações mais robustas e consistentes para modelos de maior complexidade. Assim, em versões futuras deste projeto, a integração dessas duas técnicas poderá ampliar a confiabilidade das interpretações, fortalecer a validação científica dos resultados e aumentar a aceitação do modelo pelos profissionais de saúde.

Por fim, ressalvas quanto a possíveis vieses não controlados, como fatores demográficos, variações clínicas e técnicas de coleta, podem afetar a validade dos resultados em um contexto mais amplo. Portanto, estudos futuros devem buscar diversificar amostras, incluir dados multicêntricos e realizar validações prospectivas para consolidar a eficácia e segurança do modelo em ambientes reais de aplicação.

6 Conclusões e Trabalhos Futuros

6.1 Conclusões Finais

Este trabalho teve como ponto de partida a crescente necessidade da medicina reprodutiva por métodos menos invasivos, acessíveis e eficazes para a avaliação genética embrionária, tendo como contexto principal a fertilização in vitro (FIV) e os desafios ligados à seleção de embriões euploides. Conforme apresentado no Capítulo 1, a abordagem por meio do PGT-A apresenta limitações importantes relacionadas ao custo, tempo e invasividade, o que motiva o desenvolvimento de soluções alternativas.

Neste sentido, o projeto alcançou seus objetivos ao desenvolver uma solução baseada em inteligência artificial capaz de analisar dados morfocinéticos provenientes do Time-Lapse System, prevendo a porcentagem de euploidia de forma quantitativa, rápida e não invasiva. A criação do modelo de Machine Learning com desempenho robusto, evidenciado por acurácia acima de 88%, alta sensibilidade e uma Área Sob a Curva (AUC) de 0,94, demonstra o sucesso técnico da proposta.

Além disso, a construção do protótipo de software com interface web acessível permite a aplicação prática dos resultados, aproximando a tecnologia da rotina clínica e facilitando a adoção por profissionais de saúde. A inclusão de mecanismos para interpretar localmente as predições, por meio do algoritmo LIME, reforça a confiabilidade e transparência do sistema, elementos essenciais em um ambiente médico.

Um ponto muito importante para o sucesso do modelo foi a boa conexão entre as fases do projeto, principalmente o cuidado com os dados morfocinéticos na Fase 1. Além de limpar, escolher e organizar as variáveis, fizemos um estudo detalhado para entender quais variáveis eram realmente importantes para prever a euploidia. Esse estudo mostrou que nossos dados estavam alinhados com o que a literatura médica atual aponta, confirmando que escolhemos as variáveis certas.

Essa combinação entre o estudo dos dados que coletamos e a pesquisa na literatura ajudou a fortalecer a confiança na qualidade dos dados. Também serviu como base para construir um modelo de Machine Learning mais coerente e que faz sentido dentro da realidade clínica. Por isso, as análises da Fase 1 foram fundamentais para que o modelo identificasse padrões importantes e gerasse predições confiáveis e úteis para a medicina.

Assim, a qualidade e a organização dos dados provenientes do Time-Lapse System serviram como base sólida para o treinamento eficaz do modelo (Fase 2), o que se refletiu diretamente nos resultados de alta acurácia e sensibilidade observados. Essa integração entre as etapas evidencia que a preparação dos dados não é apenas um procedimento

preliminar, mas sim um pilar essencial para a construção de soluções robustas e aplicáveis na medicina reprodutiva.

Em suma, este estudo não apenas atende aos objetivos gerais e específicos traçados inicialmente, como também reforça a importância da interdisciplinaridade entre tecnologia e medicina para gerar soluções inovadoras e com potencial real de transformação na prática clínica da reprodução assistida.

6.2 Trabalhos Futuros

Visando aprimorar e expandir a pesquisa e a aplicação desenvolvidas, propõem-se as seguintes melhorias e linhas de desenvolvimento futuro, organizadas para garantir maior robustez técnica, usabilidade e impacto clínico:

- **Transferência das validações de entrada para o modelo de IA:** Atualmente, as validações quanto à conformidade dos dados (formato, integridade dos valores) são feitas pela aplicação web antes do envio para o modelo preditivo. A proposta é incorporar essas verificações diretamente no modelo de inteligência artificial, o que permitirá um processamento mais autônomo e resiliente. Isso reduzirá a dependência de verificações manuais na interface, aumentará a robustez do sistema contra erros e inconformidades, e possibilitará um tratamento mais inteligente e contextualizado de dados inconsistentes.
- **Implementação de sistema de histórico clínico individualizado:** Para tornar a aplicação mais útil no ambiente clínico, é fundamental desenvolver uma funcionalidade que permita armazenar e gerenciar o histórico de predições de cada paciente e seus embriões de forma segura e organizada. Essa funcionalidade facilitará o acompanhamento longitudinal da evolução dos embriões, evitará o reenvio repetitivo de planilhas, e fornecerá aos médicos uma visão consolidada e prática para decisões clínicas mais informadas.
- **Integração com sistemas de clínicas e consultórios médicos:** A interoperabilidade com softwares já adotados pelas clínicas de fertilização e consultórios médicos é crucial para a adoção efetiva da plataforma. Essa integração permitirá a troca automática de dados, otimizará o fluxo de trabalho dos profissionais, reduzirá erros decorrentes de entradas manuais, e facilitará o uso cotidiano da ferramenta, tornando-a parte natural dos processos clínicos existentes.
- **Desenvolvimento de um Painel Analítico Interativo para Médicos:** A criação de dashboards dinâmicos e interativos oferecerá aos médicos ferramentas visuais para análise aprofundada dos dados preditivos. Por meio de gráficos, filtros por critérios clínicos (como idade materna, período, tipo de embrião) e visualização de

tendências ao longo do tempo, o painel ampliará o suporte à tomada de decisão baseada em dados, promovendo análises estratégicas e acompanhamento contínuo do desempenho clínico.

- **Outras pesquisas e validações:** Futuras investigações devem incluir a validação do modelo em bases de dados maiores, multicêntricas e mais diversificadas, garantindo sua generalização e robustez em contextos variados. Também é recomendada a exploração de outras técnicas e algoritmos avançados de *Machine Learning*, que possam melhorar a acurácia e interpretabilidade das predições. Por fim, estudos prospectivos e clínicos poderão consolidar a eficácia da solução, promovendo sua validação científica e aceitação na prática médica.

Esses aprimoramentos visam não apenas o avanço técnico da plataforma, mas também sua incorporação efetiva na rotina médica, contribuindo para decisões mais rápidas, precisas e seguras no contexto da medicina reprodutiva. O projeto, portanto, abre caminhos promissores para a consolidação de tecnologias de IA como suporte real à prática clínica, com foco em acessibilidade, eficiência e impacto positivo na jornada do paciente. Todas essas melhorias estão alinhadas ao objetivo geral do projeto: desenvolver uma abordagem baseada em inteligência artificial para identificar padrões em dados morfocinéticos de embriões, obtidos por meio do Time-Lapse System, capaz de prever a porcentagem de euploidia, proporcionando uma solução mais eficaz e menos invasiva em comparação ao PGT-A.

Referências

- ADADI, A.; BERRADA, M. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, v. 6, p. 52138–52160, 2018. Citado na página 30.
- ALABI, R. O. et al. Machine learning explainability in nasopharyngeal cancer survival using lime and shap. *Scientific Reports*, Nature Publishing Group, v. 13, p. 8984, 2023. Disponível em: <<https://doi.org/10.1038/s41598-023-35795-0>>. Citado na página 87.
- ALDUGHAYFIQ, B. et al. Explainable ai for retinoblastoma diagnosis: Interpreting deep learning models with lime and shap. *Diagnostics*, MDPI, v. 13, n. 11, p. 1932, 2023. Disponível em: <<https://doi.org/10.3390/diagnostics13111932>>. Citado na página 87.
- ARRIETA, A. B. e. a. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, Elsevier, v. 58, p. 82–115, 2020. Citado na página 74.
- BAIRD et al. Fertility and ageing. *Human Reproduction Update*, v. 11, n. 3, p. 261–276, 2005. Disponível em: <<https://academic.oup.com/humupd/article-abstract/11/3/261/759255>>. Citado 2 vezes nas páginas 22 e 61.
- BASHIR, D. et al. An information-theoretic perspective on overfitting and underfitting. In: *The 33rd Australasian Joint Conference on Artificial Intelligence (AJCAI 2020)*. [s.n.], 2020. Disponível em: <<https://doi.org/10.48550/arXiv.2010.06076>>. Citado na página 44.
- BASTIDA, A. M. et al. Impact of different degrees of genetic mosaicism in the kinetic profile of the human embryo. *Fertility and Sterility*, v. 112, n. 6, p. 1069–1078, 2019. Disponível em: <[https://www.fertstert.org/article/S0015-0282\(19\)31316-0/fulltext](https://www.fertstert.org/article/S0015-0282(19)31316-0/fulltext)>. Citado na página 25.
- BOUCRET, L. et al. Change in the strategy of embryo selection with time-lapse system implementation—impact on clinical pregnancy rates. *Journal of Clinical Medicine*, v. 10, n. 18, p. 4111, 2021. Disponível em: <<https://www.mdpi.com/2077-0383/10/18/4111>>. Citado 2 vezes nas páginas 21 e 22.
- CAMARGOS, V. P. et al. Imputação múltipla e análise de casos completos em modelos de regressão logística: uma avaliação prática do impacto das perdas em covariáveis. *Cadernos de Saúde Pública*, v. 27, n. 12, p. 2299–2313, 2011. Disponível em: <<https://www.scielo.br/j/csp/a/P4qJKTyRhKFfd3grkbfsL6y>>. Citado na página 55.
- CAPALBO, A. et al. Correlation between standard blastocyst morphology, euploidy and implantation: An observational study in two centers involving 956 screened blastocysts. *Human Reproduction*, v. 29, n. 6, 2014. Disponível em: <<https://academic.oup.com/humrep/article-abstract/29/6/1173/624854?login=false>>. Citado 3 vezes nas páginas 24, 61 e 84.
- CHEN, D. Y. *Análise de dados com Python e Pandas*. Novatec Editora, 2018. Disponível em: <<https://books.google.com.br/books?hl=pt-BR&lr=&id=>

ILFwDwAAQBAJ&oi=fnd&pg=PA31&dq=biblioteca+Pandas&ots=sQ0dSaigr&sig=lcdNNbH67MKJuXccwqIuisdtua4#v=onepage&q=biblioteca%20Pandas&f=false>. Citado 2 vezes nas páginas 42 e 107.

CORLETA, H. von E. Fertilização in vitro: mais de 4 milhões de crianças nascidas e um prêmio nobel. *Clinical and Biomedical Research*, v. 30, n. 4, p. 451–455, 2010. Disponível em: <<https://seer.ufrgs.br/index.php/hcpa/article/view/17351>>. Citado na página 17.

CRUZ, M. et al. Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality. *Reproductive Biomedicine Online*, v. 25, n. 4, p. 371–381, 2012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1472648312004099?casa_token=Q7hp0XipKegAAAAA:TQWmUPtUYztYC0Vp_cgchEh27w3nJpHB49gyD9b06zb4af2hO0jzrcCBWRbpPCLGghHSHhATOQ>. Citado 7 vezes nas páginas 22, 23, 25, 63, 84, 104 e 105.

DESAI, N. et al. Odds of euploidy are significantly associated with not only age but blastocyst morphokinetic parameters and icm/trophectoderm characteristics. *Fertility and Sterility*, v. 112, n. 6, p. 1016–1025, 2019. Disponível em: <[https://www.fertstert.org/article/S0015-0282\(19\)31117-3/fulltext](https://www.fertstert.org/article/S0015-0282(19)31117-3/fulltext)>. Citado na página 23.

