

FELIPE DELESTRO MATOS

**SOFTWARE BASEADO EM REDE NEURAL ARTIFICIAL
DESENVOLVIDA POR MEIO DE ALGORITMO GENÉTICO PARA A
CLASSIFICAÇÃO MORFOLÓGICA DE BLASTOCISTOS BOVINOS**

**ASSIS
2014**

FELIPE DELESTRO MATOS

**SOFTWARE BASEADO EM REDE NEURAL ARTIFICIAL
DESENVOLVIDA POR MEIO DE ALGORITMO GENÉTICO PARA A
CLASSIFICAÇÃO MORFOLÓGICA DE BLASTOCISTOS BOVINOS**

Dissertação apresentada à Faculdade de Ciências e Letras de Assis – UNESP – Universidade Estadual Paulista para a obtenção do título de Mestre em Biociências (Área de conhecimento: Caracterização e Aplicação da Diversidade Biológica).

Orientador: Marcelo Fábio Gouveia Nogueira

Co-Orientador: José Celso Rocha

ASSIS
2014

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca do Instituto Educacional de Assis – I E D A

Matos, Felipe Delestro
M433s Software baseado em rede neural artificial desenvolvido por
meio de algoritmo genético para a classificação morfológica de
blastocistos bovinos. / Felipe Delestro Matos. Assis, 2014
128 f. : il.

Dissertação de Mestrado – Faculdade de Ciências e
Letras de Assis – Universidade Estadual Paulista.
Orientador: Dr. Marcelo Fábio Gouveia Nogueira

1. Embriologia. 2. Blastocisto. I. Título.

CDD 636.20824

AGRADECIMENTO

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio à pesquisa (auxílio e bolsa), bem como os recursos computacionais, instalações, equipamentos e assistência providos pela Faculdade de Ciências e Letras UNESP de Assis. Também agradecemos à equipe da *In Vitro Brasil*, empresa parceira no fornecimento de imagens para a classificação embrionária.

SUMÁRIO

1	Introdução geral	9
1.1	Introdução	10
1.2	Classificação embrionária	13
1.3	Processamento de imagens	16
1.3.1	Transformada de Hough.....	16
1.3.2	Análises de textura	18
1.3.3	Transformada watershed.....	20
1.4	Redes neurais artificiais	23
1.4.1	Funções de transferência	24
1.4.2	Funções de treinamento.....	26
1.5	Algoritmos genéticos	26
1.6	Análises estatísticas	27
1.6.1	Análise de colinearidade	27
1.6.2	Receiver operator characteristic (ROC).....	28
1.6.3	Matriz de confusão	29
2	Metodologia.....	30
2.1	Desenvolvimento do banco de dados	30
2.2	Padronização das imagens	32
2.3	Segmentação dos embriões.....	36
2.4	Extração de informações.....	40
2.4.1	Descrição das variáveis.....	42
2.5	Análise de colinearidade	48
2.6	Algoritmos genéticos e arquitetura	51
3	Artigo.....	55
3.1	Resultados	60
3.1.1	Desenvolvimento do banco de dados	60
3.1.2	Padronização das imagens	61
3.1.3	Segmentação dos embriões.....	64
3.1.4	Extração de informações.....	66
3.1.5	Descrição das variáveis.....	66
3.1.6	Análise de colinearidade	73
3.1.7	Algoritmos genéticos e arquitetura	75
3.1.8	Classificação embrionária	77
3.2	Discussão.....	79
3.3	Métodos.....	82
3.4	Agradecimentos	83
3.5	Contribuição dos Autores	83
3.6	Referências bibliográficas	84
4	Anexos	88
4.1	Normas de publicação – Scientific Reports.....	88
4.2	Depósitos de patentes.....	94

4.3 Logo proposto para o software.....	96
4.4 Algoritmos desenvolvidos.....	97
4.5 Amostra do banco de dados.....	119
5 Discussão geral.....	120
5.1 Perspectivas Futuras.....	123
6 Referências Bibliográficas.....	125

MATOS, Felipe Delestro. **Software baseado em rede neural artificial desenvolvida por meio de algoritmo genético para a classificação morfológica de blastocistos bovinos.** 2014. 128 f. Dissertação (Mestrado em Biociências). – Faculdade de Ciências e Letras, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Assis, 2014.

RESUMO

A classificação morfológica embrionária possui grande importância para inúmeras técnicas laboratoriais (desde pesquisas básicas às aplicadas na reprodução assistida). Entretanto, o método utilizado para realizar a classificação dos embriões em diferentes graus de qualidade sempre foi baseado na subjetividade do avaliador e, por mais que sejam estabelecidos padrões de graus de qualidade e descrições das características morfológicas que categorizam um embrião em cada grau, não há atualmente um método preciso que possa gerar resultados consistentes e confiáveis. Assim, nosso trabalho resultou no desenvolvimento de um software capaz de realizar a classificação da qualidade morfológica de blastocistos bovinos. Utilizamos como base de funcionamento técnicas de inteligência artificial (mais especificamente de Redes Neurais Artificiais e Algoritmos Genéticos). Resultados indicam uma taxa de acerto global de **79,2%** na classificação de blastocistos bovinos em 3 graus de qualidade, sendo que para os blastocistos classificados como Excelentes ou Bons (Classe 1) a taxa de acerto é de **82,6%**, para os blastocistos classificados como Regulares (Classe 2) é de **16,7%** e para os blastocistos classificados como Pobres (Classe 3) a taxa de acerto é de **91,7%**.

Palavras-chaves: Embriologia, Blastocisto, Qualidade, Classificação, Software.

MATOS, Felipe Delestro. **Software based on artificial neural network developed through genetic algorithm for morphological classification of cattle blastocysts.** 2014. 128 f. Dissertação (Master in Biosciences). – Faculdade de Ciências e Letras, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Assis, 2014.

ABSTRACT

Embryonic morphological classification has great importance for numerous laboratory techniques (from basic to applied research in assisted reproduction). However, the method used to perform the classification of embryos in varying degrees of quality has always been based on the subjectivity of the evaluator. Although quality standards and descriptions of morphological characteristics that categorize an embryo in each grade are established, currently there is not an accurate method that can generate consistent and reliable results. Thus, our work resulted in the development of a software able to perform the classification of morphological quality of bovine blastocysts. Artificial Intelligence techniques (such as Artificial Neural Networks and Genetic Algorithms) were used in the development. Results indicate an overall accuracy of **79.2%** in the classification of bovine blastocysts in 3 degrees of quality. For blastocysts classified as Excellent or Good (Class 1) the hit rate is **82.6%**, for blastocysts classified as Regular (Class 2) is **16.7%** and for blastocysts classified as poor (Class 3) the hit rate is **91.7%**.

Keywords: Embryology, Blastocyst, Quality, Assessment, Software.

1 INTRODUÇÃO GERAL

Esta dissertação para o programa de Mestrado em Biociências – Caracterização e Aplicação da Diversidade Biológica – se encontra dividida em quatro sessões principais. A primeira, denominada **Introdução Geral**, aborda todas as técnicas e metodologias desenvolvidas e utilizadas durante o trabalho de maneira mais profunda e detalhada. A parte subsequente, **Artigo** corresponde a uma proposta de publicação resultante do trabalho efetuado. A sessão **Anexos** compreende documentos relacionados (como normas de publicação do periódico e depósitos de patente) bem como todos os algoritmos desenvolvidos na metodologia. Finalizando, a sessão **Discussão** apresenta uma análise crítica dos resultados atingidos, bem como sua comparação com outros trabalhos semelhantes.

1.1 INTRODUÇÃO

Desde o desenvolvimento das primeiras técnicas bem sucedidas de superovulação e transferência de embrião (TE) e de produção *in vitro* de embriões (PIVE) em mamíferos, ficou evidente a relação direta entre a qualidade embrionária e a taxa de gestação após a transferência de embriões em fêmeas receptoras. Desse modo, embriões morfológicamente classificados como de alta qualidade produzem maiores taxas de gestação (SCHNEIDER; CASTLEBERRY; GRIFFIN, 1980; TERVIT et al., 1980). Disso decorre a necessidade por um sistema que seja capaz de padronizar os elementos que categorizam os embriões em diferentes graus de acordo com sua qualidade e, portanto, também indiretamente segundo sua viabilidade.

Atualmente, para a espécie bovina, é tido como padrão o sistema classificatório em três graus de qualidade: *bom*, *regular* ou *pobre* (BÓ; MAPLETOFT, 2013), recomendado pela Sociedade Internacional de Transferência de Embriões (IETS). Maiores detalhes sobre o sistema de classificação podem ser encontrados na sessão *Classificação Embrionária*.

O sistema é baseado na análise visual (subjetiva e qualitativa) da morfologia do embrião que é comumente realizada por meio de microscopia óptica (estereomicroscópio). A técnica depende da experiência e acurácia do embriologista em analisar e categorizar desde as variáveis mais evidentes até as nuances que tornam um embrião mais ou menos apto ao desenvolvimento. Nessa análise clássica da morfologia embrionária, as variáveis não são objetivamente mensuradas, de forma que ela é subjetiva e pouco repetível (BÉNYEI et al., 2006). Assim, um mesmo embrião que seja

mensurado por diferentes especialistas pode ser classificado em distintos graus de qualidade. Tal discordância ocorre principalmente entre graus próximos (FARIN et al., 1995).

Buscando soluções para a questão da subjetividade da análise morfológica embrionária, diversos métodos alternativos estão sendo desenvolvidos. (HELD et al., 2012; HOSHI, 2003; LÓPEZ-DAMIÁN E. P, GALINA C. S, MERCHANT H, CEDILLO-PELÁEZ C, 2008; OVERSTRÖM, 1996) Dentre eles podem ser destacados o cultivo *in vitro* de embriões (OVERSTRÖM, 1996), a integridade da membrana do blastômero (OVERSTRÖM, 1996), a análise do metabolismo embrionário, (OVERSTRÖM, 1996) a mensuração da respiração celular (HOSHI, 2003), a análise mediante microscopia eletrônica (LÓPEZ-DAMIÁN E. P, GALINA C. S, MERCHANT H, CEDILLO-PELÁEZ C, 2008) o uso de índices de birrefringência da zona pelúcida. (HELD et al., 2012) e, inclusive, testes iniciais no desenvolvimento de softwares para análise semiautomática de embriões (SANTOS FILHO et al., 2012). Entretanto, nenhum método até o momento apresenta uma solução definitiva para a mensuração da qualidade do embrião, sendo estes apresentados apenas experimentais, de modo que ainda é necessário o desenvolvimento de técnicas que sejam rápidas, não invasivas e objetivas (LINDNER; JR, 1983; OVERSTRÖM, 1996). Um fator adicional é o alto custo de alguns métodos, o que impede a sua ampla utilização. Desse modo, apesar da subjetividade, a análise visual da morfologia continua sendo amplamente utilizada para a determinação da qualidade embrionária.

Com este trabalho tivemos o objetivo de testar um método que obtivesse grande precisão na análise morfológica, mediante um processo de extração automática de

informações a partir de imagens digitais bidimensionais dos embriões e sua posterior classificação em graus de qualidade, de acordo com a especificidade de cada espécie. Foram utilizados apenas embriões em estágio de blastocisto (desde blastocisto inicial a blastocisto expandido), uma vez que este grau de expansão apresenta os melhores resultados em procedimentos de transferência embrionária e atualmente é o mais utilizado pelos laboratórios de produção *in vitro* (BALABAN et al., 2000).

O processo ocorreu em duas etapas principais; a primeira sendo a extração automática das variáveis via técnicas de processamento de imagens, e a segunda sendo a etapa classificatória, na qual os dados são interpretados por um sistema de Redes Neurais Artificiais (RNA).

1.2 CLASSIFICAÇÃO EMBRIONÁRIA

Diversos parâmetros são utilizados na determinação da qualidade, como o formato do embrião, cor, quantidade e compactação das células, tamanho do espaço perivitelino, quantidade de células que sofreram extrusão ou degeneração, número de vesículas, homogeneidade no tamanho dos blastômeros, granulação citoplasmática e fragmentação dos blastômeros (BÓ; MAPLETOFT, 2013; LINDNER; JR, 1983; OVERSTRÖM, 1996). Entretanto, a análise continua sendo subjetiva e dependente da experiência do avaliador. A análise também está sujeita a variações, ou seja, diferentes avaliadores podem inferir graus diferentes de qualidade a um mesmo embrião (FARIN et al., 1995).

De modo geral, os blastocistos de embriões bovinos eram classificados em quatro distintos graus de qualidade: *Excelente*, *Bom*, *Regular* ou *Pobre*, havendo uma evidente correlação entre maiores taxas de sucesso em Fertilização *in vitro* (FIV) com o grau de qualidade dos blastocistos transferidos. (OVERSTRÖM, 1996)

Entretanto, a dificuldade de diferenciação entre graus de qualidade muito próximos (entre *Excelente* e *Bom*, por exemplo) sugere a utilização de um sistema classificatório com apenas três graus. Atualmente, o manual da Sociedade Internacional de Transferência de Embriões (IETS) sugere a classificação em um código que varia de 1 a 4 (BÓ; MAPLETOFT, 2013), sendo:

1) Excelente ou Bom: Os embriões possuem um aspecto simétrico e esférico, com blastômeros uniformes em tamanho, cor e densidade.

2) Regular: Embriões com irregularidades moderadas. Pelo menos 50% das células devem estar intactas.

3) Pobre: Possuem grandes irregularidades, mas pelo menos 25% das células devem estar intactas.

4) Morto ou degenerado: Não viáveis e devem ser descartados. Nas análises deste trabalho não foram utilizados embriões mortos.

A classificação do embrião de acordo com o seu estágio de desenvolvimento também ocorre. É utilizado um código numérico que varia de **1** (oócito não fertilizado ou embrião de uma célula) a **9** (Blastocisto eclodido ou expandido) (BÓ; MAPLETOFT, 2013). Neste trabalho foram utilizados apenas embriões no estágio de blastocisto, uma vez que este é o estágio mais indicado atualmente para a transferência embrionária, e o utilizado pela *In Vitro Brasil* (empresa que atuou como parceira no fornecimento das imagens).

A distinção entre os diferentes graus de qualidade de um blastocisto não é uma tarefa trivial, necessitando de grande experiência por parte do embriologista. A **Figura 1** demonstra casos nos quais fica mais evidente a diferença entre os graus de qualidade, sendo possível notar claras diferenças morfológicas entre os embriões, principalmente entre os Graus 1 e 3.



Figura 1 Blastocistos bovinos com diferentes graus de qualidade. Da esquerda para a direita, Excelente ou Bom (Grau 1), Regular (Grau 2) e Pobre (Grau 3).

Entretanto, exemplos como os da **Figura 2** demonstram como pode haver grande variação dentro de uma mesma classe. São apresentados 3 casos de embriões classificados como Grau 2, mas que demonstram grandes diferenças morfológicas entre si. Este exemplo demonstra a dificuldade de classificação em casos nos quais o blastocisto não se encontra claramente em um grau definido.



Figura 2 Exemplos de blastocistos Grau 2 (Regular) que, apesar de pertencerem a mesma classificação, possuem diferenças morfológicas expressivas.

Como não há uma metodologia que permita categorizar um embrião em seu devido grau de qualidade de maneira exata (BÓ; MAPLETOFT, 2013), a precisão da

avaliação morfológica fica toda a cargo da experiência do embriologista que realiza a análise.

1.3 PROCESSAMENTO DE IMAGENS

Uma vez que a base de dados é constituída por um conjunto de fotografias obtidas por meio de microscopia, técnicas de Processamento de Imagens Digitais são necessárias na interpretação das informações. O processamento de imagens digitais se constitui em um conjunto de técnicas e metodologias amplamente empregadas na resolução de diversos problemas relacionados desde ao melhoramento de fotografias até à visão computacional (GONZALEZ; WOODS, 2007).

Essa é uma etapa chave no desenvolvimento do software automatizado, uma vez que toda a interpretação das imagens deve ocorrer sem a necessidade de interferência do usuário. Descrições detalhadas de cada etapa podem ser encontradas na sessão de *Metodologia*, sendo que nos próximos tópicos abordarão os aspectos teóricos das principais técnicas utilizadas.

1.3.1 Transformada de Hough

A Transformada de Hough, através de sua adaptação denominada *Circle Hough Transform (CHT)* se tornou um dos métodos mais difundidos para detecção de círculos em processamento de imagens digitais (ATHERTON; KERBYSON, 1999). Basicamente, a transformada de Hough realiza um mapeamento entre o espaço da imagem e um espaço de parâmetros, utilizando como base uma versão binária da imagem original após ter sido submetida a um processo de detecção de bordas.

Originalmente a transformada era utilizada apenas na detecção de linhas retas, mas adaptações permitem a utilização de qualquer forma definida. O espaço de parâmetros é organizado na forma de uma matriz acumuladora, de forma que através de seus maiores valores é possível determinar os casos nos quais a forma (em nosso caso, o círculo) se encontra melhor apresentado na imagem original (BALLARD, 1981).

O funcionamento do processo pode ser compreendido através das etapas apresentadas na Figura 3. Digamos que se deseja detectar o melhor círculo representado pelos pontos vermelhos (Figura 3A), que possua um determinado círculo r . É possível traçar um círculo de raio r tendo como centro o ponto indicado (Figura 3B), sendo que os valores da circunferência são adicionados à matriz acumuladora. Após realizada a operação em todos os pontos, o valor máximo da matriz acumuladora indicará o centro do círculo a ser detectado (Figura 3C), que pode então ser estabelecido, por seu centro e raio são conhecidos (Figura 3D). Tal processo pode ser repetido para diferentes raios, e o que obtiver um maior valor da matriz acumuladora (maior score) indicará uma melhor detecção (ATHERTON; KERBYSON, 1999).

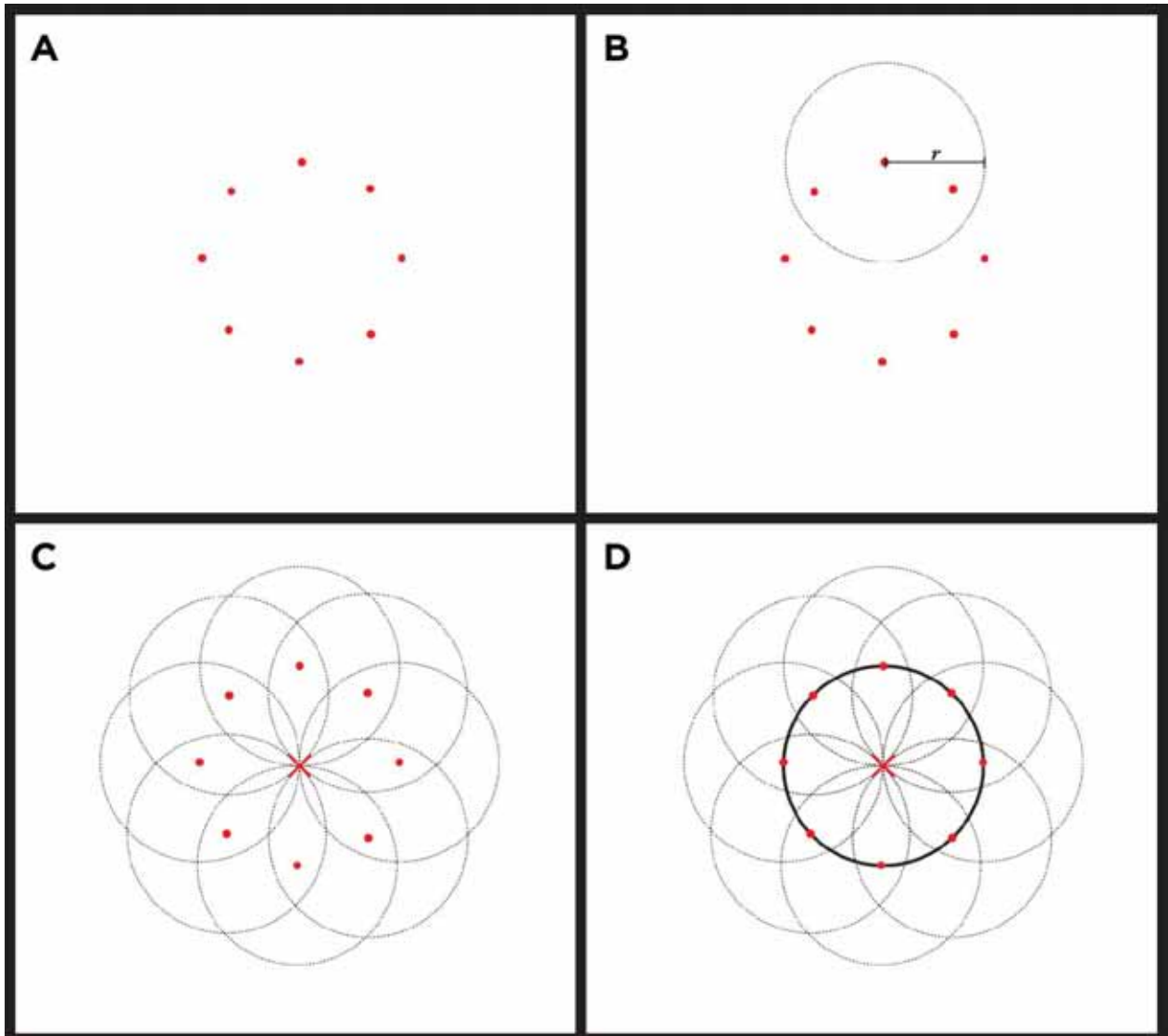


Figura 3 Etapas da criação do espaço de parâmetros na transformada circular de Hough. Pelo valor máximo obtido na matriz acumuladora é possível obter o centro para um círculo de raio r a ser detectado.

1.3.2 Análises de textura

A Textura é uma das mais importantes características usadas na identificação de objetos ou regiões de interesse em uma imagem (HARALICK; SHANMUGAM; DINSTEIN, 1973). De modo geral, uma textura pode ser definida como padrões aleatórios regulares que se repetem por uma região da imagem, de forma a prover

informações a respeito do arranjo estrutural de superfícies (ROBERTI DE SIQUEIRA; ROBSON SCHWARTZ; PEDRINI, 2013). Embora não haja um sistema único para classificação de texturas, a *Gray Level Co-occurrence Matrix* (GLCM) se demonstra como sendo um robusto método estatístico de classificação (HU; ZHAO; WANG, 2008).

A informação de textura é extraída com base na variação de intensidade entre dois pixels vizinhos, de modo a formar uma matriz de co-ocorrência, sendo que os descritores são obtidos tendo a matriz como base. O número de linhas e colunas da matriz dependerá do número de níveis de cinza da imagem, sendo uma prática comum o processo de determinação do limiar para a obtenção de uma quantidade reduzida de níveis.

Dentre as metodologias estatísticas para a análise de texturas em imagens, a utilização de GLCM se apresenta como um dos métodos mais eficientes (HU; ZHAO; WANG, 2008). A GLCM descreve a distribuição espacial dos valores de intensidade dos pixels levando-se em consideração uma determinada distância e ângulo, tornando possível o reconhecimento e classificação de texturas. Cada matriz é a probabilidade de dois pixels vizinhos (um com intensidade i e outro com intensidade j) a uma determinada distância d e ângulo θ , formando $P = (i, j, d, \theta)$. Para o cálculo da GLCM a intensidade da imagem é ajustada, de modo a possuir apenas 8 tons de cinza.

Na **Tabela 1** podemos analisar um exemplo de GLCM, resultante de uma das imagens do banco de dados. Neste caso, foi utilizada uma escala de 8 tons de cinza.

Tabela 1 *Gray Level Co-occurrence Matrix* para a imagem utilizada de exemplo.

	1	2	3	4	5	6	7	8
1	37449	1032	35	0	17	357	0	0
2	1003	5966	1541	74	2	0	0	0
3	64	1438	6766	2291	106	1	0	0
4	0	136	2080	10348	3076	96	3	0
5	84	14	231	2833	17050	3517	128	3
6	290	0	13	188	3375	19637	2233	167
7	0	0	0	5	226	2098	6114	1277
8	0	0	0	0	9	196	1242	5439

1.3.3 Transformada watershed

A segmentação de imagens é de grande importância para a interpretação independente de diferentes regiões. Duas principais abordagens podem ser utilizadas na segmentação: a detecção de bordas que delimitem uma região ou a busca de regiões que possuam similaridades em intensidade dos pixels. A metodologia de Watershed busca segmentar a imagem através da segunda estratégia (BEUCHER, 1992).

A transformada watershed propõe uma abordagem morfológica para o problema de segmentação de imagens, através de sua interpretação como sendo superfícies, na qual os níveis de cinza de cada pixel determina a altitude de uma dada região. A partir

desta noção, deseja-se então identificar bacias hidrográficas, definidas por mínimos regionais e suas regiões de domínio (KÖRBES, 2010). A inundação dessas regiões, associada à elevação de barreiras nas regiões de encontro das águas permite uma segmentação das regiões, conforme pode ser verificado na Figura 4.

