

UNIVERSIDADE DE VILA VELHA – ES
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FARMACÊUTICAS

**CLASSIFICAÇÃO DO GRAU HISTOLÓGICO DE MALIGNIDADE EM
CÂNCER DE MAMA UTILIZANDO PROGRAMAS DE ANÁLISE
AUTOMATIZADA DE IMAGENS E APRENDIZADO DE MÁQUINA**

PAULO CÉSAR RIBEIRO BOASQUEVISQUE

VILA VELHA
ABRIL / 2020

UNIVERSIDADE DE VILA VELHA – ES
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FARMACÊUTICAS

**CLASSIFICAÇÃO DO GRAU HISTOLÓGICO DE MALIGNIDADE EM
CÂNCER DE MAMA UTILIZANDO PROGRAMAS DE ANÁLISE
AUTOMATIZADA DE IMAGENS E APRENDIZADO DE MÁQUINA**

Tese apresentada a Universidade Vila Velha,
como pré-requisito do Programa de Pós-
graduação em Ciências Farmacêuticas, para
a obtenção do título de Doutor em Ciências
Farmacêuticas.

PAULO CÉSAR RIBEIRO BOASQUEVISQUE

VILA VELHA
ABRIL / 2020

Catálogo na publicação elaborada pela Biblioteca Central / UVV-ES

B662c Boasquevisque, Paulo César Ribeiro
Classificação do grau histológico de malignidade em câncer de mama utilizando programas de análise automatizada de imagens e aprendizado de máquina / Paulo César Ribeiro Boasquevisque – 2020.
62 f.; il.

Orientador: Dominik Lenz.
Tese (Doutorado em Ciências Farmacêuticas), Universidade Vila Velha, 2020.
Inclui bibliografias.

1. Farmacologia. 2. Histologia. 3. Inteligência artificial. 4. Carcinoma de células escamosas. 5. Mamas – Câncer. I. Lenz, Dominik. II. Universidade Vila Velha. III. Título.

CDD 616.99249

PAULO CÉSAR RIBEIRO BOASQUEVISQUE

**CLASSIFICAÇÃO DO GRAU HISTOLÓGICO DE MALIGNIDADE EM
CÂNCER DE MAMA UTILIZANDO PROGRAMAS DE ANÁLISE
AUTOMATIZADA DE IMAGENS E APRENDIZADO DE MÁQUINA**

Tese apresentada a Universidade Vila Velha,
como pré-requisito do Programa de Pós-
graduação em Ciências Farmacêuticas, para
a obtenção do título de Doutor em Ciências
Farmacêuticas.

Aprovada em 07 de abril de 2020,

Banca Examinadora:



Prof. Dr. Elisardo Corral Vasquez (UVV -ES)



Prof. Dr. Manuel Campos Toimil (USC – Espanha)



Prof. Dr. Márcio Fronza (UVV – ES)



Prof. Dr. Fábio Ribeiro Braga (UVV – ES)



Prof. Dr. Dominik Lenz (UVV – ES)

Orientador

Porque se chamavam homens,
também se chamavam sonhos
e sonhos não envelhecem...

(Márcio Borges, Lô Borges, Milton Nascimento)

SUMÁRIO

LISTA DE QUADROS	vii
LISTA DE FIGURAS.....	viii
LISTA DE TABELAS	xi
RESUMO	xii
ABSTRACT	xiii
1 Introdução.....	1
2. Objetivos.....	5
3. Material e métodos	6
3.1 A amostra - Critérios de inclusão e exclusão	6
3.2 Critérios Éticos	8
3.3 Digitalização das lâminas histológicas	9
3.4 O software CellProfiler.....	9
3.5 Algoritmo CellProfiler	10
3.6 Etapas do Algoritmo CellProfiler	10
3.6.1 Etapa 1 - LoadImages	10
3.6.2 Etapa 2 - ColorToGray.	11
3.6.3 Etapa 3 - ImageMath.	14
3.6.4 Etapa 4 - ApplyThreshold.	16
3.6.5 Etapa 5 IdentifyPrimaryObjects.....	17
3.6.6 Etapa 6 MeasureObjectSizeShape.	19
3.6.7 Etapa 7 FilterObjects	21
3.6.8 Etapa 8 MeasureObjectSizeShape.	22
3.6.9 Etapa 9 ExportToDatabase.	24
3.7 O Software Tanagra.....	25
4 Análise estatística	27
5 Resultados e discussão	28
5.1 Dados Epidemiológicos.....	28
5.2 Artefatos em Histologia.....	37
5.3 Planilha da Base de Dados do software Tanagra.....	37

5.4 Parâmetros quantitativos (atributos) dos objetos primários	40
5.4.1 Principais atributos para classificação automatizada do score tubular no grau histológico em câncer de mama.....	41
5.4.2 Principais atributos para classificação automatizada do score nuclear no grau histológico em câncer de mama.	42
5.4.3 Principais atributos para classificação automatizada do score índice mitótico em grau histológico de câncer de mama.	44
5.4.4 Principais atributos para classificação automatizada do grau histológico de câncer de mama.	45
5.5 Distribuição dos casos de câncer de mama quanto aos scores tubular, nuclear e índice mitótico da planilha de base de dados analisada pelo software Tanagra	47
5.6 Classificação automatizada do score tubular	48
5.7 Classificação automatizada do score nuclear	51
5.8 Classificação automatizada do score índice mitótico	52
5.9 Classificação automatizada do grau histológico de malignidade	54
6 Conclusões	57
7 Referências bibliográficas	58

LISTA DE QUADROS

Quadro 1	Aspectos componentes do grau histológico em câncer de mama.....	2
Quadro 2	Grau histológico de malignidade em câncer de mama.....	2
Quadro 3	Critérios de inclusão na amostra do estudo.....	6
Quadro 4	Critérios de exclusão na amostra do estudo.....	7
Quadro 5	Algortimo CellProfiler.....	10
Quadro 6	Distribuição dos casos de câncer de mama analisados pelo software Tanagra conforme grau histológico.....	38
Quadro 7	Distribuição dos casos por pontuação no Score Tubular.....	49
Quadro 8	Estatística comparativa da classificação do score tubular dada pelo Software Tanagra com a do médico patologista.....	50
Quadro 9	Distribuição dos casos por pontuação no score nuclear.....	51
Quadro 10	Estatística comparativa da classificação do score nuclear dada pelo software Tanagra com a do médico patologista.....	52
Quadro 11	Distribuição dos casos por pontuação no score índice mitótico.....	52
Quadro 12	Estatística comparativa da classificação do score índice mitótico dada pelo software Tanagra com a do médico patologista.....	53
Quadro 13	Distribuição dos casos de câncer de mama analisados pelo software Tanagra conforme grau histológico.....	54
Quadro 14	Estatística comparativa da classificação do grau histológico de dada pelo software Tanagra com a do médico patologista.....	55

LISTA DE FIGURAS

Figura 1	Carcinoma in situ e aspecto histológico normal da mama (Fonte: anatpat.unicamp.br).....	8
Figura 2	Algoritmo CellProfiler Etapa 1 – LoadImages.....	11
Figura 3	Imagens digitalizadas de câncer de mama aumento 40 x.....	11
Figura 4	Algoritmo CellProfiler Etapa 2 – ColorToGray.....	12
Figura 5A	Imagens originais convertidas para espectro branco-cinza preto.....	12
Figura 5B	Imagem digitalizada de câncer de mama na coloração original (imagem magnificada).....	13
Figura 5C	Imagem digitalizada magnificada de câncer de mama com a coloração original convertida para o espectro branco – cinza – preto.....	13
Figura 6	Algoritmo CellProfiler Etapa 3 – ImageMath.....	14
Figura 7A	Inversão da coloração dos núcleos para a cor branca e demais elementos para o espectro cinza – preto.....	15
Figura 7B	Núcleos celulares corados na cor preta (imagem magnificada).....	15
Figura 7C	Núcleos celulares corados na cor branca com a coloração dos demais elementos invertida para o espectro cinza – preto.....	16
Figura 8	Algoritmo CellProfiler Etapa 4 – ApplyThreshold.....	16
Figura 9	Definição dos limites dos núcleos celulares.....	17
Figura 10	Algoritmo CellProfiler Etapa 5 – IdentifyPrimaryObjects.....	18
Figura 11A	Identificação dos núcleos como Objetos Primários do estudo.....	18

Figura 11B	Identificação dos núcleos como Objetos Primários do estudo (imagem magnificada).....	19
Figura 12	Algoritmo CellProfiler Etapa 6 – MeasureObjectSizeShape.....	20
Figura 13	Relação das mensurações dos parâmetros identificados para cada objeto primário de estudo.....	20
Figura 14	Algoritmo CellProfiler Etapa 7 - FilterObjects.....	21
Figura 15A	Aplicação de filtro de imagem para eliminação de artefatos não celulares.....	21
Figura 15B	Núcleos celulares definidos após a aplicação de filtro de imagem para eliminação de artefatos não celulares (imagem magnificada).....	22
Figura 16	Algoritmo CellProfiler Etapa 8 – MeasureObjectSizeShape.....	23
Figura 17	Relação das mensurações dos atributos identificados nos núcleos celulares.....	23
Figura 18	Algoritmo CellProfiler Etapa 9 – ExportToDatabase.....	24
Figura 19	Base de Dados do software CellProfiler com 1937 imagens e 47 parâmetros.....	25
Figura 20	Exclusão de 438 imagens (linhas) realizada pelo software Tanagra na Base de Dados transferidas do software Cellprofiler.....	38
Figura 21	Planilha da Base de Dados do software Tanagra.....	39
Figura 22	Atributos dos núcleos celulares (objetos primários).....	40
Figura 23	Principais atributos - Score Tubular.....	41
Figura 24	Principais atributos para o score tubular quanto ao p valor.....	42

Figura 25	Principais atributos - Score Nuclear.....	43
Figura 26	Principais atributos para o score nuclear quanto ao <i>p</i> valor.....	43
Figura 27	Principais atributos – Índice Mitótico.....	44
Figura 28	Principais atributos para o score índice mitótico quanto ao <i>p</i> valor.....	45
Figura 29	Principais atributos – Grau Histológico.....	46
Figura 30	Principais atributos para o grau histológico quanto ao <i>p</i> valor.....	46
Figura 31	Gráfico de Dispersão de análise automatizada para o score tubular.....	47
Figura 32	Gráfico de Dispersão de análise automatizada para o score nuclear.....	48
Figura 33	Gráfico de Dispersão de análise automatizada para o índice mitótico...	48
Figura 34	Análise comparativa entre classificação dada pelo médico patologista e classificação automatizada do software Tanagra para o score tubular...	49
Figura 35	Análise comparativa entre classificação dada pelo médico patologista e classificação automatizada realizada pelo software Tanagra para o score nuclear.....	51
Figura 36	Classificação automatizada do score índice mitótico: Análise comparativa entre análise do médico patologista e do software Tanagra.....	53
Figura 37	Classificação automatizada do grau histológico de malignidade: Análise comparativa entre análise do médico patologista e do software Tanagra.....	55

LISTA DE TABELAS

Tabela 1	Estatísticas descritivas quanto a idade das pacientes.....	28
Tabela 2	Caracterização da amostra quanto ao Grau Histológico.....	28
Tabela 3	Caracterização da amostra quanto tempo entre diagnóstico e tratamento.....	28
Tabela 4	Caracterização da amostra: Dados epidemiológicos.....	30
Tabela 5	Caracterização da amostra quanto a ocupação profissional.....	31
Tabela 6	Caracterização da amostra quanto ao município de procedência.....	32
Tabela 7	Dados epidemiológicos do tumor.....	34
Tabela 8	Estadiamento clínico do tumor (TNM).....	35
Tabela 9	Estado da doença ao final do 1º tratamento.....	36
Tabela 10	Causa básica da morte e fonte de custeio do tratamento.....	36
Tabela 11	Escala de concordância do Índice <i>Kappa</i> entre analisadores.....	50

RESUMO

Boasquevisque, Paulo César Ribeiro. DSc. Universidade Vila Velha - E.S. abril 2020. **Classificação do grau histológico de malignidade em câncer de mama utilizando programas de análise automatizada de imagens celulares e o princípio de aprendizado de máquina.** Orientador: Dominik Lenz.

O câncer de mama é o tipo de câncer mais comum entre as mulheres correspondendo a 24,2% de todos os cânceres em estatísticas mundiais e a 29,7% no Brasil; é a principal causa de óbitos por câncer em mulheres brasileiras. A Inteligência Artificial tem sido cada vez mais utilizada como ferramenta de apoio no diagnóstico das doenças em geral; diversos algoritmos e protocolos tem sido propostos para o câncer de mama utilizando-se da ferramenta “machine learning”, que permite a análise automatizada de dados e imagens com diagnósticos e prognósticos seguros e confiáveis. A determinação do grau histológico do câncer de mama constitui um dos principais sistemas de graduação da malignidade do câncer de mama. O objetivo do estudo foi determinar o grau histológico de malignidade do câncer de mama de forma automatizada, baseado na associação de parâmetros histopatológicos convertidos em parâmetros quantitativos, utilizando-se do princípio de “machine learning” com os programas computacionais de livre acesso CellProfiler e Tanagra. O estudo foi composto por 224 mulheres portadoras de câncer de mama em seus tipos histológicos mas comuns, das quais foram obtidas 20 fotografias digitais no campo de grande aumento (40 X) de lâminas histológicas do tecido neoplásico obtido por cirurgia, sendo as imagens transferidas para o software CellProfiler e tratadas conforme algoritmo pré-determinado resultando em base de dados exportadas para o software Tanagra onde se procedeu a classificação automatizada do grau histológico de malignidade desses cânceres de mama que foi comparada com a classificação realizada por médico patologista. O índice *Kappa* de concordância entre os analisadores médico patologista e a análise automatizada realizada pelo software Tanagra foi de 0,91 para o score tubular, de 0,55 para o score nuclear e de 0,49 para o score índice mitótico. Para a classificação automatizada do grau histológico de malignidade o índice *Kappa* de concordância entre os analisadores foi de 0,55 correlacionando-se diretamente com a frequência de apresentação dos grupos de cada graduação na amostra analisada. O estudo representa uma das primeiras pesquisas utilizando softwares de livre acesso para diagnóstico do grau histológico de malignidade em câncer de mama e demonstrou ser factível e reproduzível a análise automatizada de parâmetros histopatológicos para esse diagnóstico, sendo a formação de uma ampla base de dados o fator determinante para adequada acurácia do método.

Palavras – chave: inteligência artificial, carcinoma, parâmetros histopatológicos, cellprofiler, tanagra.

ABSTRACT

Boasquevisque, Paulo César Ribeiro. DSc. Universidade Vila Velha - E.S. abril 2020. **Histological malignancy grade classification in breast cancer using automated image analysis programs and machine learning.** Orientador: Dominik Lenz.

Breast cancer is the most common type of cancer among women, accounting for 24.2% of all cancers in world statistics and 29.7% in Brazil; is the leading cause of cancer deaths in Brazilian women. Artificial Intelligence has been increasingly used as a support tool in the diagnosis of diseases in general; several algorithms and protocols have been proposed for breast cancer, using the machine learning tool, which allows an automated analysis of data and images with diagnosis and prognosis of insurance and tests. The determination of the histological degree of breast cancer represents one of the main breast cancer graduation systems. The objective of the study was to determine the histological degree of malignancy of breast cancer in an automated way, based on the association of histopathological tests converted into quantitative values, using the principle of machine learning with the free access computer programs CellProfiler and Tanagra. The study consisted of 224 women with breast cancer in its most common types, of which 20 were photographed with 20 digital photographs (40 X) of histological slides of neoplastic tissue used by surgery, such as images transferred to CellProfiler software and treatments according to a predetermined algorithm, resulting in the database exported to the Tanagra software, where an automated classification of the histological degree of malignancy of these breast cancers was carried out, which was compared with a classification carried out by a pathologist. The Kappa index of agreement between the medical pathologist and an automated analysis performed by the Tanagra software was 0.91 for the tubular score, 0.55 for the nuclear score and 0.49 for the mitotic score. For an automated classification of histological grade of malignancy or Kappa index of agreement between the analyzers, it was 0.55 directly correlated with the frequency of presentation of the study groups of each class in the analyzed analysis. The study represents one of the first researches that use free access software for the diagnosis of histological grade of malignancy in breast cancer and demonstrates that an automated analysis of histopathological analyzes for this diagnosis is feasible and reproducible, being the formation of a wide database or determining factor for adequate method accuracy.

Key words: artificial intelligence, carcinoma, histopathological parameters, cellprofiler, tanagra.

INTRODUÇÃO

O câncer de mama é o tipo de câncer mais comum entre as mulheres e o segundo tipo de câncer mais frequente no mundo, excluindo-se o câncer de pele não melanoma, correspondendo a 24,2% de todos os cânceres em estatísticas mundiais e a 29,7% no Brasil. O câncer de mama é raro no homem representando menos de 1% dos casos (AMERICAN CANCER SOCIETY, 2019), (INCA, 2020).

Os tipos histológicos mais frequentes são o carcinoma ductal infiltrante, carcinoma lobular invasivo e a forma mista ductal lobular infiltrante, que juntos respondem por cerca de 90% dos cânceres de mama (AMERICAN CANCER SOCIETY, 2019), (INCA, 2020).

No Brasil são estimados 66.280 novos casos de câncer de mama por ano com um risco de 61,61 casos a cada 100 mil mulheres. Para o Espírito Santo, um estado da Região Sudeste do Brasil que é a região de maior desenvolvimento econômico do país, 1.130 novos casos de câncer de mama em mulheres foram estimados para o ano de 2018 e para o ano de 2020 são estimados 790 novos casos de câncer de mama (INCA, 2017), (INCA, 2020).

O câncer de mama representa a principal causa de óbitos por câncer em mulheres brasileiras, com 15.403 óbitos no ano de 2015 e 16.724 óbitos no ano de 2017 correspondendo a 16,2% (2015) e 16,1% (2017) de todos os óbitos em mulheres no Brasil (MINISTÉRIO DA SAÚDE, 2015), (MINISTÉRIO DA SAÚDE, 2017).

Para o sucesso do tratamento e controle do câncer de mama na população feminina, torna-se fundamental a identificação e atuação nos fatores de risco para a doença, o diagnóstico precoce e o acesso imediato ao tratamento, condições determinantes para o prognóstico da evolução da doença (AMERICAN CANCER SOCIETY, 2019), (INCA, 2020).

Dentre os fatores para definição do prognóstico do câncer de mama destacam-se a classificação TNM a partir da análise do tamanho do tumor, o comprometimento dos linfonodos axilares e a presença de metástases a distância; o grau histológico de malignidade do tumor; a presença de biomarcadores no tecido tumoral e o perfil de expressão gênica do tumor (BECK et al., 2011), (CHEN et al., 2015), (CHEN et al., 2017), (RAWAT et al., 2018), (WHITNEY et al., 2018)

O grau histológico de malignidade proposto por Scarff, Bloom e Richardson e modificado por Elston e Ellis, conhecido como o Sistema de Classificação de Nottingham, é considerado um dos principais fatores para determinação do

prognóstico do câncer de mama (BECK et al., 2011), (XU et al., 2016). (CHEN et al.,2017), (ROMO-BUCHELI,2017), (LU et al., 2018).

Essa graduação histológica é realizada pelo médico patologista analisando 03 (três) aspectos: o aspecto da arquitetura tubular mamária, a morfologia nuclear e o número de células em mitose (índice mitótico) no tecido mamário comprometido, atribuindo para os scores tubular, nuclear e índice mitótico a pontuação de 1 a 3 conforme a intensidade das alterações teciduais observadas. A somatória das pontuações dos 3 scores resulta no valor final que define o grau histológico de malignidade em grau 1 (3 a 5 pontos), grau 2 (6 a 7 pontos) e grau 3 (8 a 9 pontos) correspondendo uma maior pontuação a um maior o grau de malignidade (Quadros 1 e 2).

Scores do Grau Histológico	pontuação	características
Formação tubular (score tubular)	1	> 75% túbulos
	2	10% a 75% túbulos
	3	< 10% túbulos
Morfologia nuclear (score nuclear)	1	semelhante a célula normal
	2	tamanhos intermediários
	3	núcleos grandes, pleomorfismo, nucléolos proeminentes
Índice mitótico (score índice mitótico)	1	até 10 mitoses / 10 cga (40 x)
	2	11 a 20 mitoses / 10 cga (40 x)
	3	> 20 mitoses / 10 cga (40 x)

Quadro 1. Aspectos componentes do grau histológico em câncer de mama.

Grau histológico de malignidade	pontuação
grau 1	3 - 5 pontos
grau 2	6 - 7 pontos
grau 3	8 - 9 pontos

Quadro 2. Grau histológico de malignidade em câncer de mama.

Com o objetivo de oferecer agilidade e segurança no diagnóstico das doenças em geral, a Inteligência Artificial tem sido cada vez mais utilizada como ferramenta de apoio nesse propósito (WERNICK et al., 2010), (MULRANE et al., 2008), (JONES et al., 2009), (HITCHCOCK, 2011), (SOMMER; GERLICH, 2013), (LOUKAS et al., 2013), (SINGH; CARPENTER; GENOVESIO, 2014), (BUZIN et al.,

2015), (VU et al., 2016), (DORDEA et al., 2016), (YU et al., 2016), (HENNIG et al., 2017), (EULENBERG et al., 2017), (PESAPANE; CODARI; SARDANELLI, 2018), (CHING et al., 2018).

O conceito de Inteligência Artificial foi introduzido na década de 1950 e pode ser definido como o ramo da Ciência da Computação que visa desenvolver a capacidade de programas computacionais de analisar e interpretar dados de forma autônoma, correlacionando conceitos e propondo soluções de forma proativa para ações requeridas, a partir do armazenamento de grande quantidade de informações concentradas na máquina.