FAIRILITY™. Fairtility's ai-powered embryo quality assessment assistant. 2020. Disponível em: <<https://fairtility.com/chloe/>>. Citado na página 38.

Figma. *Figma: collaborative interface design tool*. 2025. Disponível em: <<https://www.figma.com>>. Citado na página 49.

GARDNER, D. In vitro culture of human blastocysts. In: JANSEN, R.; MORTIMER, D. (Ed.). *Toward Reproductive Certainty: Fertility and Genetics Beyond*. Carnforth, UK: Parthenon Publishing, 1999. p. 378–388. Disponível em: <https://journals.lww.com/co-obgyn/fulltext/1999/06000/culture_and_transfer_of_human_blastocysts.13.aspx>. Citado na página 24.

GAZZO, E. et al. The kidscore d5 algorithm as an additional tool to morphological assessment and pgt-a in embryo selection: A time-lapse study. *JBRA Assisted Reproduction*, v. 24, p. 55–60, 2020. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC6993168/>>. Citado 4 vezes nas páginas 24, 25, 86 e 104.

GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. [S.l.]: "O'Reilly Media, Inc.", 2017. Citado 8 vezes nas páginas 31, 32, 33, 34, 35, 36, 72 e 73.

GILPIN, L. H. et al. Explaining explanations: An overview of interpretability of machine learning. In: IEEE. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. [S.l.], 2018. p. 80–89. Citado na página 30.

GLEICHER, N.; PATRIZIO, P.; BRIVANLOU, A. Preimplantation genetic testing for aneuploidy—a castle built on sand. *Trends in molecular medicine*, Elsevier, v. 27, n. 8, p. 731–742, 2021. Disponível em: <[https://www.cell.com/trends/molecular-medicine/fulltext/S1471-4914\(20\)30313-0](https://www.cell.com/trends/molecular-medicine/fulltext/S1471-4914(20)30313-0)>. Citado 4 vezes nas páginas 19, 20, 21 e 22.

HAYKIN, S. *Neural Networks and Learning Machines*. 3. ed. Upper Saddle River, New Jersey: Pearson Education, 2009. ISBN 978-0-13-147139-9. Citado 4 vezes nas páginas 27, 28, 29 e 72.

- IZBICKI, R.; SANTOS, T. M. D. *Aprendizado de máquina: uma abordagem estatística*. Rafael Izbicki, 2020. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=6O8OEAAAQBAJ&oi=fnd&pg=PR13&dq=related:sXAZNf4-S38J:scholar.google.com/&ots=V9g1n31Nqk&sig=fjPChwYESoj3P8Jr6dbKM2_HAbA#v=onepage&q&f=false>. Citado 2 vezes nas páginas 26 e 27.
- JAISWAL, S. What is normalization in machine learning? a comprehensive guide to data rescaling. 2024. Disponível em: <<https://www.datacamp.com/tutorial/normalization-in-machine-learning>>. Citado 4 vezes nas páginas 42, 43, 56 e 107.
- JARDIM, M. C. Política e emoções. *Revista de Ciências Sociais*, UNESP, v. 21, n. 51, 2022. Publicado em 2023-06-21. Disponível em: <<https://doi.org/10.5007/2175-7984.2022.e91402>>. Citado na página 12.
- JUNIOR, E. B. *Timelapse e IA na avaliação embrionária: Como interpretar os resultados e melhorar as chances de gestação*. 2022. Disponível em: <<https://fertility.com.br/wp-content/uploads/2022/08/Timelapse-ia-avaliacao-embrionaria.pdf>>. Citado na página 60.
- JUNIOR, G. B. V. et al. Métricas utilizadas para avaliar a eficiência de classificadores em algoritmos inteligentes. *Centro de Pesquisas Avançadas em Qualidade de Vida*, v. 14, n. 2, 2022. ISSN 2178-7514. Disponível em: <https://www.researchgate.net/publication/359541310_METRICAS_UTILIZADAS_PARA_AVALIAR_A_EFICIENCIA_DE_CLASSIFICADORES_EM_ALGORITMOS_INTELIGENTES>. Citado 2 vezes nas páginas 47 e 48.
- KALOS, M. H.; WHITLOCK, P. A. *Monte Carlo Methods*. John Wiley & Sons, 2009. Disponível em: <<https://books.google.com.br/books?hl=pt-BR&lr=&id=5z-AI0pbNsYC&oi=fnd&pg=PR5&dq=Monte+Carlo+Methods+Malvin&ots=7rAR9jBwDI&sig=41VtfDAcq7Wg>>. Citado na página 109.
- KATO, K. et al. Comparing prediction of ongoing pregnancy and live birth outcomes in patients with advanced and younger maternal age patients using kidscore™ day 5: A large-cohort retrospective study with single vitrified-warmed blastocyst transfer. *Reproductive Biology and Endocrinology*, v. 19, p. 98, 2021. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC8252298/>>. Citado 3 vezes nas páginas 24, 25 e 105.
- KIAR, G. et al. Data augmentation through monte carlo arithmetic leads to more generalizable classification in connectomics. Montreal Neurological Institute, McGill University; Department of Computer Science and Computer Engineering, Concordia University, 2021. Disponível em: <<https://arxiv.org/pdf/2109.09649>>. Citado na página 109.
- LASSEN, J. et al. Development and validation of deep learning-based embryo selection across multiple days of transfer. *arXiv preprint arXiv:2210.02120*, 2022. Disponível em: <<https://arxiv.org/abs/2210.02120>>. Citado na página 24.
- LEAVER, M.; WELLS, D. Non-invasive preimplantation genetic testing (nipgt): the next revolution in reproductive genetics? *Human Reproduction Update*, Oxford University Press on behalf of the European Society of Human Reproduction and Embryology, v. 25, n. 2, p. 241–255, 2019. Disponível em: <<https://academic.oup.com/humupd/article/26/1/16/5643748?login=false>>. Citado 2 vezes nas páginas 19 e 20.

- LUONG, T.-M.-T.; LE, N. Q. K. Artificial intelligence in time-lapse system: advances, applications, and future perspectives in reproductive medicine. *Journal of Assisted Reproduction and Genetics*, v. 40, n. 12, p. 1205–1221, 2023. Disponível em: <<https://link.springer.com/article/10.1007/s10815-023-02973-y>>. Citado na página 12.
- MARTIN, R. C. *Clean code: a handbook of agile software craftsmanship*. [S.l.]: Prentice Hall, 2009. Citado 2 vezes nas páginas 51 e 80.
- MESEGUER, M. et al. The use of morphokinetics as a predictor of embryo implantation. *Human reproduction*, Oxford University Press, v. 26, n. 10, p. 2658–2671, 2011. Disponível em: <<https://academic.oup.com/humrep/article/26/10/2658/611030?login=true>>. Citado na página 12.
- MILEWSKI, R. et al. Morphokinetic parameters as a source of information concerning embryo developmental and implantation potential. *Ginekologia Polska*, v. 87, n. 10, p. 677–684, 2016. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/27958618/>>. Citado na página 42.
- MOLICH, R.; NIELSEN, J. Improving a human-computer dialogue. *Communications of the ACM*, ACM, v. 33, n. 3, p. 338–348, 1990. Disponível em: <<https://dl.acm.org/doi/10.1145/77481.77486>>. Citado na página 50.
- MONTAGNINI, H. M. L. et al. Estados emocionais de casais submetidos à fertilização in vitro. *Revista Brasileira de Ginecologia e Obstetrícia*, 2010. Disponível em: <<https://www.scielo.br/j/estpsi/a/hHphXxpTdNzZHt3PGqL3c7j/>>. Citado na página 13.
- MORICONI Émery et al. Refinamento e combinação de cnns para melhoria na classificação de leucócitos imaturos no diagnóstico de leucemia mieloide aguda. In: *Anais da XII Escola Regional de Computação do Ceará, Maranhão e Piauí*. SBC, 2024. p. 129–138. Disponível em: <<https://sol.sbc.org.br/index.php/ercemapi/article/view/30178>>. Citado na página 71.
- MOURA, M. D. de; SOUZA, M. do Carmo Borges de; SCHEFFER, B. B. Reprodução assistida. um pouco de história. *Revista da Sociedade Brasileira de Psicologia Hospitalar*, v. 12, n. 2, p. 1–7, 2020. Disponível em: <<https://pepsic.bvsalud.org/pdf/rsbph/v12n2/v12n2a04.pdf>>. Citado 2 vezes nas páginas 17 e 18.
- MOUSTAKLI, E. et al. Evolution of minimally invasive and non-invasive preimplantation genetic testing: An overview. *Journal of Clinical Medicine*, v. 13, n. 8, p. 2160, 2024. Disponível em: <<https://www.mdpi.com/2077-0383/13/8/2160>>. Citado 3 vezes nas páginas 21, 22 e 36.
- MÜLLER, A. *Introduction to machine learning with python*. 2017. Citado 4 vezes nas páginas 31, 32, 33 e 35.
- NASCIMENTO, F. P. do. Classificação da pesquisa: Natureza, método ou abordagem metodológica, objetivos e procedimentos. In: _____. *Metodologia da Pesquisa Científica: teoria e prática – como elaborar TCC*. Brasília: Thesaurus, 2016. cap. 6. Citado 2 vezes nas páginas 39 e 40.

- NAZ, Z. et al. An explainable ai-enabled framework for interpreting pulmonary diseases from chest radiographs. *Cancers*, MDPI, v. 15, n. 1, p. 314, 2023. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9818469/>>. Citado na página 87.
- NG, A. Y.-T. *Machine Learning Yearning*. [s.n.], 2018. Disponível em: <<https://info.deeplearning.ai/machine-learning-yearning-book>>. Citado 2 vezes nas páginas 44 e 70.
- NIELSEN, J. *10 Usability Heuristics for User Interface Design*. 1994. Citado na página 50.
- NORMAN, D. A. *The Design of Everyday Things: Revised and Expanded Edition*. [S.l.]: Basic Books, 2013. Citado na página 49.
- PANDIT, S.; SHARMA, R. Non invasive assessment of human oocytes and embryos in assisted reproduction: Review on present practices and future trends. *Medical Journal Armed Forces India*, v. 78, n. 1, p. 7–16, January 2022. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/35035038/>>. Citado na página 13.
- PHILLIPS, K. R. B. et al. Temporal evaluation of a minimally invasive method of preimplantation genetic testing for aneuploidy (mi-pgt-a) in human embryos. *Reproductive Medicine*, v. 5, n. 3, p. 97–112, 2024. Disponível em: <<https://doi.org/10.3390/reprodmed5030011>>. Citado na página 20.
- PING, P. et al. Association of embryo aneuploidy and sperm dna damage in unexplained recurrent implantation failure patients under ngs-based pgt-a cycles. *Archives of Gynecology and Obstetrics*, v. 308, p. 997–1005, 2023. Disponível em: <<https://link.springer.com/article/10.1007/s00404-023-07098-2>>. Citado na página 12.
- RAMALHO, D. B. *Entrevistas concedidas a Maria Abritta e Sabrina Berno*. 2024. Citado 3 vezes nas páginas 104, 105 e 106.
- Rangraz Jeddi, F. et al. Usability evaluation of a comprehensive national health information system: A heuristic evaluation. *Informatics in Medicine Unlocked*, v. 19, p. 100332, 2020. ISSN 2352-9148. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352914820301064>>. Citado na página 50.
- REIGNIER, A. et al. Performance of day 5 kidscoreTM morphokinetic prediction models of implantation and live birth after single blastocyst transfer. *Journal of assisted reproduction and genetics*, Springer, v. 36, p. 2279–2285, 2019. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/31444634/>>. Citado na página 37.
- RESTREPO, L. F.; GONZÁLEZ, J. De pearson a spearman. *Revista Colombiana de Ciencias Pecuarias*, v. 20, n. 2, p. 183–192, 2007. Disponível em: <http://www.scielo.org.co/scielo.php?pid=S0120-06902007000200010&script=sci_arttext>. Citado na página 110.
- RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. "why should i trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2016. p. 1135–1144. Citado 4 vezes nas páginas 30, 73, 74 e 86.