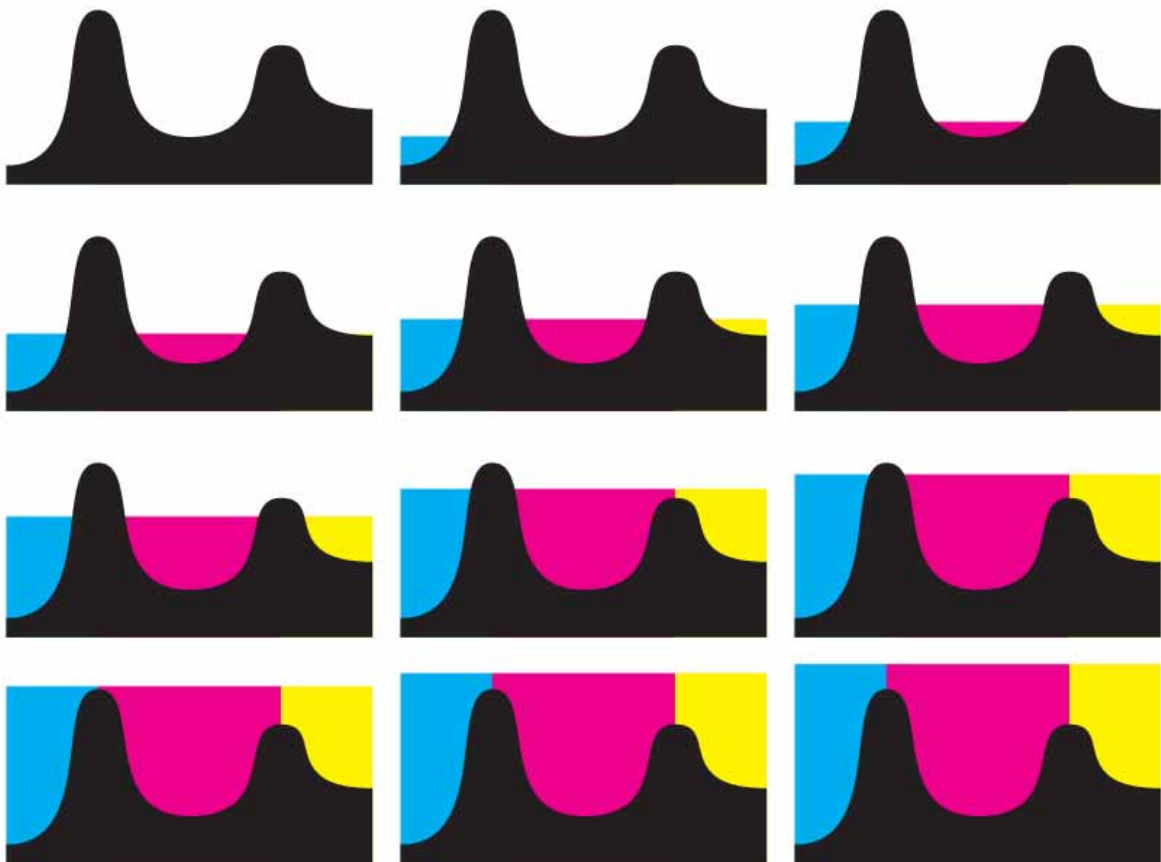


Figura 4 Etapas de um algoritmo watershed, realizando segmentação através de inundação dos vales.

Na Figura 5 podemos visualizar um exemplo de segmentação de um blastocisto por meio da transformada watershed. Na esquerda uma versão do embrião após as etapas de padronização e isolamento (maiores detalhes na sessão 2 Metodologia, página 30) e na direita a segmentação resultante da transformada Watershed. Apenas para efeito de visualização, cada região recebeu uma cor aleatória.

A Massa Celular Interna (MCI), por ser uma região escura relativamente homogênea, apresenta-se aproximadamente como a região de maior área após a segmentação. A maior área segmentada será, portanto, utilizada como máscara da MCI nas etapas de extração de variáveis (2.4 Extração de informações, página 40).

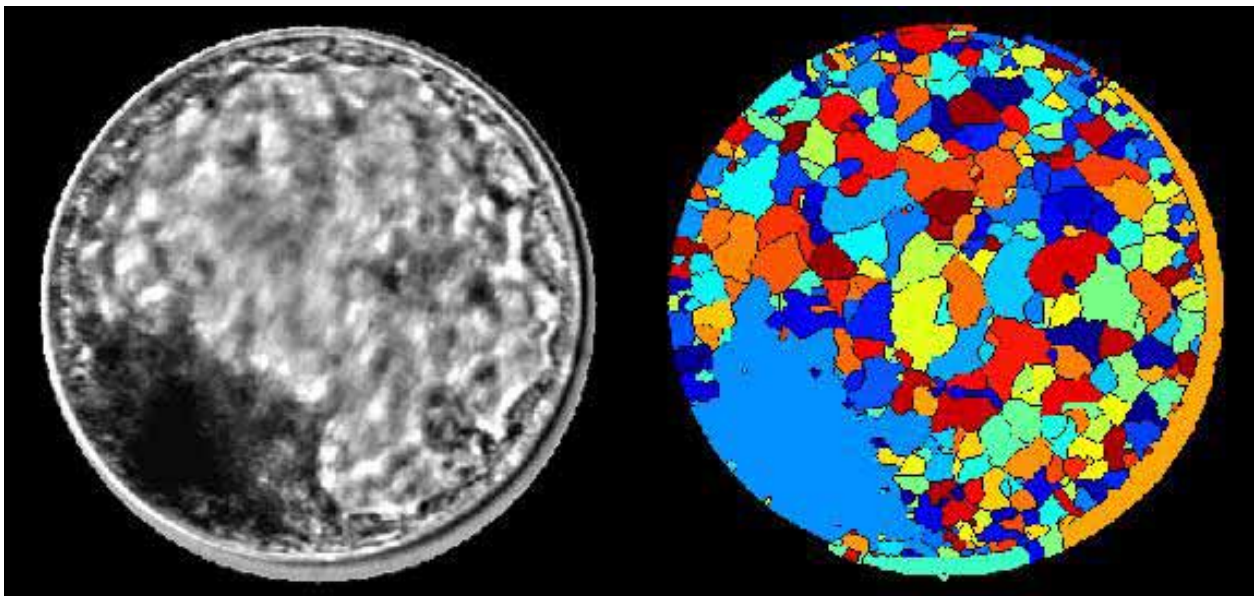


Figura 5 Exemplo de Blastocisto (esquerda) e sua segmentação via transformada watershed (direita).

1.4 REDES NEURAIS ARTIFICIAIS

O sistema classificatório foi desenvolvido utilizando-se a metodologia de Redes Neurais Artificiais (RNA), uma técnica de inteligência artificial indicada para a resolução de problemas não lineares e com variáveis interconectadas. (GOETHALS et al., 2007; LI, 1994; ZHANG; PATUWO; HU, 1998) Seu uso é recomendado em tarefas nas quais métodos computacionais tradicionais são limitados, ou seja, quando é exigido um grande poder de adaptação e generalização na classificação dos dados (JAIN; MAO; MOHIUDDIN, 1996).

As RNA vêm sendo aplicadas em uma diversidade de áreas, desde auxílios administrativos (ROCHA et al., 2011) à predição do índice da bolsa de valores. (GURESEN; KAYAKUTLU; DAIM, 2011) Basicamente, uma RNA é um sistema que atua na resolução de problemas simulando o funcionamento de um conjunto de neurônios biológicos. Tais neurônios da RNA (também chamados de *perceptrons*) precisam ser expostos aos dados de treinamento, de forma a aprenderem a generalizar uma saída a partir de um conjunto de dados de entrada. Uma vez devidamente treinada, a RNA é capaz de realizar predições sem os dados de saída (HAYKIN, 1998; ZHANG; PATUWO; HU, 1998). Portanto, a RNA é um sistema inteligente capaz de elucidar um problema complexo a partir de um aprendizado assistido.

Entretanto, existem alguns entraves no desenvolvimento de classificadores baseados em RNA, principalmente quanto ao desenvolvimento da melhor arquitetura de rede. (WALCZAK; CERPA, 1999) Para solucionar este problema, foi desenvolvida uma

metodologia baseada em Algoritmos Genéticos (AG) para a determinação da melhor arquitetura. Um exemplo de arquitetura pode ser encontrado na Figura 6.

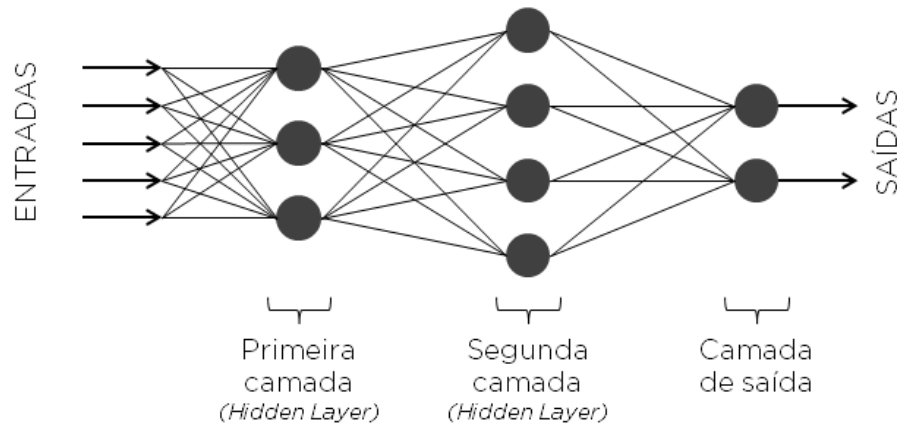


Figura 6 Exemplo de arquitetura de uma RNA, com 3 neurônios na primeira camada, 4 na segunda e 2 na camada de saída.

Redes Neurais Artificiais também vem sendo amplamente utilizadas na resolução de problemas ligados ao processamento de imagens (EGMONT-PETERSEN; DE RIDDER; HANDELS, 2002). Dessa forma, a técnica se demonstra adequada para a classificação morfológica de blastocistos a partir de imagens bidimensionais.

1.4.1 Funções de transferência

Antes de transmitir a informação processada pelos pesos à saída do neurônio, o valor final da somatória é aplicado em uma função de transferência. Diversas funções podem ser utilizadas, sendo que as escolhidas para este trabalho podem ser analisadas na **Figura 7**.

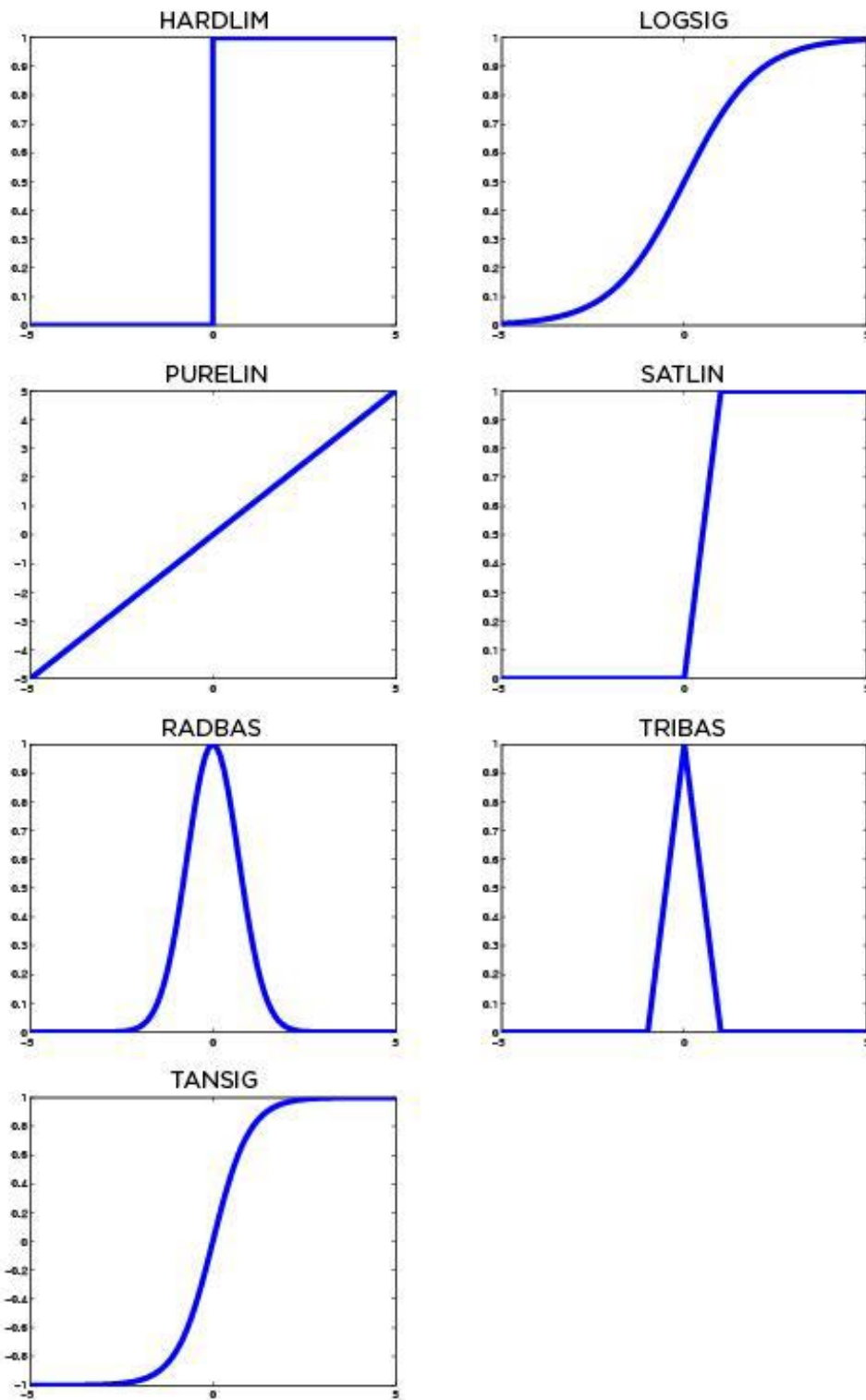


Figura 7 Funções de transferência utilizadas no desenvolvimento de arquitetura das redes neurais artificiais.

1.4.2 Funções de treinamento

O treinamento da RNA, ou seja, a adaptação dos pesos para que haja a convergência para a situação de menor erro possível, é realizado por funções de treinamento pré-definidas. Neste trabalho foram utilizadas as funções “Gradient descent with momentum and adaptive learning rate backpropagation” (**traingdx**), “Scaled conjugate gradient backpropagation” (**trainscg**) e “Gradient descent with momentum backpropagation” (**traingdm**) (BENVENUTO; PIAZZA, 1992; HECHT-NIELSEN, 1989).

1.5 ALGORITMOS GENÉTICOS

Os Algoritmos Genéticos (AG) constituem um conjunto de técnicas computacionais capazes de resolver problemas combinatórios por meio de uma estratégia evolutiva (DE, 1988; MITCHELL, 1995). Simulando indivíduos inicialmente aleatórios, que contêm uma solução para um determinado problema, ocorre um processo de cruzamento e mutação de forma que através de uma pressão de seleção são capazes de convergir para a melhor solução do problema (ou o conjunto de melhores soluções).

O correto desenvolvimento da arquitetura de uma Rede Neural Artificial é peça chave na resolução eficiente de problemas (WALCZAK; CERPA, 1999). Caso não seja corretamente determinada, a RNA pode não encontrar a melhor solução possível para o conjunto de dados apresentado, podendo ocorrer classificações errôneas ou problemas de *overfitting* quando exposta aos dados de teste.

Diversos fatores são responsáveis pela arquitetura, como a **quantidade de camadas ocultas**, o **número de neurônios em cada camada**, as **funções de transferência** entre os neurônios e a **função de treinamento** da RNA (RAMLALL, 2009). Entretanto, não há um método que possa determinar precisamente quais as melhores condições de arquitetura para a resolução de cada problema, existindo apenas algumas sugestões heurísticas que não necessariamente determinam sempre a melhor opção (WALCZAK; CERPA, 1999).

Buscando soluções para este dilema, desenvolvemos um sistema baseado em Algoritmos Genéticos que busca a melhor configuração para a resolução do problema de arquitetura. Alguns trabalhos já exploram uma abordagem semelhante (YAO; LIU, 1998), apresentando resultados interessantes e demonstrando que algoritmos evolutivos atuam com eficiência na busca de uma melhor arquitetura.

1.6 ANÁLISES ESTATÍSTICAS

Algumas técnicas específicas são necessárias para a correta interpretação dos dados obtidos durante as etapas de extração de informações e classificação do banco de dados. Aspectos teóricos de cada técnica serão demonstrados a seguir.

1.6.1 Análise de colinearidade

Realizar uma análise de colinearidade nas variáveis utilizadas como entrada em um sistema de Redes Neurais Artificiais é uma prática comum e recomendada (WALCZAK; CERPA, 1999). Através desta análise é possível eliminar variáveis muito

correlacionadas entre si, fato que poderia ocasionar uma predisposição indesejada em favorecer certos fatores.

Através da análise de colinearidade é possível determinar o *Variance Inflation Fator* (VIF), que representa o grau de multi-colinearidade da variável independente em questão quando comparada com as outras variáveis independentes analisadas. Como regra, podem ser consideradas colineares variáveis que apresentem valores de VIF maiores do que 10 (O'BRIEN, 2007). O valor de VIF pode ser calculado de acordo com a Equação (1), na qual R_i é o coeficiente de determinação de uma regressão de uma variável i .

$$VIF = \frac{1}{1-R_i^2} \quad (1)$$

1.6.2 Receiver operator characteristic (ROC)

Uma das maneiras de se analisar os resultados de uma RNA para reconhecimento de padrões é através de uma curva de Característica de Operação do Receptor (COR), mais conhecida como *Receiver Operator Characteristic* (ROC). Para que seja obtida, dois valores são necessários: 1) A relação de verdadeiros positivos do teste com o total de positivos (TVP = Taxa de Verdadeiros Positivos) também denominada *Sensibilidade* e 2) A relação de falsos positivos com o total de negativos (TFP = Taxa de Falsos Positivos) também denominada *Especificidade*. Para construir a curva ROC deve-se traçar um diagrama que represente valores de Sensibilidade em

função da proporção de falsos positivos dados por $1 - \text{Especificidade}$ (falsos positivos) para um conjunto diverso de pontos de corte (BROWN; DAVIS, 2006).

1.6.3 Matriz de confusão

Outra metodologia utilizada para a análise de resultado se sistemas multiclassés é a chamada Matriz de Confusão (ou *Confusion Matrix*). Através dela é possível observar o cruzamento entre os dados fornecidos pelo classificador (sistema de RNA) e o gabarito original de classificação. Dessa forma, obtêm-se a acurácia geral do sistema, bem como a acurácia para cada classe (MACHART; RALAIVOLA, 2012).

Através da **Figura 8** é possível observar como seria o funcionamento da matriz em um exemplo binário didático, no qual 10 amostras devem ser classificadas, sendo 7 delas são segundo o gabarito classe 1 e as 3 restantes são classe 2. O sistema

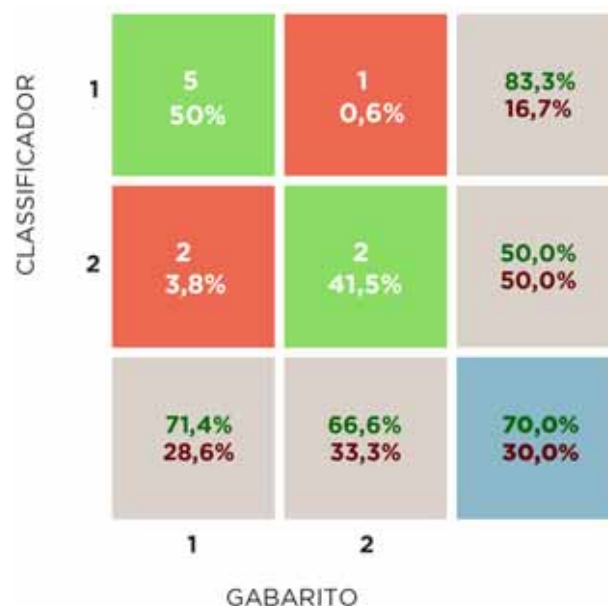


Figura 8 Exemplo de funcionamento de uma Matriz de Confusão.

classificador registrou 5 delas corretamente como classe 1, e 2 amostras foram corretamente classificadas como classe 2. Portanto, a diagonal da matriz de confusão demonstra os acertos do classificador.

2 METODOLOGIA

Nesta sessão serão detalhadas as etapas metodológicas desenvolvidas durante o projeto. Na sessão **Artigo** estas metodologias serão novamente apresentadas, embora de maneira mais resumida.

2.1 DESENVOLVIMENTO DO BANCO DE DADOS

Para que seja desenvolvido o software de análise morfológica, inicialmente foi necessária a obtenção de um banco de dados de imagens de blastocistos bovinos, obtidos em parceria com a *In Vitro Brasil*. Estas imagens serviram como base para o desenvolvimento dos algoritmos de interpretação, bem como o treinamento, validação e teste da Rede Neural Artificial.

As imagens foram capturadas por meio de microscópio invertido Olympus IX71, com magnificação de 32x e captura de imagem pelo software *LuCam Capture v6.30*. As imagens foram armazenadas em formato JPG de 8 bits (RGB), em resolução 1280x1024 pixels. Cada imagem possui apenas um blastocisto, aproximadamente centralizado. Para cada blastocisto foram capturadas duas imagens, a primeira com sua Massa Celular Interna (MCI) na parte inferior e uma segunda com a MCI lateral, conforme demonstrado na Figura 10.

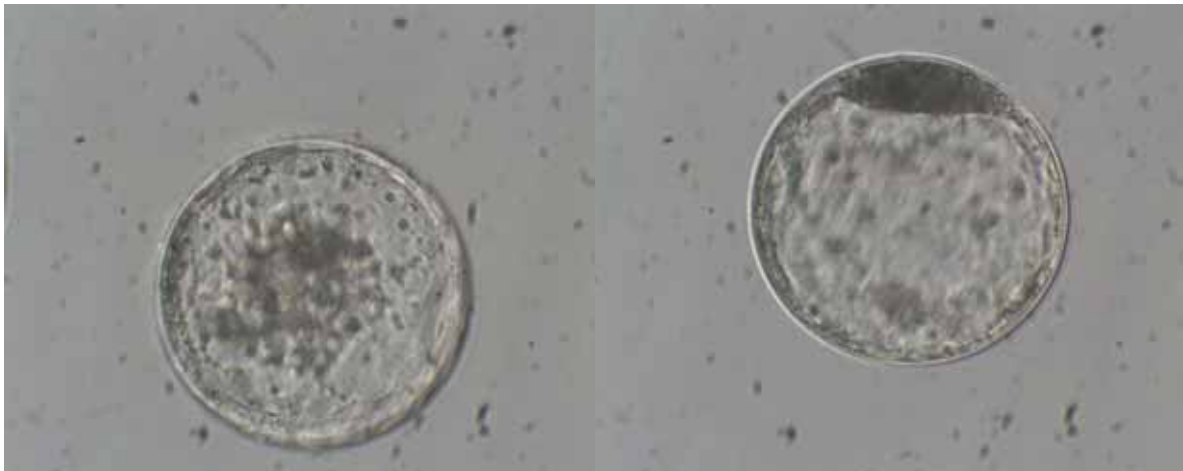


Figura 10 Imagens de um mesmo blastocisto bovino, com MCI na parte inferior (esquerda) e lateral (direita).

Os embriologistas responsáveis pela rotina laboratorial da *In Vitro* classificaram as imagens em três graus de qualidade, segundo os padrões da *International Embryo Transfer Society* (IETS). No total foram obtidas **744** imagens, correspondendo a **372** blastocistos distintos (uma vez que cada um possui duas imagens). A classificação dos blastocistos se distribui de acordo com a **Figura 9**.

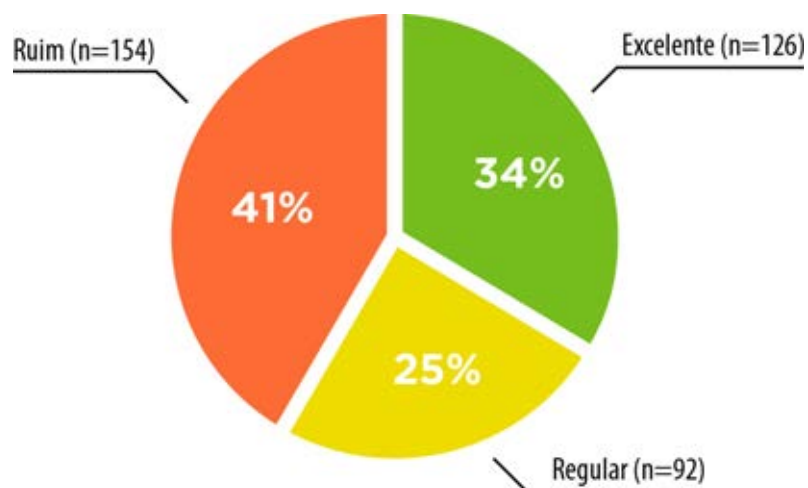


Figura 9 Distribuição do banco de dados de blastocistos bovinos nos três graus de qualidade definidos pela IETS, Excelente ou Bom (Grau 1), Regular (Grau 2) e Ruim (Grau 3).

Na sessão **Anexos** é possível encontrar uma amostra do banco de dados. Todas as imagens foram nomeadas seguindo o padrão: “*ID_Posição_Qualidade.jpg*”. “*ID*” corresponde a um número de no máximo três algarismos (de 001 até 999), em “*Posição*” adotou-se “a” para vista superior (esquerda na Figura 10) e “b” para vista lateral (direita na Figura 10). “*Qualidade*” é definida como 1 para Excelente ou Bom, 2 para Regular e 3 para Ruim. No exemplo da Figura 10, os embriões são “*152_a_1.jpg*” e “*152_b_1.jpg*”, respectivamente.

2.2 PADRONIZAÇÃO DAS IMAGENS

Cada laboratório possui um padrão diferente para a captura de imagens, de forma que se torna indispensável uma padronização das mesmas. Além das diferenças de resolução e formato, lotes diferentes de imagens (o banco de dados foi desenvolvido mediante envios parciais) possuíam características distintas de iluminação, conforme demonstrado na Figura 11. Todos esses fatores devem ser levados em consideração ao se importar uma imagem para o algoritmo, para que as futuras etapas de segmentação e extração de informações encontrem condições similares, independentemente das características iniciais da imagem. Todos os algoritmos foram desenvolvidos em ambiente Matlab (GONZALEZ, 2004) e permitem uma análise automatizada das imagens, sem a necessidade de intervenção do usuário.



Figura 11 Embriões cujas imagens foram capturadas em momentos distintos. Notar a discrepância entre a iluminação das duas imagens o que faz necessária uma padronização via processamento de imagens.

O algoritmo de padronização segue os seguintes passos:

1) Importação da Imagem: A imagem é importada no ambiente Matlab, independentemente do formato (JPG ou BMP).

2) Conversão em Escala de Cinza: Como as informações de cor não serão utilizadas na interpretação das imagens, todas são convertidas em escala de cinza 8-bit. Tal processo também torna todas as etapas subsequentes mais rápidas, uma vez que ocorre a diminuição da dimensão espectral da imagem (das três bandas RGB para apenas uma).