Os ramos de desenvolvimento da Inteligência Artificial denominados redes neurais, machine learning e deep learning representam a evolução dessa tecnologia, inicialmente utilizada no mercado financeiro, progredindo com grande sucesso para o ramo do comércio (e – commerce), atingindo as diversas áreas das atividades humanas como tarefas de percepção, detecção de objetos, reconhecimentos e identificação pela voz, visão(retina), face e digitais, tradução e reconhecimento de discursos para execução de ordens e ações de multitarefas em geral.

Em particular para o câncer de mama, diversos algoritmos e protocolos tem sido propostos utilizando-se da ferramenta “machine learning”, segmento da Inteligência Artificial que permite concentrar uma grande variedade de informações em um sistema computadorizado criando-se ambiente adequado para o aprendizado de máquina no sentido de se obter a partir da análise automatizada de dados e imagens, diagnósticos e prognósticos seguros e confiáveis que permitam orientar o tratamento da doença (ARAÚJO et al., 2017), (BECK et al., 2011), (CHEN et al., 2015), (CHEN et al.,2017), (RAWAT et al., 2018), (WHITNEY et al., 2018), (LOUKAS et al., 2013), (SINGH; CARPENTER; GENOVESIO, 2014), (XU et al., 2016), (ZHONG et al., 2016), (MONTAZERI et al., 2016), (HOLTEN – ROSSING et al., 2017), (ROMO- BUCHELI et al., 2017), (PEIKARI et al., 2017), (RADIYA; ZHU; BECK, 2017), (HAN et al., 2017), (BEJNORDI et al., 2017), (SINTHIA; MALATHI, 2018), (TOPRAK, 2018), (SHI et al., 2018), (LU et al., 2018).

O aprendizado de máquina, uma vez atingida a acurácia e a precisão adequadas para o método, apresenta como vantagens a possibilidade de se reunir em uma única ferramenta digital um grande volume de informações sobre determinada doença; suprimindo a subjetividade da avaliação humana com agilidade na análise do material a ser estudado, objetivando diagnósticos seguros e rápidos, podendo ser utilizado inclusive como uma “segunda opinião especializada” em casos

de maior complexidade (WERNICK et al., 2010), (MULRANE et al., 2008), (JONES et al., 2009), (MISSELWITZ et al, 2010).

OBJETIVOS

O objetivo geral do estudo foi determinar de forma automatizada o grau histológico de malignidade em câncer de mama baseando-se no princípio de “machine learning” para análise de imagens digitalizadas de cortes histológicos de tecido mamário de mulheres portadoras de câncer de mama, utilizando-se para tal dos programas computacionais de livre acesso CellProfiler e Tanagra.

Como objetivos específicos foram propostos realizar a determinação automatizada da graduação para os scores tubular, nuclear e índice mitótico baseado na avaliação realizada pelos programas de análise de imagens celulares CellProfiler e Tanagra; analisar a performance do software Tanagra para a classificação automatizada dos scores tubular, nuclear, índice mitótico e do grau histológico de malignidade através da avaliação estatística dos valores preditivos, da acurácia, da classificação incorreta (erro) e do índice Kappa de concordância entre analisadores correlacionando os resultados obtidos com a classificação automatizada do grau histológico de malignidade com a classificação realizada por médico patologista.

MATERIAL E MÉTODOS

3.1 - A amostra – Critérios de inclusão e exclusão.

O Hospital Santa Rita de Cássia localizado na cidade de Vitória, capital do estado do Espírito Santo foi o escolhido pelos pesquisadores para a realização do estudo por ser considerado o principal hospital de referência para tratamento oncológico no Espírito Santo.

O ano de 2015 foi o selecionado para o estudo por representar no início da pesquisa, o ano mais recente em que todos os dados das mulheres tratadas de câncer de mama no Hospital Santa Rita de Cássia estarem cadastrados no Registro Hospitalar de Câncer do Ministério da Saúde do Governo Federal.

O estudo foi composto por mulheres portadoras de câncer de mama em seus tipos histológicos mas frequentes carcinoma ductal infiltrante, carcinoma lobular invasivo e a forma mista ductal lobular infiltrante; que foram submetidas a tratamento cirúrgico para essa doença no ano de 2015 e que até o momento da cirurgia não haviam sido submetidas a tratamentos adjuvantes de quimioterapia ou radioterapia; das quais se pode obter dados completos epidemiológicos, do diagnóstico e do tratamento e que possuíam lâminas histológicas coradas pelo método Hematoxilina & Eosina com coloração preservada para a obtenção de fotografias digitais de qualidade adequada minimizando efeitos causados por artefatos histológicos. (Quadro 3)

Foram excluídos do estudo os casos de câncer de mama de outros tipos histológicos diferentes de carcinomas ductais infiltrantes, lobulares invasivos ou sua forma mista ductal infiltrante; além dos demais critérios de exclusão listados no Quadro 4.

-
1. mulheres com câncer de mama tipo carcinoma ductal infiltrante, carcinoma lobular invasivo ou ductal lobular infiltrante;
 2. submetidas a tratamento cirúrgico para a doença em 2015;
 3. não submetidas a quimioterapia e / ou radioterapia;
 4. dados completos epidemiológicos, de diagnóstico e tratamento
 5. lâminas histológicas coradas pelo método H&E com coloração preservada e sem artefatos histológicos.
-

Quadro 3. Critérios de inclusão na amostra do estudo.

-
1. mulheres com câncer de mama de tipos histológicos diferentes dos mais comuns
 2. carcinoma “in situ” de mama
 3. não operadas para câncer de mama em 2015
 4. submetidas a tratamentos coadjuvantes de quimioterapia e / ou radioterapia
 5. dados incompletos epidemiológicos, de diagnóstico e tratamento
 6. lâminas histológicas com coloração não preservada ou com artefatos histológicos identificados
-

Quadro 4. Critérios de exclusão na amostra do estudo.

Os casos de carcinoma “in situ” de mama foram excluídos por representarem um tipo específico da doença que por respeitar a membrana basal dos ductos acometidos não invadindo tecidos vizinhos e mantendo o restante do tecido mamário de aspecto histológico dentro da normalidade, poderiam induzir a erro de classificação automatizada (Figura 1).

Da mesma forma, os casos com presença de artefatos histológicos foram excluídos do estudo por representarem as alterações artefatuais causas de erros diagnósticos.

O uso da quimioterapia em câncer de mama aumenta o número de células em apoptose que podem eventualmente serem confundidas com células em mitose e a radioterapia altera a arquitetura do tecido mamário pelo processo inflamatório e fibrose local que provoca; razões pelas quais pacientes previamente submetidas a esses tratamentos coadjuvantes foram excluídas do estudo.

No ano de 2015 foram atendidas no Hospital Santa Rita de Cássia 625 mulheres portadoras de câncer de mama. Dos 276 casos que atenderam aos critérios de inclusão, foram ainda excluídas 07 pacientes por apresentarem cânceres de mama “in situ” e 45 por apresentarem alterações teciduais de coloração ou artefatos histológicos que poderiam comprometer o aprendizado de máquina e por consequência a análise automatizada das imagens desses casos; resultando no total de 224 casos integrantes desse estudo.

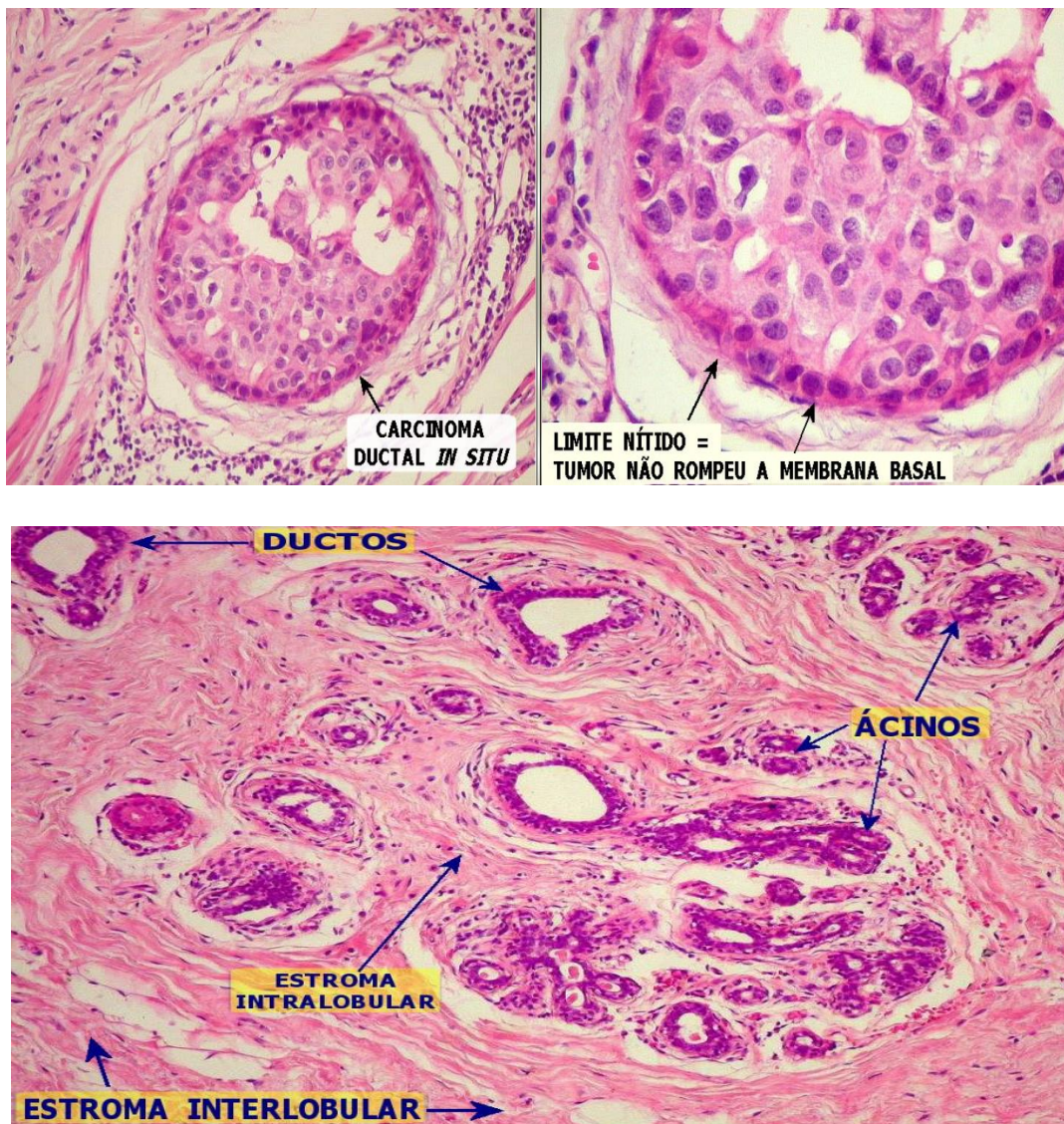


Figura 1. Carcinoma in situ e aspecto histológico normal da mama (Fonte: anatpat.unicamp.br).

3.2 - Critérios Éticos.

O Projeto de Pesquisa recebeu parecer favorável do Comitê de Ética em Pesquisa em Seres Humanos do Hospital Universitário Cassiano Antônio de Moraes da Universidade Federal do Espírito Santo sob o nº 2.014.675 de 12/04/2017 e do Comitê de Ética em Pesquisa em Seres Humanos da Universidade Vila Velha sob o nº 2.020.954 de 18/04/2017.

Foram também firmados pelos pesquisadores junto ao Hospital Santa Rita de Cássia, Termo de Compromisso de Sigilo dos Dados em 11/01/2017 e Declaração de Autorização para realizar a captação de dados para o Projeto de Pesquisa nas dependências do Hospital Santa Rita de Cássia em 13/02/2017.

3.3 - Digitalização das lâminas histológicas

Todas as lâminas histológicas dos 224 casos selecionados foram revisadas por médico patologista do corpo clínico do Hospital Santa Rita de Cássia de forma aleatória e sem acesso aos dados das pacientes, com o objetivo de selecionar para cada caso as lâminas de melhor aspecto de coloração preservada e realizar a classificação do grau histológico de malignidade para todos os casos, sendo obtidas 20 fotografias digitais de diferentes locais do tecido mamário englobando tecidos de aspecto normal e áreas comprometidas por câncer de mama nos aumentos de 4, 10 e 40 vezes totalizando 60 fotografias digitais por paciente e um total global de 13.440 fotografias digitais, utilizando para o registro fotográfico câmera digital modelo Moticam 1000 1,3 MPixel MTC 1000 acoplada em microscópio óptico de luz.

Analisadas as imagens digitalizadas, decidiu-se por utilizar em todo o estudo somente as fotografias obtidas no aumento de 40 vezes (campo de grande aumento) que é o aumento habitualmente utilizado pelo médico patologista para analisar as lâminas histológicas com vistas ao diagnóstico das alterações neoplásicas, avaliação do aspecto nuclear e contagem de células em mitose e por conseguinte realizar a classificação do grau histológico de malignidade em câncer de mama.

3.4 – O software CellProfiler

O software CellProfiler, um programa de análise automatizada de imagens celulares de livre acesso na internet concebido em 2005 por cientistas norte-americanos coordenados pela cientista Anne E. Carpenter, foi o programa selecionado pelos pesquisadores para realizar a análise das imagens digitalizadas no aumento de 40 vezes e processar as alterações dessas imagens para o ambiente computacional, transformando parâmetros qualitativos das imagens originais em parâmetros quantitativos identificados pelo software CellProfiler, de forma a permitir em etapa subsequente, a utilização do segundo programa de análise automatizada de imagens celulares de livre acesso selecionado pelos pesquisadores, o software Tanagra, a análise e classificação automatizada dessas imagens convertidas em parâmetros quantitativos (CARPENTER et al., 2006), (JONES et al., 2009), (LAMPRECHT; SABATINI; CARPENTER, 2007), (LENZ et al., 2017).

3.5 - Algoritmo CellProfiler

Para a análise das imagens digitalizadas de câncer de mama foi elaborado um protocolo de ações (algoritmo) com o objetivo de realizar o tratamento para o ambiente computacional realizando as alterações dos parâmetros qualitativos da imagem resultando em parâmetros quantitativos identificados pelo software CellProfiler para a identificação dos objetos de estudo, os núcleos celulares. O Algoritmo CellProfiler foi composto por uma sequência de 9 (nove) etapas conforme exposto no Quadro 5.

Etapa	Sequência Algoritmo Cellprofiler
1	Loadimages
2	ColorToGray
3	ImageMath
4	ApplyThreshold
5	IdentifyPrimaryObjects
6	MeasureObjectSizeShape
7	FilterObjects
8	MeasureObjectSizeShape
9	ExportToDatabase

Quadro 5. Algoritmo CellProfiler.

3.6– Etapas do Algoritmo CellProfiler

3.6.1 - Etapa 1 – LoadImages.

Todas as 4.480 imagens digitalizadas das lâminas histológicas de câncer de mama obtidas no aumento de 40 vezes foram transferidas para o software CellProfiler e foram mantidas somente as imagens reconhecidas como adequadas por esse programa de análise de imagens celulares, totalizando 1937 imagens, que submetidas as etapas seguintes do algoritmo CellProfiler geraram para cada imagem digitalizada, 47 parâmetros quantitativos chamados de atributos (Figura 2).

Os atributos são aspectos e características identificados pelo software CellProfiler que expressam o valor médio dos parâmetros quantitativos dos objetos de estudo (os núcleos celulares) e que permitem identificar e classificar cada objeto de forma automatizada.

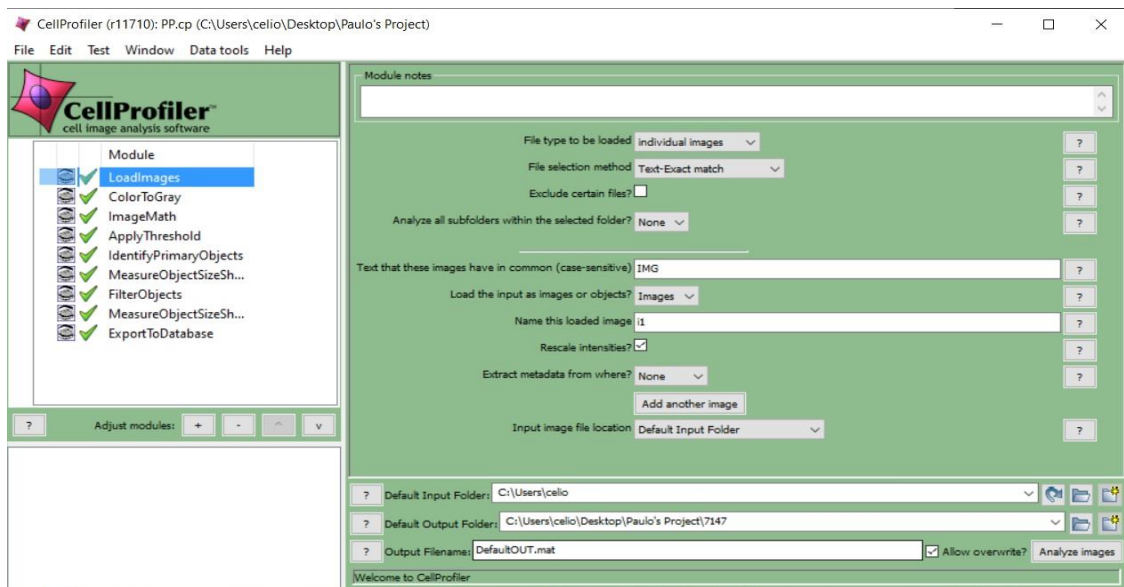


Figura 2. Algoritmo CellProfiler Etapa 1 – LoadImages.

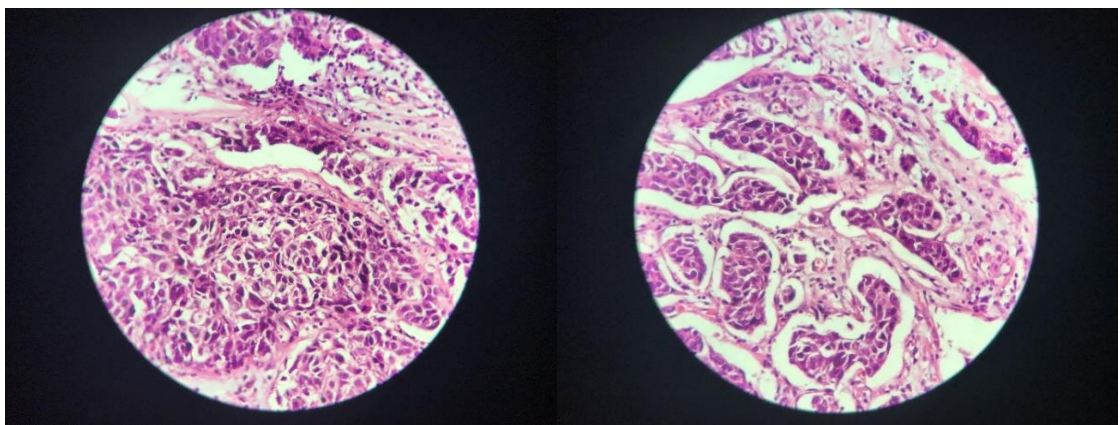


Figura 3. Imagens digitalizadas de câncer de mama aumento 40 x.

3.6.2 - Etapa 2 – ColorToGray.

As imagens digitalizadas originais foram convertidas para o espectro de coloração branco – cinza – preto com a finalidade de tornar mais evidente os objetos de estudo, os núcleos celulares. (Figura 4 e Figuras 5A, 5B e 5C).

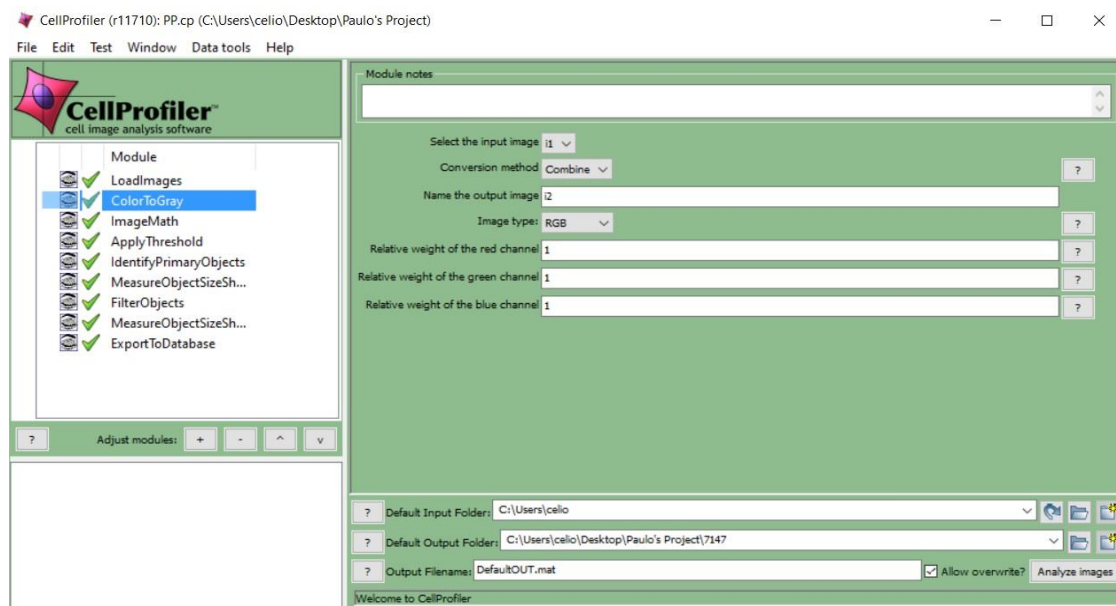


Figura 4. Algoritmo CellProfiler Etapa 2 – ColorToGray.