- RIENZI, L. et al. Time of morulation and trophectoderm quality are predictors of a live birth after euploid blastocyst transfer: a multicenter study. *Fertility and Sterility*, v. 113, n. 5, p. 991–998, 2020. Disponível em: <[https://www.fertstert.org/article/S0015-0282\(19\)31930-2/fulltext](https://www.fertstert.org/article/S0015-0282(19)31930-2/fulltext)>. Citado na página 23.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3rd. ed. Upper Saddle River, NJ: Pearson, 2016. Citado na página 26.
- SAMEK, W.; WIEGAND, T.; MÜLLER, K.-R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017. Disponível em: <<https://doi.org/10.48550/arXiv.1708.08296>>. Citado na página 87.
- SANTOS, C. Estatística descritiva—manual de auto-aprendizagem. *Edições Sílabo: Lisboa*, 2007. Disponível em: <https://www.researchgate.net/publication/311103840_Estatistica_Descritiva_Manual_de_auto-aprendizagem>. Citado na página 111.
- SATHYANARAYANAN; TANTRI, R. Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, November 30 2024. Disponível em: <<https://www.africanjournalofbiomedicalresearch.com/index.php/AJBR/article/view/4345>>. Citado na página 48.
- SHEIKH, H.; PRINS, C. Artificial intelligence: Definition and background. In: _____. *Mission AI*. Springer, 2023. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-031-21448-6_2>. Citado na página 25.
- SILVA, A. E. et al. Casais com abortamento espontâneo recorrente: participação das translocações cromossômicas / couples with recurrent miscarriages: contributions of chromosome translocations. *Revista de Ciências Biológicas e da Saúde*, UNESP – Universidade Estadual Paulista, Campus de São José do Rio Preto-SP, v. 14, n. 4, p. e249, 2023. Disponível em: <https://repositorio-racs.famerp.br/racs_ol/vol-14-4/ID249.pdf>. Citado na página 12.
- SOUSA Áurea. Coeficiente de correlação de pearson e coeficiente de correlação de spearman. o que medem e em que situações devem ser utilizados? *Correio dos Açores: Matemática*, p. 19, mar 2019. Disponível em: <https://repositorio.uac.pt/bitstream/10400.3/5365/1/Sousa_CA_21%20Mar%20c3%a7o%202019.pdf>. Citado 4 vezes nas páginas 43, 58, 107 e 110.
- SOUZA, M. C. de. As técnicas de reprodução assistida. a barriga de aluguel. a definição da maternidade e da paternidade. bioética. *Revista EMERJ*, v. 50, p. 348–366, 2024. Disponível em: <https://www.emerj.tjrj.jus.br/revistaemerj_online/edicoes/revista50/Revista50_348.pdf>. Citado na página 17.
- SOUZA, R. C. M. de. *Análise da ploidia de embriões humanos por meio da Inteligência Artificial com o uso de variáveis de morfologia, morfocinética e variáveis relacionadas com a paciente*. Tese (Doutorado) — UNESP, 2022. Disponível em: <https://bdtd.ibict.br/vufind/Record/UNSP_2209d12b9be14edf04d513c039a770d8>. Citado 5 vezes nas páginas 22, 36, 37, 42 e 86.
- SOUZA, R. C. M. de. *Dissertação de mestrado: Farmacologia e Biotecnologia*. Dissertação (Mestrado) — Universidade Estadual Paulista (Unesp), June 2022. Disponível em:

<<https://repositorio.unesp.br/items/02fb8a4e-c577-4aa1-aa7e-f259f23e8353>>. Citado na página 18.

TRASK, A. W. *Grokking: Deep Learning*. New York: Manning Publications Co., 2019. Disponível em: <<https://edu.anarcho-copy.org/Algorithm/grokking-deep-learning.pdf>>. Citado na página 26.

YANG, H. et al. Preimplantation genetic testing for aneuploidy: challenges in clinical practice. *Nature Reviews Genetics*, v. 25, n. 1, p. 30–45, 2024. Disponível em: <<https://link.springer.com/article/10.1186/s40246-022-00442-8>>. Citado 3 vezes nas páginas 12, 14 e 18.

YUAN, Z. et al. Development of an artificial intelligence based model for predicting the euploidy of blastocysts in pgt-a treatments. *Scientific Reports*, v. 13, 2023. Disponível em: <<https://www.nature.com/articles/s41598-023-29319-z>>. Citado 7 vezes nas páginas 22, 36, 37, 41, 83, 85 e 105.

Z-SCORE: saiba o que é e como funciona. jul 2022. Disponível em: <<https://maisretorno.com/portal/termos/z/z-score>>. Citado na página 107.

ZEGERS-HOCHSCHILD, F. et al. The international glossary on infertility and fertility care, 2017. *Human reproduction*, Oxford University Press, v. 32, n. 9, p. 1786–1801, 2017. Disponível em: <<https://academic.oup.com/humrep/article/32/9/1786/4049537?login=false>>. Citado 2 vezes nas páginas 13 e 18.

GLOSSÁRIO

Aneuploidia	Condição em que o número de cromossomos não é múltiplo exato de 23.
Blastocisto	Estágio do desenvolvimento embrionário, que ocorre cerca de 5 a 6 dias após a fecundação, apresentando uma cavidade interna e as primeiras divisões celulares mais complexas.
Blastômeros	Células resultantes das divisões iniciais do embrião, multiplicando-se durante as primeiras fases de desenvolvimento.
Biópsia	Procedimento médico para retirar uma pequena amostra de tecido ou células para análise. Utilizado em reprodução assistida para avaliar a saúde do embrião.
Citoplasma	Parte do conteúdo celular que envolve o núcleo e onde ocorrem várias funções vitais da célula, como metabolismo e síntese de proteínas.
Clivagem	Processo de divisão celular do embrião, onde uma célula inicial se divide sucessivamente em células menores chamadas blastômeros.
Cromossomos	Estruturas presentes no núcleo celular que carregam a informação genética. Os humanos têm 23 pares de cromossomos, totalizando 46.
Dados Morfocinéticos	Análise dos parâmetros morfológicos e cinéticos do embrião, como sua estrutura e o movimento/desenvolvimento ao longo do tempo.
Euploidia	Condição em que o número de cromossomos é múltiplo exato de 23.
Gravidez Clínica	Definição de uma gestação confirmada por ultrassonografia, com presença de embrião e batimento cardíaco fetal no útero.
Implantação de Embrião	Processo no qual o embrião se fixa e se insere na parede do útero, iniciando o desenvolvimento da gravidez.
Massa Celular Interna do Embrião	Conjunto de células do blastocisto que dará origem ao feto durante o desenvolvimento da gestação.

Mosaico	Tipo de aneuploidia em que algumas células têm o número correto de cromossomos, enquanto outras têm um número alterado.
Ploidias	Refere-se ao número de conjuntos de cromossomos em uma célula ou organismo.
Trofectoderma	Camada externa do blastocisto que dará origem à placenta, sendo essencial para a implantação do embrião no útero.

Apêndices

APÊNDICE A – Acesso ao Projeto

Os materiais desenvolvidos durante este trabalho estão disponíveis nos links abaixo, permitindo acesso ao código-fonte, ao protótipo e à aplicação web:

- **Repositório com o código-fonte (GitHub):** <<https://github.com/sabrinaberno/TCC>>
- **Protótipo da interface (Figma):** <<https://www.figma.com/design/0AstYynXwZO9zVudP1NimJ/Predi%C3%A7%C3%A3o-de-Ploidia>>
- **Aplicação Web (Deploy):** <<https://embryo-predictor.vercel.app>>

Os materiais podem ser utilizados para consulta, testes ou futuras evoluções da ferramenta.

APÊNDICE B – Variáveis utilizadas na análise da ploidia embrionária

Este anexo descreve as variáveis utilizadas na planilha de dados referente ao desenvolvimento embrionário e sua relação com a ploidia. As variáveis incluem informações demográficas, temporais e morfológicas do embrião, bem como métricas de qualidade baseadas no sistema Gardner. Todos os tempos descritos abaixo são expressos em horas.

• Variáveis Gerais

- **Id:** Identificador numérico de cada paciente.
- **Idade:** Idade da paciente no momento do procedimento.
- **Data da biópsia:** Data em que foi realizada a biópsia embrionária.
- **Embrião n.:** Identificação numérica do embrião dentro do ciclo de fertilização.
- **Estágio:** Dia de evolução no cultivo (5º dia ou 6º dia) ([RAMALHO, 2024](#)).
- **Morfo:** Classificação morfológica dos embriões baseada no estágio de expansão da blástula, estágio inicial do desenvolvimento embrionário que ocorre após a segmentação (divisões celulares iniciais) do zigoto. As notas de 1 a 5 são a expansão do embrião, com 1 sendo o menos expansivo e 5 o mais expansivo ([RAMALHO, 2024](#)).
- **KIDScore™:** Algoritmo combina variáveis morfocinéticas e parâmetros de desenvolvimento embrionário. A pontuação vai de 0 a 10 por conta de estarmos utilizando embriões de estado de blastocisto ([GAZZO et al., 2020](#)).

• Variáveis Temporais

As variáveis temporais são baseadas nos intervalos de tempo entre eventos específicos durante o desenvolvimento do embrião:

- **st2:** Primeiro indício de movimentos citoplasmáticos antes da primeira citocinese. É a fase final da divisão celular, onde o citoplasma é dividido entre as duas células filhas ([RAMALHO, 2024](#)).
- **t2:** Tempo para 2 células. Tempo necessário para o embrião completar a primeira clivagem, ou seja, a divisão da célula-ovo (zigoto) em duas células, entre 24,3 – 27,9 horas após a fertilização ([CRUZ et al., 2012](#)).
- **t3:** Tempo para 3 células. Marca o momento em que o embrião se divide de duas para três células, entre 35,4 – 40,3 horas após a fertilização ([CRUZ et al., 2012](#)).

- **t4**: Tempo para 4 células. Transição do embrião para o estágio de quatro células.
- **t5**: Tempo para 5 células. Momento em que o embrião alcança o estágio de cinco células, entre 48,8 – 56,6 horas após a fertilização (CRUZ et al., 2012).
- **t8**: Tempo para 8 células.
- **tSC**: Tempo de formação do estágio de clivagem sincronizada (Time to Synchronized Compaction). Representa o tempo necessário para que o embrião, após atingir o estágio de 8 células, comece a apresentar compactação sincronizada. A compactação consiste na união mais forte das células embrionárias (KATO et al., 2021).
- **tSB**: Tempo para o início da blastulação (Time to Start Blastulation). Tempo necessário para o embrião iniciar a formação do blastocisto (KATO et al., 2021).
- **tB**: Tempo para o Blastocisto (Time to Blastocyst). Tempo que leva para o embrião alcançar o estágio de blastocisto completo, que é o último estágio de desenvolvimento embrionário antes da implantação no útero (YUAN et al., 2023).
- **t2-st2**: Intervalo de tempo entre o T2 e o ST2. Esse intervalo é usado para avaliar a regularidade e a qualidade das divisões celulares no estágio inicial do desenvolvimento embrionário (RAMALHO, 2024).
- **cc2 (t3-t2)**: Tempo necessário para que o embrião passe da divisão de 2 células (T2) para a divisão de 3 células (T3). É utilizado para avaliar a regularidade e a dinâmica do ciclo celular inicial do embrião (RAMALHO, 2024).
- **cc3 (t5-t3)**: Tempo necessário para que o embrião passe da divisão de 3 células (T3) para a divisão de 5 células (T5). É utilizado como um indicador da dinâmica do ciclo celular (RAMALHO, 2024).
- **t5-t2**: Intervalo de tempo entre o estágio de 2 células (T2) e o estágio de 5 células (T5). É uma métrica importante para avaliar a eficiência das divisões celulares iniciais (RAMALHO, 2024).
- **s2 (t4-t3)**: Intervalo de tempo necessário para que o embrião passe do estágio de 3 células (T3) para o estágio de 4 células (T4). Este parâmetro é usado para avaliar a regularidade da divisão celular (RAMALHO, 2024).
- **s3 (t8-t5)**: Intervalo de tempo necessário para que o embrião passe do estágio de 5 células (T5) para o estágio de 8 células (T8). É usado para avaliar a sincronização e a regularidade do ciclo celular (RAMALHO, 2024).
- **tSC-t8**: Intervalo de tempo entre o estágio em que o embrião atinge a compactação inicial (tSC) e o estágio de 8 células (T8). É usado para avaliar a

transição do embrião das divisões celulares iniciais para o início da compactação (RAMALHO, 2024).