3) Ajuste de resolução e proporção: Além de resoluções diferentes, as imagens também possuem proporções (*ratios*) diferentes. Uma imagem de 640x480 pixels possui uma proporção de 0,75 enquanto que em uma imagem de 1280x1024 pixels a proporção é de 0,8. Portanto, apenas redimensionar todas as imagens para uma das resoluções iria distorcer as proporções das outras. Foi escolhida como padrão a

resolução 640x480 pixels, por ser o padrão inferior e apresentar informações suficientes para a interpretação. Utilizar a resolução menor também aumenta a eficiência das etapas subsequentes do algoritmo, além de não criar limitações de resolução mínima para o software final. Portanto, as imagens de resolução 1280x1024 foram ajustadas para a proporção de 0,75 (mediante corte de uma região superior de 64 pixels) e em seguida redimensionadas para 640x480 pixels. As proporções podem ser visualizadas na **Figura 12**.

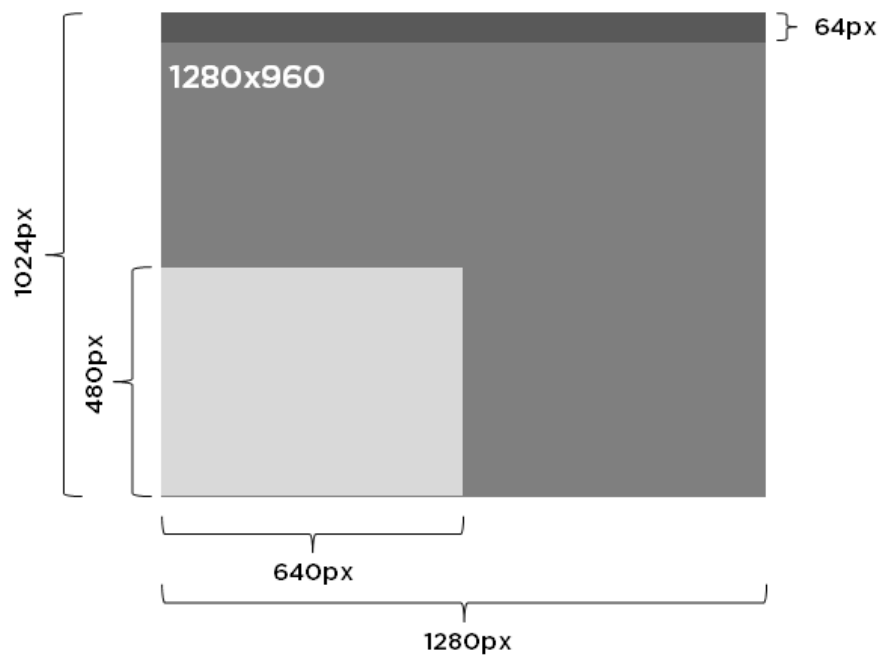


Figura 12 Demonstração das escalas e padrões do algoritmo de ajuste de resolução e proporção.

4) Ajuste de Intensidade: Para minimizar os efeitos das diferentes iluminações, as imagens são submetidas a um ajuste de histograma, de forma que 1% de toda a informação fique saturada entre pixels claros e escuros. Tal processo também atua aumentando o contraste da imagem, o que facilita a etapa subsequente de segmentação (Figura 13).

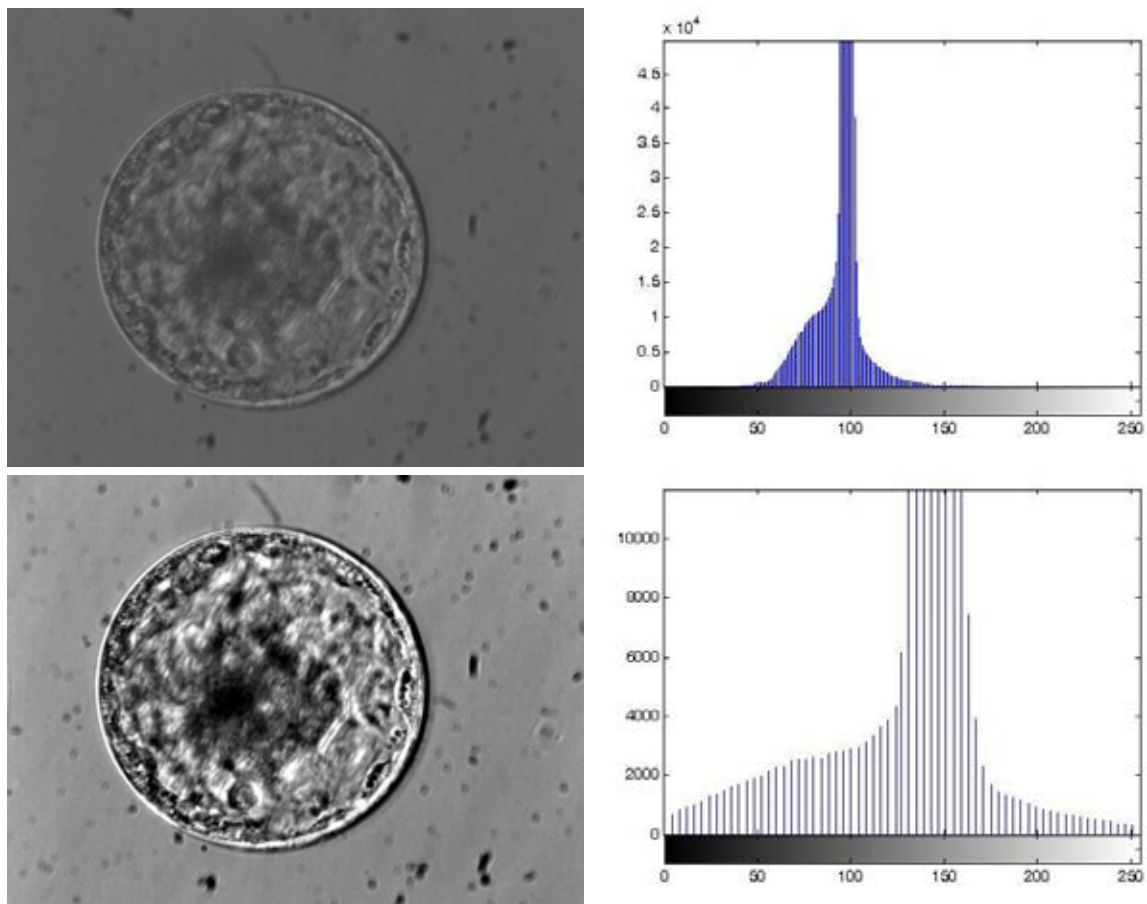


Figura 13 Imagens de um embrião bovino antes (acima) e após (abaixo) a padronização. A direita de cada imagem é possível visualizar o correspondente histograma.

2.3 SEGMENTAÇÃO DOS EMBRIÕES

Uma vez padronizadas, as imagens devem sofrer um processo de segmentação, de forma que o blastocisto seja devidamente isolado do restante da imagem e possa ser submetido às técnicas de extração de informações. As seguintes etapas são empregadas no algoritmo de segmentação:

1) Gradiente da Imagem: Inicialmente é calculada a magnitude do gradiente da imagem, de forma que as bordas sejam ressaltadas para as etapas posteriores. Esta operação detecta bordas em todas as direções, característica essencial pelo formato circular do embrião. Quanto maior a variação da intensidade dos pixels, maior será o

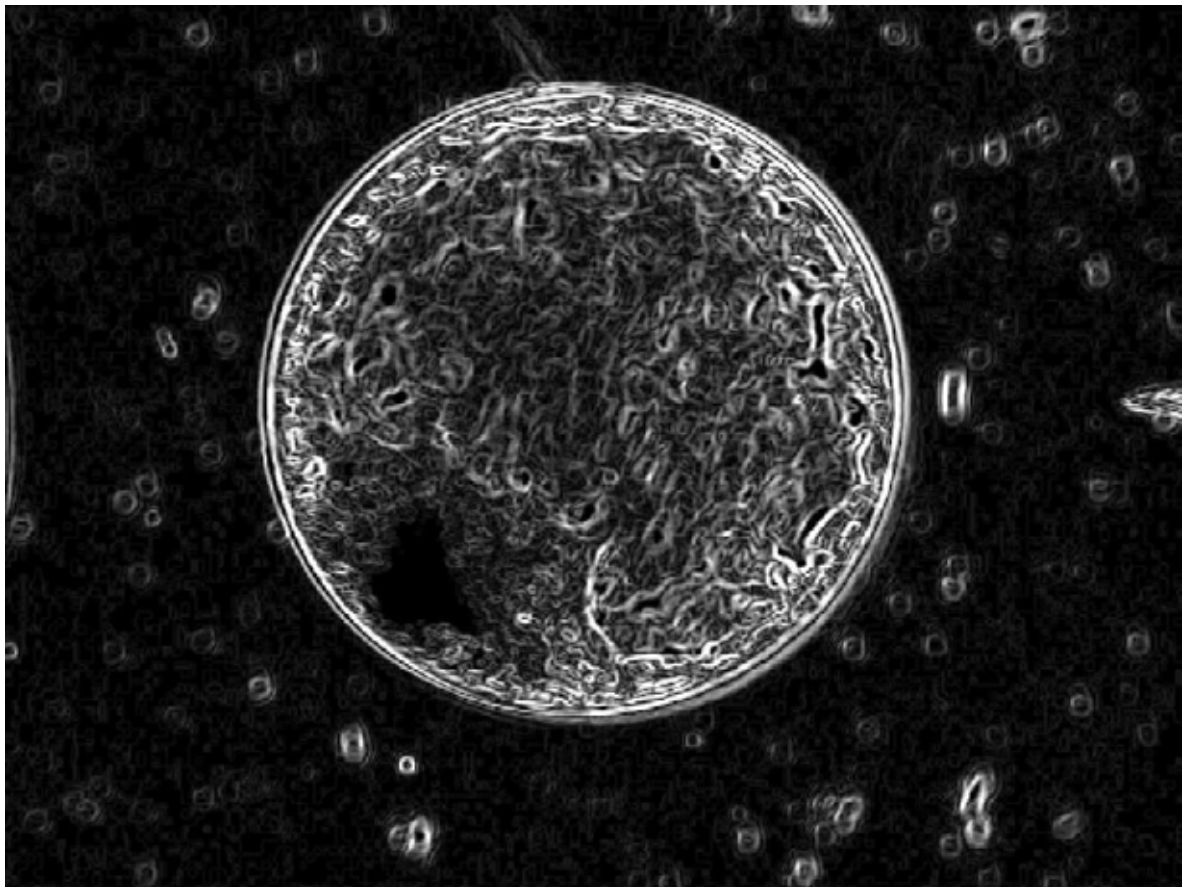


Figura 14 Imagem resultante do cálculo de gradiente, em um blastocisto bovino, após a sua padronização.

valor de magnitude resultante no gradiente final (Figura 14).

2) Imagem Binária: Nesta etapa ocorre o cálculo da imagem binária e foi utilizado como limiar de intensidade, o valor 128 (lembrando que trata-se de uma imagem 8-bit, tendo portanto 255 como o valor máximo de intensidade para cada pixel). O resultado da operação pode ser visto na Figura 15.

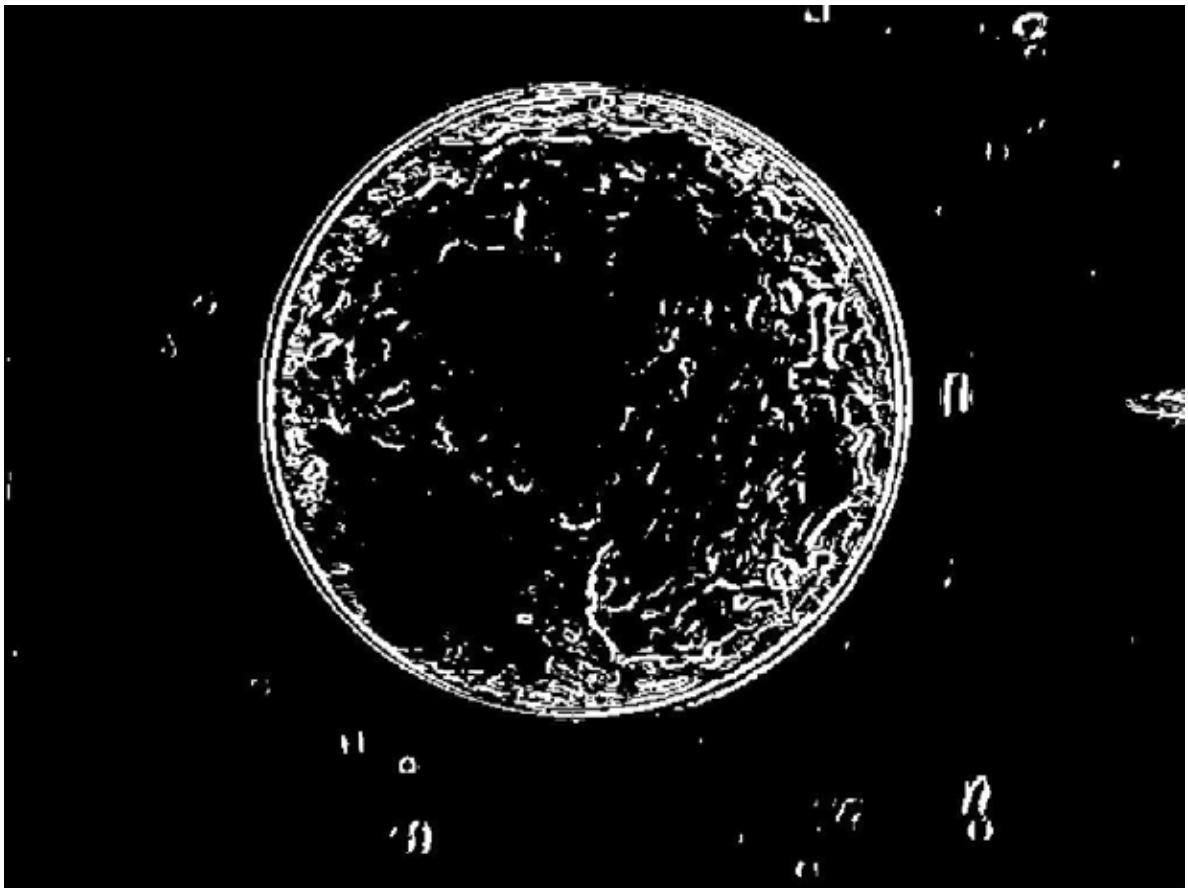


Figura 15 Imagem Binária após padronização e cálculo de gradiente. Pixels brancos possuem valor um e pixels pretos valor zero.

3) Transformada Circular de Hough: Uma vez obtida a imagem binária, círculos podem ser detectados através da Transformada Circular de Hough (ATHERTON; KERBYSON, 1999; BALLARD, 1981; YUEN et al., 1989). Para maior eficiência e precisão do algoritmo, a detecção de círculos ocorre em duas etapas: a primeira procura círculos de raio entre 100 e 150 pixels enquanto a segunda procura raios entre 150 e 250 pixels. Tais valores foram obtidos por meio de verificação do algoritmo com todo o banco de dados, de forma que tanto blastocistos iniciais (menores) como expandidos (maiores) pudessem ser corretamente detectados. O resultado de métrica das duas buscas é comparado e o maior é utilizado como melhor círculo detectado.

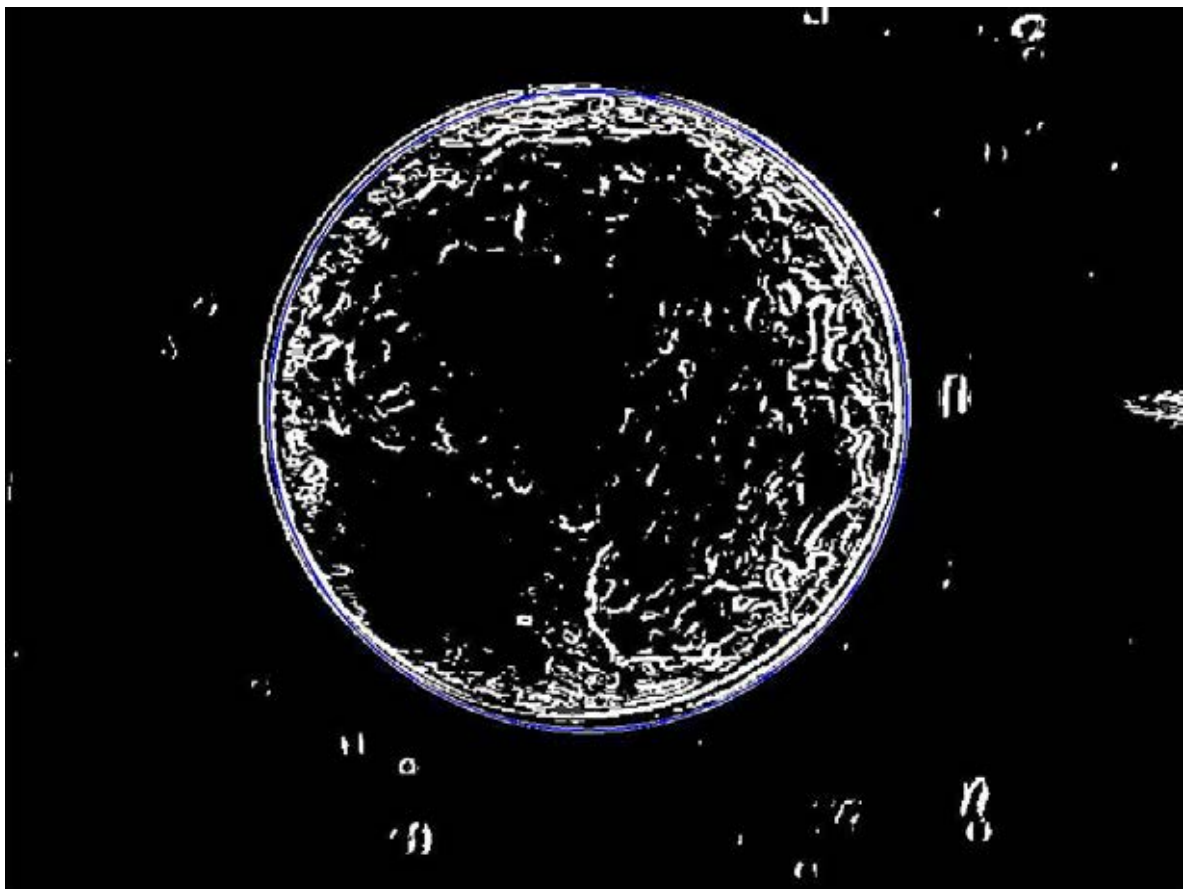


Figura 16 Melhor círculo encontrado pela Transformada de Hough, representado pela linha azul na figura.

4) Isolamento dos embriões: Uma vez detectada a circunferência que melhor representa o embrião, esta pode ser utilizada como máscara para seu isolamento. Três versões são geradas, uma primeira com o raio aumentado em 5 pixels, para que a zona pelúcida seja com certeza incluída (denominada **B+**), uma segunda com o raio reduzido em 40 pixels, de modo que a trofocotoderme seja eliminado e apenas a massa celular interna (MCI) e a blastocèle possam ser analisadas (denominada **B-**) e uma terceira, utilizando a diferença entre B+ e B-, de modo que apenas a região da trofocotoderme possa ser analisada (denominada **TE**). Tais valores de expansão e contração foram determinados por meio de análise do banco de dados. O resultado final do isolamento pode ser encontrado na Figura 17.

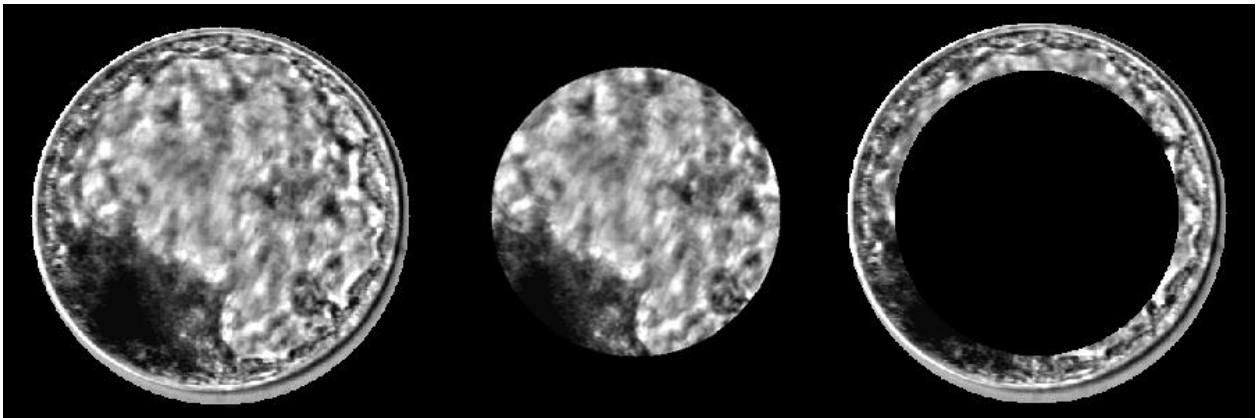


Figura 17 Versões finais isoladas de imagem de blastocisto bovino. À esquerda versão com raio dilatado em 5 pixels (B+), ao centro com raio reduzido em 40 pixels (B-) e à direita a diferença entre os dois raios (TE).

2.4 EXTRAÇÃO DE INFORMAÇÕES

Uma vez que o embrião foi devidamente padronizado e isolado, é possível realizar a etapa de extração de informações, de forma que as variáveis obtidas sejam utilizadas como entrada para o sistema de RNA. O objetivo da extração de informações é obter um vetor numérico que represente a imagem original, porém com dimensionalidade reduzida.

Diversos algoritmos atuam independentemente neste processo, cada um fornecendo uma variável de entrada para a RNA. Foram utilizadas metodologias de *Gray Level Co-Occurrence Matrix* (GLCM) (HARALICK; SHANMUGAM; DINSTEN, 1973) para a análise de texturas, *Transformada de Hough* (ATHERTON; KERBYSON, 1999; BALLARD, 1981; YUEN et al., 1989) para a detecção de círculos, *Transformada Watershed* (BEUCHER, 1992; KÖRBES, 2010) para análise de sub-regiões no blastocisto e também mensurações diretas da imagem isolada. Foram desenvolvidos algoritmos para a extração de 36 variáveis distintas, resumidas na Tabela 2.

Tabela 2 Lista das 36 variáveis desenvolvidas.

ID	Nome	Descrição
1	ContrastB-	Textura (Contraste) de B-
2	CorrelationB-	Textura (Correlação) de B-
3	EnergyB-	Textura (Energia) de B-
4	HomogeneityB-	Textura (Homogeneidade) de B-
5	ContrastTE	Textura (Contraste) de TE
6	CorrelationTE	Textura (Correlação) de TE
7	EnergyTE	Textura (Energia) de TE
8	HomogeneityTE	Textura (Homogeneidade) de TE
9	C1	Círculos escuros de 4 a 8 pixels
10	C1Media	Intensidade Média de C1
11	C2	Círculos claros de 4 a 8 pixels
12	C2Media	Intensidade Média de C2
13	C3	Círculos escuros de 9 a 15 pixels
14	C3Media	Intensidade Média de C3
15	C4	Círculos claros de 9 a 15 pixels
16	C4Media	Intensidade Média de C4
17	Raio	Raio do Blastocisto
18	Soma	Relação da soma da imagem binária com a área
19	MediaB+	Intensidade média de cinza de B+
20	DesvioB-	Desvio Padrão dos valores de cinza de B-
21	MediaB-	Valor médio de B-
22	ModaB-	Valor modal de B-
23	DarkB-	Relação da área dos 10% pixels mais escuros com área total
24	MeanCountB-	Relação de área de 20px ao redor da média com área total
25	BrightB-	Relação da área dos 10% pixels mais claros com área total
26	DesvioTE	Desvio Padrão dos valores de cinza de TE
27	MediaTE	Valor médio de TE
28	ModaTE	Valor modal de TE
29	DarkTE	Relação da área dos 10% pixels mais escuros de TE com área total
30	MeanCountTE	Relação de área de 20px ao redor da média de TE com área total
31	BrightTE	Relação da área dos 10% pixels mais claros de TE com área total
32	NumeroWater	Numero de regiões segmentadas pro Watershed
33	AreaMCI	Valor da maior área encontrada por Watershed
34	ConvexMCI	Área convexa da maior região do Watershed
35	EccenMCI	Excentricidade da maior região do Watershed
36	MeanMCI	Valor médio de intensidade da maior região do Watershed

2.4.1 Descrição das variáveis

Todas as etapas da extração de informações são executadas automaticamente e sem a necessidade de intervenção do usuário. A seguir uma descrição mais detalhada de cada variável.

1) ContrastB-

Variável determinada através de GLCM (detalhes de funcionamento em 1.3.2 Análises de textura, página 18), o contraste é a mensuração da diferença de intensidade entre um pixel e seus vizinhos por toda a imagem, sendo que uma imagem constante possui contraste zero. Ele é calculado por meio da Equação (2):

$$\sum_{i,j} |i - j|^2 p(i, j) \quad (2)$$

2) CorrelationB-

Demonstra a correlação entre os pixels da imagem, em uma determinada vizinhança por toda a imagem. Valores -1 ou 1 indicam uma imagem perfeitamente correlata, de maneira negativa ou positiva, respectivamente. Determinado por meio da Equação (3):

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (3)$$

3) Energy_{B-}

Soma dos elementos da GLCM ao quadrado. Um valor de energia igual a 1 corresponde a uma imagem constante. Representado pela Equação (4):

$$\sum_{i,j} p(i,j)^2 \quad (4)$$

4) Homogeneity_{B-}

Mensuração da proximidade de distribuição dos elementos da GLCM com a diagonal da GLCM. A homogeneidade possui valor 1 para uma GLCM diagonal. Representada pela Equação (5):

$$\sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (5)$$

5) Contrast_{TE}, 6) Correlation_{TE}, 7) Energy_{TE} e 8) Homogeneity_{TE}

Seguem o mesmo princípio as variáveis 1, 2, 3 e 4, respectivamente. Entretanto, utilizam como base a região **TE** ao invés de **B-** para o cálculo da GLCM.