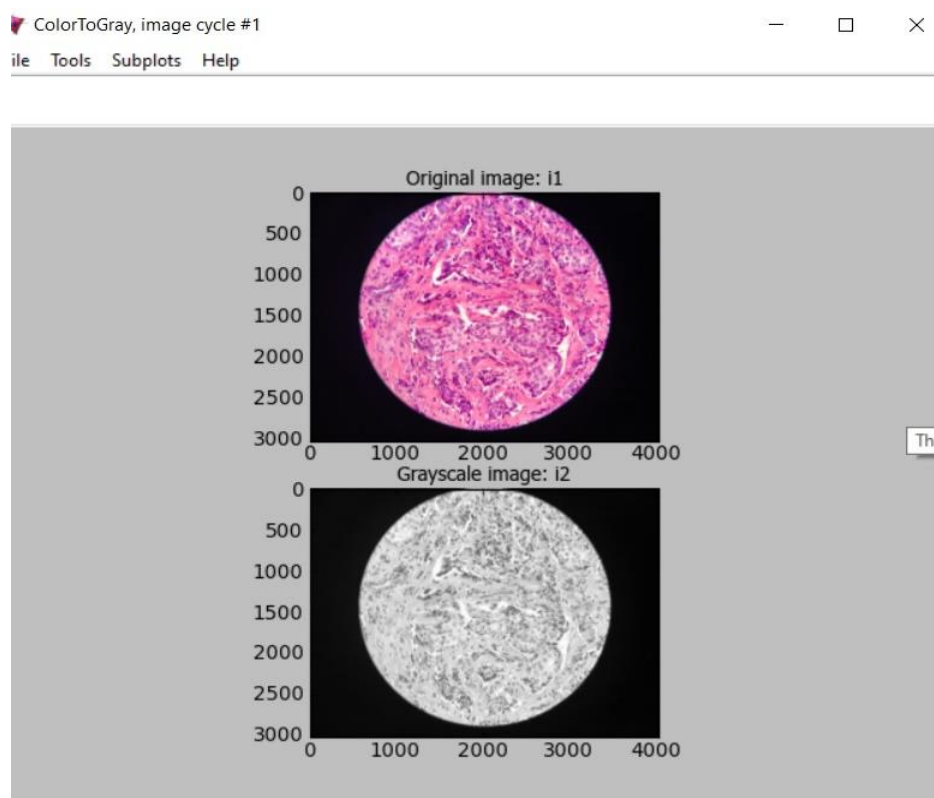


Figura 5A. Imagem digitalizada de câncer de mama com coloração original convertida para o espectro branco – cinza - preto.

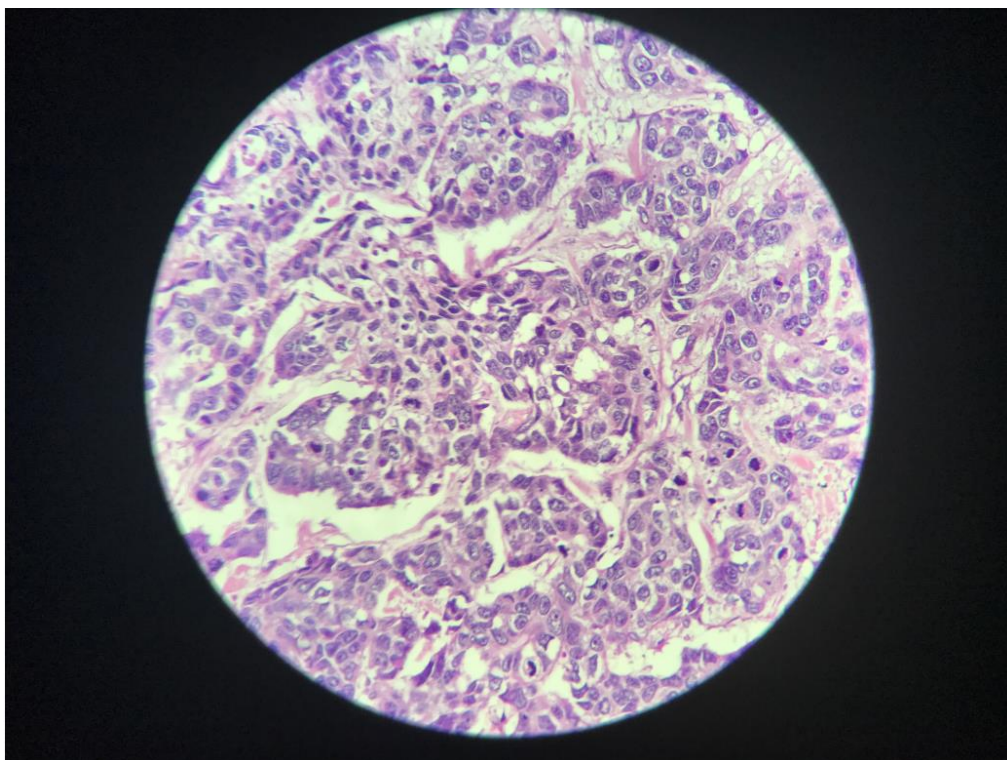


Figura 5B. Imagem digitalizada magnificada de câncer de mama na coloração original.

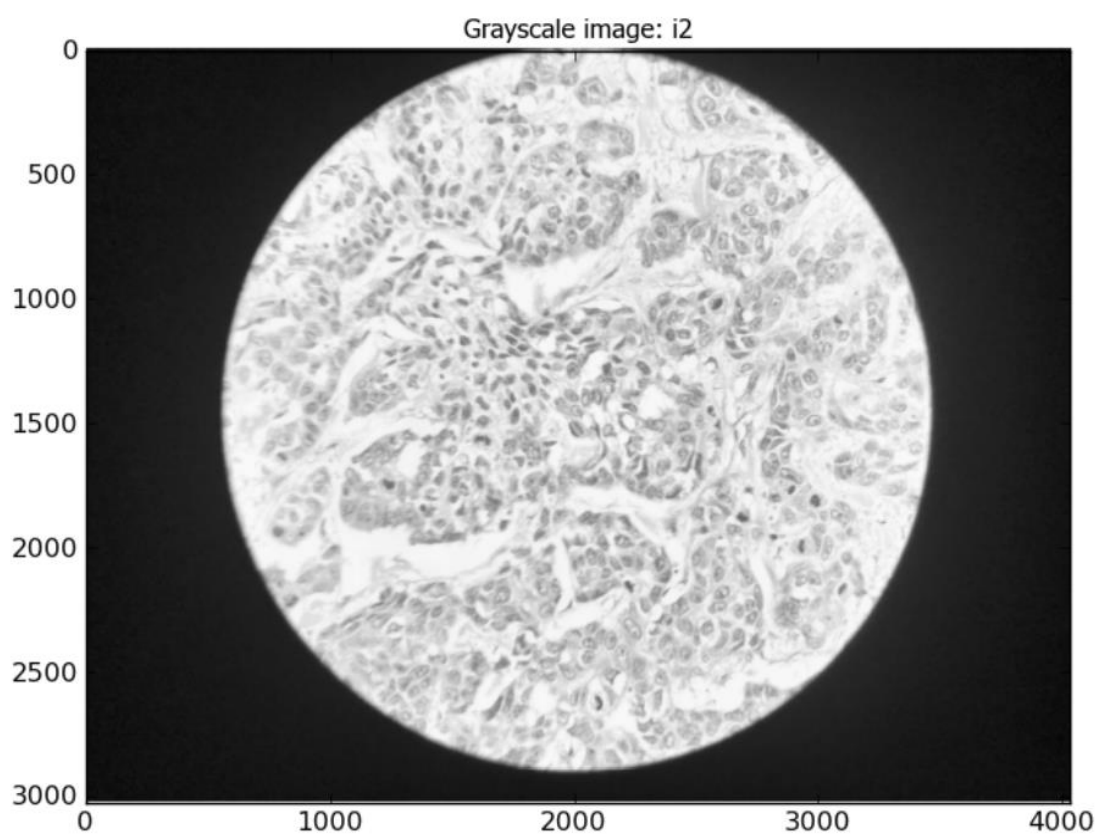


Figura 5C. Imagem digitalizada magnificada de câncer de mama com a coloração original convertida para o espectro branco – cinza – preto.

3.6.3 - Etapa 3 – ImageMath.

O software CellProfiler analisa os objetos de estudo conforme a intensidade de luz que refletem e dentro do espectro branco – cinza – preto a cor que mais reflete a luz é a cor branca.

Uma vez que os núcleos das células constituem os objetos selecionados para o estudo, faz-se necessário a inversão da coloração dos núcleos inicialmente corados em preto para a cor branca tendo os demais elementos invertidos sua coloração para a cor preta, de forma a tornar os núcleos mais evidentes e com maior intensidade de luz refletida, o que permite a sua análise de forma mais precisa. (Figura 6 e Figuras 7A, 7B e 7C).

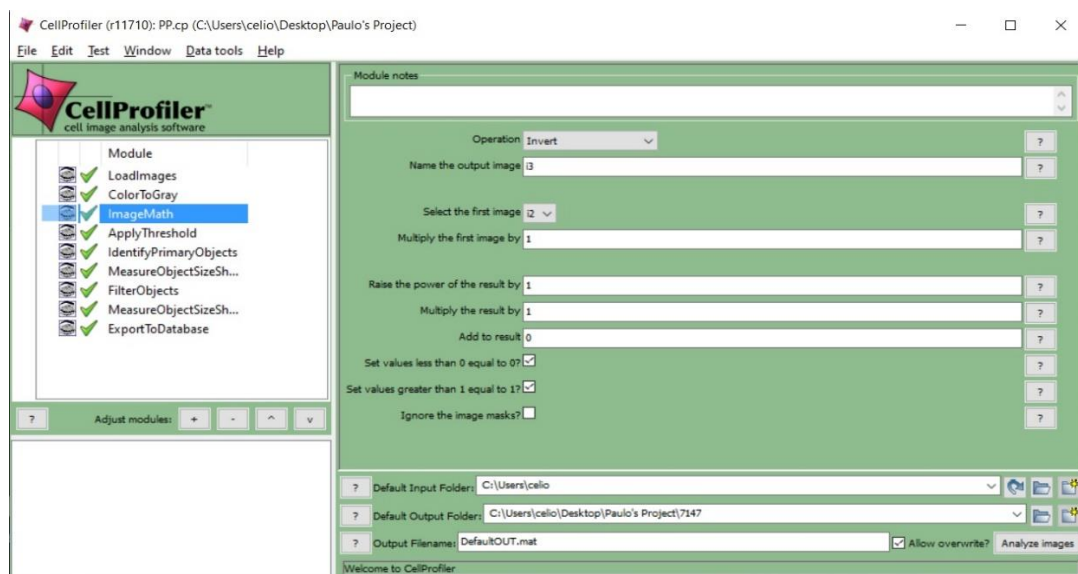


Figura 6. Algoritmo CellProfiler Etapa 3 – ImageMath.

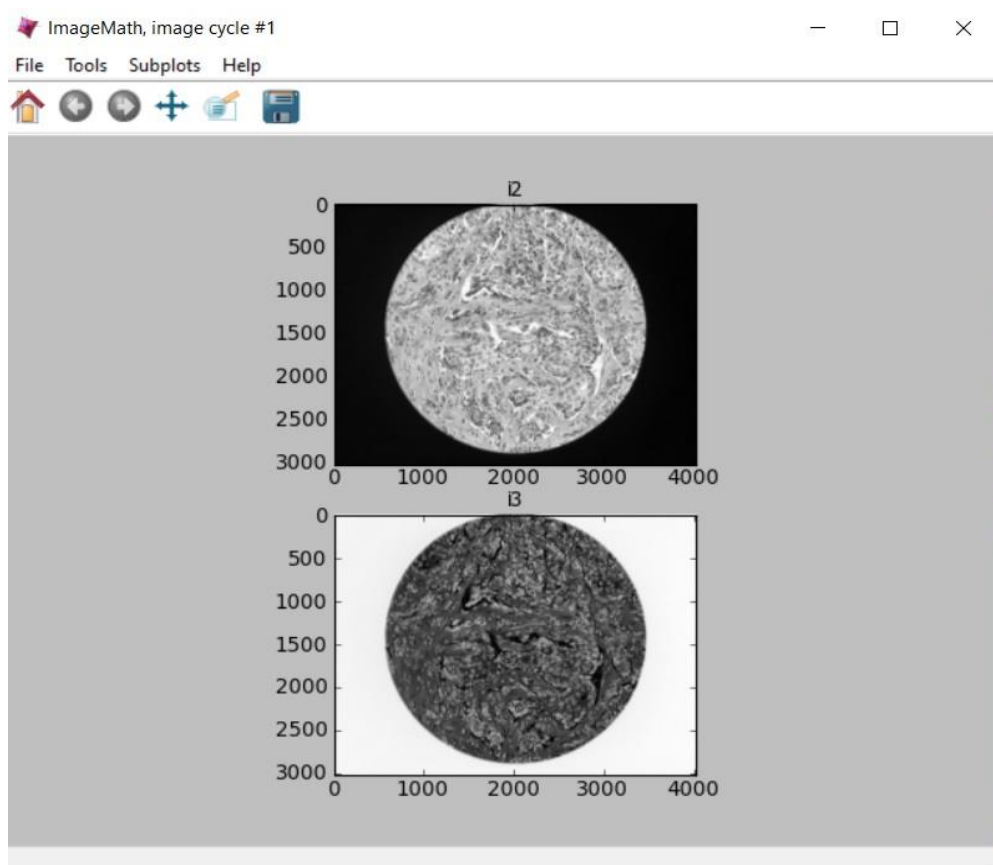


Figura 7A. Inversão da coloração dos núcleos para a cor branca e demais elementos para o espectro cinza – preto.

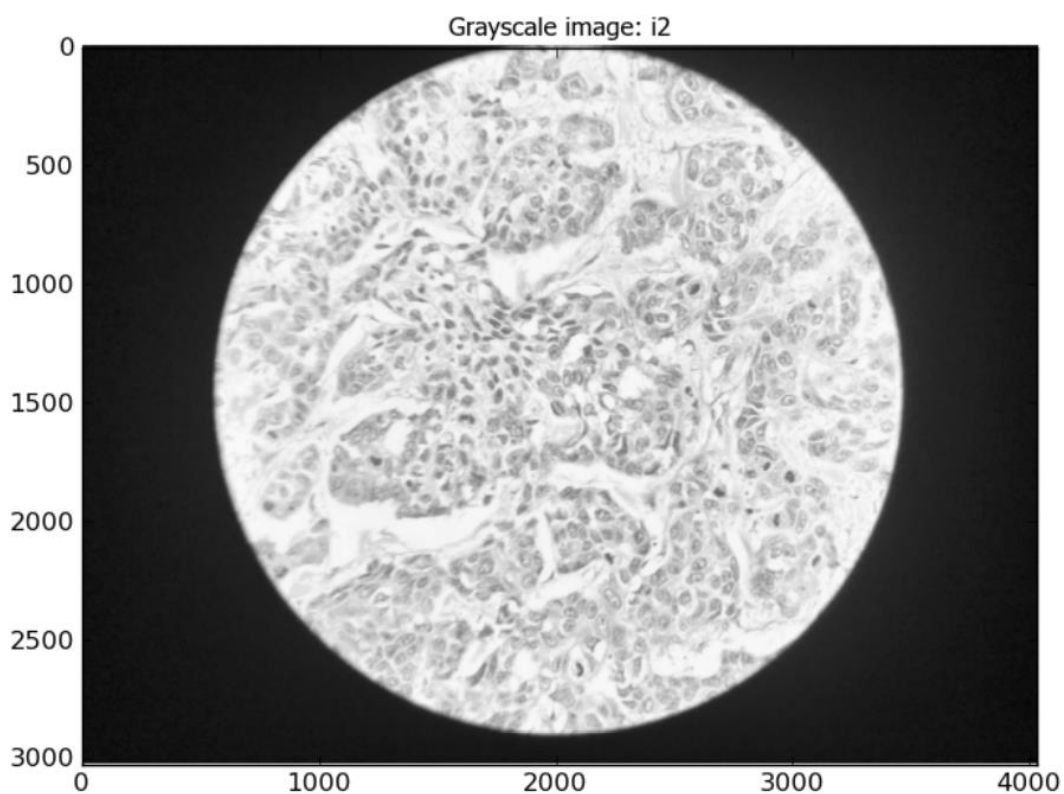


Figura 7B. Núcleos celulares corados na cor preta (imagem magnificada).

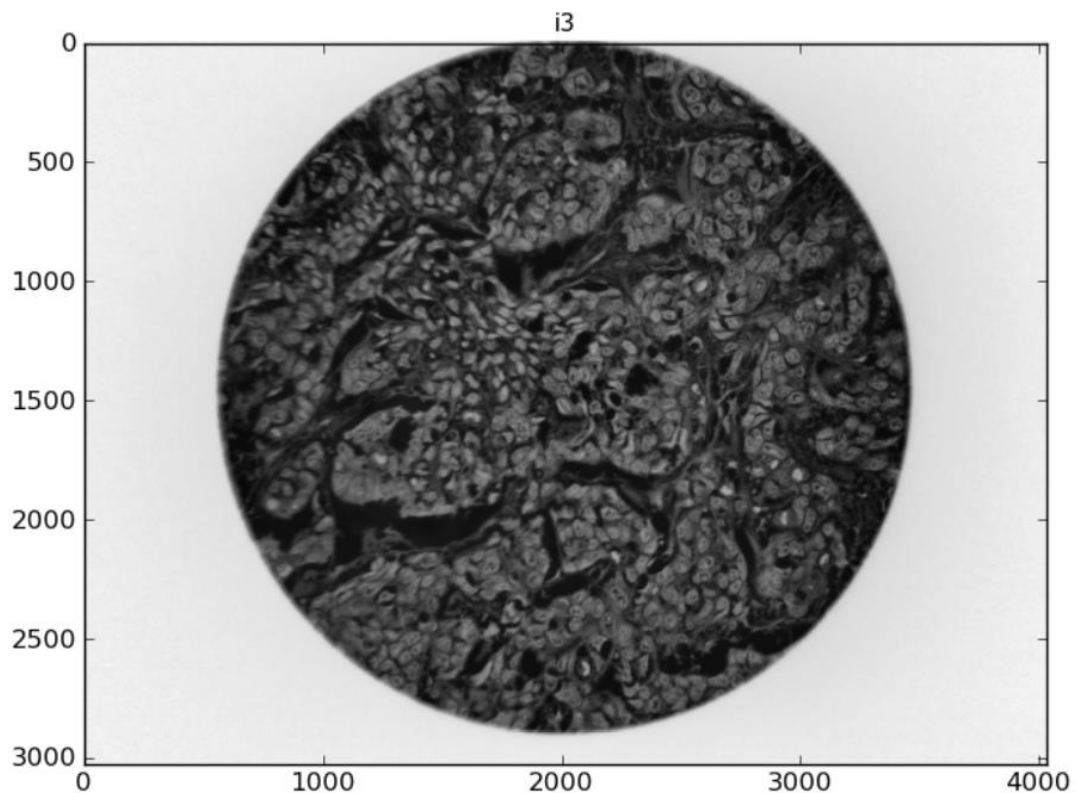


Figura 7C. Núcleos celulares corados na cor branca com a coloração dos demais elementos invertida para o espectro cinza - preto (imagem magnificada).

3.6.4 - Etapa 4 – *ApplyThreshold*.

Nesta etapa foram estabelecidos os limites dos núcleos celulares removendo os tecidos circunjacentes, tornando os núcleos mais evidentes e destacados como os objetos de estudo (Figuras 8 e 9).

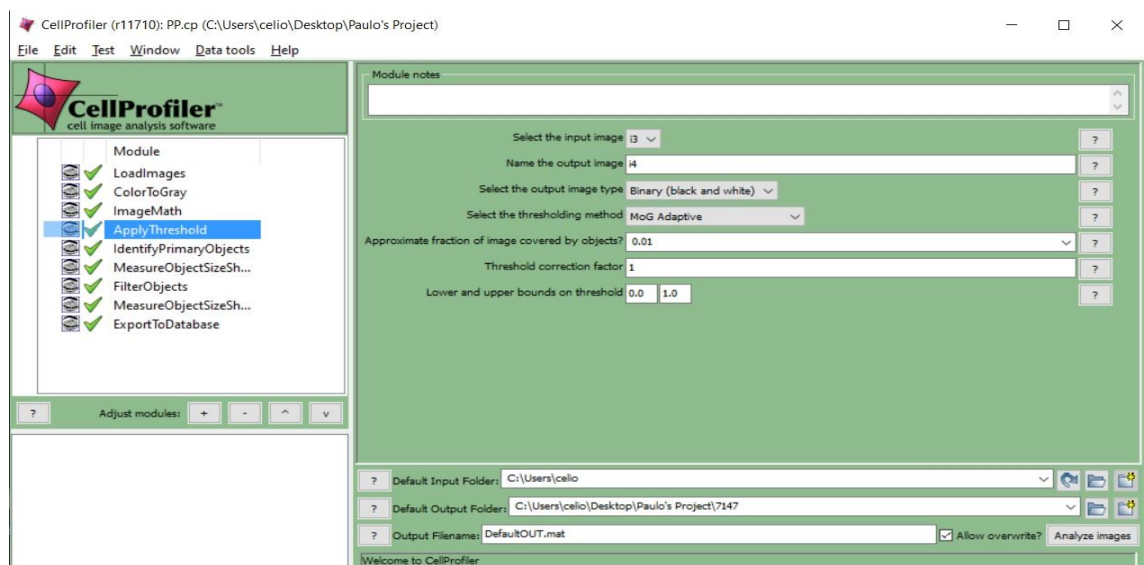


Figura 8. Algoritmo CellProfiler Etapa 4 – *ApplyThreshold*.