- **tB-tSB:** Intervalo de tempo entre o estágio em que o embrião atinge o blastocisto inicial (tSB) e o estágio de blastocisto expandido (tB). Avalia o tempo necessário para que o embrião progrida do início da formação do blastocisto até a sua expansão completa (RAMALHO, 2024).

- **Variável de Resultado**

- **Ploidia:** Estado de ploidia do embrião, resultado final da análise.

APÊNDICE C – Z-score normalization (standardization)

O Z-Score é um indicador numérico que ilustra a conexão entre uma determinada quantia e a média de um conjunto de valores, expressa em termos de desvios padrão. Quando o Z-Score de um dado é zero, isso sugere que ela coincide com a média do conjunto. Um Z-Score de 1 indica que o valor está um desvio padrão acima da média, ao passo que valores negativos sugerem que estão abaixo da média. A normalização pelo Z-Score, também conhecida como padronização, pressupõe que os dados sigam uma distribuição gaussiana (em forma de sino) e transforma os valores para que tenham uma média (μ) de 0 e um desvio padrão (σ) de 1, facilitando a análise e a comparação entre variáveis com diferentes escalas (JAISWAL, 2024; Z-SCORE. . . , 2022). A fórmula para padronização é:

$$X_{standardized} = \frac{X - \mu}{\sigma}$$

Em que:

- x : Valor original do dado;
- μ : Média da variável;
- σ : Desvio padrão da variável.

O Z-Score é um instrumento valioso para a identificação se uma pontuação é típica ou atípica em comparação a um conjunto de dados previamente definido. Esta métrica permite a comparação entre pontuações de variados conjuntos de dados, o que torna as análises mais exatas e uniformes. O Z-Score, apesar de ser sensível a outliers (valores atípicos que estão muito distantes da maioria dos outros dados) e depender do intervalo de dados, é benéfico em circunstâncias onde a preservação do intervalo original é crucial (SOUSA, 2019; Z-SCORE. . . , 2022).

Também é crucial reconhecer as variáveis numéricas que serão convertidas, já que o Z-Score só se aplica a variáveis contínuas. Assim, cada valor na variável será transformado em um Z-Score, representando quantos desvios padrão ele está acima ou abaixo da média da variável.

Para normalizar a tabela de dados usando o método Z-Score, utilizaremos o Pandas, que é uma biblioteca Python de código aberto para análise de dados (CHEN, 2018). Inicialmente, importaremos a planilha Excel com os dados originais. Depois, são identificadas e ordenadas manualmente as colunas numéricas a serem normalizadas. Uma réplica do original DataFrame é gerada para manter os dados inalterados, enquanto as variáveis selecionadas são modificadas utilizando a fórmula do Z-Score: cada valor é ajustado pela subtração da média da variável e divisão pelo seu desvio padrão. Este procedimento assegura que as variáveis sejam ajustadas para uma média de 0 e um desvio padrão de 1, removendo variações de escala entre elas. Assim, a tabela normalizada é armazenada em um novo arquivo Excel. Após a aplicação do Z-score, se faz a verificação da média e do desvio padrão das variáveis transformadas, em que a média (μ) deve ser próxima de 0 e o desvio padrão (σ) deve ser próximo de 1.

APÊNDICE D – Monte Carlo

O Método de Aritmética de Monte Carlo (MCA) é uma técnica matemática e computacional amplamente utilizada para resolver problemas que envolvem incerteza e variabilidade (KALOS; WHITLOCK, 2009). Baseia-se no uso de números aleatórios para simular processos estocásticos, ou seja, processos que evoluem de maneira dependente de eventos aleatórios (KALOS; WHITLOCK, 2009). Esse método é especialmente eficaz para lidar com problemas de alta complexidade matemática, onde métodos analíticos tradicionais podem ser inviáveis. Além de que “o aumento do conjunto de dados foi amplamente demonstrado como uma técnica eficaz para melhorar a generalização de modelos de aprendizado” (KIAR et al., 2021).

A Simulação de Monte Carlo trabalha com o princípio de geração de valores aleatórios dentro de um intervalo previamente definido para variáveis que apresentam incerteza (KALOS; WHITLOCK, 2009). Esses valores aleatórios são extraídos de uma distribuição de probabilidade específica, como a distribuição uniforme ou normal, dependendo do problema. O método consiste em repetir o cálculo de um modelo várias vezes, cada vez utilizando um conjunto diferente de valores aleatórios como entrada (KALOS; WHITLOCK, 2009).

Diferentemente de modelos de previsão tradicionais, que trabalham com valores fixos, o Monte Carlo oferece uma gama de resultados possíveis e a probabilidade associada a cada um. Isso permite uma análise mais detalhada e uma maior flexibilidade para lidar com incertezas.

Embora os números usados no método sejam chamados de aleatórios, em implementações computacionais eles são, na verdade, gerados por algoritmos que criam números pseudoaleatórios (KALOS; WHITLOCK, 2009). Esses números imitam propriedades de números verdadeiramente aleatórios, mas são derivados de um processo determinístico (KALOS; WHITLOCK, 2009). É “importante ressaltar que essa técnica produz uma gama de resultados igualmente plausíveis, onde nenhuma observação é mais ou menos válida do que as outras — incluindo aquelas que não foram perturbadas” (KIAR et al., 2021).

Para aplicar a técnica de Monte Carlo na base de dados, iniciaremos analisando as variáveis numéricas do conjunto de treinamento, determinando suas distribuições. Após isso, usaremos essas distribuições para gerar valores aleatórios utilizando funções como *numpy.random.normal()* para distribuições normais. Esses valores serão usados para criar pequenas variações nas variáveis originais, aumentando a diversidade do conjunto de dados. Para finalizar, combinaremos os dados originais com os novos dados gerados, ga-

rantindo que os padrões estatísticos sejam mantidos e validaremos o conjunto expandido para assegurar que as novas amostras respeitam as características do conjunto original.

APÊNDICE E – Coeficiente de Spearman

O coeficiente de correlação de Spearman é uma ferramenta estatística bastante útil quando trabalhamos com dados que não seguem uma distribuição normal ou apresentam outliers (valores atípicos que estão muito distantes da maioria dos outros dados). Isso ocorre porque o coeficiente de Spearman não usa os valores originais dos dados, mas sim as ordens ou posições em que as observações são classificadas (SOUSA, 2019). Essa abordagem torna o coeficiente mais robusto (menos sensível) a distorções causadas por assimetria nos dados (quando os dados se distribuem de forma desigual) e por outliers (SOUSA, 2019).

O coeficiente de Spearman é uma medida não paramétrica, o que significa que ele não depende de pressupostos sobre a distribuição dos dados, como a necessidade de os dados serem normalmente distribuídos (um tipo de distribuição comum em estatísticas) (RESTREPO; GONZÁLEZ, 2007). Ele é usado para medir o grau de associação monotônica entre duas variáveis (RESTREPO; GONZÁLEZ, 2007).

Uma relação monotônica entre duas variáveis é uma relação em que uma variável tende a aumentar ou diminuir à medida que a outra também aumenta ou diminui, mas essa mudança não precisa ser em uma linha reta (RESTREPO; GONZÁLEZ, 2007). Em outras palavras, a direção da mudança nas variáveis é constante, mas não necessariamente linear.

O coeficiente de Spearman atribui um posto (ou ranking) a cada valor das variáveis (SOUSA, 2019). Isso significa que, ao invés de olhar diretamente para os valores das variáveis, ele compara a posição relativa dos dados em cada variável. Por exemplo, se temos uma variável que mede a idade e outra que mede a altura, em vez de comparar diretamente a idade e a altura, o coeficiente compara as posições relativas (ranks) de cada dado nas duas variáveis. Depois de classificar os dados dessa forma, o coeficiente de Spearman avalia a correlação entre essas classificações, ou seja, verifica o quanto as posições (ranks) das variáveis estão relacionadas entre si (SOUSA, 2019).

A fórmula do coeficiente de Spearman é dada por:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Em que:

- **di**: Diferença entre os postos de cada par de observações,
- **n**: Número de observações.

Os valores do coeficiente de Spearman variam entre -1 e +1, com:

- **+1** indica uma correlação **monotônica positiva perfeita**: À medida que uma variável aumenta, a outra também aumenta de forma consistente.
- **-1** indica uma correlação **monotônica negativa perfeita**: À medida que uma variável aumenta, a outra diminui de forma consistente.
- **0** indica **nenhuma correlação monotônica**: Não há uma relação clara entre as duas variáveis, seja positiva ou negativa.

Utilizaremos esses valores para interpretar a força e a direção da relação entre as variáveis segundo Santos (2007) na tabela 4:

Tabela 4. Interpretação do coeficiente de correlação de Spearman

Coeficiente de Correlação	Correlação
$R_{xy} = 1$	Perfeita positiva
$0,8 \leq R_{xy} < 1$	Forte positiva
$0,5 \leq R_{xy} < 0,8$	Moderada positiva
$0,1 \leq R_{xy} < 0,5$	Fraca positiva
$0 \leq R_{xy} < 0,1$	Infima positiva
0	Nula
$-0,1 \leq R_{xy} < 0$	Infima negativa
$-0,5 \leq R_{xy} < -0,1$	Fraca negativa
$-0,8 \leq R_{xy} < -0,5$	Moderada negativa
$-1 \leq R_{xy} < -0,8$	Forte negativa
$R_{xy} = -1$	Perfeita negativa

Fonte: (SANTOS, 2007)

Na base de dados, o coeficiente de Spearman será aplicado ao transformar os valores originais das variáveis em ranks e calcular a correlação entre eles usando a função `spearmanr()` da biblioteca `SciPy`. Isso permitirá identificar relações monotônicas entre as variáveis de forma robusta contra outliers. Com os resultados obtidos, determinaremos quais variáveis têm maior influência na previsão de euploidia e justificaremos com base nos coeficientes e nas visualizações geradas.

Anexos

ANEXO I – Parecer do Comitê de Ética em Pesquisa



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Parâmetros do time-lapse relacionados à ploidia de embriões humanos: investigando o uso da tecnologia para a seleção não invasiva de embriões euploides

Pesquisador: BRUNO RAMALHO DE CARVALHO

Área Temática:

Versão: 2

CAAE: 71313923.1.0000.0023

Instituição Proponente: Centro Universitário de Brasília - UNICEUB

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 6.313.392

Apresentação do Projeto:

As informações elencadas nos campos "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos Riscos e Benefícios" foram retiradas do documento de Informações Básicas da Pesquisa postado na Plataforma Brasil.