9) C1

Transformada de Hough para a detecção de círculos escuros com raio entre 4 e 8 pixels. O resultado final é a quantidade de círculos encontrados, com uma sensibilidade de 0.935 (valor utilizado como padrão de seleção na matriz acumuladora da transformada).

10) MediaC1

Intensidade média de brilho de todos os círculos detectados em C1. Os centros dos círculos e o raio médio dos mesmos são utilizados na construção de uma máscara binária, que aplicada na imagem de original do blastocisto segmentado permite o isolamento apenas dos círculos. É então calculada a média da intensidade luminosa das regiões isoladas.

11) C2 e 12) MediaC2

Seguem o mesmo padrão das variáveis C1 e MediaC1, respectivamente. Entretanto, nas variáveis C2 e MediaC2 são detectados círculos claros.

13) C3 e 14) C3Media

Seguem os mesmos padrões estabelecidos para as variáveis C1 e C1Media, respectivamente. Entretanto são detectados círculos de raio 9 a 15 pixels, com uma sensibilidade de 0,94.

15) C4 e 16) C4Media

Seguem o mesmo padrão das variáveis C3 e MediaC3, respectivamente. Entretanto, nas variáveis C4 e MediaC4 são detectados círculos claros.

17) Raio

Calculado como a metade da largura da imagem B+.

18) Soma

Primeiramente é calculada uma imagem binária de B+, utilizando o algoritmo de Otsu (OTSU, 1975) para a detecção do limiar. Em seguida, é calculada a somatória de todos os valores da imagem binária resultante. Este valor é então dividido pela área total do Blastocisto.

19) MediaB+

Intensidade média de cinza dos pixels em B+.

20) DesvioB-

Desvio padrão da intensidade de cinza dos pixels em B-, sendo calculado segundo a Equação (6), sendo x o vetor de intensidades da imagem e n o número de elementos no vetor.

$$DesvioPadr\tilde{a}o = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \quad (6)$$

Aonde:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

21) MediaB-

Calculado da mesma maneira que MediaB+, mas utilizando a região de B-.

22) ModaB-

Valor modal de B-, ou seja, o valor de intensidade luminosa mais frequente em B-.

23) DarkB-

Inicialmente são contados os pixels que tenham intensidade luminosa menor ou igual a 25, ou seja, 10% da máxima possível (lembrando que como se trata de valores 8-bit, a intensidade luminosa varia entre 0 e 255). Em seguida, este valor é dividido pela área total do embrião.

24) MeanCountB-

São contados todos os pixels que contenham intensidade luminosa entre 10 pixels menor e 10 pixels maior do que a intensidade média. Este valor é então dividido pela área total do embrião.

25) BrightB-

Mesmo raciocínio empregado em DarkB-, mas utilizando a contagem dos pixels de intensidade maior ou igual a 230 (10% mais claros da imagem).

26) DesvioTE, 27) MediaTE, 28) ModaTE, 29) DarkTE, 30) MeanCountTE, e 31)BrightTE

Seguem o mesmo padrão das variáveis 20 a 25, respectivamente. Entretanto, a imagem TE é utilizada como base para os cálculos.

32) NumeroWater

Número total de regiões encontradas pela transformada Watershed.

33) AreaMCI

Área da maior região encontrada pela transformada Watershed, correspondendo aproximadamente à área da Massa Celular Interna (MCI).

34) ConvexMCI

Área do menor polígono convexo que contenha a maior região detectada pela transformada Watershed. Deste modo, procura-se uma representação mais fiel da real área da MCI.

35) EccenMCI

Excentricidade da maior região detectada pela transformada Watershed. Um valor zero indicaria uma área perfeitamente circular, e o valor 1 um segmento de linha. Considerando uma elipse que tenha a forma aproximada da maior região detectada pela transformada Watershed, EccenMCI é calculado como sendo a relação entre a distância do foco da elipse e o comprimento de seu maior eixo.

36) MeanMCI

Valor médio de intensidade luminosa dos pixels da maior região detectada pela transformada Watershed.

2.5 ANÁLISE DE COLINEARIDADE

Após a extração das variáveis de todos os embriões do banco de dados, foi realizada uma análise de colinearidade das mesmas. Tal análise é importante para atingir uma melhor performance de treinamento do sistema de RNA (WALCZAK; CERPA, 1999). Foi utilizado para análise o software IBM SPSS Statistics v22, e como corte para o valor de *Variance Inflation Fator* (VIF) foi utilizado o valor 10 (O'BRIEN, 2007). Foi retirada a variável que possuísse o maior valor de VIF, e tal processo se repetiu até que todas possuíssem valores menores do que 10. A Tabela 3 demonstra as etapas, sendo que foram necessárias 13 iterações até que nenhuma das variáveis possuísse valores de VIF maiores do que 10.

Tabela 3 Análise de colinearidade das variáveis extraídas, nas 13 etapas necessárias para que todas obtivessem valores de VIF menores do que 10. Cores vermelhas representam valores altos de colinearidade, enquanto cores verdes valores baixos. A coluna ID representa cada variável extraída.

ID	VIF												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	41,4	40,6	39,2	35,4	23,9	15,2	15,1	15,1	14,4	13,2			
2	37,8	37,3	35,9	34,7	14,5	14,4	14,4	14,3	13,7	13,1	5,1	5,0	5,0
3	16,6	16,6	16,5	15,0	11,6	9,8	9,8	9,7	8,7	3,3	2,7	2,7	2,6
4	33,3	33,3	33,3	32,4	32,3								
5	146,1	145,7											
6	87,8	87,2	5,5	5,4	5,3	5,3	5,3	5,2	5,1	5,0	4,8	4,3	3,6
7	197,2	25,5	25,4	22,8	22,8	22,5	22,5						
8	16,7	14,3	13,6	13,5	13,4	13,4	13,4	10,0	9,6	9,6	9,5	9,5	7,3
9	4,4	4,3	4,3	4,3	4,3	4,2	4,2	4,2	4,2	4,2	4,2	4,2	4,1
10	2,9	2,8	2,8	2,8	2,8	2,8	2,8	2,7	2,7	2,7	2,6	2,6	2,6
11	6,3	6,3	6,2	6,1	6,1	5,7	5,7	5,6	5,6	5,6	5,6	5,6	5,5
12	1,9	1,9	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8
13	3,2	3,2	3,2	3,2	3,2	3,1	3,1	3,1	3,1	3,1	3,1	3,1	3,1
14	2,8	2,8	2,8	2,8	2,7	2,7	2,7	2,7	2,7	2,6	2,6	2,6	2,6
15	2,9	2,9	2,9	2,9	2,9	2,9	2,9	2,9	2,9	2,9	2,9	2,8	2,8
16	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,5
17	31,1	26,2	24,5	21,8	21,3	21,2	21,2	20,1					
18	12,5	12,5	12,4	12,3	12,1	11,0	10,8	10,1	9,7	9,5	8,4	8,3	8,1
19	77,5	42,0	41,7	16,9	9,4	9,2	9,2	8,3	8,2	7,5	6,9	6,7	5,6
20	88,3	87,9	85,3	61,5									
21	118,5	90,5	90,5										
22	2,4	2,4	2,4	2,4	2,4	2,4	2,4	2,4	2,3	1,9	1,9	1,9	1,9
23	59,7	59,2	58,8	23,4	20,6	20,4	20,1	20,0	17,3				
24	3,7	3,7	3,7	3,7	3,3	3,2	3,2	3,2	3,1	3,1	2,4	2,4	2,4
25	8,9	8,8	8,7	8,2	4,5	4,5	4,5	4,4	4,4	4,3	3,2	3,2	3,1
26	91,9	55,7	20,2	15,9	15,9	15,9	15,9	12,3	12,3	11,3	11,3	11,0	
27	279,3												
28	2,3	2,3	2,3	2,3	2,2	2,2	2,2	2,2	2,2	2,2	2,1	1,7	1,7
29	29,2	22,6	21,1	20,3	20,1	19,7	19,7	19,6	12,4	11,8	11,6		
30	12,9	9,9	9,9	8,7	7,8	7,8	7,8	7,6	7,6	7,1	7,1	6,5	6,3
31	9,5	9,5	9,5	9,5	9,0	9,0	8,9	8,3	8,3	8,3	8,3	7,8	4,0
32	4,4	4,4	4,4	4,4	4,3	3,8	3,7	3,5	3,4	3,3	3,3	3,0	3,0
33	30,7	30,6	30,5	29,5	29,3	28,7							
34	24,4	24,4	24,3	24,1	24,1	23,7	3,5	3,3	3,2	3,0	3,0	3,0	2,8
35	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,2	1,2
36	1,9	1,9	1,8	1,8	1,8	1,8	1,7	1,7	1,7	1,7	1,7	1,7	1,7

Desse modo, ao final da análise de colinearidade, restaram **24 variáveis** para serem utilizadas no sistema, resumidas na Tabela 4.

Tabela 4 Variáveis selecionadas após a análise de colinearidade.

ID	Nome	Descrição
2	CorrelationB-	Textura (Correlação) de B-
3	EnergyB-	Textura (Energia) de B-
6	CorrelationTE	Textura (Correlação) de TE
8	HomogeneityTE	Textura (Homogeneidade) de TE
9	C1	Círculos escuros de 4 a 8 pixels
10	C1Media	Intensidade Média de C1
11	C2	Círculos claros de 4 a 8 pixels
12	C2Media	Intensidade Média de C2
13	C3	Círculos escuros de 9 a 15 pixels
14	C3Media	Intensidade Média de C3
15	C4	Círculos claros de 9 a 15 pixels
16	C4Media	Intensidade Média de C4
18	Soma	Relação da soma da imagem binária com a área
19	MediaB+	Intensidade média de cinza de B+
22	ModaB-	Valor modal de B-
24	MeanCountB-	Relação de área de 20px ao redor da média com área total
25	BrightB-	Relação da área dos 10% pixels mais claros com área total
28	ModaTE	Valor modal de TE
30	MeanCountTE	Relação de área de 20px ao redor da média de TE com área total
31	BrightTE	Relação da área dos 10% pixels mais claros de TE com área total
32	NumeroWater	Numero de regiões segmentadas pro Watershed
34	ConvexMCI	Área convexa da maior região do Watershed
35	EccenMCI	Excentricidade da maior região do Watershed
36	MeanMCI	Valor médio de intensidade da maior região do Watershed

2.6 ALGORITMOS GENÉTICOS E ARQUITETURA

O funcionamento do Algoritmo Genético (AG) utilizado na determinação de melhor arquitetura para a RNA pode ser descrito nas seguintes etapas:

1) Criação da população inicial

Cada indivíduo representará uma arquitetura diferente, cujas informações estão contidas em seu “código genético” (também denominado *string*). Cada string possui as informações para:

- Quantidade de neurônios das camadas ocultas;
- Função de transferência para as camadas ocultas;
- Função de transferência para camada de saída;
- Função de treinamento a ser utilizada;
- Quantidade de camadas ocultas.

A quantidade total de indivíduos é determinada pelo usuário. As funções de transferência e treinamento são extraídas de uma lista pré-determinada (que também pode ser editada pelo usuário). Neste caso, foram utilizadas as seguintes listas:

Funções de transferência: *logsig*, *purelin*, *tansig*, *hardlim*, *tribas*, *radbas* e *satlin*.

Funções de treinamento: *traingdx*, *trainscg* e *traingdm*.

A quantidade de neurônios é aleatoriamente sorteada dentro de um intervalo estabelecido pelo usuário. A quantidade de camada ocultas é escolhida aleatoriamente, sendo no **mínimo uma** e no **máximo três** (COSNARD, 1993).

2) Teste e seleção da população

Uma vez criados os indivíduos (geradas as redes baseadas nas *strings*) toda a população é treinada e testada, utilizando-se o banco de dados para qual se deseja encontrar a melhor arquitetura. Todos os indivíduos são então ordenados de acordo com o seu erro nos dados de **teste**. O usuário determina um limiar de corte, de modo que apenas os melhores indivíduos (ou seja, os *mais aptos*) podem passar para a geração seguinte. O valor de corte corresponde à porcentagem da população com o menor erro. Por exemplo, em uma população de 100 indivíduos, um valor de corte de 0,3 determina que apenas os 30 melhores (com menor erro) serão utilizados na geração seguinte.

3) Replicação e repovoamento

Os indivíduos que passaram pelo corte formarão a população para a nova época de iteração do algoritmo, de forma que a quantidade total de indivíduos continue constante. Seguindo o exemplo anterior, se em uma população de 100 indivíduos apenas 30 passarem para a próxima época, estes 30 serão os “genitores” dos próximos 100 indivíduos.

Dos selecionados, um par é escolhido ao acaso. Ocorre então a criação de uma nova *string*, que é uma combinação das *strings* dos genitores. Cada característica (ou cada *gene*) é aleatoriamente escolhida entre um dos genitores, até que a nova *string* esteja completa (ou seja, um novo indivíduo para a próxima iteração). Também existe a possibilidade de ocorrer uma mutação durante a criação da nova *string*, tornando

possível a criação de variabilidade genética na população. A taxa de mutação (probabilidade de que possa ocorrer) também é determinada pelo usuário.

Na sequência, um novo par é selecionado aleatoriamente (dentre os mais adaptados da etapa anterior), e o processo se repete até que a população esteja novamente completa. Portanto, teremos uma nova população, que é combinação aleatória dos indivíduos melhor adaptados da iteração anterior, acrescida da possibilidade de terem ocorrido mutações.

4) Obtenção da melhor arquitetura

O processo ocorre por quantas gerações (iterações) forem determinadas pelo usuário, sempre selecionando os indivíduos mais aptos e recombinao suas *strings* para a próxima geração. Como combinações que causam um alto erro são eliminadas, a cada iteração o algoritmo se aproxima do conjunto de arquiteturas ótimas para a resolução do problema. Ao final de todas as gerações, o algoritmo apresenta a rede que obteve o menor erro para os **dados de teste** (dentre todas as gerações testadas). Não apenas são apresentadas as características de arquitetura da melhor rede, mas também a própria rede já corretamente treinada.

Através da Figura 18, é possível observar um fluxograma que demonstra de maneira simplificada o funcionamento do processo de obtenção da melhor arquitetura por meio de algoritmos genéticos.

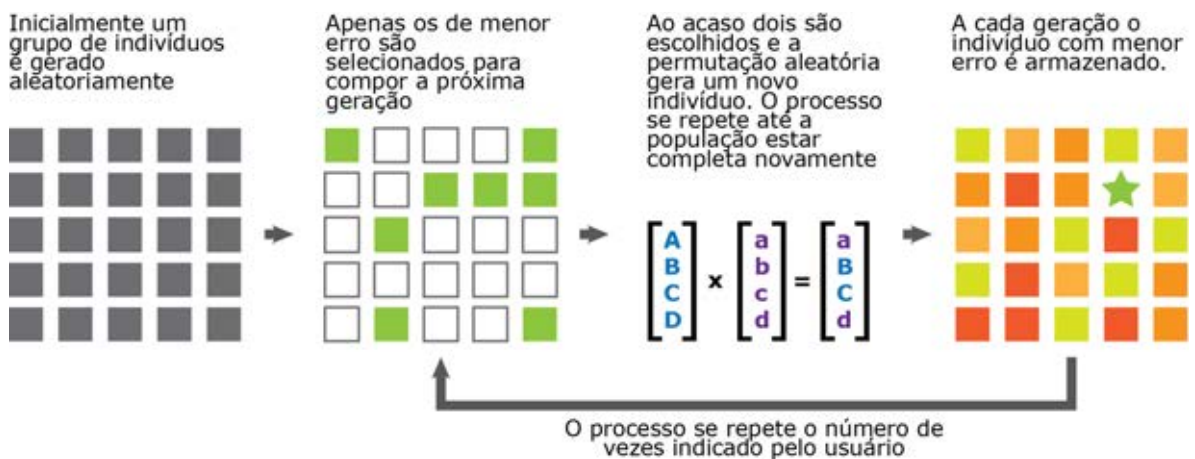


Figura 18 Fluxograma de funcionamento do sistema baseado em algoritmos genéticos para obtenção de melhor arquitetura de Rede Neural Artificial.

3 ARTIGO

Nesta sessão é apresentada uma proposta de artigo científico a ser submetido à *Scientific Reports*, periódico *open access* da *Nature Publishing Group*, atualmente com fator de impacto 2.927 e ISSN 2045-2322.

As normas de publicação podem ser encontradas na sessão 4.1 (Normas de publicação – *Scientific Reports*, página 88), sendo que o manuscrito difere das mesmas por utilizar língua portuguesa e seguir as especificidades de formatação da dissertação.

Software baseado em redes Neurais Artificiais para a classificação morfológica de blastocistos bovinos

Felipe Delestro Matos^{1,2} | Andrea Cristina Basso³ | José Celso Rocha¹ | Marcelo Fábio Gouveia Nogueira²

¹ Laboratório de Matemática Aplicada (MaAp) – Faculdade de Ciências e Letras (FCL) Universidade Estadual Paulista (Unesp), Assis, Brasil

² Laboratório de Micromanipulação Embrionária (LaMEem) – Faculdade de Ciências e Letras (FCL) Universidade Estadual Paulista (Unesp), Assis, Brasil

³ In Vitro Brasil SA - Mogi Mirim-SP, Brasil

*A classificação morfológica embrionária possui grande importância para inúmeras técnicas laboratoriais (desde pesquisas básicas às aplicadas na reprodução assistida). Entretanto, o método utilizado para realizar a classificação dos embriões em diferentes graus de qualidade sempre foi baseado na subjetividade do avaliador e, por mais que sejam estabelecidos padrões de graus de qualidade e descrições das características morfológicas que categorizam um embrião em cada grau, não há atualmente um método preciso que possa gerar resultados consistentes e confiáveis. Assim, nosso trabalho resultou no desenvolvimento de um software capaz de realizar a classificação da qualidade morfológica de blastocistos bovinos. Utilizamos como base de funcionamento técnicas de inteligência artificial (mais especificamente de Redes Neurais Artificiais e Algoritmos Genéticos). Resultados indicam uma taxa de acerto global de **79,2%** na classificação de blastocistos bovinos em 3 graus de qualidade, sendo que para os blastocistos classificados como Bons (Classe 1) a taxa de acerto é de **82,6%**, para os blastocistos classificados como Regulares (Classe 2) é de **16,7%** e para os blastocistos classificados como Pobres (Classe 3) a taxa de acerto é de **91,7%**.*

Desde o desenvolvimento das primeiras técnicas bem sucedidas de superovulação e transferência de embrião (TE) e de produção *in vitro* de embriões (PIVE) em mamíferos, ficou evidente a relação direta entre a qualidade embrionária e a taxa de gestação após a transferência de embriões em fêmeas receptoras. Desse modo, embriões morfologicamente classificados como de alta qualidade produzem maiores taxas de gestação (SCHNEIDER; CASTLEBERRY; GRIFFIN, 1980; TERVIT et al., 1980). Disso decorre a necessidade por um sistema que seja capaz de padronizar os elementos que categorizam os embriões em diferentes graus de acordo com sua qualidade e, portanto, também indiretamente segundo sua viabilidade.

Atualmente, para a espécie bovina, é tido como padrão o sistema classificatório em três graus de qualidade: *bom*, *regular* ou *pobre* (BÓ; MAPLETOFT, 2013), recomendado pela Sociedade Internacional de Transferência de Embriões (IETS). Maiores detalhes sobre o sistema de classificação podem ser encontrados na sessão *Classificação Embrionária*.

Ambos os sistemas são baseados na análise visual (subjetiva e qualitativa) da morfologia do embrião e é comumente realizada por meio de microscopia óptica (estereomicroscópio). A técnica depende da experiência e acurácia do embriologista em analisar e categorizar desde as variáveis mais evidentes até as nuances que tornam um embrião mais ou menos apto ao desenvolvimento. Nessa análise clássica da morfologia embrionária, as variáveis não são objetivamente mensuradas, de forma que ela é subjetiva e pouco repetível (BÉNYEI et al., 2006). Assim, um mesmo embrião que seja mensurado por diferentes especialistas pode ser classificado em distintos graus de

qualidade. Tal discordância ocorre principalmente entre graus próximos, como entre embriões *bons* e *excelentes* (FARIN et al., 1995).

Buscando soluções para a questão da subjetividade da análise morfológica embrionária, diversos métodos alternativos estão sendo desenvolvidos. (HELD et al., 2012; HOSHI, 2003; LÓPEZ-DAMIÁN E. P, GALINA C. S, MERCHANT H, CEDILLO-PELÁEZ C, 2008; OVERSTRÖM, 1996) Dentre eles podem ser destacados o cultivo *in vitro* de embriões (OVERSTRÖM, 1996), a integridade da membrana do blastômero (OVERSTRÖM, 1996), a análise do metabolismo embrionário (OVERSTRÖM, 1996), a mensuração da respiração celular (HOSHI, 2003), a análise mediante microscopia eletrônica (LÓPEZ-DAMIÁN E. P, GALINA C. S, MERCHANT H, CEDILLO-PELÁEZ C, 2008) e o uso de índices de birrefringência da zona pelúcida. (HELD et al., 2012) e, inclusive, testes iniciais no desenvolvimento de softwares para análise semiautomática de embriões (SANTOS FILHO et al., 2012). Entretanto, nenhum método até o momento apresenta uma solução definitiva para a mensuração da qualidade do embrião, sendo estes apresentados apenas experimentais, de modo que ainda é necessário o desenvolvimento de técnicas que sejam rápidas, não invasivas e objetivas (LINDNER; JR, 1983; OVERSTRÖM, 1996). Um fator adicional é o alto custo de alguns métodos, o que impede a sua ampla utilização. Desse modo, apesar da subjetividade, a análise visual da morfologia continua sendo amplamente utilizada para a determinação da qualidade embrionária.

Com este trabalho tivemos o objetivo de testar um método que obtivesse grande precisão na análise morfológica, mediante um processo de extração automática de informações a partir de imagens digitais bidimensionais dos embriões e sua posterior

classificação em graus de qualidade, de acordo com a especificidade de cada espécie. Foram utilizados apenas embriões em estágio de blastocisto (desde blastocisto inicial a blastocisto expandido), uma vez que este grau de expansão apresenta os melhores resultados em procedimentos de transferência embrionária e atualmente é o mais utilizado atualmente pelos laboratórios de produção *in vitro* (BALABAN et al., 2000).

O processo ocorreu em duas etapas principais; a primeira sendo a extração automática das variáveis via técnicas de processamento de imagens, e a segunda sendo a etapa classificatória, na qual os dados são interpretados por um sistema de Redes Neurais Artificiais (RNA).

3.1 RESULTADOS

3.1.1 Desenvolvimento do banco de dados

Para que seja desenvolvido o software de análise morfológica, inicialmente foi necessária a obtenção de um banco de dados de imagens de blastocistos bovinos, obtidos em parceria com a empresa de reprodução *In Vitro Brasil*. Estas imagens serviram como base para o desenvolvimento dos algoritmos de interpretação, bem como o treinamento, validação e teste da Rede Neural Artificial.

Os embriologistas responsáveis pela rotina laboratorial da *In Vitro* classificaram as imagens em três graus de qualidade, segundo os padrões da *International Embryo Transfer Society* (IETS). No total foram obtidas **372** imagens. A classificação dos blastocistos se distribui de acordo com a Figura 19.

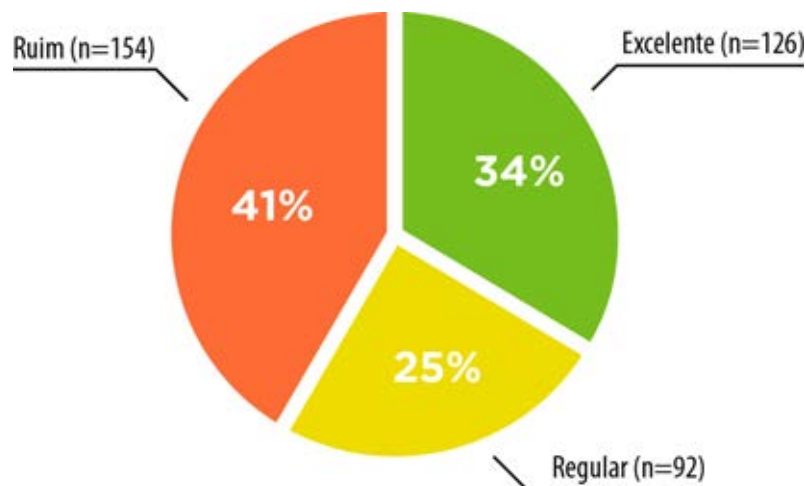


Figura 19 Distribuição do banco de dados de blastocistos bovinos nos três graus de qualidade definidos pela IETS, Excelente ou Bom (Grau 1), Regular (Grau 2) e Ruim (Grau 3).

3.1.2 Padronização das imagens

Cada laboratório possui um padrão diferente para a captura de imagens, de forma que se torna indispensável uma padronização das mesmas. Além das diferenças de resolução e formato, lotes diferentes de imagens (o banco de dados foi desenvolvido mediante envios parciais) possuíam características distintas de iluminação, conforme demonstrado na Figura 20. Todos esses fatores devem ser levados em consideração ao se importar uma imagem para o algoritmo, para que as futuras etapas de segmentação e extração de informações encontrem condições similares, independentemente das características iniciais da imagem. Todos os algoritmos foram desenvolvidos em ambiente Matlab (GONZALEZ, 2004) e permitem uma análise automatizada das imagens, sem a necessidade de intervenção do usuário.



Figura 20 Embriões cujas imagens foram capturadas em momentos distintos. Notar a discrepância entre a iluminação das duas imagens o que faz necessária uma padronização via processamento de imagens.

O algoritmo de padronização segue os seguintes passos:

1) Importação da Imagem: A imagem é importada no ambiente Matlab, independentemente do formato (JPG ou BMP).

2) Conversão em Escala de Cinza: Como as informações de cor não serão utilizadas na interpretação das imagens, todas são convertidas em escala de cinza 8-bit. Tal processo também torna todas as etapas subsequentes mais rápidas, uma vez que ocorre a diminuição da dimensão espectral da imagem (das três bandas RGB para apenas uma).