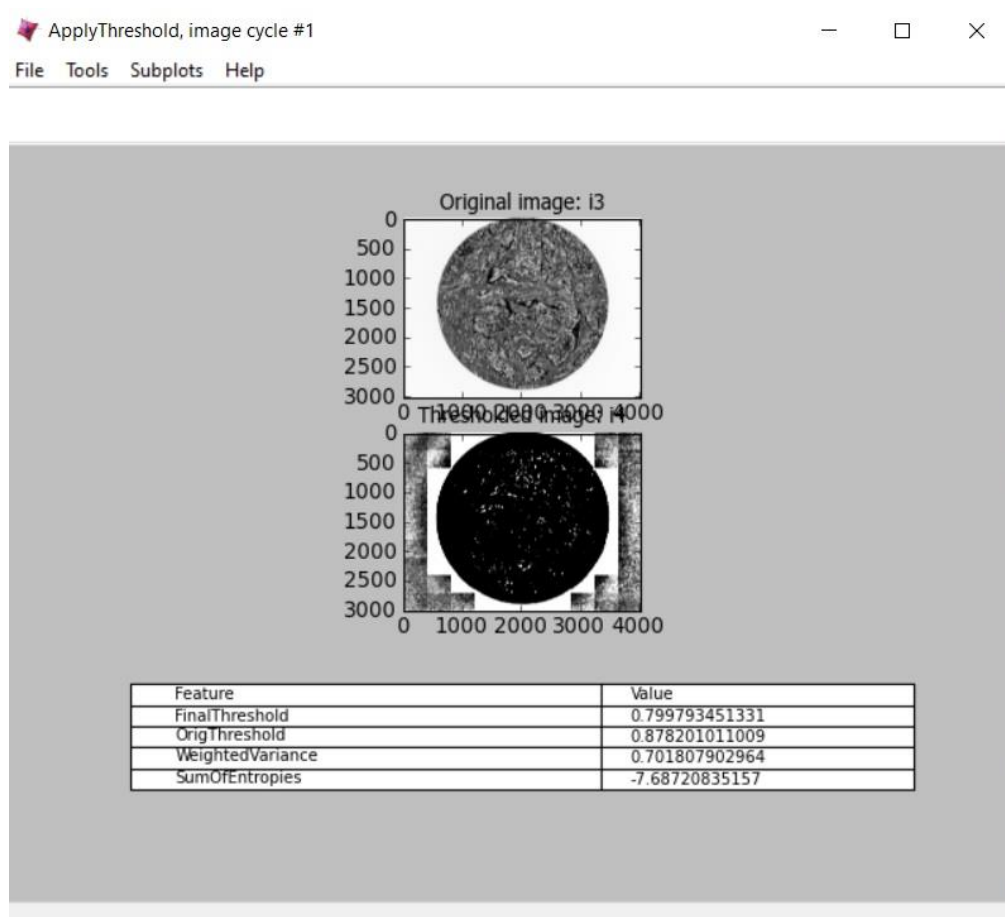


Figura 9. Definição dos limites dos núcleos celulares.

3.6.5 - Etapa 5 – *IdentifyPrimaryObjects*.

Os núcleos celulares uma vez evidenciados pela coloração branca, com seus limites bem estabelecidos e sem a interferência visual dos tecidos circunvizinhos podem ser definidos nessa etapa como os Objetos Primários do estudo (Figura 10 e Figura 11A e 11B).

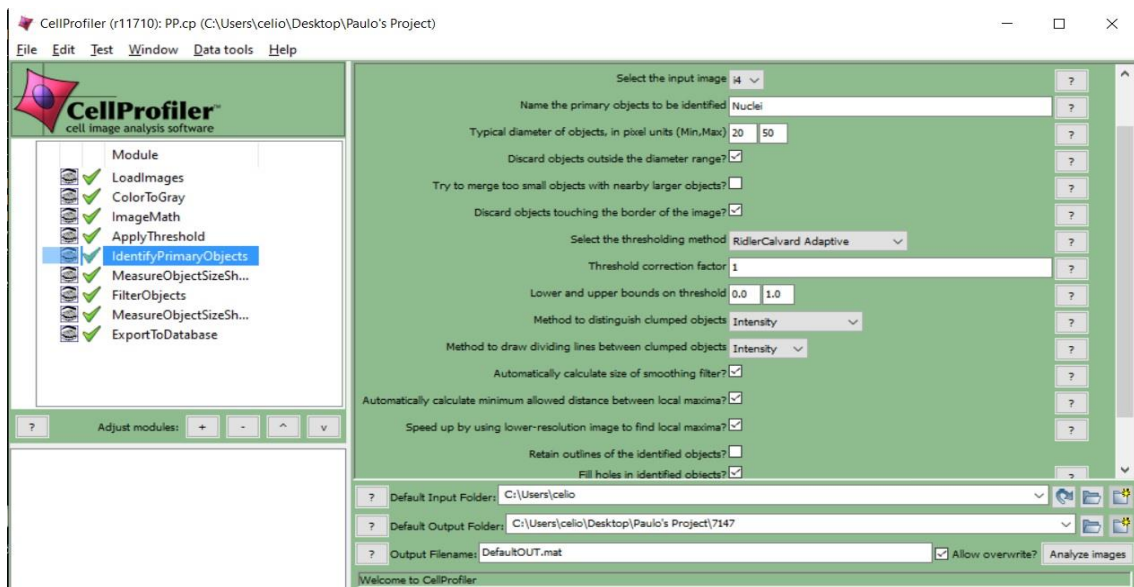


Figura 10. Algoritmo CellProfiler Etapa 5 – IdentifyPrimaryObjects.

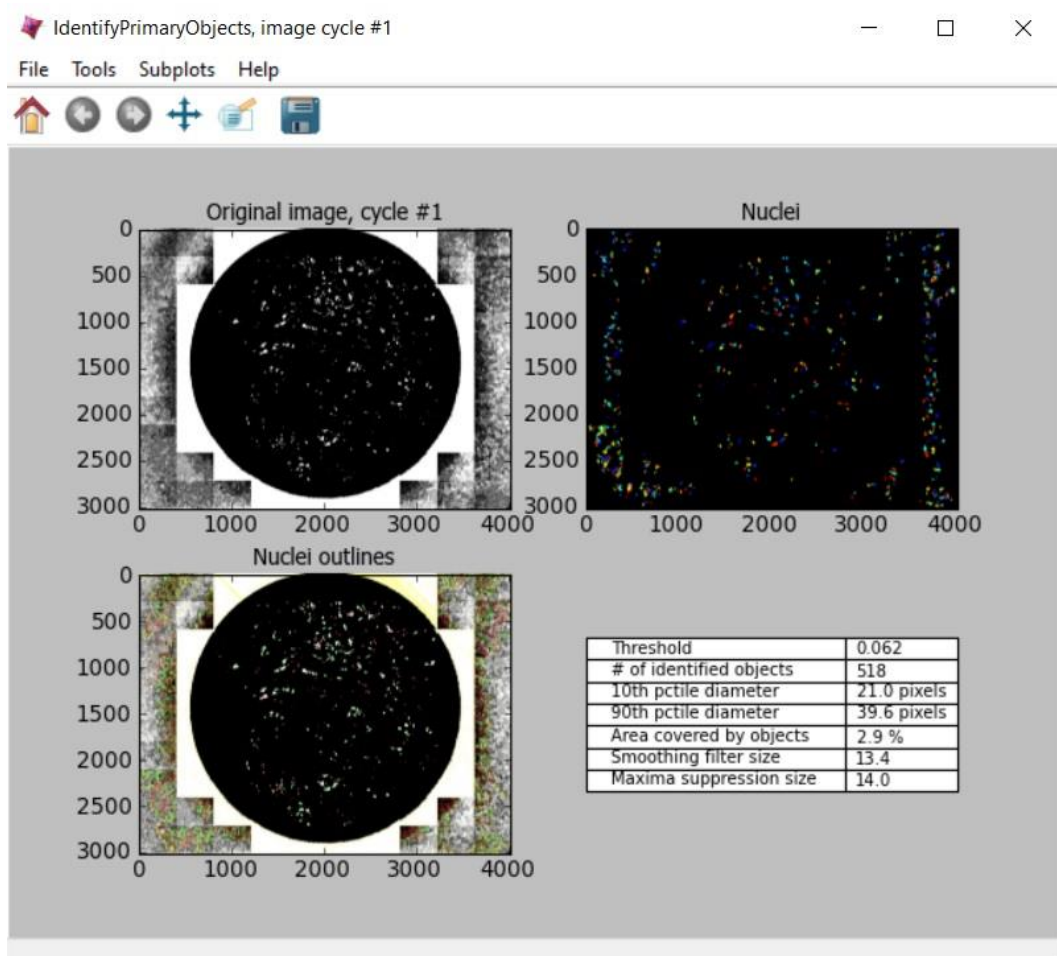


Figura 11 A. Identificação dos núcleos como Objetos Primários do estudo.

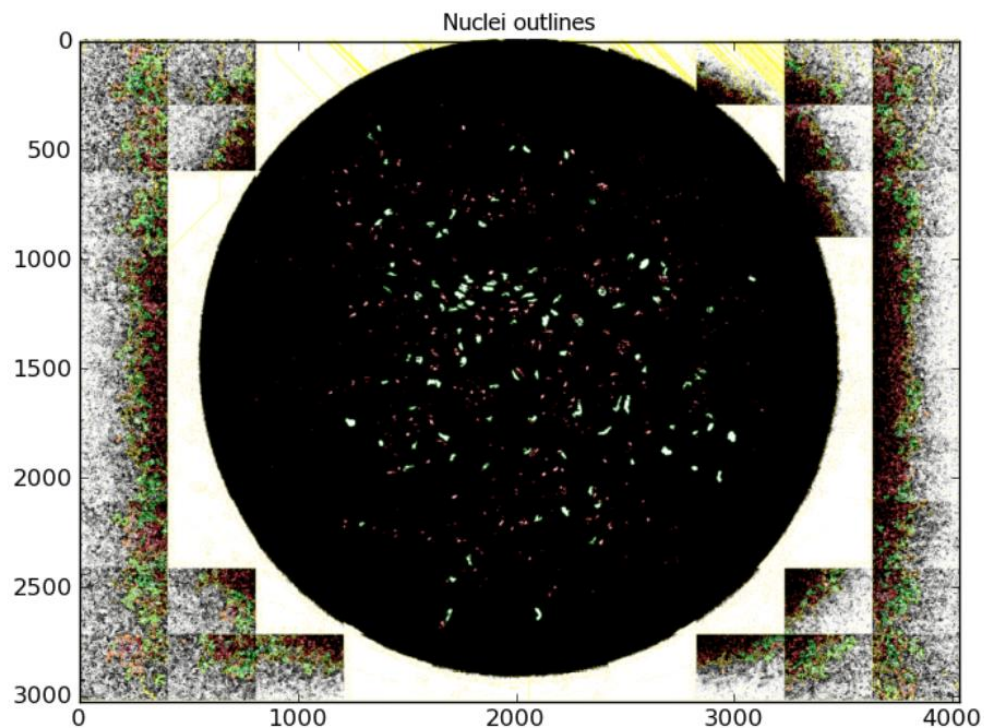


Figura 11B. Identificação dos núcleos como Objetos Primários do estudo (imagem magnificada).

3.6.6 - Etapa 6 – *MeasureObjectSizeShape*.

Uma vez identificados os núcleos celulares como o objeto primário de estudo pelo software CellProfiler é possível realizar diversas mensurações e tomadas de medidas com a definição de parâmetros que permitam identificar e classificar os objetos primários obtendo-se pela média dessas mensurações 47 parâmetros (atributos) identificados pelo software CellProfiler para cada objeto do estudo nessa etapa do algoritmo.(Figuras 12 e 13).

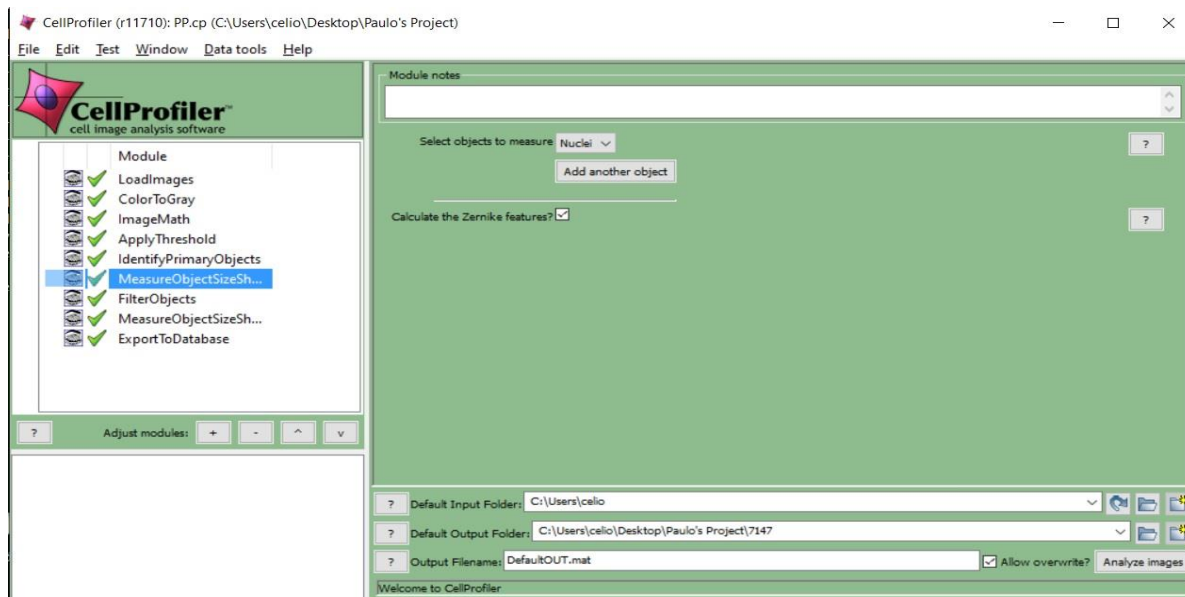


Figura 12. Algoritmo CellProfiler Etapa 6 – MeasureObjectSizeShape.

MeasureObjectSizeShape, image cycle #2

File Tools Subplots Help

Nuclei	Orientation	-1.73	-5.03	52.55
Nuclei	Compactness	1.66	1.56	0.47
Nuclei	Area	652.57	515.50	357.72
Nuclei	Center X	2350.09	3302.00	1488.25
Nuclei	Center Y	1759.37	2012.50	934.25
Nuclei	Extent	0.49	0.47	0.11
Nuclei	Perimeter	197.26	173.77	90.99
Nuclei	Solidity	0.69	0.68	0.11
Nuclei	FormFactor	0.26	0.23	0.15
Nuclei	EulerNumber	0.99	1.00	0.15
Nuclei	Zernike_0_0	0.43	0.41	0.12
Nuclei	Zernike_1_1	0.05	0.05	0.03
Nuclei	Zernike_2_0	0.12	0.13	0.03
Nuclei	Zernike_2_2	0.06	0.06	0.02
Nuclei	Zernike_3_1	0.03	0.03	0.02
Nuclei	Zernike_3_3	0.02	0.02	0.01
Nuclei	Zernike_4_0	0.02	0.02	0.01
Nuclei	Zernike_4_2	0.03	0.03	0.01
Nuclei	Zernike_4_4	0.02	0.02	0.01
Nuclei	Zernike_5_1	0.02	0.01	0.01
Nuclei	Zernike_5_3	0.02	0.02	0.01
Nuclei	Zernike_5_5	0.01	0.01	0.01
Nuclei	Zernike_6_0	0.01	0.01	0.01
Nuclei	Zernike_6_2	0.01	0.01	0.01
Nuclei	Zernike_6_4	0.01	0.01	0.01
Nuclei	Zernike_6_6	0.01	0.02	0.01
Nuclei	Zernike_7_1	0.01	0.01	0.00
Nuclei	Zernike_7_3	0.01	0.01	0.01
Nuclei	Zernike_7_5	0.01	0.01	0.01
Nuclei	Zernike_7_7	0.01	0.01	0.00
Nuclei	Zernike_8_0	0.01	0.01	0.01
Nuclei	Zernike_8_2	0.01	0.01	0.00
Nuclei	Zernike_8_4	0.01	0.01	0.00
Nuclei	Zernike_8_6	0.01	0.01	0.00
Nuclei	Zernike_8_8	0.01	0.01	0.00
Nuclei	Zernike_9_1	0.01	0.01	0.00

x=0.395161 y=0.997396

Figura 13. Relação das mensurações dos parâmetros identificados para cada objeto primário de estudo.

3.6.7 - Etapa 7 - FilterObjects

Com o objetivo de suprimir artefatos não celulares que possam interferir na análise e tomada das mensurações do objeto primário, procedeu-se nesta etapa do algoritmo CellProfiler a aplicação de filtro de imagem eliminando os artefatos e preservando apenas os núcleos celulares (Figura 14 e Figura 15A e 15B).

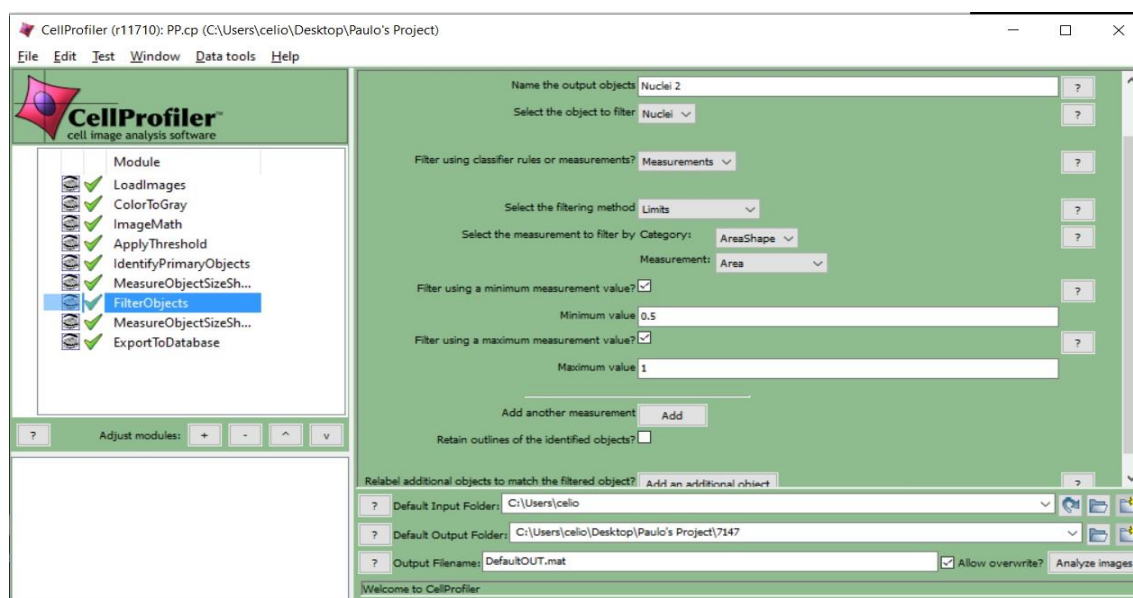


Figura 14. Algoritmo CellProfiler Etapa 7 - FilterObjects.

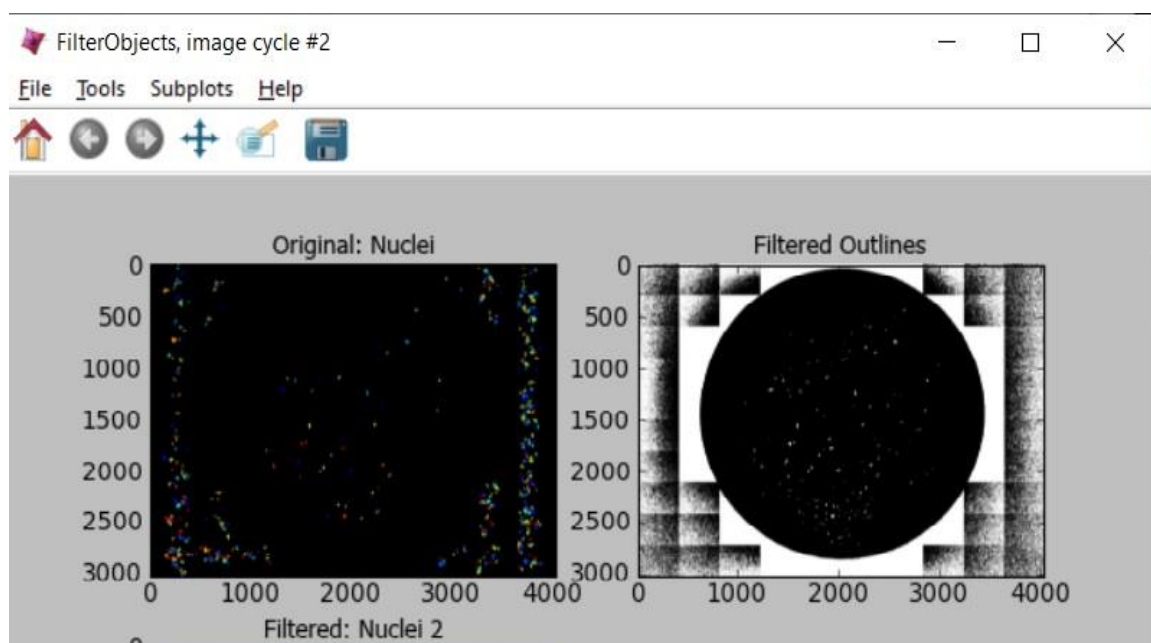


Figura 15A. Aplicação de filtro de imagem para eliminação de artefatos não celulares.