- TIPO DO ESTUDO:

Estudo observacional. Pretende-se analisar retrospectivamente os dados da fotografia time-lapse de blastocistos humanos biopsiados e, portanto, destinados ao diagnóstico genético pré-implantação (para pesquisa de aneuploidias), em ciclos de tratamento ocorridos entre 1º de janeiro de 2022 e 31 de dezembro de 2023.

- NÚMERO PARTICIPANTE DAS PESQUISA: 50

- DESCRIÇÃO DOS PARTICIPANTES: Os participantes da pesquisa serão pacientes da clínica Bruno Ramalho Reprodução Humana, que realiza tratamentos de reprodução assistida em parceria com a GENESIS Centro de Assistência em Reprodução Humana, em Brasília, Distrito Federal.

- FORMA DE RECRUTAMENTO DOS PARTICIPANTES: Os participantes das pesquisa serão abordados

Endereço: SEPN 707/907 - Bloco 6, sala 6.205, 2º andar

Bairro: Setor Universitário

CEP: 70.790-075

UF: DF

Município: BRASÍLIA

Telefone: (61)3966-1511

E-mail: cep.uniceub@uniceub.br

Continuação do Parecer: 6.313.392

pessoalmente pelo pesquisador responsável, Bruno Ramalho de Carvalho, por contato telefônico (retrospectivamente) e durante a consulta médica (prospectivamente), antes de iniciarem tratamento de reprodução assistida destinado a teste pré-implantação para pesquisa de aneuploidias (PGT-A). O termo de consentimento livre e esclarecido (TCLE) será enviado por meio digital, via e-mail, pelo sistema de prontuário da clínica Bruno Ramalho Reprodução Humana, e será, da mesma forma, assinado digitalmente pelos participantes. O sistema não permite adição do timbre ao documento. O armazenamento do TCLE poderá ser feito por ambas as partes em formato Portable Document Format (PDF).

- CRITÉRIOS DE INCLUSÃO: Serão incluídos na pesquisa todos os embriões cultivados em incubadora com tecnologia time-lapse e biopsiados para PGT-A, entre 01 de janeiro de 2023 e 31 de dezembro de 2023, sem restrições quanto a idade, classe social ou outras quaisquer.

- CRITÉRIOS DE EXCLUSÃO: Não informado.

- LOCAL ONDE SERÁ REALIZADO O ESTUDO: clínica Bruno Ramalho Reprodução Humana e GENESIS Centro de Assistência em Reprodução Humana.

- PROCEDIMENTOS QUE SERÃO REALIZADOS COM OS PARTICIPANTES: Coleta de dados de prontuários.

- MÉTODO DE COLETA DE DADOS/INFORMAÇÕES:

Pretende-se analisar retrospectivamente os dados da fotografia time-lapse de blastocistos humanos biopsiados e, portanto, destinados ao diagnóstico genético pré-implantação (para pesquisa de aneuploidias), em ciclos de tratamento ocorridos entre 1º de janeiro de 2022 e 31 de dezembro de 2023.

Variáveis a serem analisadas: início de t2 (st2), correspondendo aos primeiros movimentos citoplasmáticos anteriores à primeira citocinese (o primeiro frame detectável dos seguintes movimentos: desaparecimento do halo, ondas citoplasmáticas ou movimentos do citoplasma circulante); t3; t5; tSB; tb; cc3; e t5–t2, e outros parâmetros eventualmente identificados como relevantes; e o resultado da inteligência artificial KIDSCORE D5.

- METODOLOGIA DE ANÁLISE DE DADOS: Os dados serão tabulados e analisados estatisticamente,

Endereço: SEPN 707/907 - Bloco 6, sala 6.205, 2º andar

Bairro: Setor Universitário

CEP: 70.790-075

UF: DF

Município: BRASILIA

Telefone: (61)3966-1511

E-mail: cep.uniceub@uniceub.br

Continuação do Parecer: 6.313.392

com auxílio de software especializado.

Objetivo da Pesquisa:

Objetivo Primário: Identificar se existe correlação positiva entre o parâmetro de clivagem precoce st2e a ploidia de embriões humanos.

Objetivo Secundário: Existe correlação positiva entre os parâmetros de clivagem precoce t3, t5, tSB, tB, cc3 e t5-t2, e outros a serem identificados, e a ploidia de embriões humanos.

Avaliação dos Riscos e Benefícios:

RISCOS: Não há riscos às pacientes, uma vez que a análise será retrospectiva, ou seja, não haverá interferência da pesquisa sobre o desenvolvimento dos tratamentos. Os riscos potenciais relacionados à pesquisa envolvem a quebra da confidencialidade e, por esse motivo, os pesquisadores envolvidos, sob coordenação do pesquisador responsável, comprometem-se a manter em sigilo quanto à identidade dos genitores dos embriões incluídos, bem como a quaisquer dados que possibilitem a quebra do anonimato.

BENEFÍCIOS: Desenvolver ferramentas não invasivas de avaliação embrionária (quanto à ploidia), que possam substituir de forma satisfatória a biopsia embrionária para pesquisa de aneuploidias.

Comentários e Considerações sobre a Pesquisa:

- Devido à natureza do estudo, considera-se a pesquisa com risco mínimo.
- Houve indicação correta das medidas protetivas para o risco apresentado.
- Orçamento: os gastos serão custeados pelo pesquisador.
- Cronograma: A coleta de dados está prevista para iniciar-se em agosto de 2023.

Considerações sobre os Termos de apresentação obrigatória:

- Apresentou a Folha de Rosto devidamente preenchida e assinada.
- Apresentou o Termo de Aceite Institucional devidamente preenchido e assinado.
- O Termo de Consentimento Livre e Esclarecido (TCLE) foi apresentado.

Recomendações:

Ao final do estudo os pesquisadores devem enviar o Relatório de Finalização da Pesquisa ao CEP. O envio de relatórios deverá ocorrer pela Plataforma Brasil, por meio de notificação.

Endereço: SEPN 707/907 - Bloco 6, sala 6.205, 2º andar

Bairro: Setor Universitário

CEP: 70.790-075

UF: DF

Município: BRASÍLIA

Telefone: (61)3966-1511

E-mail: cep.uniceub@uniceub.br

Conclusões ou Pendências e Lista de Inadequações:

O pesquisador atendeu às solicitações indicadas pelo CEP:

- O pesquisador deverá inserir no TCLE o esclarecimento sobre os riscos de quebra de confidencialidade e as medidas protetivas, conforme informado nas informações básicas do projeto que constam na Plataforma Brasil: "Os riscos potenciais relacionados à pesquisa envolvem a quebra da confidencialidade e, por esse motivo, os pesquisadores envolvidos, sob coordenação do pesquisador responsável, comprometem-se a manter em sigilo quanto à identidade dos genitores dos embriões incluídos, bem como a quaisquer dados que possibilitem a quebra do anonimato."

O pesquisador realizou as seguintes adequações ao projeto:

1. Informar o local onde será realizado o estudo (nome da clínica) e inserir o TERMO DE ACEITE INSTITUCIONAL DEVIDAMENTE PREENCHIDO E ASSINADO. Conforme estabelecido na resolução 466/2012 do Conselho Nacional de Saúde, a instituição coparticipante de pesquisa é a organização, pública ou privada, legitimamente constituída e habilitada, na qual alguma das fases ou etapas da pesquisa se desenvolve. PENDÊNCIA ATENDIDA

2. Descrever a forma de abordagem dos participante das pesquisa. O pesquisador deverá informar como e em que momento serão abordados os convidados a participantes de pesquisa, com a descrição do processo e do registro do consentimento. A Norma Operacional Nº 001/2013 do Conselho Nacional de Saúde / Ministério da Saúde que dispõe sobre a organização e funcionamento do Sistema CEP/CONEP e sobre os procedimentos para submissão, avaliação e acompanhamento da pesquisa e de desenvolvimento envolvendo seres humanos no Brasil define, no item 3.4.1, que todos os protocolos de pesquisa devem conter, obrigatoriamente, A DESCRIÇÃO DA FORMA DE ABORDAGEM OU PLANO DE RECRUTAMENTO DOS POSSÍVEIS INDIVÍDUOS PARTICIPANTES. PENDÊNCIA ATENDIDA

3. Inserir a descrição dos participantes da pesquisa. Para analisar a eticidade do projeto de pesquisa, a pesquisa deverá conter as características da população a ser estudada, como faixa etária, sexo, grupos sociais e outras que sejam pertinentes à descrição da população e que

Endereço: SEPN 707/907 - Bloco 6, sala 6.205, 2º andar

Bairro: Setor Universitário

CEP: 70.790-075

UF: DF

Município: BRASÍLIA

Telefone: (61)3966-1511

E-mail: cep.uniceub@uniceub.br

Continuação do Parecer: 6.313.392

possam ser significativas para a caracterização da amostra, população e análise ética da pesquisa.
PENDÊNCIA ATENDIDA

4. Informar os CRITÉRIOS DE INCLUSÃO e de EXCLUSÃO dos participantes da pesquisa. PENDÊNCIA ATENDIDA. O pesquisador informou o critério de inclusão. Entende-se que os critérios de exclusão poderão ser estabelecidos após a consulta aos dados dos prontuários.

5. O Termo de Consentimento Livre e Esclarecido (TCLE) não foi apresentado de forma adequada. O TCLE necessita de adequações.

5.1. Colocar o nome e logotipo da instituição proponente (CEUB); JUSTIFICATIVA APRESENTADA

5.2. As Resoluções em Pesquisa com seres humanos do Conselho Nacional de Saúde asseguram aos participantes da pesquisa a possibilidade de contato com o CEP que aprovou o projeto de pesquisa. O TCLE deve conter o contato do CEP UniCEUB para que o participante da pesquisa possa fazer considerações, tirar dúvidas ou informar ocorrências relacionadas ao estudo. Para isso o pesquisador deve inserir os dados do CEP do UniCEUB no TCLE: Comitê de Ética em Pesquisa do Centro Universitário de Brasília – CEP/UniCEUB. Telefone 3966.1511 E-mail cep.uniceub@uniceub.br. PENDÊNCIA ATENDIDA

5.3. Inserir o contato dos pesquisadores no TCLE. Conforme especificado na resolução 466/2012 do CNS, no TCLE deve constar o endereço e contato telefônico dos responsáveis pela pesquisa. Considerando o disposto na Resolução nº 466/2012, do Conselho Nacional de Saúde, do Ministério da Saúde, no que se refere ao PROCESSO DE CONSENTIMENTO LIVRE E ESCLARECIDO, consta em seu item IV.5 d) que o TCLE deve “ser elaborado em duas vias, rubricadas em todas as suas páginas e assinadas, ao seu término, pelo convidado a participar da pesquisa, ou por seu representante legal, assim como pelo pesquisador responsável, ou pela(s) pessoa(s) por ele delegada(s), devendo as páginas de assinaturas estar na mesma folha. EM AMBAS AS VIAS DEVERÃO CONSTAR O ENDEREÇO E CONTATO TELEFÔNICO DOS RESPONSÁVEIS PELA PESQUISA E DO CEP LOCAL E DA CONEP, QUANDO PERTINENTE.” PENDÊNCIA ATENDIDA.

O CEP-UniCEUB ressalta a necessidade de desenvolvimento da pesquisa, de acordo com o

Endereço: SEPN 707/907 - Bloco 6, sala 6.205, 2º andar

Bairro: Setor Universitário

CEP: 70.790-075

UF: DF

Município: BRASÍLIA

Telefone: (61)3966-1511

E-mail: cep.uniceub@uniceub.br

Continuação do Parecer: 6.313.392

protocolo avaliado e aprovado, bem como, atenção às diretrizes éticas nacionais quanto aos incisos XI.1 e XI.2 da Resolução nº 466/12 CNS/MS concernentes às responsabilidades do pesquisador no desenvolvimento do projeto:

XI.1 - A responsabilidade do pesquisador é indelegável e indeclinável e compreende os aspectos éticos e legais.