3) Ajuste de resolução e proporção: Além de resoluções diferentes, as imagens também possuem proporções (*ratios*) diferentes. Uma imagem de 640x480 pixels possui uma proporção de 0,75 enquanto que em uma imagem de 1280x1024 pixels a proporção é de 0,8. Portanto, apenas redimensionar todas as imagens para uma das resoluções iria distorcer as proporções das outras. Foi escolhida como padrão a resolução 640x480 pixels, por ser o padrão inferior e apresentar informações suficientes para a interpretação. Utilizar a resolução menor também aumenta a eficiência das etapas subsequentes do algoritmo, além de não criar limitações de resolução mínima para o software final. Portanto, as imagens de resolução 1280x1024 foram ajustadas para a proporção de 0,75 (mediante corte de uma região superior de 64 pixels) e em seguida redimensionadas para 640x480 pixels.

4) Ajuste de Intensidade: Para minimizar os efeitos das diferentes iluminações, as imagens são submetidas a um ajuste de histograma, de forma que 1% de toda a informação fique saturada entre pixels claros e escuros. Tal processo também atua

aumentando o contraste da imagem, o que facilita a etapa subsequente de segmentação (Figura 21).

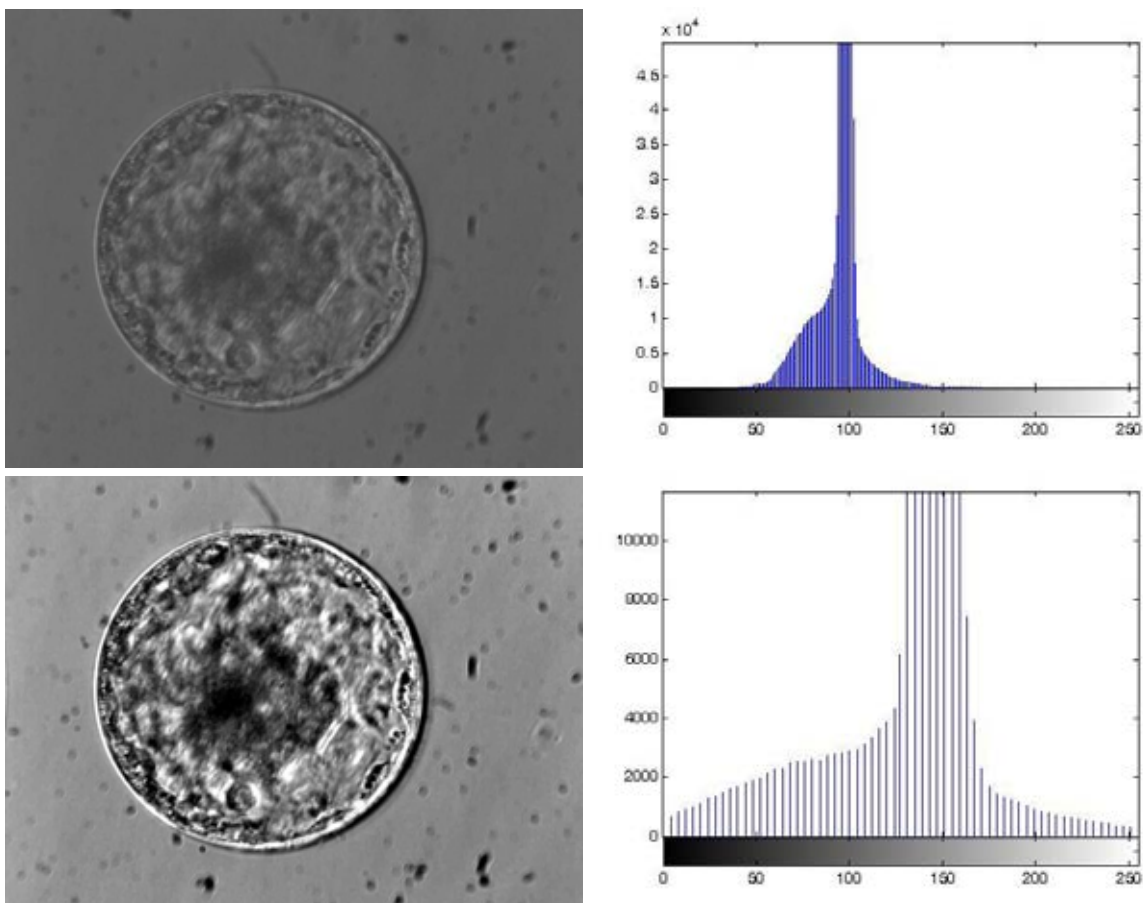


Figura 21 Imagens de um embrião bovino antes (acima) e após (abaixo) a padronização. A direita de cada imagem é possível visualizar o correspondente histograma.

3.1.3 Segmentação dos embriões

Uma vez padronizadas, as imagens devem sofrer um processo de segmentação, de forma que o blastocisto seja devidamente isolado do restante da imagem e possa ser submetido às técnicas de extração de informações. As seguintes etapas são empregadas no algoritmo de segmentação:

1) Gradiente da Imagem: Inicialmente é calculada a magnitude do gradiente da imagem, de forma que as bordas sejam ressaltadas para as etapas posteriores. Esta operação detecta bordas em todas as direções, característica essencial pelo formato circular do embrião. Quanto maior a variação da intensidade dos pixels, maior será o valor de magnitude resultante no gradiente final.

2) Imagem Binária: Nesta etapa ocorre o cálculo da imagem binária e foi utilizado como limiar de intensidade, o valor 128 (lembrando que trata-se de uma imagem 8-bit, tendo portanto 255 como o valor máximo de intensidade para cada pixel).

3) Transformada Circular de Hough: Uma vez obtida a imagem binária, círculos podem ser detectados através da Transformada Circular de Hough (ATHERTON; KERBYSON, 1999; BALLARD, 1981; YUEN et al., 1989). Para maior eficiência e precisão do algoritmo, a detecção de círculos ocorre em duas etapas: a primeira procura círculos de raio entre 100 e 150 pixels enquanto a segunda procura raios entre 150 e 250 pixels. Tais valores foram obtidos por meio de verificação do algoritmo com todo o banco de dados, de forma que tanto blastocistos iniciais (menores) como expandidos (maiores) pudessem ser corretamente detectados. O resultado de métrica das duas buscas é comparado e o maior é utilizado como melhor círculo detectado.

4) Isolamento dos embriões: Uma vez detectada a circunferência que melhor representa o embrião, esta pode ser utilizada como máscara para seu isolamento. Três versões são geradas, uma primeira com o raio aumentado em 5 pixels, para que a zona pelúcida seja com certeza incluída (denominada **B+**), uma segunda com o raio reduzido em 40 pixels, de modo que a trofocotoderme seja eliminado e apenas a massa celular interna (MCI) e a blastocela possam ser analisadas (denominada **B-**) e uma terceira, utilizando a diferença entre B+ e B-, de modo que apenas a região da trofocotoderme possa ser analisada (denominada **TE**). Tais valores de expansão e contração foram determinados por meio de análise do banco de dados. O resultado final do isolamento pode ser encontrado na Figura 22.

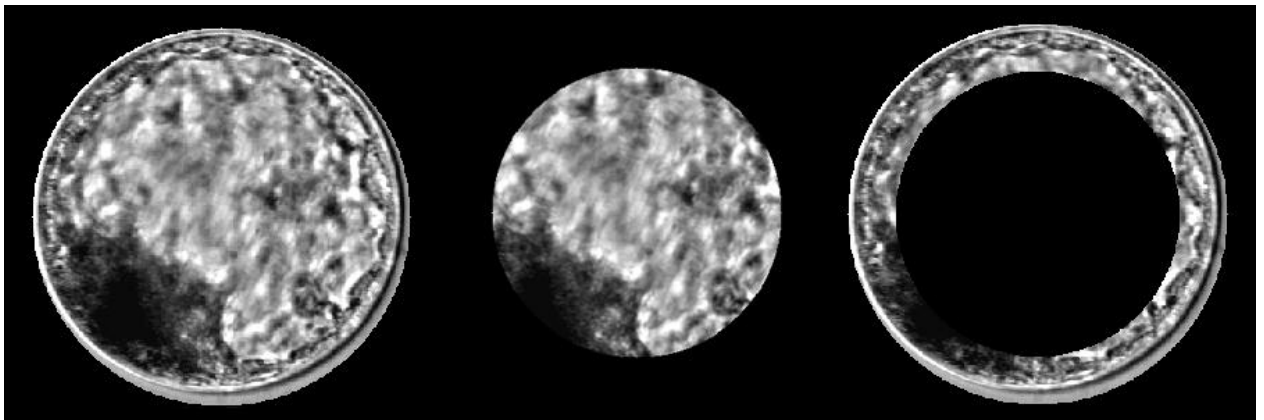


Figura 22 Versões finais isoladas de imagem de blastocisto bovino. À esquerda versão com raio dilatado em 5 pixels (B+), ao centro com raio reduzido em 40 pixels (B-) e à direita a diferença entre os dois raios (TE).

3.1.4 Extração de informações

Uma vez que o embrião foi devidamente padronizado e isolado, é possível realizar a etapa de extração de informações, de forma que as variáveis obtidas sejam utilizadas como entrada para o sistema de RNA. O objetivo da extração de informações é obter um vetor numérico que represente a imagem original, porém com dimensionalidade reduzida.

Diversos algoritmos atuam independentemente neste processo, cada um fornecendo uma variável de entrada para a RNA. Foram utilizadas metodologias de *Gray Level Co-Occurrence Matrix* (GLCM) (HARALICK; SHANMUGAM; DINSTEIN, 1973) para a análise de texturas, *Transformada de Hough* (ATHERTON; KERBYSON, 1999; BALLARD, 1981; YUEN et al., 1989) para a detecção de círculos, *Transformada Watershed* (BEUCHER, 1992; KÖRBES, 2010) para análise de sub-regiões no blastocisto e também mensurações diretas da imagem isolada.

3.1.5 Descrição das variáveis

Todas as etapas da extração de informações são executadas automaticamente e sem a necessidade de intervenção do usuário. A seguir uma descrição detalhada de cada variável:

1) ContrastB-

Variável determinada através de GLCM (detalhes de funcionamento em 1.3.2 Análises de textura, página 18), o contraste é a mensuração da diferença de

intensidade entre um pixel e seus vizinhos por toda a imagem, sendo que uma imagem constante possui contraste zero. Ele é calculado por meio da Equação (8):

$$\sum_{i,j} |i - j|^2 p(i, j) \quad (8)$$

2) CorrelationB-

Demonstra a correlação entre os pixels da imagem, em uma determinada vizinhança por toda a imagem. Valores -1 ou 1 indicam uma imagem perfeitamente correlata, de maneira negativa ou positiva, respectivamente. Determinado por meio da Equação (9):

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (9)$$

3) Energy_{B-}

Soma dos elementos da GLCM ao quadrado. Um valor de energia igual a 1 corresponde a uma imagem constante. Representado pela Equação (10):

$$\sum_{i,j} p(i,j)^2 \quad (10)$$

4) Homogeneity_{B-}

Mensuração da proximidade de distribuição dos elementos da GLCM com a diagonal da GLCM. A homogeneidade possui valor 1 para uma GLCM diagonal. Representada pela Equação (11):

$$\sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (11)$$

5) Contrast_{TE}, 6) Correlation_{TE}, 7) Energy_{TE} e 8) Homogeneity_{TE}

Seguem o mesmo princípio as variáveis 1, 2, 3 e 4, respectivamente. Entretanto, utilizam como base a região **TE** ao invés de **B-** para o cálculo da GLCM.

9) C1

Transformada de Hough para a detecção de círculos escuros com raio entre 4 e 8 pixels. O resultado final é a quantidade de círculos encontrados, com uma sensibilidade de 0.935 (valor utilizado como padrão de seleção na matriz acumuladora da transformada).

10) MediaC1

Intensidade média de brilho de todos os círculos detectados em C1. Os centros dos círculos e o raio médio dos mesmos são utilizados na construção de uma máscara binária, que aplicada na imagem de original do blastocisto segmentado permite o isolamento apenas dos círculos. É então calculada a média da intensidade luminosa das regiões isoladas.

11) C2 e 12) MediaC2

Seguem o mesmo padrão das variáveis C1 e MediaC1, respectivamente. Entretanto, nas variáveis C2 e MediaC2 são detectados círculos claros.

13) C3 e 14) C3Media

Seguem os mesmos padrões estabelecidos para as variáveis C1 e C1Media, respectivamente. Entretanto são detectados círculos de raio 9 a 15 pixels, com uma sensibilidade de 0,94.

15) C4 e 16) C4Media

Seguem o mesmo padrão das variáveis C3 e MediaC3, respectivamente. Entretanto, nas variáveis C4 e MediaC4 são detectados círculos claros.

17) Raio

Calculado como a metade da largura da imagem B+.

18) Soma

Primeiramente é calculada uma imagem binária de B+, utilizando o algoritmo de Otsu (OTSU, 1975) para a detecção do limiar. Em seguida, é calculada a somatória de todos os valores da imagem binária resultante. Este valor é então dividido pela área total do Blastocisto.

19) MediaB+

Intensidade média de cinza dos pixels em B+.

20) DesvioB-

Desvio padrão da intensidade de cinza dos pixels em B-, sendo calculado segundo a Equação (12), sendo x o vetor de intensidades da imagem e n o número de elementos no vetor.

$$DesvioPadr\tilde{a}o = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \quad (12)$$

Onde:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (13)$$

21) MediaB-

Calculado da mesma maneira que MediaB+, mas utilizando a região de B-.

22) ModaB-

Valor modal de B-, ou seja, o valor de intensidade luminosa mais frequente em B-

23) DarkB-

Inicialmente são contados os pixels que tenham intensidade luminosa menor ou igual a 25, ou seja, 10% da máxima possível (lembrando que como se trata de valores 8-bit, a intensidade luminosa varia entre 0 e 255). Em seguida, este valor é dividido pela área total do embrião.

24) MeanCountB-

São contados todos os pixels que contenham intensidade luminosa entre 10 pixels menor e 10 pixels maior do que a intensidade média. Este valor é então dividido pela área total do embrião.

25) BrightB-

Mesmo raciocínio empregado em DarkB-, mas utilizando a contagem dos pixels de intensidade maior ou igual a 230 (10% mais claros da imagem).

26) DesvioTE, 27) MediaTE, 28) ModaTE, 29) DarkTE, 30) MeanCountTE, e 31)BrightTE

Seguem o mesmo padrão das variáveis 20 a 25, respectivamente. Entretanto, a imagem TE é utilizada como base para os cálculos.

32) NumeroWater

Número total de regiões encontradas pela transformada Watershed.

33) AreaMCI

Área da maior região encontrada pela transformada Watershed, correspondendo aproximadamente à área da Massa Celular Interna (MCI).

34) ConvexMCI

Área do menor polígono convexo que contenha a maior região detectada pela transformada Watershed. Deste modo, procura-se uma representação mais fiel da real área da MCI.

35) EccenMCI

Excentricidade da maior região detectada pela transformada Watershed. Um valor zero indicaria uma área perfeitamente circular, e o valor 1 um segmento de linha. Considerando uma elipse que tenha a forma aproximada da maior região detectada pela transformada Watershed, EccenMCI é calculado como sendo a relação entre a distância do foco da elipse e o comprimento de seu maior eixo.

36) MeanMCI

Valor médio de intensidade luminosa dos pixels da maior região detectada pela transformada Watershed.

3.1.6 Análise de colinearidade

Após a extração das variáveis de todos os embriões do banco de dados, foi realizada uma análise de colinearidade das mesmas. Tal análise é importante para atingir uma melhor performance de treinamento do sistema de RNA (WALCZAK; CERPA, 1999). Foi retirada a variável que possuísse o maior valor de VIF, e tal processo se repetiu até que todas possuíssem valores menores do que 10. A Tabela 3 demonstra as etapas, sendo que foram necessárias 13 iterações até que nenhuma das variáveis possuísse valores de *Variance Inflation Fator* (VIF) maiores do que 10.

Desse modo, ao final da análise de colinearidade, restaram **24 variáveis** para serem utilizadas no sistema, resumidas na Tabela 5.

Tabela 5 Variáveis após a análise de colinearidade.

ID	Nome	Descrição
2	CorrelationB-	Textura (Correlação) de B-
3	EnergyB-	Textura (Energia) de B-
6	CorrelationTE	Textura (Correlação) de TE
8	HomogeneityTE	Textura (Homogeneidade) de TE
9	C1	Círculos escuros de 4 a 8 pixels
10	C1Media	Intensidade Média de C1
11	C2	Círculos claros de 4 a 8 pixels
12	C2Media	Intensidade Média de C2
13	C3	Círculos escuros de 9 a 15 pixels
14	C3Media	Intensidade Média de C3
15	C4	Círculos claros de 9 a 15 pixels
16	C4Media	Intensidade Média de C4
18	Soma	Relação da soma da imagem binária com a área
19	MediaB+	Intensidade média de cinza de B+
22	ModaB-	Valor modal de B-
24	MeanCountB-	Relação de área de 20px ao redor da média com área total
25	BrightB-	Relação da área dos 10% pixels mais claros com área total
28	ModaTE	Valor modal de TE
30	MeanCountTE	Relação de área de 20px ao redor da média de TE com área total
31	BrightTE	Relação da área dos 10% pixels mais claros de TE com área total
32	NumeroWater	Numero de regiões segmentadas pro Watershed
34	ConvexMCI	Área convexa da maior região do Watershed
35	EccenMCI	Excentricidade da maior região do Watershed
36	MeanMCI	Valor médio de intensidade da maior região do Watershed

3.1.7 Algoritmos genéticos e arquitetura

Para a determinação da melhor arquitetura para as redes neurais artificiais foi utilizada uma metodologia de Algoritmos Genéticos (AG), que permite a convergência para uma arquitetura que proporcione um menor erro final nos dados de teste. Foi estabelecida uma RNA com 2 camadas ocultas contendo 9 neurônios e função de transferência *satlin*, e uma camada de saída contendo função de transferência *purelin*. Como função de treinamento para a RNA foi escolhida “Scaled conjugate gradient backpropagation” (trainscg) (BENVENUTO; PIAZZA, 1992; HECHT-NIELSEN, 1989), com uma separação dos dados em 70% para treinamento, 15% para validação e 15% para teste.

A performance do treinamento da RNA pode ser averiguada na Figura 23.

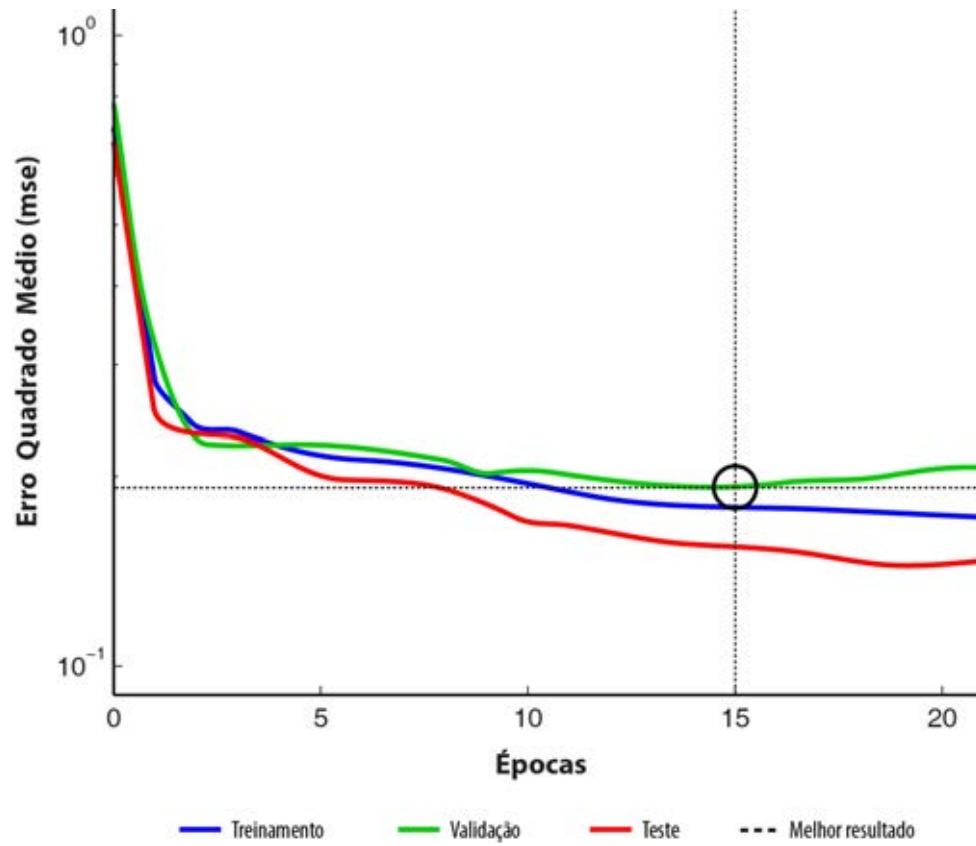


Figura 23 Performance de treinamento da rede neural artificial, separada entre dados de Treinamento (linha azul), Validação (linha verde) e Teste (linha vermelha).

3.1.8 Classificação embrionária

A melhor arquitetura definida pela evolução do AG permite a classificação do banco de dados. Os resultados finais para os dados de teste do treinamento podem ser encontrados na Matriz de Confusão da Figura 24.

DADOS DE TESTE

CLASSE DE SAÍDA DA RNA	1	19 35,8%	3 5,7%	2 3,8%	79,2% 20,8%
	2	0 0,0%	1 1,9%	0 0,0%	100% 0,0%
	3	4 7,5%	2 3,8%	22 41,5%	78,6% 21,4%
		82,6% 17,4%	16,7% 83,3%	91,7% 8,3%	79,2% 20,8%
	1	2	3	CLASSE ALVO	

Figura 24 Matriz de confusão para os dados de teste da classificação.

Outro modo de analisar os dados é através de um gráfico de *Receiver Operator Characteristic* (ROC), apresentado na Figura 25.

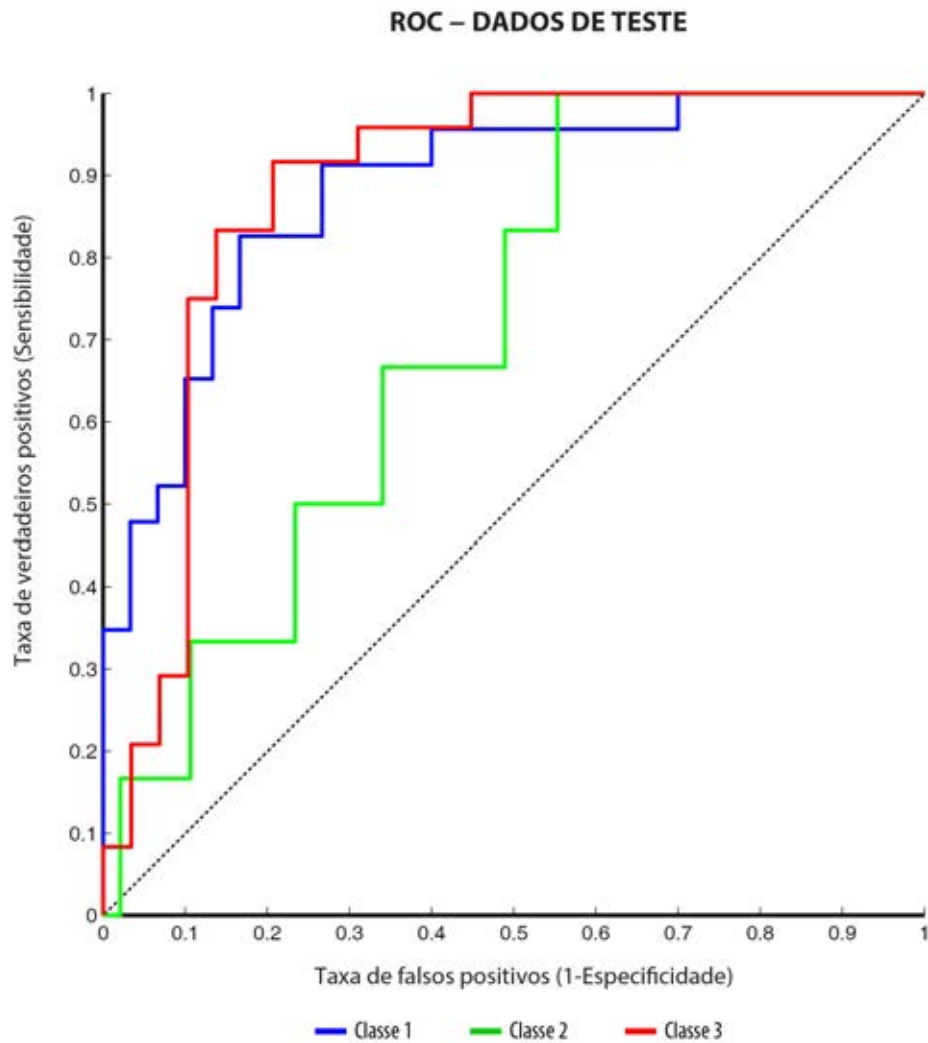


Figura 25 Receiver Operator Characteristic (ROC).

3.2 DISCUSSÃO

O desenvolvimento de um sistema capaz de atuar de maneira automática na interpretação e classificação de embriões se demonstra um grande desafio, não existindo até o momento nenhuma metodologia que atinja resultados equivalentes. O trabalho mais semelhante ao nosso avalia apenas embriões humanos, por meio de um método semiautomático (SANTOS FILHO et al., 2012), o que apresenta limitações para a sua execução na prática laboratorial.

A análise de embriões bovinos apresentou resultados promissores, com uma taxa de acerto global de **79,2%** na classificação de blastocistos bovinos em 3 graus de qualidade, sendo que para os blastocistos classificados como Excelentes ou Bons (Classe 1) a taxa de acerto é de **82,6%**, para os blastocistos classificados como Regulares (Classe 2) é de **16,7%** e para os blastocistos classificados como Pobres (Classe 3) a taxa de acerto é de **91,7%**. A função de *performance* do treinamento da RNA se apresenta estritamente decrescente durante todas as iterações, indicando que a função de treinamento atuou de maneira efetiva. O caráter automatizado da metodologia desenvolvida permite a sua real implementação em um ambiente laboratorial, permitindo que o software atue de maneira efetiva como ferramenta auxiliar na avaliação morfológica de blastocistos.