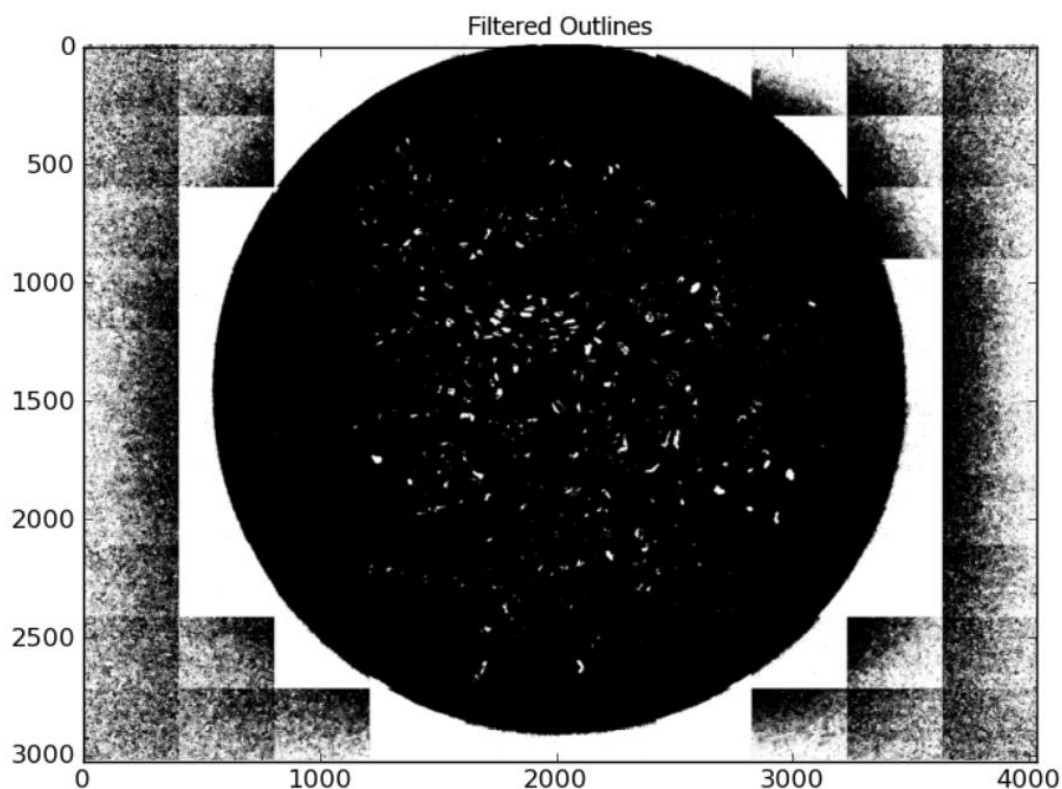


Figura 15B. Núcleos celulares, objetos de estudo, definidos após a aplicação de filtro de imagem para eliminação de artefatos não celulares (imagem magnificada).

3.6.8 - Etapa 8 – *MeasureObjectSizeShape*.

Após a aplicação do filtro de imagens com eliminação dos artefatos não celulares, foi realizada nova mensuração dos parâmetros dos núcleos celulares (objetos primários) no sentido da confirmação de suas medidas e parâmetros sem interferências dos artefatos (Figuras 16 e 17).

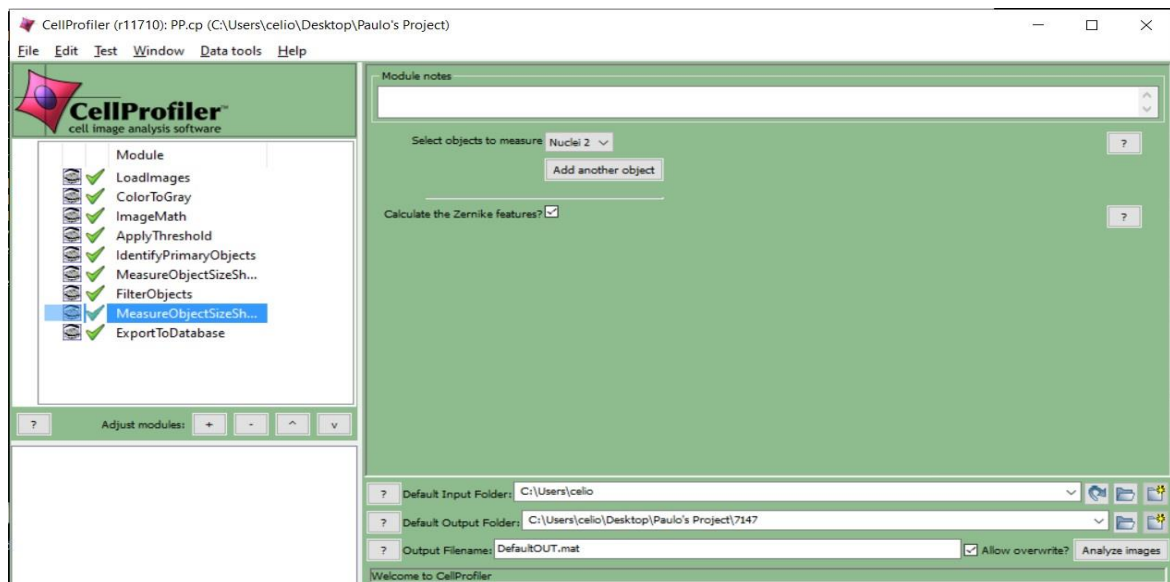


Figura 16. Algoritmo CellProfiler Etapa 8 – MeasureObjectSizeShape

MeasureObjectSizeShape, image cycle #2

File Tools Subplots Help

Nuclei	Orientation	-1.73	-5.03	52.55
Nuclei	Compactness	1.66	1.56	0.47
Nuclei	Area	652.57	515.50	357.72
Nuclei	Center X	2350.09	3302.00	1488.25
Nuclei	Center Y	1759.37	2012.50	934.25
Nuclei	Extent	0.49	0.47	0.11
Nuclei	Perimeter	197.26	173.77	90.99
Nuclei	Solidity	0.69	0.68	0.11
Nuclei	FormFactor	0.26	0.23	0.15
Nuclei	EulerNumber	0.99	1.00	0.15
Nuclei	Zernike_0_0	0.43	0.41	0.12
Nuclei	Zernike_1_1	0.05	0.05	0.03
Nuclei	Zernike_2_0	0.12	0.13	0.03
Nuclei	Zernike_2_2	0.06	0.06	0.02
Nuclei	Zernike_3_1	0.03	0.03	0.02
Nuclei	Zernike_3_3	0.02	0.02	0.01
Nuclei	Zernike_4_0	0.02	0.02	0.01
Nuclei	Zernike_4_2	0.03	0.03	0.01
Nuclei	Zernike_4_4	0.02	0.02	0.01
Nuclei	Zernike_5_1	0.02	0.01	0.01
Nuclei	Zernike_5_3	0.02	0.02	0.01
Nuclei	Zernike_5_5	0.01	0.01	0.01
Nuclei	Zernike_6_0	0.01	0.01	0.01
Nuclei	Zernike_6_2	0.01	0.01	0.01
Nuclei	Zernike_6_4	0.01	0.01	0.01
Nuclei	Zernike_6_6	0.01	0.02	0.01
Nuclei	Zernike_7_1	0.01	0.01	0.00
Nuclei	Zernike_7_3	0.01	0.01	0.01
Nuclei	Zernike_7_5	0.01	0.01	0.01
Nuclei	Zernike_7_7	0.01	0.01	0.00
Nuclei	Zernike_8_0	0.01	0.01	0.01
Nuclei	Zernike_8_2	0.01	0.01	0.00
Nuclei	Zernike_8_4	0.01	0.01	0.00
Nuclei	Zernike_8_6	0.01	0.01	0.00
Nuclei	Zernike_8_8	0.01	0.01	0.00
Nuclei	Zernike_9_1	0.01	0.01	0.00

x=0.395161 y=0.997396

Figura 17. Relação das mensurações dos atributos identificados nos núcleos celulares.

3.6.9 - Etapa 9 – ExportToDatabase.

Concluídas as etapas do algoritmo CellProfiler, foram identificados 47 parâmetros quantitativos (atributos) para cada objeto primário estudado, obtidos a partir de dados qualitativos das 1.937 imagens digitalizadas de câncer de mama em mulheres, e que foram definidos como os parâmetros que permitem identificar e analisar de forma individualizada cada um dos objetos primários (Figuras 18).

A Base de Dados construída pelo software CellProfiler foi composta de 1.937 imagens e de 47 parâmetros foi disposta em uma planilha do aplicativo Excel do Sistema Operacional Windows onde cada linha corresponde a 1 imagem e cada coluna corresponde a 1 parâmetro, sendo que os valores obtidos para cada parâmetro representam a média das mensurações de todos os núcleos celulares contidos na imagem (Figura 19).

Essa Base de Dados foi transferida para o software de análise de dados e classificação de imagens celulares, o software Tanagra, responsável pela realização da segunda etapa do estudo que foi a classificação automatizada das imagens celulares para os scores tubular, nuclear, índice mitótico e o grau histológico de malignidade.

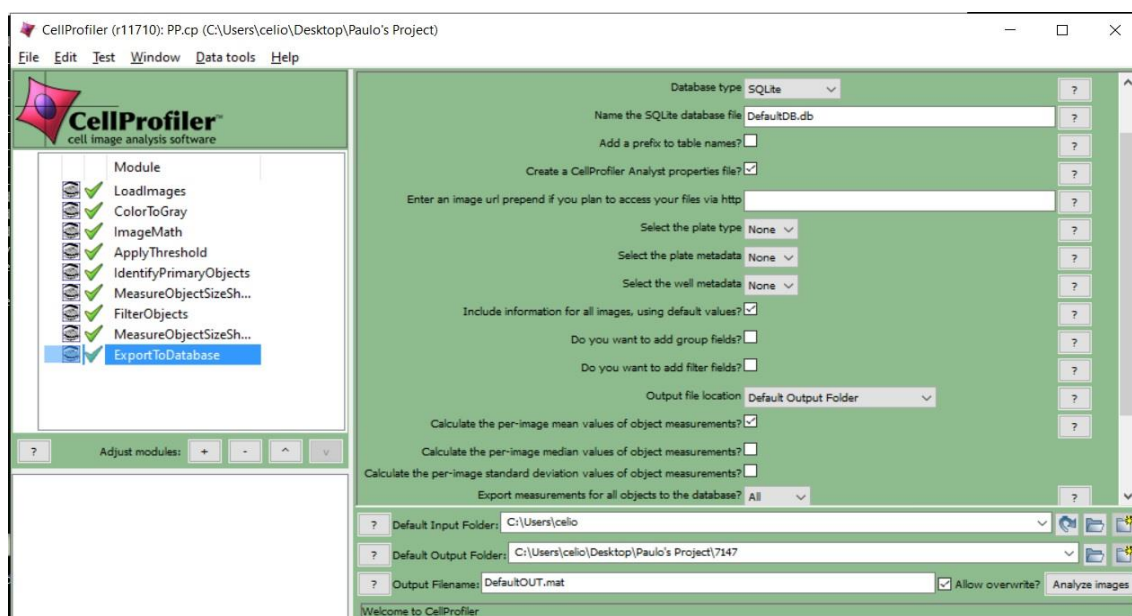


Figura 18. Algoritmo CellProfiler Etapa 9 – ExportToDatabase.

caso	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	46	2	6	3	2	1	7.481.992	2055.002	1627.373	1.6400939	0.7753838	0.9902344	0.4936868	0.2832074	46.202198	26.210161	0.6221365	196.03305	0.7097046	0.4247353	0.0502453	0.1294023	0.059
2	46	2	6	3	2	1	720.0122	2215.672	1433.894	1.6048133	0.7882495	0.9877800	0.4990199	0.2976315	45.401596	25.368277	0.5987541	185.52567	0.7182702	0.4247724	0.0532020	0.1297078	0.060
3	46	2	6	3	2	1	737.10986	19.839.735	1571.8296	1.614263	0.7719976	0.9886364	0.4995327	0.2813829	45.668877	26.192661	0.6359959	196.05801	0.7070021	0.4292709	0.0532256	0.1297066	0.058
4	46	2	6	3	2	1	727.8902	2031.6205	1598.5569	1.6188813	0.7629762	0.9942196	0.4923899	0.2795578	45.238216	26.336695	0.5158581	195.98212	0.7044799	0.4305555	0.0491079	0.1292783	0.057
5	46	2	6	3	2	1	734.6869	2070.3562	1481.1837	1.6391894	0.7824085	0.9920128	0.4896129	0.2850445	46.12088	25.798557	0.6163849	194.27106	0.7076329	0.4261944	0.0509435	0.1287664	0.060
6	46	2	6	3	2	1	717.53876	1923.8257	1631.2306	1.6039797	0.7784439	0.9911971	0.4951377	0.2880537	45.13214	25.68448	0.695087	187.51144	0.7143471	0.4257783	0.0514849	0.1301787	0.059
7	46	2	6	3	2	1	708.3752	2042.3263	1611.1941	1.6021938	0.7775518	0.9885807	0.5004241	0.3007682	44.831417	25.51689	0.6071014	184.50075	0.7190063	0.4281069	0.0526770	0.1304332	0.058
8	46	2	6	3	2	1	717.942	2070.6362	1657.6011	1.6113104	0.7871018	0.9789104	0.4979329	0.2993685	45.482094	25.23518	0.6088487	187.02448	0.720436	0.4284062	0.0504986	0.1297126	0.061
9	46	2	6	3	2	1	712.6189	2044.9283	1676.9882	1.6280221	0.7840951	0.9867924	0.4956317	0.2868243	45.231956	25.384272	0.4629381	188.20096	0.7096605	0.4240461	0.0509445	0.1295519	0.058
10	46	2	6	3	2	1	749.2998	2077.7289	1739.7163	1.6030642	0.7670494	0.983842	0.4915148	0.2700787	45.7666	26.49845	0.4861875	196.41243	0.7022196	0.4285212	0.0515559	0.1296870	0.057
11	46	2	6	3	2	1	723.4323	2032.1927	1569.6493	1.5956812	0.7684307	0.9861111	0.5029617	0.2945948	44.941063	25.909857	0.217893	188.32608	0.7148291	0.4336890	0.0510583	0.1317570	0.058
12	46	2	6	3	2	1	734.8889	1894.57	1528.7742	1.5956719	0.7788153	0.9874552	0.5057727	0.3077955	45.732372	25.908194	0.586332	190.48567	0.7201609	0.4367378	0.0508347	0.1313206	0.061
13	46	2	6	3	2	1	731.9033	2020.8154	1434.3934	1.6018896	0.7605274	0.9912088	0.4997035	0.2855478	45.039436	26.380817	0.2192257	195.12798	0.7054345	0.4411740	0.0515636	0.1273921	0.058
14	46	2	6	3	2	1	750.39954	1820.4366	1492.5153	1.6422238	0.7756295	0.9934498	0.4868748	0.2621314	46.624073	26.512833	0.2572892	202.84834	0.6958266	0.4233554	0.0508982	0.1283339	0.058
15	46	2	6	3	2	1	745.38043	2041.0592	1605.1572	1.6050323	0.7728845	0.9908883	0.4906529	0.2747624	45.64785	26.532507	0.056620	199.46709	0.7026891	0.4330431	0.0509435	0.1309485	0.058
16	46	2	6	3	2	1	741.4492	2079.0627	1465.082	1.6800073	0.7688570	0.9935205	0.4831221	0.2607510	46.65493	26.46181	0.063362	204.19536	0.6917491	0.4195878	0.0509616	0.1262946	0.056
17	46	2	6	3	2	1	725.6751	1960.8549	1601.0807	1.6702482	0.7827207	0.9815668	0.4815180	0.2595733	46.275505	26.103453	0.879139	200.54967	0.6915756	0.4191999	0.0497034	0.1251011	0.058
18	46	2	6	3	2	1	716.8862	2072.5623	1532.105	1.61145	0.770809	0.9824945	0.4900144	0.2704728	45.039513	26.205315	0.235310	198.16367	0.6963936	0.4297920	0.0525652	0.1277088	0.057
19	46	2	6	3	2	1	750.05096	1986.1093	1518.6019	1.641517	0.7736955	0.9830097	0.4879251	0.2629265	46.477535	26.79106	0.2664228	205.0933	0.6901526	0.4234374	0.0509435	0.1270913	0.058
20	46	2	6	3	2	1	716.8862	2072.5623	1532.105	1.61145	0.770809	0.9824945	0.4900144	0.2704728	45.039513	26.205315	0.235310	198.16367	0.6963936	0.4297920	0.0525652	0.1277088	0.057
1937	7015	2	6	3	2	1	744.0612	2020.9093	1700.22	1.6571178	0.7883754	0.9818594	0.4869903	0.2661222	46.83466	25.9193	0.758952	197.33472	0.6959604	0.4197617	0.0517995	0.124673	0.06
47	7015	2	6	3	2	1	724.4044	1962.04	1754.1355	1.5713758	0.7667734	0.9866667	0.4970266	0.2838507	44.604115	26.052858	1.7087245	191.5172	0.7092609	0.4385261	0.0525185	0.1304505	0.059

Figura 19. Base de Dados do software CellProfiler com 1937 imagens e 47 parâmetros.

→ seta verde indica que cada linha corresponde a 1 imagem.

→ seta vermelha indica que cada coluna corresponde a 1 parâmetro.

3.7 – O Software Tanagra

O software Tanagra é um software de livre acesso concebido no ano de 2004 pelo cientista Ricco Rakotomalala e desenvolvido sob a concepção do princípio de "machine learning" com capacidade para realizar análise de imagens celulares a partir de Base de Dados, proceder classificações automatizadas de

imagens celulares e realizar simultaneamente análises estatísticas (RAKOTOMALALA, 2005)

No presente estudo, o software Tanagra foi utilizado para realizar a classificação automatizada do grau de malignidade dos cânceres de mama para os scores tubular, nuclear e índice mitótico, bem como para o grau histológico analisando os 3 (três) parâmetros utilizados na definição do grau histológico em cânceres de mama que são o aspecto tubular, a morfologia nuclear e a contagem de células em mitose, a partir da análise da base de dados contendo 47 parâmetros quantitativos das 1.937 imagens digitalizadas de câncer de mama em mulheres no aumento de 40 vezes.

ANÁLISE ESTATÍSTICA

A análise dos dados epidemiológicos deu-se por análise descritiva, sendo as variáveis categóricas expressas pelas frequências absolutas e relativas e a distribuição das variáveis quantitativas contínuas foi avaliada mediante a determinação de suas medidas de posição central e variabilidade (mediana, média e desvio padrão).

Foram determinadas para os scores tubular, nuclear e índice mitótico que em conjunto definem o grau histológico de malignidade do câncer de mama, os parâmetros estatísticos dos Valores Preditivos Positivos, Acurácia, Erro (classificação incorreta) e o Índice *Kappa* de concordância entre os analisadores médico patologista e o programa de análise automática de imagens celulares Tanagra, utilizando as análises estatísticas realizadas pelo software Tanagra e o software estatístico MedCalc.

Esses mesmos tratamentos estatísticos foram aplicados para os resultados obtidos para a classificação automatizada do grau histológico de malignidade realizada pelo software Tanagra.

RESULTADOS E DISCUSSÃO

5 – Dados epidemiológicos da amostra.

A faixa etária mais afetada pelo câncer de mama na amostra estudada correspondeu a esperada para a doença que aumenta a sua incidência com aumento da idade, com uma mediana no estudo de 55 anos de idade (Tabela 1).

Tabela 1. Estatísticas descritivas quanto a idade das pacientes.

Variáveis	Menor Valor	Maior Valor	Mediana	Média	Desvio-padrão
Idade (em anos)	28	102	55,00	56,06	13,96

As 224 mulheres participantes do estudo apresentavam classificação do grau histológico de seus cânceres de mama distribuídos nos 3 graus conforme disposto na Tabela 2.

Tabela 2. Caracterização da amostra quanto ao Grau Histológico.

Grau Histológico do Tumor	n	%
Grau 1	7	3,13
Grau 2	136	60,71
Grau 3	81	36,16
Total	224	100,0

O tempo transcorrido entre a primeira consulta, o diagnóstico e o início do tratamento foram em sua mediana inferiores a 60 dias, atendendo o que preconiza a legislação brasileira (Tabela 3) (BRASIL L.F. N° 12.732 /2012).

Tabela 3. Caracterização da amostra quanto tempo entre diagnóstico e tratamento.

Variáveis	Menor Valor	Maior Valor	Mediana	Média	Desvio-padrão
Idade (em anos)	28	102	55,00	56,06	13,96
Tempo entre diagnóstico e início do tratamento (em dias)	0	598,00	38,00	46,06	54,01
Tempo entre tratamento e primeira consulta (em dias)	0	388,00	42,00	50,14	44,16

O Estado do Espírito Santo possui 4.018.650 habitantes com leve preponderância da população feminina com 50,75% de mulheres e 49,25% de homens, predominando na população a cor parda (48,6%) seguida da branca (42,2%) e preta (8,5%) (IBGE, 2019). Nesse estudo, a grande maioria das mulheres eram pardas (76,3%), de baixa escolaridade, casadas e sem histórico de etilismo ou tabagismo (Tabela 4).

O histórico familiar de câncer, considerado importante fator de risco quando há relato de ocorrências em parentes consanguíneos ascendentes ou colaterais até a segunda geração (pais, avós, tios e irmãos), não predominou de forma destacada embora se deva ressaltar o alto número do registro “sem informação” que correspondeu a 35% dos dados relatados. (Tabela 4).

As principais ocupações profissionais das mulheres participantes do estudo encontram-se descritas na Tabela 5 seguindo a Classificação Brasileira de Ocupações (CBO); destacando-se que a procedência dessas pacientes foram dos diversos municípios do estado do Espírito Santo e até de estados vizinhos, embora com predominância dos municípios que compõem a chamada Grande Vitória (Vitória, Vila Velha, Serra, Cariacica, Viana, Guarapari), reafirmando o carácter de referência hospitalar em oncologia do Hospital Santa Rita de Cássia. (Tabela 6)

Tabela 4. Caracterização da amostra: Dados epidemiológicos.

Variáveis	n	%
Raça / Cor		
Branca	43	19,1
Preta	4	1,8
Amarela	2	0,9
Parda	171	76,3
Sem informação	4	1,8
Escolaridade		
Nenhuma	17	7,6
Fundamental incompleto	112	50,0
Fundamental completo	16	7,1
Nível médio	49	21,9
Superior incompleto	3	1,3
Superior completo	23	10,2
Sem informação	4	1,8
Estado conjugal atual		
Solteira	61	27,2
Casada	103	46,0
Viúva	38	17,0
Separada judicialmente	18	8,0
União consensual	3	1,3
Sem informação	1	0,4
Histórico familiar de câncer		
Sim	74	33,0
Não	72	32,0
Sem informação	78	35,0
Histórico de consumo de bebida alcoólica		
Nunca	143	63,8
Ex - consumidor	13	5,8
Sim	13	5,8
Sem informação	55	24,5
Histórico de consumo de tabaco		
Nunca	133	59,3
Ex - consumidor	22	9,9
Sim	18	8,0
Sem informação	51	22,7
Total	224	100,0

Tabela 5. Caracterização da amostra quanto a ocupação profissional.