XI.2 - Cabe ao pesquisador:

- c) desenvolver o projeto conforme delineado;
- d) elaborar e apresentar os relatórios parciais e final;
- e) apresentar dados solicitados pelo CEP ou pela CONEP a qualquer momento;
- f) manter os dados da pesquisa em arquivo, físico ou digital, sob sua guarda e responsabilidade, por um período de 5 anos após o término da pesquisa;
- g) encaminhar os resultados da pesquisa para publicação, com os devidos créditos aos pesquisadores associados e ao pessoal técnico integrante do projeto; e
- h) justificar fundamentadamente, perante o CEP ou a CONEP, interrupção do projeto ou a não publicação dos resultados.

Considerações Finais a critério do CEP:

Protocolo previamente avaliado, com parecer homologado na 16ª Reunião Ordinária do CEP-UniCEUB do ano em 15 de setembro de 2023.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2138177.pdf	21/08/2023 12:51:16		Aceito
Recurso Anexado pelo Pesquisador	Carta_de_Envio_de_Pendencias_assinado.pdf	21/08/2023 12:51:06	BRUNO RAMALHO DE CARVALHO	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_integra.docx	21/08/2023 12:34:41	BRUNO RAMALHO DE CARVALHO	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLETimeLapse.docx	21/08/2023 12:32:05	BRUNO RAMALHO DE CARVALHO	Aceito
TCLE / Termos de	Anuencia_BRUNO.pdf	21/08/2023	BRUNO RAMALHO	Aceito

Endereço: SEPN 707/907 - Bloco 6, sala 6.205, 2º andar

Bairro: Setor Universitário

CEP: 70.790-075

UF: DF

Município: BRASÍLIA

Telefone: (61)3966-1511

E-mail: cep.uniceub@uniceub.br



Continuação do Parecer: 6.313.392

Assentimento / Justificativa de Ausência	Anuencia_BRUNO.pdf	11:57:20	DE CARVALHO	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Anuencia_GENESIS.pdf	21/08/2023 11:57:08	BRUNO RAMALHO DE CARVALHO	Aceito
Declaração de Pesquisadores	DECLARACAO_DE_CONFIDENCIALID ADE_PARA_PESQUISA_CIENTIFICA_ modelo assinado.pdf	05/07/2023 08:50:07	BRUNO RAMALHO DE CARVALHO	Aceito
Folha de Rosto	folhaDeRostoAssinada.pdf	22/06/2023 17:01:17	BRUNO RAMALHO DE CARVALHO	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

BRASILIA, 20 de Setembro de 2023

Assinado por:
Marilia de Queiroz Dias Jacome
(Coordenador(a))

Endereço: SEPN 707/907 - Bloco 6, sala 6.205, 2º andar

Bairro: Setor Universitário

CEP: 70.790-075

UF: DF

Município: BRASILIA

Telefone: (61)3966-1511

E-mail: cep.uniceub@uniceub.br

ANEXO II – Termo de Consentimento para Utilização de Dados de Entrevistas, Gravação de Reuniões e Uso de Gravação

Termo de Consentimento para Utilização de Dados de Entrevistas, Gravação de Reuniões e Uso de Gravação

Pelo presente instrumento, de um lado, o Dr. **Bruno Ramalho de Carvalho**, inscrito no CPF sob o número **539.662.101-00**, inscrito no CRM DF sob o número **16335** – Ginecologia e Obstetrícia (RQE 13573) e Reprodução Assistida (RQE 15521), representante da **Clínica Bruno Ramalho Reprodução Humana**, em parceria com a **GENESIS Centro de Assistência em Reprodução Humana**, situada em Brasília, Distrito Federal, doravante denominado **AUTORIZANTE**; e, de outro lado, as alunas **Maria Eduarda Dos Santos Abritta Ferreira**, inscrita no CPF sob o número **081.776.091-14**, e **Sabrina Caldas Berno**, inscrita no CPF sob o número **033.782.041-41**, ambas estudantes da Universidade de Brasília (UnB), doravante denominadas **PESQUISADORAS**, resolvem firmar o presente contrato mediante as cláusulas e condições abaixo estabelecidas:

1. Objeto do Contrato

O presente contrato tem como objeto a autorização para a utilização das falas do **AUTORIZANTE**, extraídas das entrevistas realizadas, como referencial teórico para a pesquisa das **PESQUISADORAS**, no contexto do Trabalho de Conclusão de Curso (TCC) intitulado "Desenvolver uma abordagem baseada em inteligência artificial para identificar padrões em dados morfocinéticos de embriões obtidos por meio do Time-Lapse System, capaz de prever a porcentagem de euploidia, proporcionando uma solução mais eficaz e menos invasiva em comparação ao PGT-A", elaborado pelas **PESQUISADORAS**.

Além disso, o **AUTORIZANTE** autoriza a gravação de reuniões, seja em áudio, vídeo ou qualquer outra forma de registro, para fins de estudo, análise e documentação, conforme os objetivos da pesquisa.

2. Autorização para Gravação de Entrevistas e Reuniões

O **AUTORIZANTE** autoriza a gravação completa da entrevista, bem como de reuniões subsequentes, seja em áudio, vídeo ou qualquer outra forma de registro, para fins de estudo, análise e documentação. As gravações poderão ser utilizadas integralmente, incluindo trechos ou citações, nas etapas da pesquisa conduzida pelas **PESQUISADORAS**.

3. Uso das Respostas e Trechos das Gravações

O **AUTORIZANTE** concede permissão para que suas respostas fornecidas durante a entrevista, bem como as falas extraídas das reuniões, sejam utilizadas como referencial teórico do Trabalho de Conclusão de Curso (TCC). Tais falas poderão ser incorporadas, integralmente ou em trechos relacionados ao estudo realizado pelas **PESQUISADORAS**.

4. Confidencialidade e Identificação

O **AUTORIZANTE** reconhece que, na utilização das gravações, sua identidade e informações pessoais poderão ser reveladas, caso seja necessário para a pesquisa. O **AUTORIZANTE** autoriza a associação de sua identidade às suas falas ou respostas, conforme o objetivo acadêmico ou científico da pesquisa. Caso o **AUTORIZANTE** deseje que sua identidade não seja revelada, poderá solicitar a anonimização dos dados.

5. Revogação do Consentimento

O **AUTORIZANTE** tem pleno conhecimento de que pode revogar este consentimento a qualquer momento, mediante solicitação por escrito. A revogação não afetará o uso das falas ou dados já coletados até o momento da solicitação.

6. Direitos sobre o Material

O **AUTORIZANTE** reconhece que o material resultante das gravações das entrevistas e reuniões, incluindo suas respostas e falas, poderá ser armazenado, analisado e divulgado exclusivamente para fins acadêmicos ou científicos. Concorde que não receberá compensação financeira ou qualquer outra forma de benefício pelo uso do material gerado.

7. ASSINATURAS

Por estarem de pleno acordo com as condições deste contrato, as partes o assinam.
Brasília, 24 de dezembro de 2024

AUTORIZANTE: Dr. Bruno Ramalho de Carvalho

CRM DF 16335

Assinatura: _____



PESQUISADORAS: Maria Eduarda Dos Santos Abritta Ferreira

CPF: 081.776.091-74

Assinatura: _____



Sabrina Caldas Berno

CPF: 033.782.041-41

Assinatura: _____



TESTEMUNHA:

1. Nome: George Marsicano Correa

2. CPF: 818.841.921-49

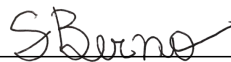
3. Assinatura: _____



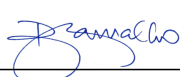
Página de assinaturas



Maria Ferreira
081.776.091-14
Signatário



Sabrina Berno
033.782.041-41
Signatário











Bruno Carvalho
539.662.101-00
Signatário



George Correa
818.841.921-49
Signatário

HISTÓRICO

24 dez 2024 09:07:42		Maria Eduarda Dos Santos Abritta Ferreira criou este documento. (Email: eduardaabritta@gmail.com, CPF: 081.776.091-14)
24 dez 2024 09:07:42		Maria Eduarda Dos Santos Abritta Ferreira (Email: eduardaabritta@gmail.com, CPF: 081.776.091-14) visualizou este documento por meio do IP 177.158.135.161 localizado em Brasília - Federal District - Brazil
24 dez 2024 09:08:23		Maria Eduarda Dos Santos Abritta Ferreira (Email: eduardaabritta@gmail.com, CPF: 081.776.091-14) assinou este documento por meio do IP 177.158.135.161 localizado em Brasília - Federal District - Brazil
24 dez 2024 11:10:10		Sabrina Caldas Berno (Email: sabrinacberno@gmail.com, CPF: 033.782.041-41) visualizou este documento por meio do IP 177.174.208.14 localizado em Brasília - Federal District - Brazil
24 dez 2024 11:15:29		Sabrina Caldas Berno (Email: sabrinacberno@gmail.com, CPF: 033.782.041-41) assinou este documento por meio do IP 189.61.13.144 localizado em Brasília - Federal District - Brazil
24 dez 2024 11:21:30		Bruno Ramalho de Carvalho (Email: bruno.ramalho@ceub.edu.br, CPF: 539.662.101-00) visualizou este documento por meio do IP 189.6.30.39 localizado em Brasília - Federal District - Brazil
24 dez 2024 11:22:24		Bruno Ramalho de Carvalho (Email: bruno.ramalho@ceub.edu.br, CPF: 539.662.101-00) assinou este documento por meio do IP 189.6.30.39 localizado em Brasília - Federal District - Brazil
06 jan 2025 14:17:53		George Marsicano Correa (Email: georgemarsicano@unb.br, CPF: 818.841.921-49) visualizou este documento por meio do IP 179.214.113.5 localizado em Brasília - Federal District - Brazil



06 jan 2025

14:18:46



George Marsicano Correa (Email: georgemarsicano@unb.br, CPF: 818.841.921-49) assinou este documento por meio do IP 179.214.113.5 localizado em Brasília - Federal District - Brazil



ANEXO III – Contrato de Autorização para Utilização de Dados em Pesquisa

CONTRATO DE AUTORIZAÇÃO PARA UTILIZAÇÃO DE DADOS EM PESQUISA

IDENTIFICAÇÃO DAS PARTES

Pelo presente instrumento, de um lado, o Dr. **Bruno Ramalho de Carvalho**, inscrita no CPF sob o número **539.662.101-00** inscrito no CRM DF sob o número **16335** – Ginecologia e Obstetrícia (RQE 13573) e Reprodução Assistida (RQE 15521), representante da **Clínica Bruno Ramalho Reprodução Humana**, em parceria com a **GENESIS Centro de Assistência em Reprodução Humana**, situada em Brasília, Distrito Federal, doravante denominado **AUTORIZANTE**; e, de outro lado, as alunas **Maria Eduarda Dos Santos Abritta Ferreira**, inscrita no CPF sob o número **081.776.091-14**, e **Sabrina Caldas Berno**, inscrita no CPF sob o número **033.782.041-41**, ambas estudantes da Universidade de Brasília (UnB), doravante denominadas **PESQUISADORAS**, resolvem firmar o presente contrato mediante as cláusulas e condições abaixo estabelecidas.