A análise dos resultados (Figura 24 e Figura 25) torna evidente uma maior taxa de acerto para as classes 1 e 3 (82,6 e 91,7%, respectivamente). Embora a taxa global de acerto seja satisfatória (79,2%) tal fato ocorre apenas devido ao reduzido número de

embriões classe 2 sorteados para o grupo de teste. Caso essa quantidade fosse maior, certamente a taxa de acerto global diminuiria.

Dois principais fatores podem ocasionar a baixa taxa de acerto da classe 2: a ausência de variáveis que possam distinguir corretamente esta classe das demais ou a classificação errônea do banco de dados. A extração de variáveis foi exaustivamente desenvolvida, até o ponto em que várias delas apresentaram alto grau de colinearidade. Portanto, acreditamos que uma reavaliação do banco de dados possa vir a aumentar a eficácia de análise da classe 2 pela RNA. Para analisar essa alternativa, foi realizada a reavaliação de algumas imagens que a RNA julgou com altos valores para as classes 1, 2 ou 3. Os resultados da reavaliação podem ser encontrados na Tabela 6, na qual é possível perceber que a reavaliação concordou em apenas 50% com a avaliação original. Os erros de reavaliação foram menores para as classes 1 e 3 (33% para ambas) do que para a classe 2 (54,4% de discordância entre a reavaliação e a avaliação original). Tal resultado reforça a hipótese de que a baixa taxa de acerto da RNA para embriões de classe 2 seja devido a problemas de classificação do banco de dados.

Através da curva de *Receiver Operator Characteristic* (Figura 25), é possível confirmar que o aprendizado foi mais efetivo para as classes 1 e 3 do que para a classe 2. A linha verde, correspondente à classe 2, se aproxima mais da diagonal da curva ROC, demonstrando sua distribuição mais aleatória.

Tabela 6 Reavaliação de uma amostra do banco de dados e comparação com avaliação original. A coluna “Imagem” mostra o código da imagem escolhida do banco de dados, “Reav.” A classificação realizada na reavaliação, “Orig.” a classificação original utilizada no treinamento da RNA e “Classe” indica a classificação dada pela RNA. As colunas 1, 2 e 3 indicam os valores de saída da RNA para as respectivas classes. As colunas de erros indicam 1 para erro e 0 para concordância.

Imagem	Reav.	Orig.	Classe	Saídas RNA			Erros		
				1	2	3	Reav X Orig	Reav X C	Orig X C
86	3	2	3	0,02	0,22	0,82	1	0	1
92	3	2	3	0,02	0,27	0,82	1	0	1
94	2	2	3	0,04	0,22	0,80	0	1	1
152	1	1	1	0,47	0,41	0,18	0	0	0
155	1	1	1	0,64	0,50	0,11	0	0	0
165	2	2	3	0,12	0,26	0,78	0	1	1
166	2	2	1	0,53	0,31	0,10	0	1	1
170	2	2	1	0,55	0,27	0,12	0	1	1
173	3	2	3	0,10	0,24	0,77	1	0	1
174	1	2	1	0,58	0,39	0,10	1	0	1
178	3	3	1	0,53	0,35	0,12	0	1	1
180	3	3	1	0,62	0,40	0,21	0	1	1
219	2	1	1	0,59	0,47	0,22	1	1	0
234	2	1	1	0,54	0,39	0,14	1	1	0
237	1	1	1	0,59	0,48	0,18	0	0	0
240	1	1	1	0,45	0,44	0,19	0	0	0
257	3	2	3	0,02	0,27	0,85	1	0	1
261	1	2	1	0,56	0,41	0,12	1	0	1
268	3	2	1	0,55	0,44	0,15	1	1	1
368	2	3	1	0,46	0,40	0,21	1	1	1
				Soma Erros			10	10	14
				Concordância			50%	50%	30%

É de nosso conhecimento apenas outra invenção registrada que atue na seleção embrionária (LOEWKE; SURAJ, 2014) que, no entanto, difere por utilizar um sistema de *timelapse* para a determinação da qualidade do embrião baseada na cinética e na simetria de suas clivagens. Entretanto, a sua implementação laboratorial implicaria em grandes gastos com equipamentos adicionais.

O nosso trabalho está protegido pelo pedido de proteção intelectual nacional BR102012031953-5 e internacional PCT-BR2013-000506, realizados juntamente com a Agência Unesp de Inovação (AUIN).

3.3 MÉTODOS

Todos os algoritmos foram desenvolvidos em ambiente Matlab R2013a (8.1.0.604), com auxílio das *Image Processing Toolbox* e *Neural Network Toolbox*. Cada uma das etapas se apresenta como uma função independente, de forma a tornar mais eficiente a adaptação do código a diferentes situações.

As imagens foram capturadas por meio de microscópio invertido Olympus IX71, com magnificação de 32x e captura de imagem pelo software *LuCam Capture* v6.30. As imagens foram armazenadas em formato JPG de 8 bits (RGB), em resolução 1280x1024 pixels. Cada imagem possui apenas um blastocisto, aproximadamente centralizado. Para cada blastocisto foram capturadas duas imagens, a primeira com sua Massa Celular Interna (MCI) na parte inferior e uma segunda com a MCI lateral.

Foi utilizado para análise o software IBM SPSS Statistics v22, e como corte para o valor de *Variance Inflation Fator* (VIF) foi utilizado o valor 10 (O'BRIEN, 2007).

Para a descoberta de melhor arquitetura para a RNA foram utilizadas técnicas de algoritmos genéticos (DE, 1988; MITCHELL, 1995).

3.4 AGRADECIMENTOS

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio à pesquisa (processo nº2012/20110-2). As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não necessariamente refletem a visão da FAPESP.

Agradecemos também os recursos computacionais, instalações, equipamentos e assistência providos pela Faculdade de Ciências e Letras UNESP de Assis. Os autores também agradecem à equipe da *In Vitro Brasil*, empresa parceira no fornecimento de imagens para a classificação embrionária.

3.5 CONTRIBUIÇÃO DOS AUTORES

F.D.M desenvolveu o sistema de AG para determinação da arquitetura da RNA, bem como o método de extração de variáveis de blastocistos. F.D.M e M.F.G.N escreveram o manuscrito. F.D.M e J.C.R desenvolveram a RNA e validaram os resultados. F.D.M e A.C.B desenvolveram o banco de dados de imagens. Todos os autores revisaram o manuscrito.

3.6 REFERÊNCIAS BIBLIOGRÁFICAS

ATHERTON, T.; KERBYSON, D. Size invariant circle detection. **Image and Vision computing**, v. 17, n. February 1997, p. 795–803, 1999.

BALABAN, B. et al. Blastocyst quality affects the success of blastocyst-stage embryo transfer. **Fertility and sterility**, v. 74, n. 2, p. 282–7, ago. 2000.

BALLARD, D. Generalizing the Hough transform to detect arbitrary shapes. **Pattern recognition**, 1981.

BENVENUTO, N.; PIAZZA, F. On the complex backpropagation algorithm. **IEEE Transactions on Signal Processing**, v. 40, n. 4, p. 967–969, abr. 1992.

BÉNYEI, B. et al. The effect of internal and external factors on bovine embryo transfer results in a tropical environment. **Animal reproduction science**, v. 93, n. 3-4, p. 268–79, jul. 2006.

BEUCHER, S. The watershed transformation applied to image segmentation. **SCANNING MICROSCOPY-SUPPLEMENT-**, p. 1–26, 1992.

BÓ, G.; MAPLETOFT, R. Evaluation and classification of bovine embryos. **Anim. Reprod**, v. 54, n. 935, p. 344–348, 2013.

BROWN, C. D.; DAVIS, H. T. Receiver operating characteristics curves and related decision measures: A tutorial. **Chemometrics and Intelligent Laboratory Systems**, v. 80, n. 1, p. 24–38, jan. 2006.

COSNARD, M. A step towards the frontier between one-hidden-layer and two-hidden-layer neural networks. **Neural Networks, 1993**. ..., p. 2292–2295, 1993.

DE, K. Learning with Genetic Algorithms : An Overview. p. 121–138, 1988.

EGMONT-PETERSEN, M.; DE RIDDER, D.; HANDELS, H. Image processing with neural networks—a review. **Pattern Recognition**, v. 35, n. 10, p. 2279–2301, out. 2002.

FARIN, P. et al. Agreement among evaluators of bovine embryos produced in vivo or in vitro. **Theriogenology**, n. 95, p. 339–349, 1995.

GOETHALS, P. L. M. et al. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. **Aquatic Ecology**, v. 41, n. 3, p. 491–508, 9 maio 2007.

GONZALEZ, R. **Digital Image processing using MATLAB**. Upper Saddle River N.J.: Pearson Prentice Hall, 2004.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. [s.l.: s.n.]. p. 976

GURESEN, E.; KAYAKUTLU, G.; DAIM, T. U. Using artificial neural network models in stock market index prediction. **Expert Systems with Applications**, v. 38, n. 8, p. 10389–10397, ago. 2011.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural Features for Image Classification. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 3, n. 6, p. 610–621, nov. 1973.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation (2nd Edition)**. [s.l.] Prentice Hall, 1998. p. 842

HECHT-NIELSEN, R. Theory of the backpropagation neural network. **Neural Networks, 1989. IJCNN., International ...**, p. 593–605, 1989.

HELD, E. et al. Zona pellucida birefringence correlates with developmental capacity of bovine oocytes classified by maturational environment, COC morphology and G6PDH activity. **Reproduction, Fertility and Development**, v. 24, n. 4, p. 568, 2012.

HOSHI, H. In vitro production of bovine embryos and their application for embryo transfer. **Theriogenology**, v. 59, n. 2, p. 675–85, 15 jan. 2003.

HU, Y.; ZHAO, C.; WANG, H. Directional Analysis of Texture Images Using Gray Level Co-Occurrence Matrix. **2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application**, p. 277–281, dez. 2008.

JAIN, A.; MAO, J.; MOHIUDDIN, K. Artificial neural networks: A tutorial. **Computer**, v. 29, p. 31–44, 1996.

KÖRBES, A. **Análise de Algoritmos da Transformada Watershed**. [s.l.] Universidade Estadual de Campinas, 2010.

LI, E. Y. Artificial neural networks and their business applications. **Information & Management**, v. 27, n. 5, p. 303–313, nov. 1994.

LINDNER, G.; JR, R. W. Bovine embryo morphology and evaluation. **Theriogenology**, v. 20, n. 0313, p. 407–416, 1983.

LOEWKE, K. E.; SURAJ, V. **IN VITRO EMBRYO BLASTOCYST PREDICTION METHODS**, 2014.

LÓPEZ-DAMIÁN E. P, GALINA C. S, MERCHANT H, CEDILLO-PELÁEZ C, A. M. Assessment of Bos taurus embryos comparing stereoscopic microscopy and

transmission electron microscopy. **Journal of Cell and Animal Biology**, v. 2, n. 3, p. 72–78, 2008.

MACHART, P.; RALAIVOLA, L. Confusion matrix stability bounds for multiclass classification. **arXiv**, n. Section 5, 2012.

MITCHELL, M. Genetic algorithms: an overview. **Complexity**, 1995.

O'BRIEN, R. M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. **Quality & Quantity**, v. 41, n. 5, p. 673–690, 13 mar. 2007.

OTSU, N. A threshold selection method from gray-level histograms. **Automatica**, v. C, n. 1, p. 62–66, 1975.

OVERSTRÖM, E. In vitro assessment of embryo viability. **Theriogenology**, v. 45, n. 95, p. 3–16, 1996.

RAMLALL, I. Artificial intelligence: neural networks simplified. **International Research Journal of Finance and ...**, p. 1–20, 2009.

ROBERTI DE SIQUEIRA, F.; ROBSON SCHWARTZ, W.; PEDRINI, H. Multi-scale gray level co-occurrence matrices for texture description. **Neurocomputing**, v. 120, p. 336–345, nov. 2013.

ROCHA, C. et al. Utilização de redes neurais artificiais para a determinação do número de refeições diárias de um restaurante universitário Use of artificial neural networks to determine. **Revista de Nutrição**, v. 24, n. 5, p. 735–742, 2011.

SANTOS FILHO, E. et al. A method for semi-automatic grading of human blastocyst microscope images. **Human reproduction (Oxford, England)**, v. 27, n. 9, p. 2641–8, set. 2012.

SCHNEIDER, H. J.; CASTLEBERRY, R. S.; GRIFFIN, J. L. Commercial aspects of bovine embryo transfer. **Theriogenology**, v. 13, n. 1, p. 73–85, jan. 1980.

TERVIT, H. et al. Non-surgical embryo transfer in cattle. **Theriogenology**, v. 13, n. 1, p. 63–71, 1980.

WALCZAK, S.; CERPA, N. Heuristic principles for the design of artificial neural networks. **Information and Software Technology**, v. 41, n. 2, p. 107–117, jan. 1999.

YAO, X.; LIU, Y. Towards designing artificial neural networks by evolution. **Applied Mathematics and Computation**, v. 91, n. 1, p. 83–90, abr. 1998.

YUEN, H. K. et al. A Comparative Study of Hough Transform Methods for Circle Finding. **Proceedings of the Alvey Vision Conference 1989**, p. 29.1–29.6, 1989.

ZHANG, G.; PATUWO, B. E.; HU, M. Forecasting with artificial neural networks:: The state of the art. **International journal of forecasting**, v. 14, p. 35–62, 1998.

4 ANEXOS

4.1 NORMAS DE PUBLICAÇÃO – SCIENTIFIC REPORTS

Disponível em <http://www.nature.com/srep/authors/index.html>

Aims and scope of the journal

Scientific Reports is an online, open access, primary research publication covering all areas of the natural sciences. Papers published by *Scientific Reports* will be technically sound, and of interest to specialists within their field. The main results and conclusions must not have been published or submitted elsewhere.

The scope of the journal reflects Nature Publishing Group's desire to ensure there is a home for all scientifically sound research in the natural sciences. To ensure each field is adequately represented editorially, *Scientific Reports* is constantly recruiting new [Editorial Board Members](#) to manage the peer-review process. We welcome recommendations of suitable Editorial Board Members at any time: please contact scientificreports@nature.com

Scientific Reports is committed to providing an efficient service for both authors and readers. A streamlined peer-review system managed by the Editorial Board - together with the support of an Editorial Advisory Panel and an in-house publishing team — allows for rapid and fair publication decisions. Prompt dissemination of accepted papers to Nature Publishing Group's wide readership and beyond is achieved through a programme of continuous online publication. Published manuscripts are enhanced by the tools and technology found on nature.com.

Transfer service

Scientific Reports is editorially independent, and Editorial Board Members make decisions independently from other Nature Publishing Group journals. It is for authors alone to decide where to submit their manuscripts. For papers that could satisfy the scope of more than one Nature Publishing Group journal, the choice of which journal to submit to first lies with the authors.

If a paper is rejected from a Nature Publishing Group journal, the authors can use an automated manuscript transfer service to submit the paper to *Scientific Reports*. Alternatively, authors may choose to submit afresh, in which case they should not use the automated transfer link, and the Editorial Board Member will evaluate the paper without reference to the previous process.

Costs

Scientific Reports is an open access publication. To provide this service, all expenses, including peer review, production, and online hosting and archiving, are recovered via an article-processing charge (APC).

The APC is as follows, plus VAT or local taxes where applicable:

\$1,350 (The Americas)

€1,046 (Europe)

¥142,500 (Japan)

£890 (UK and Rest of World)

¥9,350 (China)

Editorial board

Editorial decisions on *Scientific Reports* are made by the [Editorial Board](#) — scientists from a broad spectrum of the scientific community — who will manage the peer-review process and decide which manuscripts should be published.

Contact information

For general editorial enquiries relating to *Scientific Reports*, including manuscript submission queries, and for enquiries relating to the Editorial Advisory Panel or the Editorial Board, please contact scientificreports@nature.com.

Format of articles

Scientific Reports publishes original research in one format, Articles, which may range in length from short communications through to more in-depth studies. In most cases we do not impose strict limits on word and page lengths, however we encourage authors to write concisely and suggest authors adhere to the guidelines below.

Articles should be no more than 11 typeset pages in length. The main text (not including Abstract, Methods, References and figure legends) should be no more than 4,500 words. The maximum title length is 20 words. The Abstract — which must be no more than 200 words long and contain no references — should serve both as a general introduction to the topic and as a brief, non-technical summary of the main results and their implications.

The main text of an Article should begin with an Introduction (without heading) of referenced text that expands on the background of the work (some overlap with the Abstract is acceptable), followed by sections headed Results, Discussion and Methods. The Results and Methods sections may be divided by topical subheadings; the Discussion should be succinct and may not contain subheadings. The Methods section should be limited to 1,500 words. Figure legends are limited to 350 words. References should be limited to 60. Footnotes are not used.

Depending on the word count, Articles may have up to 8 display items (figures and/or tables). In addition, a limited number of uncaptioned molecular structure graphics and numbered mathematical equations may be included if necessary. To enable typesetting of papers, the number of display items should be commensurate with the word length — those with word counts less than 2,000 should have no more than 4 figures/tables. Please note that schemes are not used; these should be presented as figures.

Articles are peer-reviewed and include received/accepted dates. Authors must provide a competing financial interests statement within the manuscript file.

Authors should visit the [Submission Guidelines](#) for more detailed information.

Online submission

All submissions must be made via our [online submission system](#). Using this system, authors can upload manuscript files (text, figures, videos) directly to our office and check on the status of their manuscripts during the review process. In addition, reviewers can access the manuscript (in a highly secure fashion that maintains referee anonymity) over a direct internet link, which speeds up the review process. Please consult our technical information on file formats and tips for using the system effectively.

When submitting your manuscript, please note that there are restrictions on the types of files that you are able to upload. This will enable a faster production process. Accepted article and figure types are as follows:

For Article Text — txt, doc, docx, tex

For Figures — eps, tif, and jpeg

If your paper does not include formulas, we strongly encourage you to submit your paper in txt, doc or docx rather than tex.

Submission policies

Submission to *Scientific Reports* is taken to imply that the submitted manuscript has not already been published elsewhere. If similar or related work has been published or submitted elsewhere, then the authors must provide a copy with the submitted article. Authors may not submit elsewhere while the manuscript is under consideration at *Scientific Reports*.

The primary affiliation for each author should be the institution where the majority of their work was done. If an author has subsequently moved, the current address may also be stated.

If the manuscript includes personal communications, please provide a written statement of permission from any person who is quoted. E-mail permission messages are acceptable.

Please visit the [Guide to referees](#) for further information on the peer review process.

Contributors are sent proofs, and will be able to correct major scientific errors or inaccuracies in the title or author list, but *Scientific Reports* reserves the right to limit the scope of changes.

Scientific Reports reserves the right to reject a paper even after it has been accepted if it becomes apparent that there are serious problems with the scientific content or with violations of our publishing policies.

For information relating to submission of a competing financial interests statement, pre-publication publicity, deposition of data as a condition of publication, availability of data and reagents after publication, human and animal subjects, digital image integrity, biosecurity, corrections and retractions, duplicate publication, confidentiality and plagiarism, please visit [Editorial and publishing policies](#) of *Scientific Reports*.

How to submit

For detailed information about how to submit to *Scientific Reports*, please view the [Submission Guidelines](#).

Initial submission

Manuscripts should be submitted via the [online submission system](#). Copies of any papers containing similar or related work that are under consideration or in press at other journals should be included with the submission as additional supplementary information.

Each new submission is assigned to an Editorial Board Member, who will read the paper and evaluate the appropriateness for the journal. We strongly suggest that the cover letter of your manuscript includes suggestions for [Editorial Board Members](#) who would be able to handle your submission. Manuscripts that are within scope and seem, on initial assessment, to be technically sound, will be peer reviewed.

If a paper had previously been submitted to another Nature Publishing Group journal, authors can use an automated manuscript transfer service to transfer their paper to *Scientific Reports* via a link in their decision letter. Alternatively, authors may choose to submit afresh, in which case they should not use the automated transfer link.

Peer review

The corresponding author will be notified by e-mail when the Editorial Board Member decides whether or not the paper will be peer reviewed. At this point the Editorial Board Member has two options:

The Editorial Board Member may choose to contact one or more referee(s) to conduct the peer review.

The Editorial Board Member may choose to conduct the peer review themselves, based on their own experience and expertise.

After consideration the Editorial Board Member will make one of the following decisions:

Accept, with very minor editorial revisions

Request a minor revision, where authors have one month to revise their manuscript to address specific concerns

Request a major revision, where authors have three months to revise their manuscript to address significant concerns and perhaps undertake additional work

Reject outright

At the submission stage, authors may indicate a limited number of scientists who should not review the paper. Excluded scientists must be identified by name. Authors may also suggest potential reviewers; these suggestions are often helpful, although they are not always followed. By policy, referees are not identified to the authors, except at the request of the referee.

Decision after review

In cases where the referee has requested well-defined changes to the manuscript that do not appear to require extensive further experiments, the Editorial Board Member may request a revised manuscript that addresses the

referees' concerns. The decision letter will specify a deadline, and revisions that are returned within this period will retain their original submission date.

In cases where the referees' concerns are more wide-ranging, the Editorial Board Member will normally reject the manuscript. If the Editorial Board Member feels the work is of potential interest to the journal, however, they may express interest in seeing a future resubmission. The resubmitted manuscript may be sent back to the original referees or to new referees, at the Editorial Board Member's discretion. In such cases, revised manuscripts will not retain their earlier submission date.

In either case, the revised manuscript should be accompanied by a cover letter explaining how the manuscript has been changed.

An invited revision should be submitted via the revision link to the online submission system provided in the decision letter, not as a new manuscript.

Final submission and acceptance

When all editorial issues are resolved, the paper is formally accepted. The received date is the date on which the editors received the original manuscript. The accepted date is when the Editorial Board Member sends the acceptance letter.

Contributors are sent proofs; however, the production process does not allow minor changes. Only changes in the title, author list or major scientific errors will be permitted at this stage. All corrections will be approved by the publishing team. *Scientific Reports* reserves the right to make the final decision about matters of style and the size of figures.

Appeals

Even in cases where *Scientific Reports* does not invite resubmission, some authors may ask the Editorial Board to reconsider a rejection decision. These are considered appeals, which, by policy, must take second place to the normal workload. In practice, this means that decisions on appeals often take several weeks. Only one appeal is permitted for each manuscript, and appeals can only take place after peer review. Final decisions on appeals will be made by the Editorial Board Member handling the paper.

Decisions are reversed on appeal only if the relevant Editorial Board Member is convinced that the original decision was a serious mistake, not merely a borderline call that could have gone either way. Further consideration may be merited if a referee made substantial errors of fact or showed evidence of bias, but only if a reversal of that referee's opinion would have changed the original decision. Similarly, disputes on factual issues need not be resolved unless they were critical to the outcome. Thus, after careful consideration of the authors' points, most appeals will be rejected.

If an appeal merits further consideration, the Editorial Board Member may send the authors' response or the revised paper to the original referee/s.

Submission Template

Preparing a manuscript for submission to *Scientific Reports*

It is **essential** that manuscripts are submitted with all of the sections below, and in the order described.

It is **suggested** that manuscripts are formatted into the *Scientific Reports* style, using the [Scientific Reports submission template](#).

Accepted file types:

For article text: txt, doc, docx, tex

For figures: eps, tiff, jpg

If your paper does not include formulas, we strongly encourage you to submit your paper in txt, doc or docx rather than tex.

Scientific Reports sections and ordering of manuscripts

Please ensure your manuscript follows the order below.

Cover letter

All submissions must be accompanied by a [cover letter](#).

Title

The title must be no longer than 20 words.

Author list and affiliations

The corresponding author should be identified with an asterisk.

Abstract (heading not used)

The abstract must be no longer than 200 words, and must not have a heading. The Abstract should serve both as a general introduction to the topic and as a brief, non-technical summary of the main results and their implications. Abstracts must not contain references or subheadings.

Introduction (heading not used)

The Introduction section, of referenced text, expands on the background of the work (some overlap with the Abstract is acceptable). The introduction must not have a heading or subheading.

Results

Topical subheadings are allowed.

Discussion

The Discussion should be succinct and must not contain subheadings.

The recommended length for the Introduction, Results, and Discussion section combined is 4,500 words. The Results and Discussion must not be combined.

Methods

Topical subheadings are allowed. We recommend that authors limit their [Methods](#) section to 1,500 words.

References

We recommend that references are limited to 60. [References](#) should be numbered sequentially, first throughout the text, then in tables, followed by figures and, finally, boxes. *Scientific Reports* uses standard Nature referencing style, for example:

1. Schott, D. H., Collins, R. N. & Bretscher, A. Secretory vesicle transport velocity in living cells depends on the myosin V lever arm length. *J. Cell Biol.* **156**, 35-39 (2002).

Acknowledgements

Acknowledgements should be brief, and should not include thanks to anonymous referees and editors, or effusive comments. Grant or contribution numbers may be acknowledged.

Author Contributions Statement

All manuscripts must include an Author Contributions Statement. For example, "AB and CD wrote the main manuscript text and EF prepared figures 1-3. All authors reviewed the manuscript."

Additional Information

To include, in this order: **Accession codes** (where applicable); **Competing financial interests** (mandatory statement).

The corresponding author is responsible for submitting a [competing financial interests statement](#) on behalf of all authors of the paper. This statement must be included in the submitted article file.

Figure Legends

Each figure legend must be no longer than 350 words. Text for [figure legends](#) should be provided in numerical order after the references.

Display Items

Depending on the word count, Articles may have up to 8 display items ([figures](#) and/or [tables](#)). In addition, a limited number of uncaptioned molecular structure graphics and numbered mathematical equations may be included if necessary. To enable typesetting of papers, the number of display items should be commensurate with the word length - those with word counts less than 2,000 should have no more than 4 figures/tables. Please note that schemes are not used; these should be presented as figures.

Tables

Please submit tables at the end of your text document (in Word or TeX/LaTeX, as appropriate).