Ocupação	n	%
Trabalhadores de serventia (domicílios e hotéis) e trabalhadores assemelhados (doméstica, copeiro, baba, caseiro)	28	12,5
Trabalhadores agropecuários polivalentes e trabalhadores assemelhados (lavrador)	21	9,4
Cozinheiros e trabalhadores assemelhados	15	6,7
Professores não classificados sob outras epígrafes	10	4,5
Técnicos de enfermagem e trabalhadores assemelhados (exceto enfermeiros)	9	4,0
Trabalhadores de serviços de conservação, manutenção, limpeza de edifícios, empresas comerciais, etc.	9	4,0
Agentes administrativos, assistentes administrativos e trabalhadores assemelhados (secretaria)	6	2,7
Cabeleireiros, especialistas em tratamentos de beleza e trabalhadores assemelhados (manicure)	5	2,2
Comerciantes (comercio atacadista e varejista)	3	1,4
Trabalhadores agrícolas especializados não classificados sob outras epígrafes (agricultor)	3	1,4
Costureiros (confecção em série)	3	1,4
Assistentes sociais	2	0,9
Contadores	2	0,9
Funcionários públicos superiores	2	0,9
Vendedores de comercio atacadista, varejista e trabalhadores assemelhados	1	0,4
Nutricionistas e trabalhadores assemelhados	1	0,4
Advogados	1	0,4
Bibliotecários, arquivologistas e museólogos	1	0,4
Analistas de ocupações e trabalhadores assemelhados	1	0,4
Gerentes financeiros, comerciais, de marketing e de publicidade	1	0,4
Agentes de administração de empresas públicas e privadas não classificados sob outras epígrafes	1	0,4
Auxiliares de escritório e trabalhadores assemelhados (secretaria)	1	0,4
Receptionistas	1	0,4
Vendedores praticistas, representantes comerciais e trabalhadores assemelhados (caixeiro viajante)	1	0,4
Mordomos, governantas e trabalhadores assemelhados	1	0,4
Policiais e trabalhadores assemelhados	1	0,4
Trabalhadores de preparação de alimentos e bebidas não classificados sob outras epígrafes (merendeira)	1	0,4
Não se aplica	73	32,6
Sem informação	20	8,9
Total	224	100,0

Tabela 6. Caracterização da amostra quanto ao Município de procedência.

Procedência	n	%
Vitória	34	15,2
Serra	29	13,0
Cariacica	25	11,2
Vila Velha	23	10,3
Guarapari	11	5,0
Viana	8	3,6
Afonso Cláudio	6	2,7
Linhares	6	2,7
São Mateus	6	2,7
Anchieta	5	2,2
Aracruz	5	2,2
Nova Venécia	5	2,2
Venda Nova do Imigrante	4	1,8
Ibatiba	3	1,3
Itarana	3	1,3
Pedro Canário	3	1,3
Santa Teresa	3	1,3
São Domingos do Norte	3	1,3
Vila Pavão	3	1,3
Água Doce do Norte	2	0,9
Baixo Guandu	2	0,9
Barra de São Francisco	2	0,9
Colatina	2	0,9
Conceição da Barra	2	0,9
Domingos Martins	2	0,9
João Neiva	2	0,9
Montanha	2	0,9
São Gabriel da Palha	2	0,9
Sooretama	2	0,9
Teixeira de Freitas-BA	1	0,4
Águia Branca	1	0,4
Alegre	1	0,4
Brejetuba	1	0,4
Cachoeiro de Itapemirim	1	0,4
Fundão	1	0,4
Governador Lindenberg	1	0,4
Ibitirama	1	0,4
Iúna	1	0,4
Jaguare	1	0,4
Laranja da Terra	1	0,4
Mantenópolis	1	0,4
Marechal Floriano	1	0,4
Pancas	1	0,4
Pinheiros	1	0,4
Piúma	1	0,4
Santa Maria de Jetibá	1	0,4
São Roque do Canaã	1	0,4
Vila Valério	1	0,4
Total	224	100,0

A grande maioria das pacientes (72,3%) chegaram ao Hospital Santa Rita de Cássia em Vitória, Espírito Santo sem diagnóstico firmado ou qualquer tipo de tratamento iniciado, demonstrando que a Rede Básica de Atendimento Público precisa ainda ser melhor estruturada e as ações que visem o diagnóstico precoce da doença necessitam ser mais efetivas, uma vez que algumas dessas mulheres foram diagnosticadas com câncer de mama em seu tipo mais comum (adenocarcinoma ductal infiltrante) em estágios mais avançados, já com comprometimento linfonodal axilar ou mesmo metástases a distância para fígado, ossos, pulmão ou mediastino (Tabelas 7 e 8).

Apesar disso, quando tiveram a oportunidade de serem atendidas em uma estrutura hospitalar adequada para o tratamento oncológico, o diagnóstico e o início do tratamento ocorreram de forma geral em tempo ainda suficiente para se obter o controle inicial da doença na maioria das pacientes conforme Tabela 9, destacando-se que a base mais importante para o diagnóstico e para o tratamento foi a determinação do tipo histológico do tumor (Tabela 7).

Desafortunadamente, a despeito do tratamento instituído, houve progressão da doença em 3 pacientes (1,4%) e 4 pacientes (1,8%) evoluíram para óbito no período do estudo (Tabelas 9 e 10).

Tabela 7. Dados epidemiológicos do Tumor.

Variáveis	n	%
Diagnóstico e tratamento anteriores		
Sem diagnóstico / Sem tratamento	162	72,3
Com diagnóstico / Sem tratamento	60	26,8
Com diagnóstico / Com tratamento	2	0,9
Base mais importante para o diagnóstico do tumor		
Histologia do tumor primário	223	99,6
Sem informação	1	0,4
Localização do tumor primário		
C50.9	183	81,7
C50.4	34	15,2
C50.8	3	1,4
C50.0	2	0,9
C50.3	1	0,4
C50.5	1	0,4
Tipo histológico do tumor primário		
8010/3	1	0,4
8500/2	1	0,4
8500/3	204	91,1
8503/3	1	0,4
8520/3	17	7,6
TNM		
100	24	10,7
110	5	2,3
120	2	0,9
200	52	23,2
201	1	0,4
210	19	8,5
220	1	0,4
230	1	0,4
300	5	2,3
310	3	1,4
400	1	0,4
421	1	0,4
430	1	0,4
999	108	48,3
Total	224	100,0

Tabela 8. Estadiamento Clínico do tumor (TNM).

Variáveis	n	%
Estadiamento clínico do tumor (TNM)		
1	47	21,0
2A	81	36,1
2B	24	10,7
3A	28	12,5
3B	2	0,9
3C	6	2,7
4	11	5,0
99	25	11,1
pTNM		
100	33	14,8
110	10	4,5
120	3	1,4
200	46	20,5
210	26	11,6
211	1	0,4
220	13	5,8
230	3	1,4
300	4	1,8
310	6	2,7
320	7	3,2
321	1	0,4
330	3	1,4
410	1	0,4
411	1	0,4
420	1	0,4
421	1	0,4
430	1	0,4
431	1	0,4
888	2	0,9
9	1	0,4
999	59	26,4
Localização de metástase à distância		
C22	1	0,4
C22 C41	1	0,4
C41	6	2,7
C41 C34	2	0,9
Sem informação	213	95,1
Total	224	100,0

Tabela 9. Estado da doença ao final do 1º tratamento.

Variáveis	n	%
Estado da doença ao final do primeiro tratamento no hospital		
Sem evidência da doença (remissão completa)	28	12,5
Remissão parcial	2	0,9
Doença estável	142	63,4
Doença em progressão	3	1,4
Óbito	4	1,8
Não se aplica	1	0,4
Sem informação	44	19,7
Óbito por câncer		
Sim	4	1,8
Não	220	98,2
Exames relevantes		
Exame clínico e patologia clínica	224	100,0
Exames por imagem	224	100,0
Endoscopia e cirurgia exploradora	223	99,6
Anatomia patológica	224	100,0
Marcadores tumorais	208	92,9
Lateralidade do tumor		
Direita	101	45,1
Esquerda	120	53,6
Bilateral	1	0,4
Não se aplica	2	0,9
Ocorrência de mais de um tumor primário		
Sim	-	-
Não	224	100,0
Total	224	100,0

Tabela 10. Causa básica da morte e fonte de custeio do tratamento.

Variáveis	n	%
Custeio do diagnóstico do tumor no Hospital		
Público (SUS)	155	69,2
Plano de saúde	2	0,9
Particular	1	0,4
Não se aplica	66	29,5
Custeio do tratamento do tumor no Hospital		
Público (SUS)	220	98,2
Plano de saúde	2	0,9
Causa básica da morte do paciente		
C50.X	4	1,8
Sem óbito	220	98,2
Total	224	100,0

5.1 - Artefatos em Histologia.

A revisão das lâminas histológicas dos casos selecionados para o estudo realizada por médico patologista teve dentre os objetivos, identificar as de melhor aspecto quanto a preservação da coloração e eliminar as que apresentavam artefatos histológicos, fatores que poderiam interferir no processo de aprendizado de máquina e por consequência na análise automatizada realizada pelo softwares CellProfiler e Tanagra.

Artefatos em Histologia são causas de dificuldades para interpretação de imagens e podem resultar em erros diagnósticos, tendo como causas mais frequentes problemas no processamento do tecido histológico gerando cortes histológicos de espessura inadequada (muito fina ou muito grossa), dobras, queima ou destacamentos de parte dos cortes histológicos, alterações na coloração do tecido pelo uso de corantes com preparo inadequado com intensidade fraca ou forte interferindo no aspecto de sua coloração (CHARTTERJEE, 2014), (SAMAR et al., 2014), (JIMSON et al., 2016).

5.2 - Planilha da Base de Dados do software Tanagra.

O software Tanagra ao receber a planilha de Base de Dados transferida do software CellProfiler contendo 1.937 imagens e 47 parâmetros quantitativos, identificou inconsistências e perdas de dados em 438 imagens que foram excluídas pelo software Tanagra de sua Base de Dados sendo confirmados os 47 parâmetros quantitativos (atributos), consistindo a Base de Dados analisada pelo software Tanagra de 1499 imagens e 47 parâmetros (Figura 20).

As 438 imagens excluídas pelo software Tanagra representaram uma perda de 24 casos dos 224 casos da amostra inicial; entretanto não houve alteração significativa em relação a frequência da distribuição dos casos quanto a classificação do grau histológico de malignidade mantendo-se a distribuição percentual semelhante a amostra original (Quadro 6).

Download information

Datasource processing	
Computation time	15 ms
Allocated memory	367 KB
Missing or inconsistent values	
#excluded rows	438

Dataset description

1499 example(s)

Figura 20. Exclusão de 438 imagens (linhas) realizada pelo software Tanagra na Base de Dados transferidas do software Cellprofiler.

Grau Histológico	n	%
1	6	3
2	124	62
3	70	35
Total	200	100

Quadro 6. Distribuição dos casos de câncer de mama analisados pelo software Tanagra conforme grau histológico.

Assim a Base de Dados final analisada pelo software Tanagra foi composta de 1499 imagens e 47 atributos correspondendo a 200 casos de câncer de mama em mulheres com uma média de 7 a 8 imagens para cada caso, distribuídos nos 3 graus histológicos de malignidade conforme Quadro 6 e Figura 21.

Salvamento Automático

Pasta1 - Excel

Paulo Boasquevisque

Arquivo

Página Inicial

Inserir

Layout da Página

Fórmulas

Dados

Revisão

Exibir

Suplementos

Ajuda

Tanagra

Execute Tanagra

About...

A1

</

A1 caso

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
337	5823	b	f	C	b	Q	0,601	676,96	2052,67	1529,09	1,62	0,78	0,99	0,50	0,31	44,20	24,94	0,83	187,04	0,72	0,43	0,05	0,13
338	5823	b	f	C	b	Q	0,601	677,23	1920,30	1608,61	1,67	0,79	1,00	0,49	0,29	44,93	25,32	1,59	191,45	0,70	0,42	0,05	0,12
339	5823	b	f	C	b	Q	0,601	676,15	2082,22	1548,55	1,63	0,78	0,99	0,50	0,31	44,23	25,03	0,90	184,47	0,72	0,43	0,05	0,13
340	6083	b	f	C	b	Q	0,557	738,77	1748,60	1443,24	1,61	0,79	0,97	0,49	0,28	45,94	25,72	1,56	191,61	0,71	0,43	0,05	0,13
341	6083	b	f	C	b	Q	0,557	738,80	1537,60	1505,78	1,58	0,76	0,99	0,50	0,29	45,16	26,50	0,62	189,25	0,71	0,44	0,05	0,13
342	6083	b	f	C	b	Q	0,557	749,62	1646,23	1477,91	1,55	0,76	0,98	0,51	0,28	44,90	26,42	0,64	191,44	0,72	0,44	0,05	0,13
343	6083	b	f	C	b	Q	0,557	719,98	1720,71	1493,35	1,58	0,77	0,99	0,50	0,28	44,73	25,79	1,82	188,60	0,71	0,44	0,05	0,13
344	6083	b	f	C	b	Q	0,557	746,37	1686,12	1504,28	1,62	0,77	0,98	0,50	0,29	45,70	26,31	1,44	190,56	0,71	0,45	0,05	0,13
345	6083	b	f	C	b	Q	0,557	735,94	1545,74	1568,03	1,59	0,77	0,98	0,49	0,28	45,34	26,16	0,64	191,68	0,71	0,43	0,05	0,13
346	6083	b	f	C	b	Q	0,557	738,68	1945,64	1609,67	1,56	0,77	0,98	0,51	0,28	44,66	26,39	0,58	190,51	0,71	0,44	0,05	0,13
347	6083	b	f	C	b	Q	0,557	743,95	1759,36	1547,29	1,55	0,77	0,99	0,50	0,30	45,08	25,98	0,62	187,19	0,72	0,44	0,05	0,13
348	6244	b	f	C	b	Q	0,623	723,01	1799,89	1368,43	1,48	0,75	0,98	0,54	0,35	42,93	25,65	1,11	173,87	0,75	0,48	0,05	0,14
349	6244	b	f	C	b	Q	0,623	765,88	1744,23	1429,43	1,57	0,77	0,99	0,51	0,31	46,12	26,50	0,90	192,95	0,72	0,45	0,05	0,13
350	6244	b	f	C	b	Q	0,623	744,23	1626,77	1500,26	1,54	0,77	0,98	0,52	0,32	44,82	26,01	0,03	181,82	0,73	0,45	0,05	0,13
351	6244	b	f	C	b	Q	0,623	725,68	1797,90	1519,22	1,49	0,77	0,99	0,53	0,34	43,23	25,46	0,47	173,33	0,75	0,46	0,05	0,14
352	6244	b	f	C	b	Q	0,623	746,61	1655,29	1502,56	1,52	0,77	0,98	0,52	0,31	44,64	26,03	0,86	184,88	0,73	0,45	0,05	0,14
353	6244	b	f	C	b	Q	0,623	694,52	1771,27	1545,20	1,52	0,77	1,00	0,53	0,34	42,98	24,88	0,39	173,12	0,74	0,46	0,05	0,14
354	6244	b	f	C	b	Q	0,623	716,13	1846,04	1478,70	1,53	0,77	1,00	0,51	0,31	43,53	25,69	0,47	176,13	0,73	0,45	0,05	0,14
355	6244	b	f	C	b	Q	0,623	757,64	1777,73	1548,39	1,52	0,76	0,98	0,52	0,31	44,75	26,45	0,62	186,77	0,73	0,45	0,05	0,14

A1	caso																							
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1476	6844	C	I	C	C	C	0,59	741,62	1897,41	1623,00	1,53	0,76	0,99	0,51	0,29	44,30	26,25	0,62	187,40	0,72	0,45	0,05	0,14	
1477	6844	C	I	C	C	C	0,59	784,17	1545,25	1470,49	1,57	0,77	0,99	0,49	0,27	46,23	27,02	0,17	197,80	0,71	0,43	0,05	0,13	
1478	6844	C	I	C	C	C	0,59	753,94	1582,14	1539,00	1,56	0,76	0,98	0,51	0,28	45,32	26,46	2,59	192,58	0,71	0,44	0,05	0,13	
1479	6844	C	I	C	C	C	0,59	739,95	1819,76	1558,87	1,56	0,77	0,98	0,50	0,28	45,10	25,99	1,54	188,85	0,71	0,44	0,05	0,13	
1480	6844	C	I	C	C	C	0,59	731,75	2063,56	1473,24	1,57	0,76	0,99	0,50	0,28	44,80	26,27	1,56	190,67	0,71	0,44	0,05	0,13	
1481	6940	C	I	C	C	C	0,617	651,75	1984,26	1510,23	1,56	0,78	0,99	0,53	0,35	42,49	24,02	0,91	171,55	0,74	0,46	0,05	0,13	
1482	6940	C	I	C	C	C	0,617	666,30	2032,90	1350,62	1,52	0,78	0,99	0,54	0,35	42,22	24,15	1,85	168,57	0,75	0,47	0,05	0,14	
1483	6940	C	I	C	C	C	0,617	630,53	2167,49	1530,98	1,48	0,77	1,00	0,56	0,39	40,78	23,35	0,08	156,67	0,77	0,48	0,05	0,14	
1484	6940	C	I	C	C	C	0,617	692,38	1992,65	1475,78	1,60	0,78	0,98	0,50	0,28	44,33	25,26	2,76	183,83	0,71	0,43	0,05	0,13	
1485	6940	C	I	C	C	C	0,617	702,22	2168,38	1492,89	1,58	0,77	0,98	0,51	0,30	44,30	25,51	5,19	185,05	0,72	0,45	0,05	0,13	
1486	6940	C	I	C	C	C	0,617	651,64	1739,95	1676,24	1,52	0,75	1,00	0,54	0,34	41,49	24,46	2,60	167,55	0,74	0,47	0,05	0,13	
1487	6940	C	I	C	C	C	0,617	692,78	2052,89	1451,80	1,55	0,75	0,97	0,51	0,30	43,11	25,70	0,43	182,35	0,72	0,45	0,05	0,13	
1488	6940	C	I	C	C	C	0,617	688,97	1898,33	1546,16	1,61	0,78	0,98	0,50	0,29	44,40	25,12	1,98	185,24	0,71	0,43	0,05	0,13	
1489	6940	C	I	C	C	C	0,617	697,29	1703,90	1492,43	1,61	0,79	0,99	0,48	0,25	45,37	25,64	2,76	195,81	0,69	0,41	0,05	0,12	
1490	6940	C	I	C	C	C	0,617	669,44	1850,44	1537,89	1,67	0,79	0,99	0,49	0,27	43,32	25,16	0,70	187,42	0,70	0,43	0,05	0,13	
1491	6943	C	I	C	C	C	0,566	728,76	2032,16	1592,22	1,56	0,77	0,99	0,50	0,29	44,72	26,11	4,84	185,29	0,71	0,44	0,05	0,13	
1492	6943	C	I	C	C	C	0,566	767,29	1946,87	1503,39	1,57	0,76	0,99	0,50	0,27	45,95	26,60	1,29	195,57	0,71	0,44	0,05	0,13	
1493	6943	C	I	C	C	C	0,566	778,38	1796,60	1478,78	1,55	0,77	0,97	0,51	0,29	46,07	26,73	1,43	192,46	0,72	0,45	0,05	0,13	
1494	6943	C	I	C	C	C	0,566	768,53	2036,72	1513,75	1,53	0,76	0,98	0,51	0,29	45,45	26,64	1,20	192,76	0,72	0,45	0,05	0,13	
1495	6943	C	I	C	C	C	0,566	734,90	1965,28	1557,50	1,54	0,77	0,99	0,51	0,29	44,65	26,13	2,11	186,28	0,72	0,45	0,05	0,13	
1496	6943	C	I	C	C	C	0,566	739,67	2333,74	1566,63	1,55	0,78	0,99	0,51	0,29	44,83	26,03	0,24	184,98	0,72	0,44	0,05	0,13	
1497	6943	C	I	C	C	C	0,566	744,98	2007,34	1348,50	1,56	0,77	0,98	0,50	0,29	45,13	26,08	4,19	186,94	0,71	0,44	0,05	0,13	
1498	6943	C	I	C	C	C	0,566	747,23	1964,17	1608,31	1,51	0,76	0,99	0,52	0,31	44,18	26,11	6,17	182,94	0,73	0,46	0,05	0,14	
1499	6943	C	I	C	C	C	0,566	731,46	1990,61	1600,25	1,57	0,77	0,99	0,51	0,30	44,83	25,86	1,16	184,42	0,72	0,44	0,05	0,13	
1500	6943	C	I	C	C	C	0,566	795,98	1857,86	1538,13	1,54	0,76	0,97	0,51	0,29	46,20	26,84	1,00	193,36	0,72	0,45	0,05	0,13	

5.3 – Parâmetros quantitativos (atributos) dos núcleos celulares.

O software CellProfiler produziu a Base de Dados com 47 parâmetros quantitativos denominados de atributos, que são mensurações e aspectos que caracterizam e definem os objetos primários de estudo, os núcleos celulares, e que permitem a partir deles a identificação e classificação das imagens celulares.

Esses atributos foram confirmados pela análise do software Tanagra que reconheceu os 47 parâmetros de identificação das imagens para realizar a partir deles a classificação automatizada das imagens de câncer de mama (Figura 22).