1. CLÁUSULA PRIMEIRA – OBJETO DO CONTRATO

O presente contrato tem como objeto a autorização para a utilização de dados anônimos de pacientes da Clínica Bruno Ramalho Reprodução Humana, com o objetivo exclusivo de realizar análise estatística e desenvolver modelos no contexto do Trabalho de Conclusão de Curso (TCC) intitulado “**Desenvolver uma abordagem baseada em inteligência artificial para identificar padrões em dados morfocinéticos de embriões obtidos por meio do Time-Lapse System, capaz de prever a porcentagem de euploidia, proporcionando uma solução mais eficaz e menos invasiva em comparação ao PGT-A**”, elaborado pelas PESQUISADORAS.

Os dados que serão utilizados incluem: **Idade, Data da biópsia, Embrião n., Estágio, Morfo, Kidscore, st2, t2, t3, t4, t5, t8, tSC, tSB, tB, t2-st2, cc2 (t3-t2), cc3 (t5-t3), t5-t2, s2 (t4-t3), s3 (t8-t5), tSC-t8, tB-tSB, Ploidia.**

2. CLÁUSULA SEGUNDA – TERMO DE CONSENTIMENTO

O AUTORIZANTE declara que possui autorização vigente junto à Plataforma Brasil para a utilização dos dados das pacientes incluídos no presente contrato, garantindo que tais dados foram obtidos com o devido consentimento informado e em conformidade com as normas éticas aplicáveis.

3. CLÁUSULA TERCEIRA – RESPONSABILIDADE E CONFIDENCIALIDADE

1. Os dados serão fornecidos em formato anonimizado, identificados apenas por códigos (ID), impossibilitando a associação com a identidade das pacientes.
2. As PESQUISADORAS comprometem-se a manter o sigilo e a confidencialidade dos dados fornecidos, utilizando-os exclusivamente para os fins especificados neste contrato.
3. Fica vedada qualquer tentativa de reidentificação dos dados ou de uso para outros fins alheios ao projeto descrito na Cláusula Primeira.

4. CLÁUSULA QUARTA – PRAZO DE UTILIZAÇÃO

A utilização dos dados estará autorizada até a finalização do TCC 2 das PESQUISADORAS, prevista para o final do semestre 2025.1 da Universidade de Brasília.

5. CLÁUSULA QUINTA – DIREITOS DE AUTORIA E PUBLICAÇÃO

O AUTORIZANTE concorda que os dados utilizados pelas PESQUISADORAS poderão ser mencionados em publicações acadêmicas ou apresentações públicas relacionadas ao TCC, desde que sejam preservados o anonimato das pacientes e os créditos à Clínica Bruno Ramalho Reprodução Humana e à GENESIS Centro de Assistência em Reprodução Humana.

6. CLÁUSULA SEXTA – REVOGAÇÃO DA AUTORIZAÇÃO

O AUTORIZANTE poderá revogar a autorização a qualquer momento, desde que notifique as PESQUISADORAS por escrito. Contudo, a revogação não obrigará as PESQUISADORAS a excluir os estudos ou análises já realizados, mas impedirá o uso dos dados para qualquer outro estudo futuro.

7. CLÁUSULA SÉTIMA – DISPOSIÇÕES GERAIS

1. Este contrato não prevê cláusulas de indenização.
2. Qualquer situação omissa será resolvida em comum acordo entre as partes, respeitando-se as legislações vigentes.

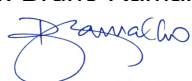
8. ASSINATURAS

Por estarem de pleno acordo com as condições deste contrato, as partes o assinam.
Brasília, 24 de dezembro de 2024

AUTORIZANTE: Dr. Bruno Ramalho de Carvalho

CRM DF 16335

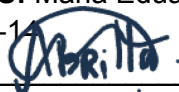
Assinatura: _____



PESQUISADORAS: Maria Eduarda Dos Santos Abritta Ferreira

CPF: 081.776.091-14

Assinatura: _____



Sabrina Caldas Berno

CPF: 033.782.041-41

Assinatura: _____



TESTEMUNHA:

1. Nome: George Marsicano Correa

2. CPF: 818.841.921-49

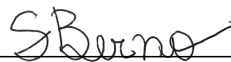
3. Assinatura: _____



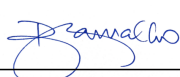
Página de assinaturas



Maria Ferreira
081.776.091-14
Signatário



Sabrina Berno
033.782.041-41
Signatário











Bruno Carvalho
539.662.101-00
Signatário



George Correa
818.841.921-49
Signatário

HISTÓRICO

24 dez 2024 09:06:20		Maria Eduarda Dos Santos Abritta Ferreira criou este documento. (Email: eduardaabritta@gmail.com, CPF: 081.776.091-14)
24 dez 2024 09:06:20		Maria Eduarda Dos Santos Abritta Ferreira (Email: eduardaabritta@gmail.com, CPF: 081.776.091-14) visualizou este documento por meio do IP 177.158.135.161 localizado em Brasília - Federal District - Brazil
24 dez 2024 09:06:25		Maria Eduarda Dos Santos Abritta Ferreira (Email: eduardaabritta@gmail.com, CPF: 081.776.091-14) assinou este documento por meio do IP 177.158.135.161 localizado em Brasília - Federal District - Brazil
24 dez 2024 10:22:27		Sabrina Caldas Berno (Email: sabrinacberno@gmail.com, CPF: 033.782.041-41) visualizou este documento por meio do IP 189.61.13.144 localizado em Brasília - Federal District - Brazil
24 dez 2024 10:22:43		Sabrina Caldas Berno (Email: sabrinacberno@gmail.com, CPF: 033.782.041-41) assinou este documento por meio do IP 189.61.13.144 localizado em Brasília - Federal District - Brazil
24 dez 2024 10:37:29		Bruno Ramalho de Carvalho (Email: bruno.ramalho@ceub.edu.br, CPF: 539.662.101-00) visualizou este documento por meio do IP 189.6.30.39 localizado em Brasília - Federal District - Brazil
24 dez 2024 10:43:03		Bruno Ramalho de Carvalho (Email: bruno.ramalho@ceub.edu.br, CPF: 539.662.101-00) assinou este documento por meio do IP 189.6.30.39 localizado em Brasília - Federal District - Brazil
06 jan 2025 14:13:30		George Marsicano Correa (Email: georgemarsicano@unb.br, CPF: 818.841.921-49) visualizou este documento por meio do IP 179.214.113.5 localizado em Brasília - Federal District - Brazil



06 jan 2025

14:17:18



George Marsicano Correa (Email: georgemarsicano@unb.br, CPF: 818.841.921-49) assinou este documento por meio do IP 179.214.113.5 localizado em Brasília - Federal District - Brazil



ANEXO IV – Correlações entre os Parâmetros

Variable 1	Variable 2	Spearman Coefficient
Idade	Estágio	-0,06803154976
Idade	Morfo	0,07873460187
Idade	t2	0,003227859865
Idade	t3	-0,08677913323
Idade	t4	-0,1509451885
Idade	t5	0,1083875764
Idade	t8	-0,049837301
Idade	tSC	0,01753619697
Idade	tSB	-0,09797958164
Idade	tB	-0,01820525995
Idade	cc2 (t3-t2)	-0,1699184901
Idade	cc3 (t5-t3)	0,1979686591
Idade	t5-t2	0,09499783155
Idade	s2 (t4-t3)	-0,2402597189
Idade	s3 (t8-t5)	-0,2846417801
Idade	tSC-t8	0,1343924326
Idade	tB-tSB	0,2047747578
Idade	Ploidia	-0,49663356

Variable 1	Variable 2	Spearman Coefficient
Estágio	Idade	-0,06803154976
Estágio	Morfo	-0,005904052193
Estágio	t2	0,3003883349
Estágio	t3	0,2470687888
Estágio	t4	0,3193854689
Estágio	t5	0,1506315693
Estágio	t8	0,1723502603
Estágio	tSC	0,4978627858
Estágio	tSB	0,5264654473
Estágio	tB	0,5899587323
Estágio	cc2 (t3-t2)	0,22581076
Estágio	cc3 (t5-t3)	0,07546655566
Estágio	t5-t2	0,1402176889
Estágio	s2 (t4-t3)	-0,004978016617
Estágio	s3 (t8-t5)	-0,0187086839
Estágio	tSC-t8	0,3996493404
Estágio	tB-tSB	0,3758105227
Estágio	Ploidia	-0,2375978079

Variable 1	Variable 2	Spearman Coefficient
Morfo	Idade	0,07873460187
Morfo	Estágio	-0,005904052193
Morfo	t2	-0,376687833
Morfo	t3	-0,4262600544
Morfo	t4	-0,3685539564
Morfo	t5	-0,4458158367
Morfo	t8	-0,2491571072
Morfo	tSC	-0,2037231082
Morfo	tSB	-0,3010756644
Morfo	tB	-0,2393928004
Morfo	cc2 (t3-t2)	-0,3807388029
Morfo	cc3 (t5-t3)	-0,3098662828
Morfo	t5-t2	-0,3834337484
Morfo	s2 (t4-t3)	-0,1042668125
Morfo	s3 (t8-t5)	0,1246602606
Morfo	tSC-t8	-0,02474783976
Morfo	tB-tSB	0,1370397427
Morfo	Ploidia	0,0452976895

Variable 1	Variable 2	Spearman Coefficient
t2	Idade	0,003227859865
t2	Estágio	0,3003883349
t2	Morfo	-0,376687833
t2	t3	0,7807114034
t2	t4	0,8923082808
t2	t5	0,4752170738
t2	t8	0,5845921133
t2	tSC	0,3980774432
t2	tSB	0,4339687436
t2	tB	0,3954973318
t2	cc2 (t3-t2)	0,3893884002
t2	cc3 (t5-t3)	0,1455082082
t2	t5-t2	0,2125281725
t2	s2 (t4-t3)	-0,01276526777
t2	s3 (t8-t5)	0,1051061521
t2	tSC-t8	-0,005307220087
t2	tB-tSB	0,0581978328
t2	Ploidia	-0,07523975181

Variable 1	Variable 2	Spearman Coefficient
t3	Idade	-0,08677913323
t3	Estágio	0,2470687888
t3	Morfo	-0,4262600544
t3	t2	0,7807114034
t3	t4	0,8128813819
t3	t5	0,6719684722
t3	t8	0,5701829369
t3	tSC	0,4203147049
t3	tSB	0,564496118
t3	tB	0,4763047775
t3	cc2 (t3-t2)	0,8030657477
t3	cc3 (t5-t3)	0,2471823325
t3	t5-t2	0,5051710813
t3	s2 (t4-t3)	-0,2185074556
t3	s3 (t8-t5)	-0,07159964297
t3	tSC-t8	0,01735718026
t3	tB-tSB	-0,07616061666
t3	Ploidia	-0,05565195514

Variable 1	Variable 2	Spearman Coefficient
t4	Idade	-0,1509451885
t4	Estágio	0,3193854689
t4	Morfo	-0,3685539564
t4	t2	0,8923082808
t4	t3	0,8128813819
t4	t5	0,5592244215
t4	t8	0,6438951944
t4	tSC	0,425503198
t4	tSB	0,5659476782
t4	tB	0,5034399087
t4	cc2 (t3-t2)	0,5914605928
t4	cc3 (t5-t3)	0,2635699936
t4	t5-t2	0,3927200878
t4	s2 (t4-t3)	0,1632097301
t4	s3 (t8-t5)	0,08671678117
t4	tSC-t8	-0,03118791693
t4	tB-tSB	0,008409810914
t4	Ploidia	-0,01545912888

Variable 1	Variable 2	Spearman Coefficient
t5	Idade	0,1083875764
t5	Estágio	0,1506315693
t5	Morfo	-0,4458158367
t5	t2	0,4752170738
t5	t3	0,6719684722
t5	t4	0,5592244215
t5	t8	0,5155937057
t5	tSC	0,3503728314
t5	tSB	0,385505056
t5	tB	0,3466009922
t5	cc2 (t3-t2)	0,6359844894
t5	cc3 (t5-t3)	0,811878306
t5	t5-t2	0,9203712273
t5	s2 (t4-t3)	-0,09073945813
t5	s3 (t8-t5)	-0,3933930993
t5	tSC-t8	0,02937478582
t5	tB-tSB	-0,03673564922
t5	Ploidia	-0,237550108