Figures

Figures should be uploaded upon submission via our [online submission system](#), in one of our accepted formats (eps, tiff, jpg). For first submissions (i.e. not revised manuscripts), authors may choose to incorporate the manuscript text and figures into a single file up to 3 MB in size - the figures may be inserted within the text at the appropriate positions, or grouped at the end.

Authors submitting revised manuscripts, or final version figures should see our [figure guidelines](#).

Supplementary Information

Any [supplementary information](#) must be uploaded as a separate file, preferably as a single PDF, clearly labelled with the manuscript title and author list.



Submissions should not include: running titles, keywords, columns, line numbers or footnotes.

Submission Template

To help authors prepare their manuscripts for submission we have made available a [word.dot style template](#) that includes the preferred styles for manuscripts submitted to *Scientific Reports*.

4.2 DEPÓSITOS DE PATENTES



 INSTITUTO NACIONAL DE PROPRIEDADE INDUSTRIAL PROTOCOLO GERAL 14/12/2012 018120046308 14:43 DESP  BR 10 2012 031953 5	< Uso exclusivo do INPI >
	Espaço para etiqueta

DEPÓSITO DE PEDIDO DE PATENTE OU DE CERTIFICADO DE ADIÇÃO

Ao Instituto Nacional da Propriedade Industrial:

O requerente solicita a concessão de um privilégio na natureza e nas condições abaixo indicadas

1. Depositante (71):

- 1.1 Nome: Universidade Estadual Paulista "Julio de Mesquita Filho" - Unesp
 1.2 Qualificação: Autarquia estadual de regime especial
 1.3 CNPJ/CPF: 48031918/0001-24
 1.4 Endereço Completo: Rua Quirino de Andrade, 215 - Centro - São Paulo - SP
 1.5 CEP: 01049-010 1.6 Telefone: (11) 33937903 1.7 Fax: (11) 33937900
 1.8 E-mail: auin@unesp.br

continua em folha anexa

- 2. Natureza:** Invenção Modelo de Utilidade Certificado de Adição

Escreva, obrigatoriamente, e por extenso, a Natureza desejada: Privilégio de Invenção

3. Título da Invenção ou Modelo de Utilidade ou Certificado de Adição(54):

MÉTODO PARA DETERMINAÇÃO DA VIABILIDADE E QUALIDADE DE EMBRIÕES

continua em folha anexa

- 4. Pedido de Divisão:** do pedido N° _____ Data de Depósito: _____

- 5. Prioridade:** interna unionista

O depositante reivindica a(s) seguinte(s):

Pais ou organização de origem	Número de depósito	Data do depósito

6. Inventor (72):

Assinale aqui se o(s) mesmo(s) requer(em) a não divulgação de seu(s) nome(s)

- 6.1 Nome: FELIPE DELESTRO MATOS 6.3 CPF: 356.099.468-32
 6.2 Qualificação: Estudante
 6.4 Endereço completo: Av. Dom Antonio, nº 2100, Parque Universitário - Asis -SP
 6.5 CEP: 19.806-900 6.6 Telefone: (11) 33937901 6.7 Fax: (11) 33937900
 6.8 E-Mail: auin@reitoria.unesp.br

continua em folha anexa



PCT

REQUERIMENTO

solicita que o presente pedido
seja processado de acordo com o
Tratado de Cooperação em matéria de Patentes

Reservado para o Organismo receptor

PCT/BR 2013 / 000506
22 NOV 2013 22 11 13
Nome do Organismo receptor e "Pedido internacional PCT"
Referência do processo do requerente ou do mandatário (facultativo) (12 caracteres no máximo) 12AUIN019

Quadro No. I TÍTULO DA INVENÇÃO	
MÉTODO PARA DETERMINAÇÃO DA VIABILIDADE E QUALIDADE DE EMBRIÕES	
Quadro No. II REQUERENTE <input type="checkbox"/> Esta pessoa é também inventor.	
Nome e endereço: (Apelido seguido do nome próprio; no caso de uma pessoa jurídica a designação oficial completa. O endereço deve incluir o código postal e o nome do país. O país do endereço indicado neste quadro é o Estado de domicílio do requerente se não for indicado nenhum domicílio abaixo).	No. de telefone +55 11 3393-7903
Universidade Estadual Paulista "Júlio de Mesquita Filho" CNPJ: 48.031.918/0001-24 Rua Quirino de Andrade, 215, Centro, São Paulo/SP, Brasil CEP: 01049-010	No. de fax
	No. do registo do requerente junto do Organismo
Autorização relativa ao correio eletrônico: o facto de marcar uma das casas abaixo autoriza o Organismo receptor, a Autoridade responsável pela pesquisa internacional, a Secretaria Internacional e a Autoridade responsável pelo exame preliminar internacional, a utilizar o endereço de correio eletrónico indicado neste quadro para enviar, se tal desejarem esses organismos, notificações relativas a este pedido internacional.	
<input checked="" type="checkbox"/> como notificações preliminares seguidas por notificações em papel; ou Correio eletrónico: <u>auin@unesp.br</u>	<input type="checkbox"/> exclusivamente em forma eletrónica (não será enviada nenhuma notificação em papel)
Nacionalidade (siglas do país): BR	Domicílio (siglas do país): BR
Esta pessoa é requerente para: <input checked="" type="checkbox"/> todos os Estados designados <input type="checkbox"/> os Estados indicados no quadro suplementar	
Quadro No. III OUTRO(S) REQUERENTE(S) E/OU OUTRO(S) INVENTOR(ES)	
<input checked="" type="checkbox"/> Outros requerentes e/ou outros inventores são indicados numa folha de continuação.	
Quadro No. IV MANDATÁRIO OU REPRESENTANTE COMUM; OU ENDEREÇO PARA A CORRESPONDÊNCIA	
A pessoa identificada a seguir é/foi nomeada para actuar em nome do(s) requerente(s) perante as autoridades internacionais competentes, como: <input checked="" type="checkbox"/> mandatário <input type="checkbox"/> representante comum	
Nome e endereço: (Apelido seguido do nome próprio; no caso de uma pessoa jurídica, a designação oficial completa. O endereço deve incluir o código postal e o nome do país).	No. de telefone +55 11 3393-7903
Fabiola de Moraes Spindorelio Rua Dr. Bento Teobaldo Ferraz, 271, Bloco 2, Barra Funda - São Paulo/SP, Brasil CEP:01140-070	No. de fax +55 11 3393-7909
	No. do registo do mandatário junto do Organismo
Autorização relativa ao correio eletrônico: o facto de marcar uma das casas abaixo autoriza o Organismo receptor, a Autoridade responsável pela pesquisa internacional, a Secretaria Internacional e a Autoridade responsável pelo exame preliminar internacional, a utilizar o endereço de correio eletrónico indicado neste quadro para enviar, se tal desejarem esses organismos, notificações relativas a este pedido internacional.	
<input checked="" type="checkbox"/> como notificações preliminares seguidas por notificações em papel; ou Correio eletrónico: <u>fabiola@reitoria.unesp.br</u>	<input type="checkbox"/> exclusivamente em forma eletrónica (não será enviada nenhuma notificação em papel)
<input type="checkbox"/> Endereço para a correspondência: marcar esta casa se não for/tiver sido nomeado nenhum mandatário ou representante comum e o espaço acima for utilizado para indicar um endereço especial para o qual deve ser enviada a correspondência.	



Formulário PCT/RO/101 (primeira folha) (16 de Setembro de 2012)

Ver as notas sobre a formulário de requerimento

4.3 LOGO PROPOSTO PARA O SOFTWARE

Visando a distribuição da metodologia desenvolvida como um produto final o **logo**, **nome** e **subtítulo** abaixo foram elaborados como identidade visual para o software final.



4.4 ALGORITMOS DESENVOLVIDOS

```

function [ inputs ] = gerainputs( formato, display )
%Função mestre, lê todas as imagens do diretório e cria matriz de saída
%para a RNA
%formato = extensão do arquivo (ex: '*.jpg')
%display: 0=silencioso , 1=mostra
if display == 0
tic
end

list = dir(formato);%carrega lista dos arquivos no diretório atual

%pega tamanho da lista
tamanho = size(list);
tamanho = tamanho(1,1);

disp ('Iniciando processamento das imagens do diretório...')
for i=1:tamanho;
mostra = [num2str(round(i*1000/tamanho)/10),' %'];
disp(mostra)

%Padroniza imagem
I = padroniza(list(i).name); %carrega nome do arquivo na posição i
if display == 1
subplot(2,2,1)
imshow(I)
end

%Isola Embrião
[ Mplus,Mminus, Bplus, Bminus ] = isola5(I);
if display == 1
subplot(2,2,2)
imshow(Mplus)
subplot(2,2,3)
imshow(Mminus)
subplot(2,2,4)
imshow(Bplus)
pause
end

%Extrai variáveis
[ contrast, correlation, energy, homo] = grayco2(Mplus);
[ C1, C2, C3 ] = houghVar( Mminus );
[ raio, soma] = props2( Bplus );

%Extrai ID do nome do arquivo
[~, name, ~]= fileparts(list(i).name);
name = strsplit(name, '_');
ID = name(1);

inputs(i,:) = [ID,contrast, correlation, energy, homo, C1, C2, C3, raio, soma];

end

if display == 0
toc
end

end

```

```

function [ targets ] = geratargets(lab,formato)
%Gera a matriz de saída para a RNA
%lab define o laboratório de origem:
%hunt = Clínica Huntington
%invitro = InVitro Brasil
%engravidada = Clínica Engravidada
%
%formato define a extensão das imagens (ex: '*.jpg')

list = dir(formato);%carrega lista dos arquivos no diretório atual
%pega tamanho da lista
tamanho = size(list);
tamanho = tamanho(1,1);
if tamanho == 0
    disp('Não há arquivos no diretório atual')
    disp('(Verifique a extensão utilizada!')
    return
end

%%%%% HUNTINGTON %%%%%
if strcmp (lab,'hunt') == 1
    disp('Gerando targets para arquivos da Huntington')
    disp(' ')
    disp('A = Excelente, Bom, Regular, Ruim')
    disp('B = Código de duas letras (Ex: AA, BC, AC etc)')
    tipo = input('Indique o tipo de imagem: ','s') ;
    disp(tipo)

    if strcmp (tipo, 'A') == 1
        for i = 1:tamanho
            [~, name, ~]= fileparts(list(i).name);
            target = strsplit(name, '_');
            targets(i,1) = target(1);
            targets(i,2) = target(2);
            if strcmp (target(3), 'EXCELENTE')
                targets(i,3) = {'1'};
            end
            if strcmp (target(3), 'BOM')
                targets(i,3) = {'2'};
            end
            if strcmp (target(3), 'REGULAR')
                targets(i,3) = {'3'};
            end
            if strcmp (target(3), 'RUIM')
                targets(i,3) = {'4'};
            end
            targets(i,4) = target (4);

        end
    end

    if strcmp (tipo, 'B') == 1

        for i = 1:tamanho
            [~, name, ~]= fileparts(list(i).name);
            target = strsplit(name, '_');
            targets(i,1) = target(1);
            targets(i,2) = target(2);
            if strcmp (target(3), 'A')
                targets(i,3) = {'1'};
            end
            if strcmp (target(3), 'B')
                targets(i,3) = {'2'};
            end
            if strcmp (target(3), 'C')
                targets(i,3) = {'3'};
            end
        end
    end
end

```

```

        end

        if strcmp (target(4), 'A')
            targets{i,4} = {'1'};
        end
        if strcmp (target(4), 'B')
            targets{i,4} = {'2'};
        end
        if strcmp (target(4), 'C')
            targets{i,4} = {'3'};
        end

        targets(i,5) = target (5);

    end

end

if strcmp (tipo, 'A') == 0 && strcmp (tipo, 'B') == 0
    disp('Por favor selecione um tipo válido de imagem')

end

end

end

%%%%% INVITRO %%%%%
if strcmp (lab,'invitro') == 1
    disp('Gerando targets para arquivos da InVitro')

    for i = 1:tamanho
        [~, name, ~]= fileparts(list(i).name);
        target = strsplit(name, '_');
        targets(i,1) = target(1);
        targets(i,2) = target(2);
        targets(i,3) = target(3);

    end

end

end

%%%%% ERRO %%%%%
if strcmp (lab,'hunt') == 0 && strcmp (lab,'invitro') == 0
    disp('Laboratório não cadastrado, escolha uma das opções:')
    disp('hunt = Clínica Huntington')
    disp('invitro = InVitro Brasil')
end

end

```

```
function [ contrast, correlation, energy, homo] = grayco2(crop)
%Extrai informações de gray-level co-occurrence matrix

glcm = graycomatrix(crop);

stats = graycoprops(glcm, {'Contrast', 'Correlation', 'Energy', 'Homogeneity'});
contrast = stats.Contrast;
correlation = stats.Correlation;
energy = stats.Energy;
homo = stats.Homogeneity;

end
```

```
function [ C1, C2, C3 ] = houghVar( I )

%%Variáveis com Hough Transform%%
%Estruturas circulares médias

% imshow(I)
% pause

[centro, raio] = imfindcircles(I,[8 15], 'ObjectPolarity','dark', 'Sensitivity',0.94 );
% h=viscircles(centro, raio,'EdgeColor','b');
% pause
qntd = size(raio);
C1 = qntd(1);

%Estruturas circulares grandes
[centro, raio] = imfindcircles(I,[15 25], 'ObjectPolarity','dark', 'Sensitivity',0.94 );
% delete(h)
% h=viscircles(centro, raio,'EdgeColor','b');
% pause
qntd = size(raio);
C2 = qntd(1);

%Estruturas circulares gigantes
[centro, raio] = imfindcircles(I,[45 60], 'ObjectPolarity','dark', 'Sensitivity',0.98 );
% delete(h)
% h=viscircles(centro, raio,'EdgeColor','b');
qntd = size(raio);
C3 = qntd(1);

end
```

```

function [ Mplus,Mminus, Bplus, Bminus ] = isola5( I )
%Função de isolamento
%A ser utilizada após a função de padronização
%I deve ser em grayscale, 640x480px

original = I;
% imshow(I)
% pause

I = imgradient(I);
I = I/max(max(I));
I = imadjust(I);
% imshow(I)
% pause

I = im2bw(I, 0.5);
B = I;
% imshow(I)
% pause

% %Detecta círculo pequeno
[centroP, raioP, metricP] = imfindcircles(I,[100 150], 'ObjectPolarity', 'bright', 'Sensitivity', 0.99 );
centroP = centroP(1,:);
raioP = raioP(1);
metricP = metricP(1);

%Detecta círculo Grande
[centroG, raioG, metricG] = imfindcircles(I,[150 250], 'ObjectPolarity', 'bright', 'Sensitivity', 0.99 );
centroG = centroG(1,:);
raioG = raioG(1);
metricG = metricG(1);

%Compara metrics
if metricP > metricG
    raio = raioP;
    centro = centroP;
else
    raio = raioG;
    centro = centroG;
end

%Raio final
raioPlus = raio*1.08; %aumenta para pegar toda a zona pelúcida !
raioMinus = raio*0.8; %diminu para eliminar TE

%Plota no segundo quadrante
% viscircles(centro, raio, 'EdgeColor', 'b');
% pause

%%%% raioPlus %%%
%Cria máscara
centro = round(centro);
raio = round(raioPlus);
% disp(centro)
% disp(raio)

mask=zeros(size(I));
mask(centro(2),centro(1))=1;
Mplus = bwdist(mask) <= raio;

% %Aplica máscara
Mplus = immultiply(Mplus,original);
% imshow(Mplus)
% pause

%Faz o crop (xmin ymin width height)
x = centro(1);

```

```
y = centro(2);
Mplus = imcrop(Mplus,[x-raio y-raio 2*raio 2*raio]);
Bplus = imcrop(B,[x-raio y-raio 2*raio 2*raio]);
% imshow(Mplus)
% pause

%%% raioMinus %%%
% Cria máscara
centro = round(centro);
raio = round(raioMinus);
% disp(centro)
% disp(raio)

mask=zeros(size(I));
mask(centro(2),centro(1))=1;
Mminus = bwdist(mask) <= raio;

% %Aplica máscara
Mminus = immultiply(Mminus,original);
% imshow(Mminus)
% pause

% Faz o crop (xmin ymin width height)
x = centro(1);
y = centro(2);
Mminus = imcrop(Mminus,[x-raio y-raio 2*raio 2*raio]);
Bminus = imcrop(B,[x-raio y-raio 2*raio 2*raio]);
% imshow(Mminus)
% pause

end
```

```

function [ I ] = padroniza( filename )
%Padroniza as Imagens

%Carrega Imagem
I = imread(filename);

% imshow(I);
% pause

%Converte em escala de cinza
I = rgb2gray (I);

%Pega tamanho e proporção da imagem
sizeI = size(I); %Atenção: [Y(vertical) X(horizontal)]
ratio = sizeI(1)/sizeI(2);

%Corrige proporção da imagem
if ratio ~= 0.75
    NewSizeY = sizeI(2)*0.75; %multiplica o valor de X da imagem original pela proporção correta, obtendo o novo valor de Y
    NewSizeX = sizeI(2);

    dif = sizeI(1) - NewSizeY; %Pega a diferença de tamanho, para fazer o crop centralizado

    I = imcrop(I, [0 dif NewSizeX NewSizeY+dif]); %realiza o crop, centralizado em Y
end

%Tamanho após correção de proporção
sizeI = size(I);

%Corrige tamanho se necessário
if sizeI(1) ~= 480
    I = imresize (I, [480 640]);
end

%Ajuste de histograma
I = imadjust (I); %1% of data is saturated at low and high intensities

% imshow(I);
% pause

end

```

```
function [ raio, area, convex, eccen] = props( B )
% Utiliza como entrada a saída da isola5.m

%Variáveis diretas do crop
raio = size(B);
raio = round(raio(1)/2);

%Cria imagem binária, sem parte externa branca
mask=zeros(size(B));
mask(raio, raio)=1;
M = bwdist(mask) <= raio-1;
% imshow(M)
% pause
% imshow(B)
% pause
M = immultiply(M,B);
% imshow(M)
% pause

propsM = regionprops(M, 'Area','ConvexArea','Eccentricity','EulerNumber');
area = propsM.Area;
convex = propsM.ConvexArea;
eccen = propsM.Eccentricity;

end
```

```
function [ raio, soma] = props2( B )
% Utiliza como entrada a saída da isola5.m

%Variáveis diretas do crop
raio = size(B);
raio = round(raio(1)/2);

%Cria imagem binária, sem parte externa branca
mask=zeros(size(B));
mask(raio, raio)=1;
M = bwdist(mask) <= raio-1;
% imshow(M)
% pause
% imshow(B)
% pause
M = immultiply(M,B);
% imshow(M)
% pause

soma = sum(sum(M));

% propsM = regionprops(M, 'Area','ConvexArea','Eccentricity','EulerNumber');
% area = propsM.Area;
% convex = propsM.ConvexArea;
% eccen = propsM.Eccentricity;

end
```

```

function [ inputs ] = salvaCrops( formato, display )
%Salva crops do diretório atual
%CUIDADO COM OVERWRITE
%formato = extensão do arquivo (ex: '*.jpg')
%display: 0=silencioso , 1=mostra

if display == 0
tic
end

list = dir(formato);%carrega lista dos arquivos no diretório atual

%pega tamanho da lista
tamanho = size(list);
tamanho = tamanho(1,1);

disp ('Iniciando processamento das imagens do diretório...')
for i=1:tamanho;
mostra = [num2str(round(i*1000/tamanho)/10), ' %'];
disp(mostra)

%Padroniza imagem
I = padroniza(list(i).name); %carrega nome do arquivo na posição i

if display == 1
subplot(1,3,1)
imshow(I)
end

%Isola Embrião
[ M, B ] = isola5(I);

if display == 1
subplot(1,3,2)
imshow(M)
subplot(1,3,3)
imshow(B)
end

% if display == 1
% pause
% end

%Escreve arquivo do crop
cmap = colormap('gray');
filename = [list(i).name, 'CROP.jpg'];
imwrite(M,cmap,filename,'jpg','Quality',100)

end

if display == 0
toc
end

end

```

```

function [ finalRNA, erroMinList, erroMaxList, maiorErroList ] = matrixRNA( qntd,min,max,inputs, targets, transferList,trainList,
fitPercent, mutation, epochs,hidden )
%qntd = tamanho da população
%min = número mínimo de neuronios
%max = número máximo de neurônios
%inputs = entradas para a RNA
%targets = alvos para a RNA. Lembrando que o erro é medido por confusion
%transferList = lista de funções de transferência
%trainList = lista de funções de treinamento
%fitPercent = Porcentagem dos mais aptos, para seleção (Padrão = 0.5)
%mutation = taxa de mutação (Padrão = 0.05)
%hidden = quantidade máxima de Hidden Layers
tic

%Inicia operações
initialPopulation = geraPopulation( qntd,min,max,transferList,trainList,hidden ); %Função que gera a população inicial

nextPopulation = initialPopulation;

%Inicia menor erro como 1 (erro máximo)
menorErro = 1;

%Cria Waitbar
waitwindow = waitbar(0,'Iniciando população...');

%Inicia épocas
for i = 1:epochs

%Atualiza waitbar
dispwait = [num2str(round((i/epochs)*100)),'% Concluído. | Menor Erro: ',num2str(menorErro)];
waitbar(i/epochs,waitwindow,dispwait)

mostra = ['Iniciando época ', num2str(i)];
disp(mostra)
%Cria as RNAs de acordo com as Strings, e testa com os inputs e targets
Population = testaPopulation( nextPopulation, inputs, targets, transferList, trainList );

%Survival of the fittest !
%Seleciona apenas os mais aptos para a próxima iteração
%porcentagem determinada por fitPercent (padrão = 0.5)
[ newPopulation, erroMax, erroMin, maiorErro ] = selecionaPopulation(Population, fitPercent);

%Repovoamento
nextPopulation = replicaPopulation( newPopulation, mutation, transferList, trainList);

erroMinList(i) = erroMin;
erroMaxList(i) = erroMax;
maiorErroList(i) = maiorErro;

%Pega RNA com menor erro

for j = 1:qntd
erro = Population{j}.erro;
if erro <= menorErro
finalRNA = Population{j};
menorErro = erro;
end
end

%Gera gráfico dos erros
h = plot(erroMinList,-'ob'); %Plot do menor erro
hold all;
h = plot(erroMaxList,'-og'); %Plot do erro de corte
hold all;

```

```

h = plot(maiorErroList, '-or'); %Plot do erro máximo
hold off;
h = legend('Menor erro','Erro de corte', 'Maior erro');
axis([1 epochs 0 1])

```

```

pause (0.1)%Pausa para gerar o gráfico

```

```

%Mostra arquitetura da rede com menor erro
%Obs: Objetos Java !
jframe(i) = view(finalRNA.net);
jframe(i).setLocation(0,0)
jframe(i).setTitle('Arquitetura da melhor rede')

```

```

if i > 1 %Pra funcionar no i=1
%Fecha view anterior
jframe(i-1).setVisible(false);
jframe(i-1).dispose();
end

```

```

%Fim de todas as épocas
end

```

```

%Fecha waitbar
close(waitwindow)

```

```

%Mostra confusion da melhor rede (Dados de TESTE)
testTargets = targets .* finalRNA.tr.testMask{1};
outputs = finalRNA.net(inputs);
figure,plotconfusion(testTargets,outputs);

```

```

%Mostra plot de performance da melhor rede
figure,plotperform(finalRNA.tr)

```

```

%Mostra resultados finais
clc

```

```

qntdLayersFinal = size(finalRNA.net.layers);
qntdLayersFinal = qntdLayersFinal(1); %Pega a quantidade de camadas totais

```

```

mostra = ['Quantidade de camadas ocultas: ', num2str(qntdLayersFinal-1)];
disp(mostra) %Quantidade de camadas ocultas = Total - 1
disp(' ')

```

```

for k = 1:qntdLayersFinal %Mostra dados de cada camada
Funcao = finalRNA.net.layers{k}.transferFcn;
mostra = ['Camada ', num2str(k), ':'];
disp(mostra)
mostra = ['Função de transferência: ', Funcao];
disp(mostra)
sizeLayer = finalRNA.net.layers{k}.size;
mostra = ['Quantidade de neurônios: ', num2str(sizeLayer)];
disp(mostra)
disp(' ')
end

```

```

disp(' ')
mostra = ['Menor erro: ', num2str(menorErro)];
disp(mostra) %Mostra o erro da melhor rede de todas as épocas

```

```

disp(' ')
%toc;
mostra = ['Tempo total de processamento: ', num2str(toc), ' segundos.'];

```

```
disp(mostra)
%Fim da função
end
```

```
function [ Population ] = geraPopulation( qntd,min,max,transferList,trainList,hidden )
%Gera uma população de Strings aleatórias
for i = 1:qntd
    Population{i}.String = geraString(min,max,transferList,trainList,hidden );
    Population{i}.sobrevive = 1; %Coloca todos como sobreviventes, pois é a primeira época !
end

end
```

```

function [ net, tr, erro, erroVal, erroTrain ] = GeraRNA( inputs, targets, string, transferList, trainList )
%Sequencia de Vars da String:
%1)Quantidade de Neurônios na camada oculta
%2)Função de Transferência na camada oculta
%3)Função de Transferência na camada de saída
%(logsig, purelin, tansig, hardlim, tribas, poslin, radbas)
%4)Função de Treinamento (traingdx, trainscg, traingdm)

qntdN = string(1); %Número de neurônios
transferHidden = transferList(string(2)); %Função de transferência Hidden Layers
transferOutput = transferList(string(3)); %Função de transferência Output Layer
trainFun = trainList(string(4));
qntdHidden = string(5); %Quantidade de Hidden Layers

%Cria a estrutura de rede, com a quantidade de camadas
% Obs: A quantidade de neurônios (qntdN) é dividida igualmente entre as camadas ocultas
if qntdHidden == 1
net = patternnet(qntdN);
end

if qntdHidden == 2
q = round(qntdN/qntdHidden);
net = patternnet([q,q]);
end

if qntdHidden == 3
q = round(qntdN/qntdHidden);
net = patternnet([q,q,q]);
end

if qntdHidden > 3
disp('Não recomendado o uso de mais do que 3 camadas ocultas !')
disp('Utilizando apenas 3 camadas ocultas...')
q = round(qntdN/qntdHidden);
net = patternnet([q,q,q]);
end

% Choose Input and Output Pre/Post-Processing Functions
% For a list of all processing functions type: help nnprocess
net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
net.outputs{2}.processFcns = {'removeconstantrows','mapminmax'};

% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
net.divideFcn = 'dividerand'; % Divide data randomly
net.divideMode = 'sample'; % Divide up every sample
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

%Escolhe função de treinamento
net.trainFcn = trainFun;

% Choose a Performance Function
net.performFcn = 'mse'; % Mean squared error

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
%net.plotFcns = {'plotperform','plottrainstate','ploterrhist','plotregression','plotfit'};

%Escolhe funções de transferência
%Obs: depende de quantas camadas foram escolhidas !
if qntdHidden == 1
net.layers{1}.transferFcn = transferHidden;
net.layers{2}.transferFcn = transferOutput;