Download information		
Datasource processing		
Computation time 16 ms		
Allocated memory 339 KB		
Dataset description		
47 attribute(s)		
1499 example(s)		
Attribute	Category	Informations
Intensity	Continue	-
Mean_Nuclei_AreaShape_Area	Continue	-
Mean_Nuclei_AreaShape_Center_X	Continue	-
Mean_Nuclei_AreaShape_Center_Y	Continue	-
Mean_Nuclei_AreaShape_Compactness	Continue	-
Mean_Nuclei_AreaShape_Eccentricity	Continue	-
Mean_Nuclei_AreaShape_EulerNumber	Continue	-
Mean_Nuclei_AreaShape_Extent	Continue	-
Mean_Nuclei_AreaShape_FormFactor	Continue	-
Mean_Nuclei_AreaShape_Zernike_5_3	Continue	-
Mean_Nuclei_AreaShape_Zernike_5_5	Continue	-
Mean_Nuclei_AreaShape_Zernike_6_0	Continue	-
Mean_Nuclei_AreaShape_Zernike_6_2	Continue	-
Mean_Nuclei_AreaShape_Zernike_6_4	Continue	-
Mean_Nuclei_AreaShape_Zernike_6_6	Continue	-
Mean_Nuclei_AreaShape_Zernike_7_1	Continue	-
Mean_Nuclei_AreaShape_Zernike_7_3	Continue	-
Mean_Nuclei_AreaShape_Zernike_7_5	Continue	-
Mean_Nuclei_AreaShape_Zernike_7_7	Continue	-
Mean_Nuclei_AreaShape_Zernike_8_0	Continue	-
Mean_Nuclei_AreaShape_Zernike_8_2	Continue	-
Mean_Nuclei_AreaShape_Zernike_8_4	Continue	-
Mean_Nuclei_AreaShape_Zernike_8_6	Continue	-
Mean_Nuclei_AreaShape_Zernike_8_8	Continue	-
Mean_Nuclei_AreaShape_Zernike_9_1	Continue	-
Mean_Nuclei_AreaShape_Zernike_9_3	Continue	-
Mean_Nuclei_AreaShape_Zernike_9_5	Continue	-
Mean_Nuclei_AreaShape_Zernike_9_7	Continue	-
Mean_Nuclei_AreaShape_Zernike_9_9	Continue	-
Mean_Nuclei_Children_Nuclei_2_Count	Continue	-
Mean_Nuclei_Location_Center_X	Continue	-
Mean_Nuclei_Location_Center_Y	Continue	-

Figura 22. Atributos dos núcleos celulares (objetos primários).

Desses 47 parâmetros quantitativos, o software Tanagra definiu os atributos de maior expressão estatística para a análise dos scores tubular, nuclear, índice mitótico e para o grau histológico de malignidade.

5.3.1 – Principais atributos para classificação automatizada do score tubular.

Dos 47 parâmetros quantitativos definidos para cada objeto primário do estudo, foram identificados pelo software Tanagra 8 (oito) atributos com maior expressão estatística (valor $p < \text{ou} = 0,001$) em relação ao score tubular e que são os de maior importância na definição automatizada do grau histológico de malignidade do câncer de mama em relação ao score tubular (Figuras 23 e 24).

Parameters	
Active parameter	0
Parameter	Value
p-value threshold	0,001000
best attributes	10

Results	
INPUT attribute selection	
INPUT selection	
Before filtering	47
After filtering	8
Keeped into INPUT selection	
Attributes	
1	Mean_Nuclei_AreaShape_Zernike_2_0
2	Mean_Nuclei_AreaShape_Extent
3	Mean_Nuclei_AreaShape_Solidity
4	Mean_Nuclei_AreaShape_Zernike_0_0
5	Mean_Nuclei_AreaShape_FormFactor
6	Mean_Nuclei_AreaShape_Perimeter
7	Mean_Nuclei_AreaShape_Compactness
8	intensity

Figura 23. Principais atributos - score tubular.

Calculations details				
N°	Attribute	F	F (max normalized)	p-value (1,1497)
1	Mean_Nuclei_AreaShape_Zernike_2_0	33,08		0,000000
2	Mean_Nuclei_AreaShape_Extent	32,55		0,000000
3	Mean_Nuclei_AreaShape_Solidity	32,03		0,000000
4	Mean_Nuclei_AreaShape_Zernike_0_0	31,52		0,000000
5	Mean_Nuclei_AreaShape_FormFactor	30,22		0,000000
6	Mean_Nuclei_AreaShape_Perimeter	26,12		0,000000
7	Mean_Nuclei_AreaShape_Compactness	23,04		0,000002
8	intensity	12,41		0,000439
9	Mean_Nuclei_AreaShape_Zernike_5_3	9,56		0,002025
10	Mean_Nuclei_AreaShape_MajorAxisLength	7,26		0,007138
11	Mean_Nuclei_AreaShape_MinorAxisLength	5,77		0,016433
12	Mean_Nuclei_AreaShape_Zernike_4_0	3,26		0,071315
13	Mean_Nuclei_AreaShape_Zernike_9_9	2,95		0,085989
14	Mean_Nuclei_AreaShape_Zernike_2_2	2,86		0,090850
15	Mean_Nuclei_AreaShape_Zernike_3_1	2,69		0,101443
16	Mean_Nuclei_AreaShape_Zernike_5_1	2,34		0,126346
17	Mean_Nuclei_AreaShape_Center_X	1,51		0,219210

Figura 24. Principais atributos para o score tubular quanto ao p valor.

5.3.2 - Principais atributos para classificação automatizada do score nuclear.

Em relação ao score nuclear, dos 47 atributos quantitativos definidos para os objetos primários do estudo, foram identificados 3 atributos com maior expressão estatística (valor $p < \text{ou} = 0,001$) e que são os de maior importância na definição automatizada do grau histológico de malignidade em relação ao score nuclear (Figuras 25 e 26).

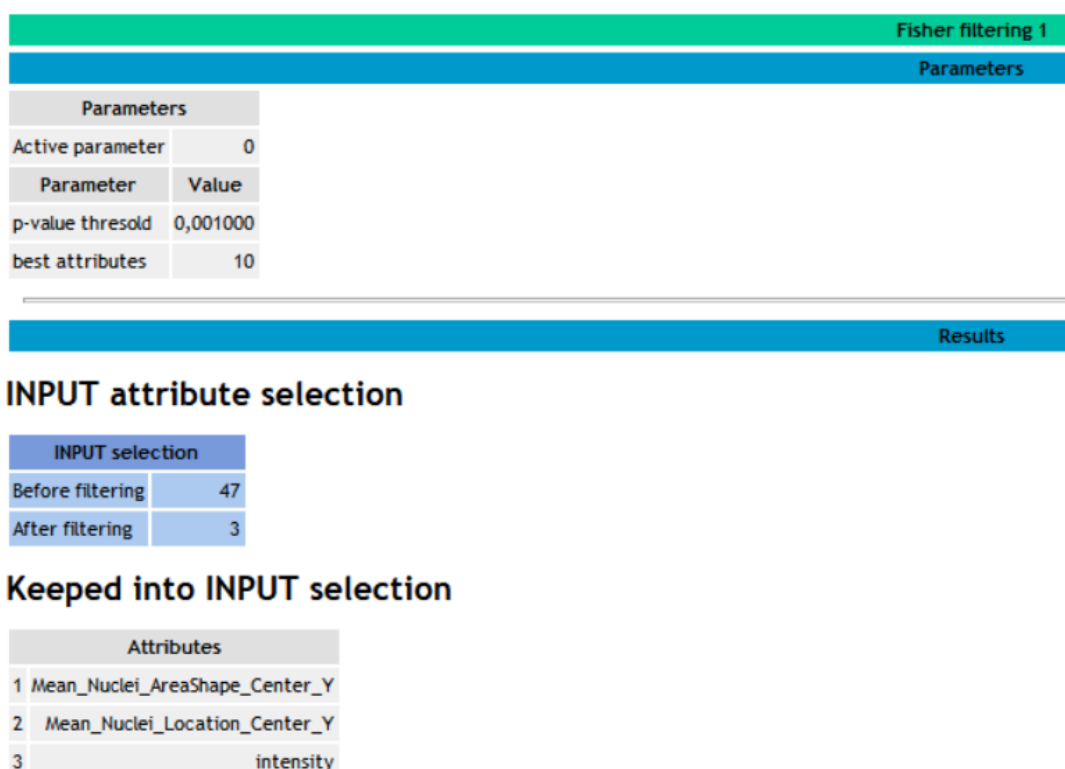


Figura 25. Principais atributos - score nuclear.

Keeped into INPUT selection

Attributes	
1	Mean_Nuclei_AreaShape_Center_Y
2	Mean_Nuclei_Location_Center_Y
3	intensity

Calculations details

N°	Attribute	F	F (max normalized)	p-value (1,1497)
1	Mean_Nuclei_AreaShape_Center_Y	26,36		0,000000
2	Mean_Nuclei_Location_Center_Y	26,24		0,000000
3	intensity	24,05		0,000001
4	Mean_Nuclei_AreaShape_MajorAxisLength	8,88		0,002922
5	Mean_Nuclei_AreaShape_Zernike_5_3	8,05		0,004620
6	Mean_Nuclei_AreaShape_Zernike_5_1	5,72		0,016899
7	Mean_Nuclei_AreaShape_EulerNumber	4,65		0,031270
8	Mean_Nuclei_AreaShape_MinorAxisLength	4,47		0,034592
9	Mean_Nuclei_AreaShape_Perimeter	4,34		0,037459
10	Mean_Nuclei_AreaShape_Zernike_0_0	3,92		0,047981
11	Mean_Nuclei_AreaShape_Zernike_4_0	3,78		0,051994

Figura 26. Principais atributos para o score nuclear quanto ao p valor.

5.3.3 – Principais atributos para classificação automatizada do score índice mitótico.

Dos 47 atributos quantitativos definidos para os núcleos celulares (objetos primários do estudo), foram identificados 2 atributos com maior expressão estatística em relação ao score índice mitótico (valor $p < \text{ou} = 0,001$) e que são os de maior importância na definição automatizada do grau histológico de malignidade em relação ao score índice mitótico (Figuras 27 e 28).

Fisher filtering 1	
Parameters	
Parameters	
Active parameter	0
Parameter	Value
p-value threshold	0,001000
best attributes	10
Results	
INPUT attribute selection	
INPUT selection	
Before filtering	47
After filtering	2
Keeped into INPUT selection	
Attributes	
1	intensity
2	Mean_Nuclei_AreaShape_Zernike_2_0

Figura 27. Principais atributos – índice mitótico.

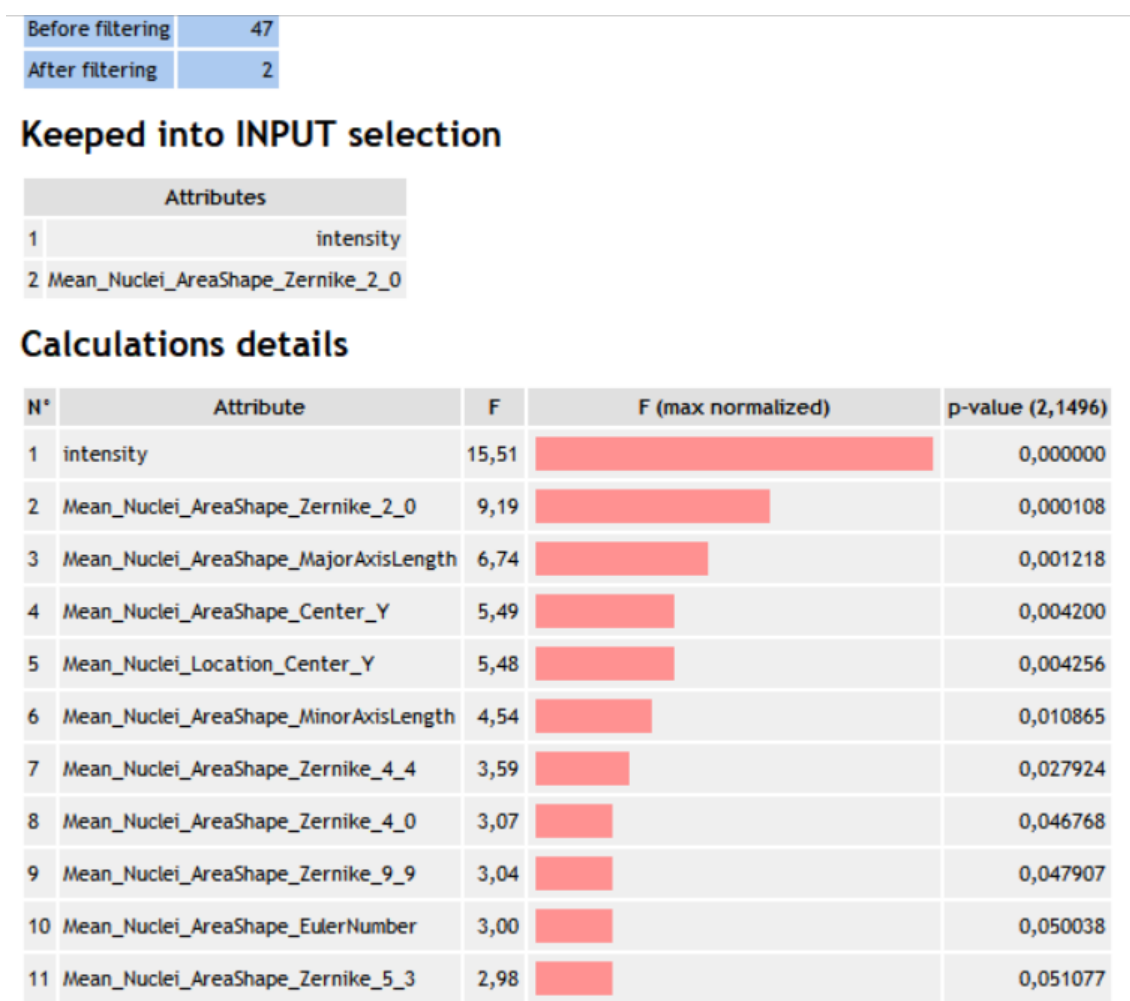


Figura 28. Principais atributos para o score índice mitótico quanto ao p valor.

5.3.4 – Principais atributos para classificação automatizada do grau histológico de câncer de mama.

Dos 47 atributos quantitativos definidos para os núcleos celulares, foram identificados 7 (sete) atributos com maior expressão estatística (valor $p < \text{ou} = 0,001$) e que são os de maior importância para a classificação automatizada do grau histológico de malignidade (Figuras 29 e 30).

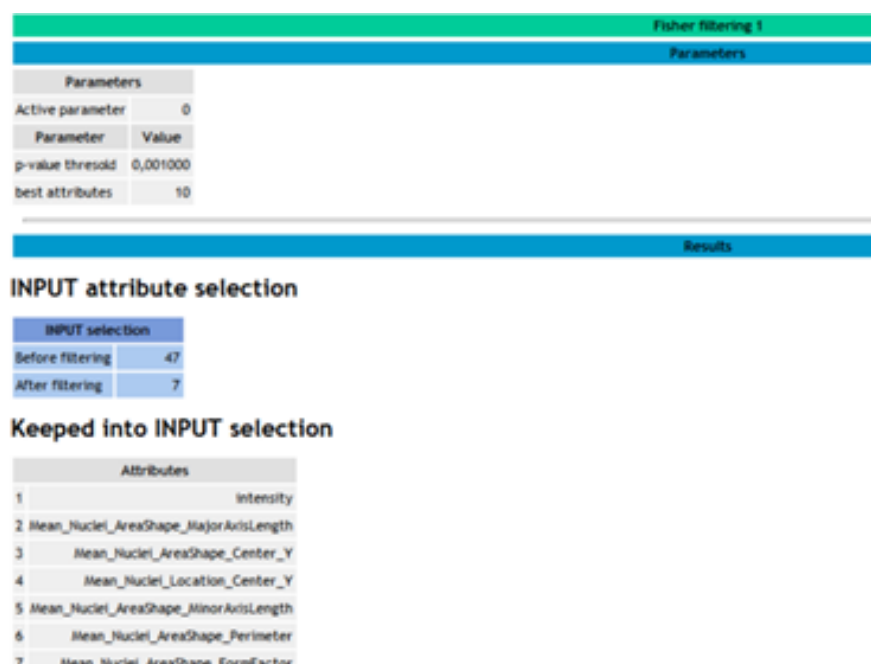


Figura 29. Principais atributos – grau histológico.

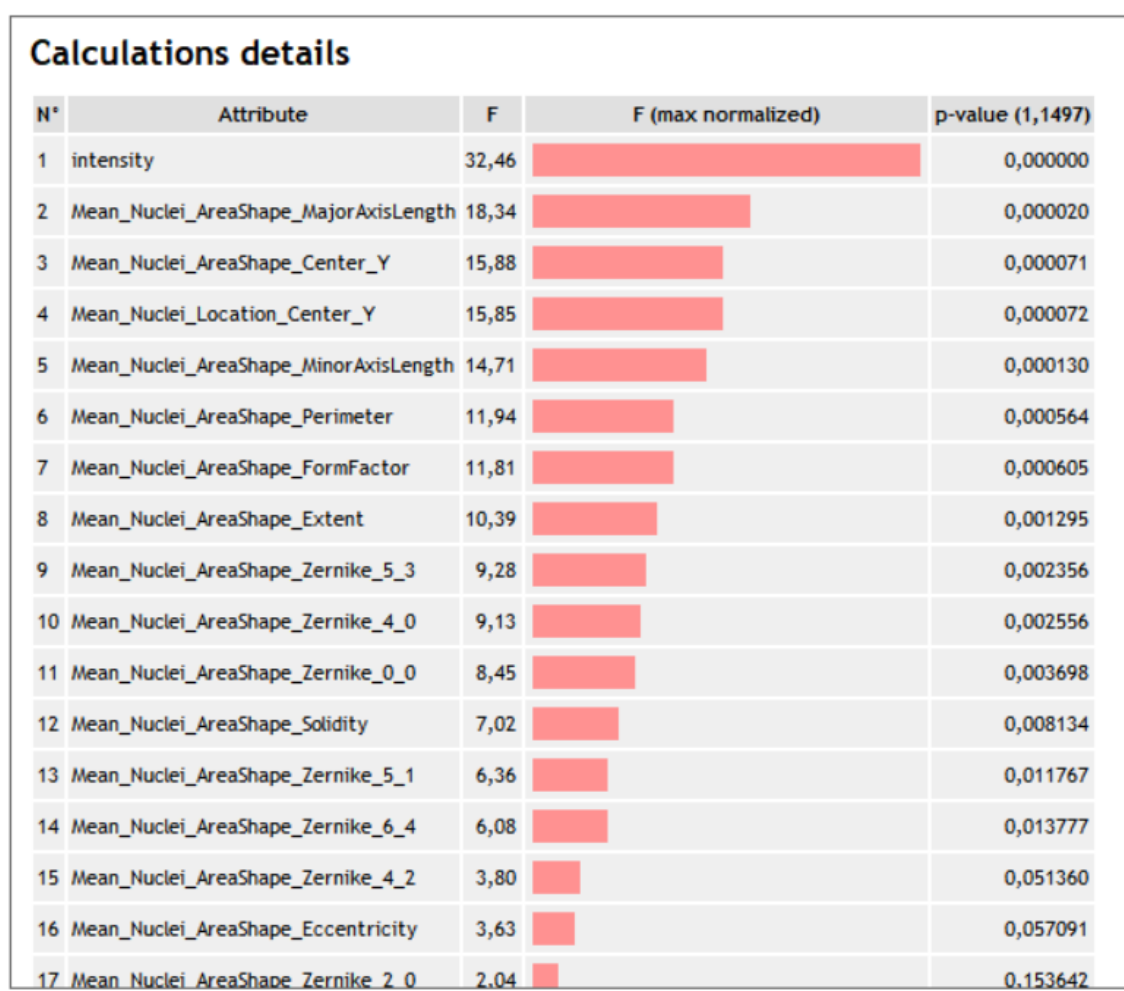


Figura 30. Principais atributos para o grau histológico quanto ao p valor.

5.4 – Distribuição dos casos de câncer de mama quanto aos scores tubular, nuclear e índice mitótico da planilha de Base de Dados analisada pelo software Tanagra.

A análise automatizada dos parâmetros quantitativos da planilha da Base de Dados realizada pelo software Tanagra em relação aos scores tubular, nuclear e índice mitótico e apresentada em gráficos de dispersão, demonstra que o software Tanagra foi capaz de identificar os 3 (três) grupos de classificação patológica para os scores tubular, nuclear e índice mitótico em relação as suas pontuações (1 ponto = a, 2 pontos = b, 3 pontos = c) de acordo com a frequência dos casos e o número de imagens para cada grupo de pontuação, conforme evidenciado nas Figuras 31, 32 e 33.

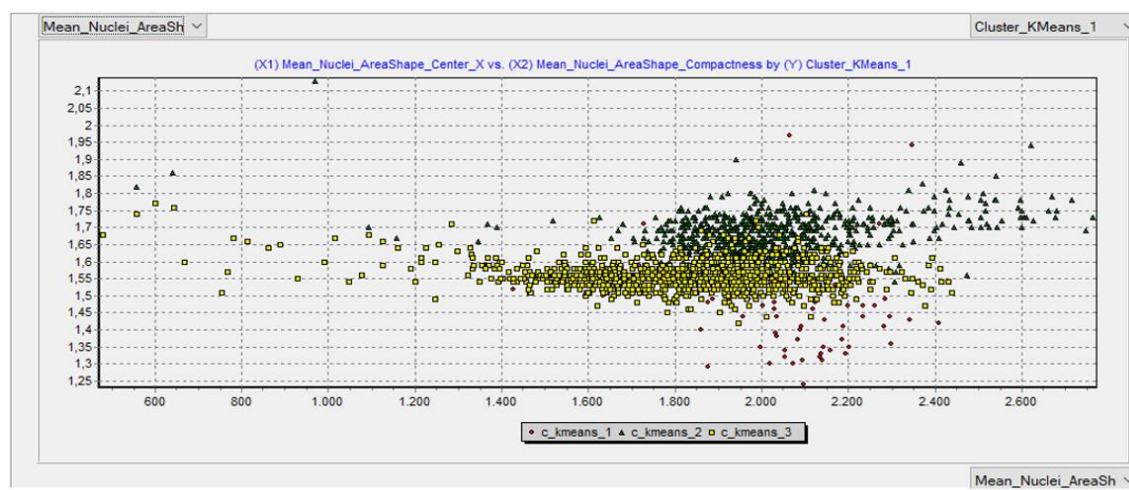


Figura 31. Gráfico de Dispersão de análise automatizada para o score tubular.