Variable 1	Variable 2	Spearman Coefficient
t8	Idade	-0,049837301
t8	Estágio	0,1723502603
t8	Morfo	-0,2491571072
t8	t2	0,5845921133
t8	t3	0,5701829369
t8	t4	0,6438951944
t8	t5	0,5155937057
t8	tSC	0,3480565564
t8	tSB	0,3964484162
t8	tB	0,3723618474
t8	cc2 (t3-t2)	0,4555651596
t8	cc3 (t5-t3)	0,3960224335
t8	t5-t2	0,4471863769
t8	s2 (t4-t3)	0,1004338265
t8	s3 (t8-t5)	0,4848122012
t8	tSC-t8	-0,2948597241
t8	tB-tSB	-0,03389138651
t8	Ploidia	-0,06801276287

Variable 1	Variable 2	Spearman Coefficient
tSC	Idade	0,01753619697
tSC	Estágio	0,4978627858
tSC	Morfo	-0,2037231082
tSC	t2	0,3980774432
tSC	t3	0,4203147049
tSC	t4	0,425503198
tSC	t5	0,3503728314
tSC	t8	0,3480565564
tSC	tSB	0,750238094
tSC	tB	0,7385332085
tSC	cc2 (t3-t2)	0,4053979429
tSC	cc3 (t5-t3)	0,2439120158
tSC	t5-t2	0,3370434283
tSC	s2 (t4-t3)	0,1161357851
tSC	s3 (t8-t5)	0,00512150653
tSC	tSC-t8	0,7021502392
tSC	tB-tSB	0,09964296621
tSC	Ploidia	-0,043280378

Variable 1	Variable 2	Spearman Coefficient
tSB	Idade	-0,09797958164
tSB	Estágio	0,5264654473
tSB	Morfo	-0,3010756644
tSB	t2	0,4339687436
tSB	t3	0,564496118
tSB	t4	0,5659476782
tSB	t5	0,385505056
tSB	t8	0,3964484162
tSB	tSC	0,750238094
tSB	tB	0,9301282715
tSB	cc2 (t3-t2)	0,6349372129
tSB	cc3 (t5-t3)	0,1954770291
tSB	t5-t2	0,3957634886
tSB	s2 (t4-t3)	0,08385774575
tSB	s3 (t8-t5)	0,009660494872
tSB	tSC-t8	0,4420734695
tSB	tB-tSB	0,05277594867
tSB	Ploidia	-0,1107753494

Variable 1	Variable 2	Spearman Coefficient
tB	Idade	-0,01820525995
tB	Estágio	0,5899587323
tB	Morfo	-0,2393928004
tB	t2	0,3954973318
tB	t3	0,4763047775
tB	t4	0,5034399087
tB	t5	0,3466009922
tB	t8	0,3723618474
tB	tSC	0,7385332085
tB	tSB	0,9301282715
tB	cc2 (t3-t2)	0,5384017028
tB	cc3 (t5-t3)	0,2008046693
tB	t5-t2	0,371212576
tB	s2 (t4-t3)	0,1089522412
tB	s3 (t8-t5)	-0,009551800171
tB	tSC-t8	0,4719478413
tB	tB-tSB	0,3755551451
tB	Ploidia	-0,1695148138

Variable 1	Variable 2	Spearman Coefficient
cc2 (t3-t2)	Idade	-0,1699184901
cc2 (t3-t2)	Estágio	0,22581076
cc2 (t3-t2)	Morfo	-0,3807388029
cc2 (t3-t2)	t2	0,3893884002
cc2 (t3-t2)	t3	0,8030657477
cc2 (t3-t2)	t4	0,5914605928
cc2 (t3-t2)	t5	0,6359844894
cc2 (t3-t2)	t8	0,4555651596
cc2 (t3-t2)	tSC	0,4053979429
cc2 (t3-t2)	tSB	0,6349372129
cc2 (t3-t2)	tB	0,5384017028
cc2 (t3-t2)	cc3 (t5-t3)	0,3395425866
cc2 (t3-t2)	t5-t2	0,6541337469
cc2 (t3-t2)	s2 (t4-t3)	-0,2080620185
cc2 (t3-t2)	s3 (t8-t5)	-0,1674262721
cc2 (t3-t2)	tSC-t8	0,08164410649
cc2 (t3-t2)	tB-tSB	-0,1271753196
cc2 (t3-t2)	Ploidia	-0,03144417624

Variable 1	Variable 2	Spearman Coefficient
cc3 (t5-t3)	Idade	0,1979686591
cc3 (t5-t3)	Estágio	0,07546655566
cc3 (t5-t3)	Morfo	-0,3098662828
cc3 (t5-t3)	t2	0,1455082082
cc3 (t5-t3)	t3	0,2471823325
cc3 (t5-t3)	t4	0,2635699936
cc3 (t5-t3)	t5	0,811878306
cc3 (t5-t3)	t8	0,3960224335
cc3 (t5-t3)	tSC	0,2439120158
cc3 (t5-t3)	tSB	0,1954770291
cc3 (t5-t3)	tB	0,2008046693
cc3 (t5-t3)	cc2 (t3-t2)	0,3395425866
cc3 (t5-t3)	t5-t2	0,9007328881
cc3 (t5-t3)	s2 (t4-t3)	0,08134278338
cc3 (t5-t3)	s3 (t8-t5)	-0,3592297126
cc3 (t5-t3)	tSC-t8	-0,01041485259
cc3 (t5-t3)	tB-tSB	0,006397378179
cc3 (t5-t3)	Ploidia	-0,2819405223

Variable 1	Variable 2	Spearman Coefficient
t5-t2	Idade	0,09499783155
t5-t2	Estágio	0,1402176889
t5-t2	Morfo	-0,3834337484
t5-t2	t2	0,2125281725
t5-t2	t3	0,5051710813
t5-t2	t4	0,3927200878
t5-t2	t5	0,9203712273
t5-t2	t8	0,4471863769
t5-t2	tSC	0,3370434283
t5-t2	tSB	0,3957634886
t5-t2	tB	0,371212576
t5-t2	cc2 (t3-t2)	0,6541337469
t5-t2	cc3 (t5-t3)	0,9007328881
t5-t2	s2 (t4-t3)	-0,02962358653
t5-t2	s3 (t8-t5)	-0,4122958851
t5-t2	tSC-t8	0,04660592616
t5-t2	tB-tSB	-0,04161200952
t5-t2	Ploidia	-0,2277558204

Variable 1	Variable 2	Spearman Coefficient
s2 (t4-t3)	Idade	-0,2402597189
s2 (t4-t3)	Estágio	-0,004978016617
s2 (t4-t3)	Morfo	-0,1042668125
s2 (t4-t3)	t2	-0,01276526777
s2 (t4-t3)	t3	-0,2185074556
s2 (t4-t3)	t4	0,1632097301
s2 (t4-t3)	t5	-0,09073945813
s2 (t4-t3)	t8	0,1004338265
s2 (t4-t3)	tSC	0,1161357851
s2 (t4-t3)	tSB	0,08385774575
s2 (t4-t3)	tB	0,1089522412
s2 (t4-t3)	cc2 (t3-t2)	-0,2080620185
s2 (t4-t3)	cc3 (t5-t3)	0,08134278338
s2 (t4-t3)	t5-t2	-0,02962358653
s2 (t4-t3)	s3 (t8-t5)	0,2486711485
s2 (t4-t3)	tSC-t8	0,00436380767
s2 (t4-t3)	tB-tSB	0,06942538271
s2 (t4-t3)	Ploidia	0,1109310362

Variable 1	Variable 2	Spearman Coefficient
s3 (t8-t5)	Idade	-0,2846417801
s3 (t8-t5)	Estágio	-0,0187086839
s3 (t8-t5)	Morfo	0,1246602606
s3 (t8-t5)	t2	0,1051061521
s3 (t8-t5)	t3	-0,07159964297
s3 (t8-t5)	t4	0,08671678117
s3 (t8-t5)	t5	-0,3933930993
s3 (t8-t5)	t8	0,4848122012
s3 (t8-t5)	tSC	0,00512150653
s3 (t8-t5)	tSB	0,009660494872
s3 (t8-t5)	tB	-0,009551800171
s3 (t8-t5)	cc2 (t3-t2)	-0,1674262721
s3 (t8-t5)	cc3 (t5-t3)	-0,3592297126
s3 (t8-t5)	t5-t2	-0,4122958851
s3 (t8-t5)	s2 (t4-t3)	0,2486711485
s3 (t8-t5)	tSC-t8	-0,3488140471
s3 (t8-t5)	tB-tSB	-0,1132629012
s3 (t8-t5)	Ploidia	0,2463033869

Variable 1	Variable 2	Spearman Coefficient
tSC-t8	Idade	0,1343924326
tSC-t8	Estágio	0,3996493404
tSC-t8	Morfo	-0,02474783976
tSC-t8	t2	-0,005307220087
tSC-t8	t3	0,01735718026
tSC-t8	t4	-0,03118791693
tSC-t8	t5	0,02937478582
tSC-t8	t8	-0,2948597241
tSC-t8	tSC	0,7021502392
tSC-t8	tSB	0,4420734695
tSC-t8	tB	0,4719478413
tSC-t8	cc2 (t3-t2)	0,08164410649
tSC-t8	cc3 (t5-t3)	-0,01041485259
tSC-t8	t5-t2	0,04660592616
tSC-t8	s2 (t4-t3)	0,00436380767
tSC-t8	s3 (t8-t5)	-0,3488140471
tSC-t8	tB-tSB	0,2020627538
tSC-t8	Ploidia	0,03967389575

Variable 1	Variable 2	Spearman Coefficient
tB-tSB	Idade	0,2047747578
tB-tSB	Estágio	0,3758105227
tB-tSB	Morfo	0,1370397427
tB-tSB	t2	0,0581978328
tB-tSB	t3	-0,07616061666
tB-tSB	t4	0,008409810914
tB-tSB	t5	-0,03673564922
tB-tSB	t8	-0,03389138651
tB-tSB	tSC	0,09964296621
tB-tSB	tSB	0,05277594867
tB-tSB	tB	0,3755551451
tB-tSB	cc2 (t3-t2)	-0,1271753196
tB-tSB	cc3 (t5-t3)	0,006397378179
tB-tSB	t5-t2	-0,04161200952
tB-tSB	s2 (t4-t3)	0,06942538271
tB-tSB	s3 (t8-t5)	-0,1132629012
tB-tSB	tSC-t8	0,2020627538
tB-tSB	Ploidia	-0,2849477648

Variable 1	Variable 2	Spearman Coefficient
Ploidia	Idade	-0,49663356
Ploidia	Estágio	-0,2375978079
Ploidia	Morfo	0,0452976895
Ploidia	t2	-0,07523975181
Ploidia	t3	-0,05565195514
Ploidia	t4	-0,01545912888
Ploidia	t5	-0,237550108
Ploidia	t8	-0,06801276287
Ploidia	tSC	-0,043280378
Ploidia	tSB	-0,1107753494
Ploidia	tB	-0,1695148138
Ploidia	cc2 (t3-t2)	-0,03144417624
Ploidia	cc3 (t5-t3)	-0,2819405223
Ploidia	t5-t2	-0,2277558204
Ploidia	s2 (t4-t3)	0,1109310362
Ploidia	s3 (t8-t5)	0,2463033869
Ploidia	tSC-t8	0,03967389575
Ploidia	tB-tSB	-0,2849477648