```

```

end

if qntdHidden == 2
net.layers{1}.transferFcn = transferHidden;
net.layers{2}.transferFcn = transferHidden;
net.layers{3}.transferFcn = transferOutput;
end

if qntdHidden >= 3
net.layers{1}.transferFcn = transferHidden;
net.layers{2}.transferFcn = transferHidden;
net.layers{3}.transferFcn = transferHidden;
net.layers{4}.transferFcn = transferOutput;
end

%Não mostra treinamento
net.trainParam.showWindow = 0;
net.trainParam.showCommandLine = 0;

% Train the Network
[net,tr] = train(net,inputs,targets);

% Test the Network
outputs = net(inputs);
%errors = gsubtract(targets,outputs)
%performance = perform(net,targets,outputs);

% Recalculate Training, Validation and Test Performance
trainTargets = targets .* tr.trainMask{1};
valTargets = targets .* tr.valMask{1};
testTargets = targets .* tr.testMask{1};
%trainPerformance = perform(net,trainTargets,outputs)
%valPerformance = perform(net,valTargets,outputs)
%testPerformance = perform(net,testTargets,outputs);

%Calcula ERRO
erro = confusion(testTargets, outputs);
erroVal = confusion(valTargets, outputs);
erroTrain = confusion(trainTargets,outputs);

% Plots Confusion dos dados de Teste
%figure, plotconfusion(testTargets,outputs)

end

```

```
function [ string ] = geraString(min,max,transferList,trainList,hidden )
% gera a string aleatória para a função geraRNA

%Valor aleatório para número de neurônios, entre o máximo e mínimo
string(1) = randi([min,max]);

%Escolhe funções de transferência
sizeTransfer = size(transferList);
sizeTransfer = sizeTransfer(2); %Pega tamanho da lista, para randomizar
string(2) = randi([1,sizeTransfer]); %Aleatório para camadas ocultas
string(3) = randi([1,sizeTransfer]); %Aleatório para camda de saída

%Escolhe função de treinamento
sizeTrain = size(trainList);
sizeTrain = sizeTrain(2); %Tamanho da lista
string(4) = randi([1,sizeTrain]); %Escolhe aleatório

%Escolhe número de Hidden Layers
string(5) = randi([1,hidden]);

end
```

```

function [ newPopulation ] = replicaPopulation( Population, mutation, transferList, trainList)
%Repopoamento utilizando apenas os sobreviventes de "selecionaPopulation.m"

%Pega o tamanho total
sizePopulation = size(Population);
sizePopulation = sizePopulation(2);

for i = 1:sizePopulation

    %Encontra vaga. Se o individuo estava morto, será substituído
    %(a quantidade total de indivíduos não varia !)
    if Population{i}.sobrevive == 0
        %Seleciona progenitores
        p = 1;
        while p <= 2
            r = randi([1 sizePopulation]);
            if Population{r}.sobrevive == 1
                Progenitor{p} = Population{r};
                p = p+1;
            end
        end

        pai = Progenitor{1};
        pai = pai.String;
        mae = Progenitor{2};
        mae = mae.String;

        %pega tamanho da string
        tamanhoString = size(pai);
        tamanhoString = tamanhoString(2);

        %Percorre a String realizando o cruzamento
        for k = 1:tamanhoString
            doador = randi([1 2]); % 1=pai, 2=mae
            if doador == 1
                newString(k) = pai(k);
            else
                newString(k) = mae(k);
            end
        end

        %Gerador de mutações
        r = rand;
        if r < mutation
            infoMutation = ['Mutação no gene ', num2str(k), ' do indivíduo ', num2str(i)];
            disp(infoMutation) %Mostra quando e com quem ocorreu uma mutação !

            %Mutações específicas para cada tipo de gene
            if k == 1 %Mutação no número de neurônios
                UpOrDown = randi([-10 10]); %Pode variar de -10 até +10
                newString(k) = newString(k) + UpOrDown;
                if newString(k) <= 0
                    newString(k) = 1; %Evita quntd negativa de neurônios !
                end
            end

            if k == 2 %Mutação na função de transferencia da 1a camada
                sizeTransfer = size(transferList);
                sizeTransfer = sizeTransfer(2);
                randGene = randi ([1 sizeTransfer]);
                newString(k) = randGene;
            end

            if k == 3 %Mutação na função de transferencia da 2a camada
                sizeTransfer = size(transferList);
                sizeTransfer = sizeTransfer(2);
                randGene = randi ([1 sizeTransfer]);
                newString(k) = randGene;
            end
        end
    end
end

```

```

end

if k == 4 %Mutaçao na função de treinamento
sizeTrain = size(trainList);
sizeTrain = sizeTrain(2);
randGene = randi ([1 sizeTrain]);
newString(k) = randGene;
end

%Fim das mutações
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

Population{i}.sobrevive = 1; %Coloca o novo indivíduo como vivo
Population{i}.String = newString; %Finaliza o cruzamento
%Final do cruzamento
end

%Toda a população percorrida
end

newPopulation = Population; %Gera população final
end

```

```

function [ newPopulation, erroMax, erroMin, maiorErro ] = selecionaPopulation( Population, fitPercent )
%Survival of the fittest !
% fitPercent = fração que sobrevive para próxima iteração (ex: 0.2 = 20%)
sizePopulation = size(Population);
sizePopulation = sizePopulation(2);

for i = 1:sizePopulation
    erros(i) = Population{i}.erro;
end
erros = sort(erros); %alinha erros em ordem crescente
% bar(erros)
% pause
qntdSobrevive = size(erros);
qntdSobrevive = qntdSobrevive(2);
maiorErro = erros(qntdSobrevive); %Aqui ainda está a população inteira "viva"

%Pega a quantidade máxima de sobreviventes
qntdSobrevive = round(qntdSobrevive * fitPercent); %erros que passam no fit.
%round para pegar uma posição inteira do vetor !

%Erro máximo é o erro do último da lista de sobreviventes
erroMax = erros(qntdSobrevive); %erro de threshold para proxima geração
%Erro mínimo é do primeiro da lista
erroMin = erros(1);

%Seleciona os aptos
for i= 1:sizePopulation

    if Population{i}.erro > erroMax
        Population{i}.sobrevive = 0; %Zero significa morto
    else
        Population{i}.sobrevive = 1; %Um significa vivo
    end

end

newPopulation = Population; %Próxima geração feita
end

```

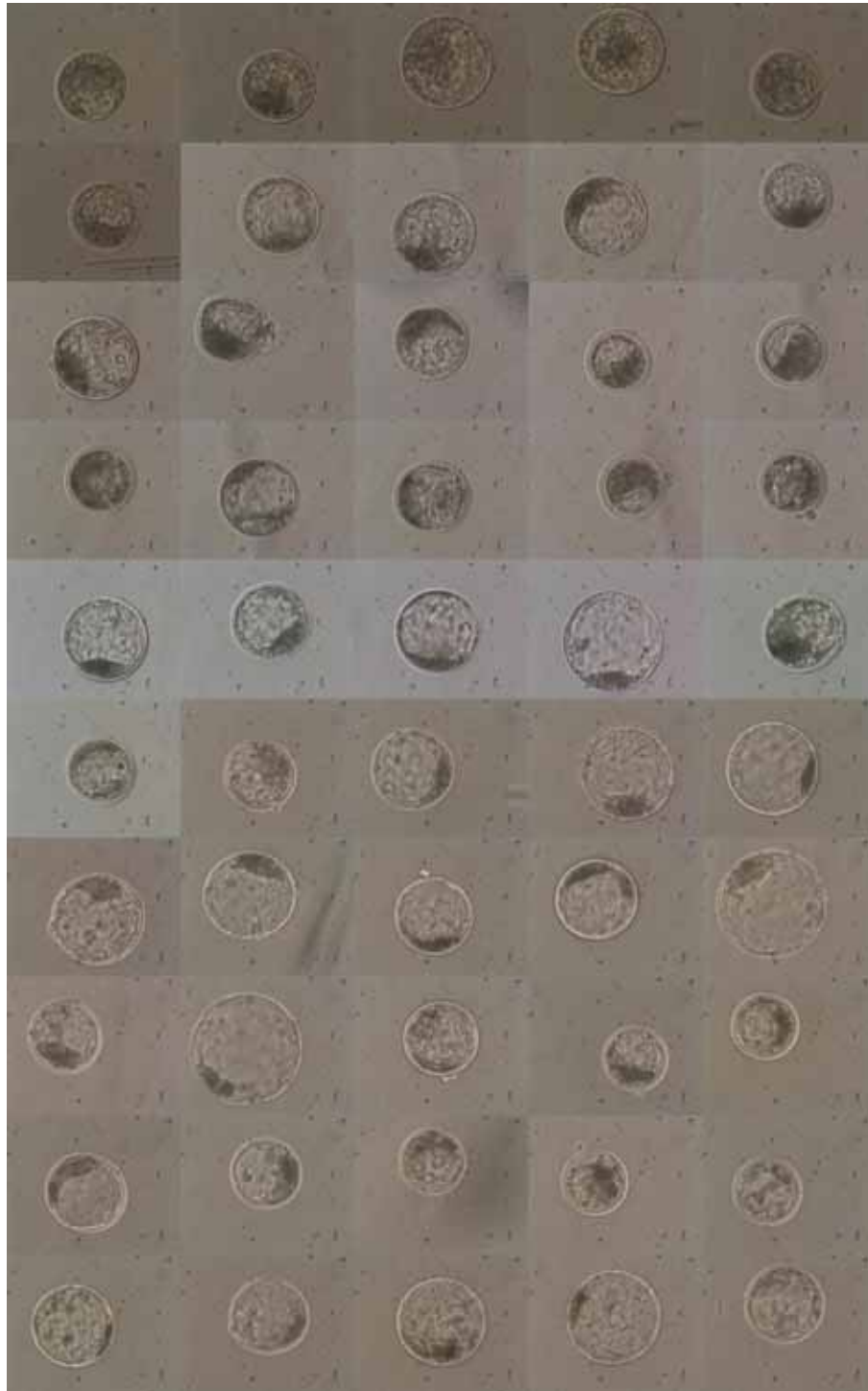
```
function [ Population ] = testaPopulation( Population, inputs, targets, transferList, trainList )

%Pega tamanho da população
tamanho = size(Population);
tamanho = tamanho(2);

for i = 1:tamanho
    percent = [num2str((i/tamanho)*100) , '%'];
    disp(percent) %Mostra porcentagem concluída
    string = Population{i}.String;
    [ net, tr, erro, erroVal, erroTrain ] = GeraRNA( inputs, targets, string, transferList, trainList );
    Population{i}.net = net;
    Population{i}.tr = tr;
    Population{i}.erro = erro;
    Population{i}.erroVal = erroVal;
    Population{i}.erroTrain = erroTrain;
end

end
```

4.5 AMOSTRA DO BANCO DE DADOS



5 DISCUSSÃO GERAL

O desenvolvimento de um sistema capaz de atuar de maneira automática na interpretação e classificação de embriões se demonstra um grande desafio, não existindo até o momento nenhuma metodologia que atinja resultados equivalentes. O trabalho mais semelhante ao nosso avalia apenas embriões humanos, por meio de um método semiautomático (SANTOS FILHO et al., 2012), o que apresenta limitações para a sua execução na prática laboratorial.

A análise de embriões bovinos apresentou resultados promissores, com uma taxa de acerto global de **79,2%** na classificação de blastocistos bovinos em 3 graus de qualidade, sendo que para os blastocistos classificados como Excelentes ou Bons (Classe 1) a taxa de acerto é de **82,6%**, para os blastocistos classificados como Regulares (Classe 2) é de **16,7%** e para os blastocistos classificados como Pobres (Classe 3) a taxa de acerto é de **91,7%**. A função de *performance* do treinamento da RNA se apresenta estritamente decrescente durante todas as iterações, indicando que a função de treinamento atuou de maneira efetiva. O caráter automatizado da metodologia desenvolvida permite a sua real implementação em um ambiente laboratorial, permitindo que o software atue de maneira efetiva como ferramenta auxiliar na avaliação morfológica de blastocistos.

A análise dos resultados (Figura 24 e Figura 25) torna evidente uma maior taxa de acerto para as classes 1 e 3 (82,6 e 91,7%, respectivamente). Embora a taxa global de acerto seja satisfatória (79,2%) tal fato ocorre apenas devido ao reduzido número de

embriões classe 2 sorteados para o grupo de teste. Caso essa quantidade fosse maior, certamente a taxa de acerto global diminuiria.

Dois principais fatores podem ocasionar a baixa taxa de acerto da classe 2: a ausência de variáveis que possam distinguir corretamente esta classe das demais ou a classificação errônea do banco de dados. A extração de variáveis foi exaustivamente desenvolvida, até o ponto em que várias delas apresentaram alto grau de colinearidade. Portanto, acreditamos que uma reavaliação do banco de dados possa vir a aumentar a eficácia de análise da classe 2 pela RNA. Para analisar essa alternativa, foi realizada a reavaliação de algumas imagens que a RNA julgou com altos valores para as classes 1, 2 ou 3. Os resultados da reavaliação podem ser encontrados na Tabela 6, na qual é possível perceber que a reavaliação concordou em apenas 50% com a avaliação original. Os erros de reavaliação foram menores para as classes 1 e 3 (33% para ambas) do que para a classe 2 (54,4% de discordância entre a reavaliação e a avaliação original). Tal resultado reforça a hipótese de que a baixa taxa de acerto da RNA para embriões de classe 2 seja devido a problemas de classificação do banco de dados.

Através da curva de *Receiver Operator Characteristic* (Figura 25), é possível confirmar que o aprendizado foi mais efetivo para as classes 1 e 3 do que para a classe 2. A linha verde, correspondente à classe 2, se aproxima mais da diagonal da curva ROC, demonstrando sua distribuição mais aleatória.

Tabela 6 Reavaliação de uma amostra do banco de dados e comparação com avaliação original. A coluna “Imagem” mostra o código da imagem escolhida do banco de dados, “Reav.” a classificação realizada na reavaliação, “Orig.” a classificação original utilizada no treinamento da RNA e “Classe” ou “C” indica a classificação fornecida pela RNA. As colunas 1, 2 e 3 indicam os valores de saída da RNA para as respectivas classes. As colunas de erros indicam 1 para erro e 0 para concordância.

Imagem	Reav.	Orig.	Classe	Saídas RNA			Erros		
				1	2	3	Reav X Orig	Reav X C	Orig X C
86	3	2	3	0,02	0,22	0,82	1	0	1
92	3	2	3	0,02	0,27	0,82	1	0	1
94	2	2	3	0,04	0,22	0,80	0	1	1
152	1	1	1	0,47	0,41	0,18	0	0	0
155	1	1	1	0,64	0,50	0,11	0	0	0
165	2	2	3	0,12	0,26	0,78	0	1	1
166	2	2	1	0,53	0,31	0,10	0	1	1
170	2	2	1	0,55	0,27	0,12	0	1	1
173	3	2	3	0,10	0,24	0,77	1	0	1
174	1	2	1	0,58	0,39	0,10	1	0	1
178	3	3	1	0,53	0,35	0,12	0	1	1
180	3	3	1	0,62	0,40	0,21	0	1	1
219	2	1	1	0,59	0,47	0,22	1	1	0
234	2	1	1	0,54	0,39	0,14	1	1	0
237	1	1	1	0,59	0,48	0,18	0	0	0
240	1	1	1	0,45	0,44	0,19	0	0	0
257	3	2	3	0,02	0,27	0,85	1	0	1
261	1	2	1	0,56	0,41	0,12	1	0	1
268	3	2	1	0,55	0,44	0,15	1	1	1
368	2	3	1	0,46	0,40	0,21	1	1	1
				Soma Erros			10	10	14
				Concordância			50%	50%	30%

A metodologia de Algoritmos Genéticos, desenvolvida para descoberta da melhor arquitetura de RNA se apresenta eficaz e prática, de modo que o problema de ser encontrada a melhor arquitetura de rede fica completamente a cargo do algoritmo. Este, por conta da sua estratégia evolutiva, converge para as melhores arquiteturas em um número relativamente baixo de iterações. Tal metodologia apresenta um caráter inovador, uma vez que foge de nosso conhecimento aplicações práticas de AG na descoberta de arquitetura de RNA em aplicações com dados reais.

É de nosso conhecimento apenas outra invenção registrada que atue na seleção embrionária (LOEWKE; SURAJ, 2014) que, no entanto, difere por utilizar um sistema de *timelapse* para a determinação da qualidade do embrião baseada na cinética e na simetria de suas clivagens. Entretanto, a sua implementação laboratorial implicaria em grandes gastos com equipamentos adicionais.

O nosso trabalho está protegido pelo pedido de proteção intelectual nacional BR102012031953-5 e internacional PCT-BR2013-000506, realizados juntamente com a Agência Unesp de Inovação (AUIN).

5.1 PERSPECTIVAS FUTURAS

Muito ainda pode ser feito quanto à validação e divulgação da metodologia apresentada, sendo que um aprimoramento contínuo é essencial para que esta tecnologia venha a se tornar a ferramenta auxiliar padrão na seleção embrionária. A reavaliação do banco de dados, por meio de pesquisas on-line (por exemplo, através do sistema “Surveymonkey”) e com especialistas da área, pode vir a sanar a baixa taxa de acerto obtida para a classe 2.

Também vislumbramos a possibilidade de desenvolvimento de um sistema semelhante, que analise diretamente a possibilidade de um embrião resultar em uma prenhez com sucesso. Entretanto, para que haja viabilidade de tal análise, seria necessária a incorporação de variáveis descritoras da fêmea receptora, que desempenha um importante papel no sucesso de uma gestação.

Em última instância, apesar de todas as tentativas de otimização do software, os resultados obtidos acabaram mimetizando o ser humano, ao menos em parte, na complexa atividade que é categorizar uma estrutura tridimensional baseada em uma imagem bidimensional. O fato de haver um déficit de acertos na categoria intermediária (blastocistos grau 2) e da RNA apresentar resultados expressivamente melhores nas categorias extremas (blastocistos de graus 1 ou 3) sugere que o software está se comportando tal qual um embriologista, tendo uma maior facilidade para classificar os casos extremos e apresentando uma possibilidade maior de erro ao se deparar com os graus intermediários de qualidade. Superar essa dificuldade é a etapa necessária para tornar a metodologia desenvolvida uma ferramenta prática e que possa auxiliar o embriologista em sua rotina laboratorial.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ATHERTON, T.; KERBYSON, D. Size invariant circle detection. **Image and Vision computing**, v. 17, n. February 1997, p. 795–803, 1999.

BALABAN, B. et al. Blastocyst quality affects the success of blastocyst-stage embryo transfer. **Fertility and sterility**, v. 74, n. 2, p. 282–7, ago. 2000.

BALLARD, D. Generalizing the Hough transform to detect arbitrary shapes. **Pattern recognition**, 1981.

BENVENUTO, N.; PIAZZA, F. On the complex backpropagation algorithm. **IEEE Transactions on Signal Processing**, v. 40, n. 4, p. 967–969, abr. 1992.

BÉNYEI, B. et al. The effect of internal and external factors on bovine embryo transfer results in a tropical environment. **Animal reproduction science**, v. 93, n. 3-4, p. 268–79, jul. 2006.

BEUCHER, S. The watershed transformation applied to image segmentation. **SCANNING MICROSCOPY-SUPPLEMENT-**, p. 1–26, 1992.

BÓ, G.; MAPLETOFT, R. Evaluation and classification of bovine embryos. **Anim. Reprod**, v. 54, n. 935, p. 344–348, 2013.

BROWN, C. D.; DAVIS, H. T. Receiver operating characteristics curves and related decision measures: A tutorial. **Chemometrics and Intelligent Laboratory Systems**, v. 80, n. 1, p. 24–38, jan. 2006.

COSNARD, M. A step towards the frontier between one-hidden-layer and two-hidden-layer neural networks. **Neural Networks, 1993**. ..., p. 2292–2295, 1993.

DE, K. Learning with Genetic Algorithms : An Overview. p. 121–138, 1988.

EGMONT-PETERSEN, M.; DE RIDDER, D.; HANDELS, H. Image processing with neural networks—a review. **Pattern Recognition**, v. 35, n. 10, p. 2279–2301, out. 2002.

FARIN, P. et al. Agreement among evaluators of bovine embryos produced in vivo or in vitro. **Theriogenology**, n. 95, p. 339–349, 1995.

GOETHALS, P. L. M. et al. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. **Aquatic Ecology**, v. 41, n. 3, p. 491–508, 9 maio 2007.

GONZALEZ, R. **Digital Image processing using MATLAB**. Upper Saddle River N.J.: Pearson Prentice Hall, 2004.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. [s.l.: s.n.]. p. 976

GURESEN, E.; KAYAKUTLU, G.; DAIM, T. U. Using artificial neural network models in stock market index prediction. **Expert Systems with Applications**, v. 38, n. 8, p. 10389–10397, ago. 2011.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural Features for Image Classification. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 3, n. 6, p. 610–621, nov. 1973.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation (2nd Edition)**. [s.l.] Prentice Hall, 1998. p. 842

HECHT-NIELSEN, R. Theory of the backpropagation neural network. **Neural Networks, 1989. IJCNN., International ...**, p. 593–605, 1989.

HELD, E. et al. Zona pellucida birefringence correlates with developmental capacity of bovine oocytes classified by maturational environment, COC morphology and G6PDH activity. **Reproduction, Fertility and Development**, v. 24, n. 4, p. 568, 2012.

HOSHI, H. In vitro production of bovine embryos and their application for embryo transfer. **Theriogenology**, v. 59, n. 2, p. 675–85, 15 jan. 2003.

HU, Y.; ZHAO, C.; WANG, H. Directional Analysis of Texture Images Using Gray Level Co-Occurrence Matrix. **2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application**, p. 277–281, dez. 2008.

JAIN, A.; MAO, J.; MOHIUDDIN, K. Artificial neural networks: A tutorial. **Computer**, v. 29, p. 31–44, 1996.

KÖRBES, A. **Análise de Algoritmos da Transformada Watershed**. [s.l.] Universidade Estadual de Campinas, 2010.

LI, E. Y. Artificial neural networks and their business applications. **Information & Management**, v. 27, n. 5, p. 303–313, nov. 1994.

LINDNER, G.; JR, R. W. Bovine embryo morphology and evaluation. **Theriogenology**, v. 20, n. 0313, p. 407–416, 1983.

LOEWKE, K. E.; SURAJ, V. **IN VITRO EMBRYO BLASTOCYST PREDICTION METHODS**, 2014.

- LÓPEZ-DAMIÁN E. P, GALINA C. S, MERCHANT H, CEDILLO-PELÁEZ C, A. M. Assessment of *Bos taurus* embryos comparing stereoscopic microscopy and transmission electron microscopy. **Journal of Cell and Animal Biology**, v. 2, n. 3, p. 72–78, 2008.
- MACHART, P.; RALAIVOLA, L. Confusion matrix stability bounds for multiclass classification. **arXiv**, n. Section 5, 2012.
- MITCHELL, M. Genetic algorithms: an overview. **Complexity**, 1995.
- O'BRIEN, R. M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. **Quality & Quantity**, v. 41, n. 5, p. 673–690, 13 mar. 2007.
- OTSU, N. A threshold selection method from gray-level histograms. **Automatica**, v. C, n. 1, p. 62–66, 1975.
- OVERSTRÖM, E. In vitro assessment of embryo viability. **Theriogenology**, v. 45, n. 95, p. 3–16, 1996.
- RAMLALL, I. Artificial intelligence: neural networks simplified. **International Research Journal of Finance and ...**, p. 1–20, 2009.
- ROBERTI DE SIQUEIRA, F.; ROBSON SCHWARTZ, W.; PEDRINI, H. Multi-scale gray level co-occurrence matrices for texture description. **Neurocomputing**, v. 120, p. 336–345, nov. 2013.
- ROCHA, C. et al. Utilização de redes neurais artificiais para a determinação do número de refeições diárias de um restaurante universitário Use of artificial neural networks to determine. **Revista de Nutrição**, v. 24, n. 5, p. 735–742, 2011.
- SANTOS FILHO, E. et al. A method for semi-automatic grading of human blastocyst microscope images. **Human reproduction (Oxford, England)**, v. 27, n. 9, p. 2641–8, set. 2012.
- SCHNEIDER, H. J.; CASTLEBERRY, R. S.; GRIFFIN, J. L. Commercial aspects of bovine embryo transfer. **Theriogenology**, v. 13, n. 1, p. 73–85, jan. 1980.
- TERVIT, H. et al. Non-surgical embryo transfer in cattle. **Theriogenology**, v. 13, n. 1, p. 63–71, 1980.
- WALCZAK, S.; CERPA, N. Heuristic principles for the design of artificial neural networks. **Information and Software Technology**, v. 41, n. 2, p. 107–117, jan. 1999.
- YAO, X.; LIU, Y. Towards designing artificial neural networks by evolution. **Applied Mathematics and Computation**, v. 91, n. 1, p. 83–90, abr. 1998.

YUEN, H. K. et al. A Comparative Study of Hough Transform Methods for Circle Finding. **Proceedings of the Alvey Vision Conference 1989**, p. 29.1–29.6, 1989.

ZHANG, G.; PATUWO, B. E.; HU, M. Forecasting with artificial neural networks:: The state of the art. **International journal of forecasting**, v. 14, p. 35–62, 1998.