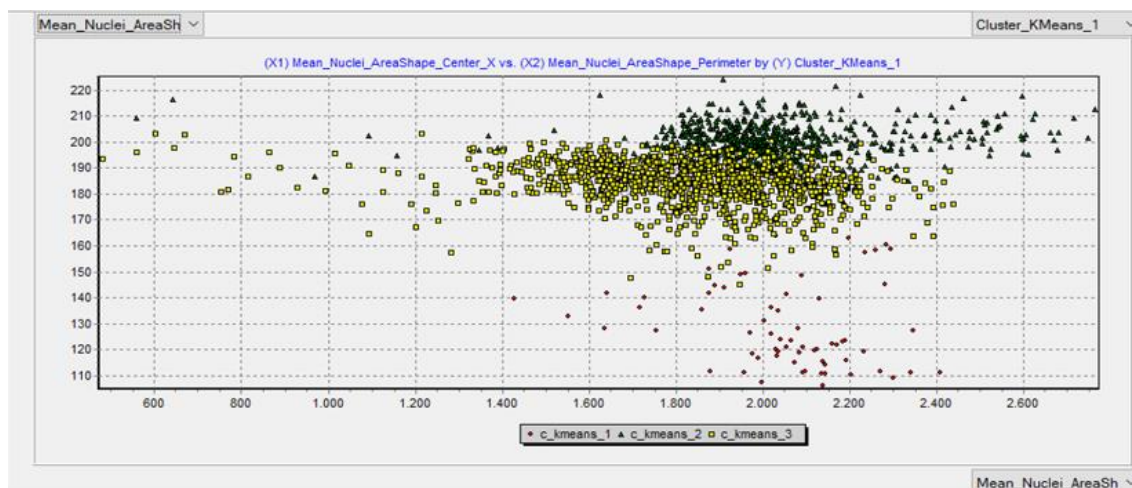


Figura 32. Gráfico de Dispersão de análise automatizada para o score nuclear.

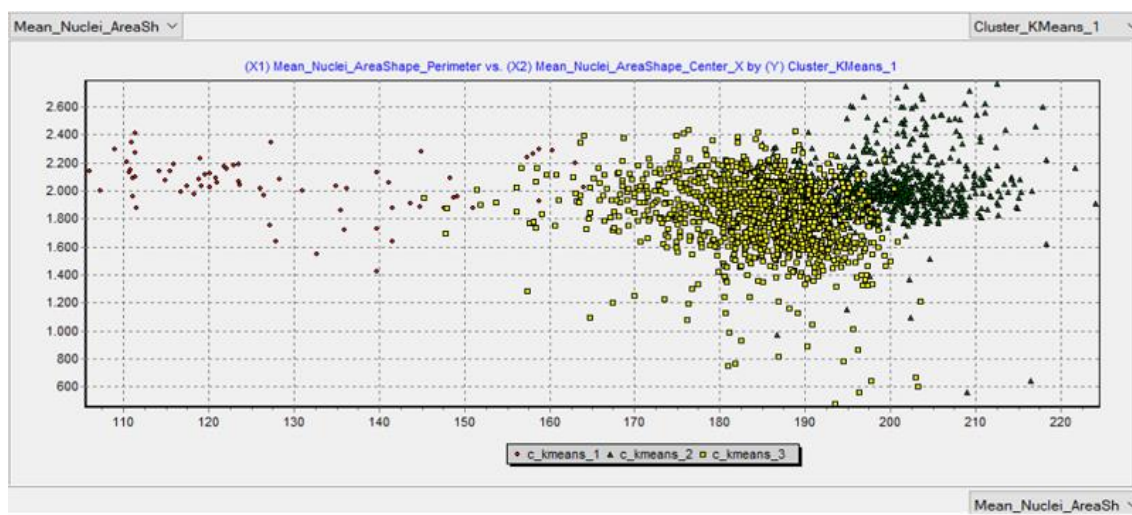


Figura 33. Gráfico de Dispersão de análise automatizada para o índice mitótico.

5.5 – Classificação automatizada do score tubular.

A Base de Dados analisada pelo software Tanagra para o score tubular quanto distribuição dos casos pela frequência das pontuações encontra-se relacionada no Quadro 6.

Pontuação Score Tubular	Número de casos	%
1 (a)	01	0,5
2 (b)	45	22,5
3 (c)	154	77,0
Total	200	100

Quadro 7. Distribuição dos casos por pontuação no Score Tubular.

O princípio de machine learning com aprendizado de máquina supervisionado baseia-se na assertiva de quanto maior o número de informações e dados fornecidos a máquina melhor se torna a sua performance; existindo uma relação diretamente proporcional a quantidade de informações fornecidas a máquina e a sua performance na ação solicitada, confirmada em nosso estudo na análise dos parâmetros de classificação do grau histológico em câncer de mama (WERNICK et al., 2010), (MULRANE et al., 2008), (JONES et al., 2009), (MISSELWITZ et al., 2010).

Assim, é possível entender a razão pela qual o software Tanagra em sua classificação automatizada do score tubular ter identificado os casos de pontuações mais frequentes na amostra analisada 3 (c) e 2 (b) com expressivos indicadores de acertos diagnósticos para esses casos; não tendo identificado nenhum caso com pontuação menor 1 (a) pela baixa frequência na amostra analisada em relação ao score tubular (Quadro 7 e Figura 34).

Results						
Classifier performances						
Error rate			0,0280			
Values prediction			Confusion matrix			
Value	Recall	1-Precision		c	b	Sum
c	0,9942	0,0284	c	1198	7	1205
b	0,8810	0,0263	b	35	259	294
			Sum	1233	266	1499

Figura 34. Análise comparativa entre classificação dada pelo médico patologista e classificação automatizada do software Tanagra para o score tubular.

A análise estatística dos resultados obtidos pela classificação automatizada realizada pelo software Tanagra para o score tubular está demonstrada no Quadro 8.

O índice de concordância *Kappa* entre os analisadores médico patologista e software Tanagra em relação ao score tubular foi de 0,91 correspondendo a uma alta concordância entre os analisadores e considerada “ótima” na tabela de concordância do Índice *Kappa* (Quadro 8 e Tabela 11).

Indicadores estatísticos	Resultados
Valor Preditivo Positivo c	0,9942
Valor Preditivo Positivo b	0,8810
Acurácia	0,9720
Classificação Incorreta (Erro)	0,0280
Índice <i>Kappa</i> de concordância (<i>K</i>)	0,91

Quadro 8. Estatística comparativa da classificação automatizada do score tubular dada pelo software Tanagra.

Tabela 11. Escala de concordância do Índice *Kappa* entre analisadores.

Kappa	Concordância
<0,00	Nenhuma
0,00-0,20	Fraca
0,21-0,40	Sofrível
0,41-0,60	Regular
0,61-0,80	Boa
0,81-0,99	Ótima
1,00	Perfeita

Adaptado de Landis & Koch, Biometrics, 1977

(Fonte: <https://posstrictusensu.iptsp.ufg.br>)

5.6 – Classificação automatizada do score nuclear.

A Base de Dados analisada pelo software Tanagra em relação ao score nuclear quanto a frequência de casos para cada graduação de pontuação encontra-se descrita no Quadro 9.

Pontuação Score Nuclear	Número de casos	%
1 (a)	03	1,5
2 (b)	108	54
3 (c)	89	44,5
Total	200	100

Quadro 9. Distribuição dos casos por pontuação no score nuclear.

A baixa frequência de casos de pontuação 1 (a) justifica a não identificação de nenhum desses casos pelo software Tanagra em sua classificação automatizada do score nuclear, identificando os casos de pontuações mais presentes na amostra analisada 2 (b) e 3 (c) com expressivo indicador de acerto para os casos de maior frequência 2 (b) e menor acerto em relação aos casos 3 (c), refletindo as suas frequências percentuais na amostra analisada (Quadro 9 e Figura 35).

A análise estatística dos resultados obtidos pela classificação automatizada realizada pelo software Tanagra para o score nuclear encontra-se demonstrada no Quadro 10.

Results						
Classifier performances						
Error rate		0,2188				
Values prediction			Confusion matrix			
Value	Recall	1-Precision		b	c	Sum
b	0,9138	0,2580	b	742	70	812
c	0,6245	0,1403	c	258	429	687
			Sum	1000	499	1499

Figura 35. Análise comparativa entre classificação dada pelo médico patologista e classificação automatizada realizada pelo software Tanagra para o score nuclear.

O índice de concordância *Kappa* entre os analisadores médico patologista e software Tanagra em relação ao score nuclear foi de 0,55 expressando uma concordância considerada “regular” na tabela de concordância do Índice *Kappa* e refletindo a variação de frequências entre os casos de cada pontuação na amostra analisada (Quadro 10 e Tabela 11).

Indicadores estatísticos	Resultados
Valor Preditivo Positivo b	0,9138
Valor Preditivo Positivo c	0,6245
Acurácia	0,7811
Classificação Incorreta (Erro)	0,2188
Índice Kappa de concordância (<i>K</i>)	0,55

Quadro 10. Estatística comparativa da classificação do score nuclear dada pelo software Tanagra com a do médico patologista.

5.7 – Classificação automatizada do score índice mitótico.

Em relação ao score índice mitótico a Base de Dados analisada pelo software Tanagra apresentou uma melhor distribuição dos casos entre as pontuações, o que permitiu que o software Tanagra em sua classificação automatizada do score índice mitótico identificasse os casos das 3 pontuações 1(a), 2 (b) e 3 (c). (Quadro 11)

Pontuação Score Índice Mitótico	Número de casos	%
1 (a)	71	35,5
2 (b)	101	50,5
3 (c)	28	14
Total	200	100

Quadro 11. Distribuição dos casos por pontuação no score índice mitótico.

O aprendizado de máquina para a análise e contagem automatizada de células em mitose pode ser considerada uma tarefa com maior grau de dificuldade, uma vez que o ciclo da mitose apresenta 4 fases distintas (prófase, metáfase, anáfase

e telófase) e os núcleos podem se apresentar com diferentes aspectos morfológicos conforme a fase da mitose em que se encontrava no momento da análise.

Isso representa um fator de maior complexidade para o aprendizado de máquina, exigindo um maior número de casos para a capacitação da máquina refletindo diretamente nos resultados da análise automatizada para o índice mitótico (XU et al., 2016), (ROMO-BUCHELI, 2017).

Dessa forma, destaca-se o expressivo acerto na classificação automatizada pelo software Tanagra em relação aos casos de pontuação 2 (b); sendo que o índice de concordância *Kappa* entre os analisadores médico patologista e software Tanagra em relação ao score índice mitótico foi de 0,49 refletindo uma concordância considerada “regular” na tabela de concordância do Índice *Kappa* (Figura 36, Quadro 12 e Tabela 11).

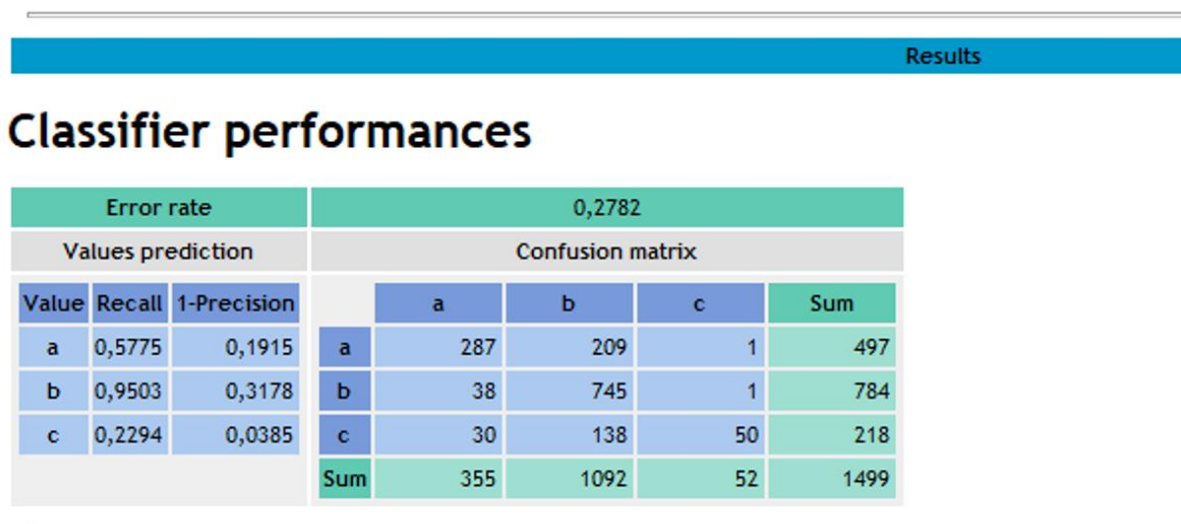


Figura 36. Classificação automatizada do score índice mitótico: Análise comparativa entre análise do médico patologista e do software Tanagra.

A análise estatística dos resultados obtidos pela classificação automatizada realizada pelo software Tanagra para o índice mitótico encontra-se demonstrada no Quadro 12.

Indicadores estatísticos	Resultados
Valor Preditivo Positivo a	0,5775
Valor Preditivo Positivo b	0,9503
Valor Preditivo Positivo c	0,2294
Acurácia	0,7218
Classificação Incorreta (Erro)	0,2782
Índice <i>Kappa</i> de concordância (<i>K</i>)	0,49

Quadro 12. Estatística comparativa da classificação do score índice mitótico dada pelo software Tanagra com a do médico patologista.

5.8 - Classificação automatizada do grau histológico de malignidade.

O grau histológico de malignidade na Base de Dados analisada pelo software Tanagra apresenta uma distribuição de frequência dos 200 casos de câncer de mama nos graus 1, 2 e 3 conforme descrito no Quadro 12.

Grau Histológico	n	%
1 (a)	6	3
2 (b)	124	62
3 (c)	70	35
Total	200	100

Quadro 13. Distribuição dos casos de câncer de mama analisados pelo software Tanagra conforme grau histológico.

A baixa frequência de casos do Grau 1 (a) reflete-se na não identificação de nenhum desses casos pelo software Tanagra em sua classificação automatizada do grau histológico de malignidade, identificando os casos de graus 2 (b) e 3 (c) que foram os de maior frequência na amostra analisada com expressivo indicador de acertos para os casos de maior frequência Grau 2 (b) e menor acerto em relação aos casos de Grau 3 (c), refletindo o aprendizado de máquina diretamente proporcional ao volume de informações oferecidas à máquina relacionado as frequências em que os casos apareceram na amostra analisada (Figura 37 e Quadro 13).

Classifier performances

Error rate			0,1921			
Values prediction			Confusion matrix			
Value	Recall	1-Precision		b	c	Sum
b	0,9682	0,2201	b	914	30	944
c	0,5351	0,0917	c	258	297	555
			Sum	1172	327	1499

Figura 37. Classificação automatizada do grau histológico de malignidade: Análise comparativa entre análise do médico patologista e do software Tanagra.

A análise estatística dos resultados obtidos pela classificação automatizada realizada pelo software Tanagra em relação ao grau histológico de malignidade encontra-se demonstrada no Quadro 14.

O índice de concordância *Kappa* entre os analisadores médico patologista e software Tanagra em relação ao grau histológico de malignidade foi de 0,55 expressando uma concordância considerada “regular” na tabela de concordância do Índice *Kappa* (Quadro 14 e Tabela 11).

Indicadores estatísticos	Resultados
Valor Preditivo Positivo b (Grau 2)	0,9682
Valor Preditivo Positivo c (Grau 3)	0,5351
Acurácia = 0,8078	0,8078
Classificação Incorreta (Erro)	0,1921
Índice <i>Kappa</i> de concordância (<i>K</i>)	0,55

Quadro 14. Estatística comparativa da classificação do grau histológico dada pelo software Tanagra com a do médico patologista.

Da análise dos resultados obtidos na classificação automatizada pelo software Tanagra, percebe-se que os índices de maior acerto e concordância ocorreram em relação ao score tubular onde os programas de análise de imagens CellProfiler e Tanagra identificaram um maior número de parâmetros quantitativos

(atributos) de expressão estatística com quantidade de casos e imagens suficientes para um aprendizado de máquina adequado.

Da mesma forma, em relação aos scores nuclear, score índice mitótico e para o grau histológico de malignidade, quando a frequência de casos nos grupos gerou uma quantidade de dados suficientes para um aprendizado de máquina adequado, os índices de acertos da classificação automatizada para esses grupos específicos podem ser considerados igualmente expressivos.

CONCLUSÃO

O câncer de mama, uma doença de grande impacto na vida das mulheres, ainda é diagnosticado em estágios mais avançados em relação ao grau de malignidade, fato este também observado na análise da amostra desse estudo.

O objetivo dos pesquisadores que decidem trabalhar com programas de análise automatizada de imagens celulares deve ser construir bancos de dados com o maior número de informações relevantes com vistas a capacitar a máquina para realizar tarefas com segurança, oferecendo importante ferramenta de auxílio no diagnóstico, tratamento e análise de fatores prognósticos para as doenças em geral.

O presente estudo representa uma das primeiras pesquisas utilizando softwares de análise de imagens celulares de livre acesso para diagnóstico do grau histológico de malignidade em câncer de mama e como tal oferece a contribuição de demonstrar ser possível a realização de análises automatizadas de parâmetros histopatológicos.

Dos resultados obtidos no estudo, pode-se concluir que imagens digitalizadas de lâminas histológicas de câncer de mama permitem a análise automatizada e reprodutível de parâmetros histopatológicos qualitativos convertidos em parâmetros quantitativos com vistas ao diagnóstico, graduação de seus scores e definição do grau histológico de malignidade, sendo necessária a ampliação do banco de dados com intuito de fornecer a máquina, informações e dados em quantidade suficiente para formação de conceitos e conhecimentos sólidos que possam subsidiar as suas análises em todos os aspectos solicitados.

Observou-se no estudo, uma relação diretamente proporcional entre a distribuição percentual dos casos e imagens entre os graus de pontuação dos scores tubular, nuclear, índice mitótico e do grau histológico de malignidade com o número de acertos e índices de concordância entre a classificação automatizada e a classificação dada pelo médico patologista.

Conclui-se portanto, que o uso de programas de análise de imagens celulares são capazes de realizar análise automatizadas e reprodutíveis de imagens digitalizadas de câncer de mama e constituem um campo de perspectivas promissoras.

Nesse sentido, novos estudos nas áreas conjuntas de “machine learning” e câncer de mama em mulheres poderão trazer novas e importantes contribuições.

REFERÊNCIAS BIBLIOGRÁFICAS

AMERICAN CANCER SOCIETY. Breast Cancer Facts & Figures 2019-2020. American Cancer Society INC. Atlanta, 2019. www.cancer.org acessado em 30/01/2020.

ARAÚJO, T.; ARESTA, G.; CASTRO, E.; ROUCO, J.; AGUIAR, P.; ELOY, C.; POLÓNIA, A.; CAMPILHO, A. *Classification of breast cancer histology images using Convolutional Neural Networks*. PLoS One. 2017 Jun 1;12(6):e0177544.

BEJNORDI, B.E.; VETA, M.; VAN DIEST, P.J.; VAN GINNEKEN, B.; KARSSEMEIJER, N.; LITJENS, G.; VAN DER LAAK, J.A.W.M. *Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer*. the CAMELYON16 Consortium JAMA. 2017 Dec 12; 318(22): 2199–2210.

BECK, A.H.; SANGOI, A.R.; LEUNG, S.; MARINELLI, R.J.; NIELSEN, T.O.; VAN DE VIJVER, M.J.; WEST, R.B.; VAN DE RIJN, M.; KOLLER, D. *Systematic analysis of breast cancer morphology uncovers stromal features associated with survival*. Science Translational Medicine. 2011 Nov 9;3(108):108ra113.

BRASIL. Lei Federal nº 12.732, de 22 de novembro de 2012.

BRAY, M.A.; CARPENTER, A.E. *Quality Control for High-Throughput Imaging Experiments Using Machine Learning in Cellprofiler*. Methods in Molecular Biology. 2018; 1683:89-112.

BUZIN, A.R.; PINTO, F.E.; NIESCHK, E K.; MITTAG, A.; DE ANDRADE, T.U.; ENDRINGER, D. C.; TARNOK, A.; LENZ, D. *Replacement of specific markers for apoptosis and necrosis by nuclear morphology for affordable cytometry*. Journal of Immunological Methods, v. 1, p. 1-6, 2015.

CARPENTER, A.E.; JONES, T.R.; LAMPRECHT, M.R.; CLARKE, C.; KANG, I.H.; FRIMAN, O.; GUERTIN, D.A.; CHANG, J.H.; LINDQUIST, R.A.; MOFFAT, J.; GOLLAND, P.; SABATINI, D.M. *CellProfiler: image analysis software for identifying and quantifying cell phenotypes*. Genome Biology. 2006;7(10):R100. Epub 2006 Oct 31.

CHATTERJEE, S. *Artefacts in histopathology*. Journal of Oral and Maxillofacial Pathology 2014; 18:111-6.

CHEN, J.M.; QU, A.P.; WANG, L.W.; YUAN, J.P.; YANG, F.; XIANG, Q.M.; MASKEY, N.; YANG, G.F.; LIU, J.; LI, Y. *New breast cancer prognostic factors identified by computer-aided image analysis of HE stained histopathology images*. Scientific Reports. 2015 May 29; 5:10690.

CHEN, J.M.; LI, Y.; XU, J.; GONG, L.; WANG, L.W.; LIU, W.L.; LIU, J. *Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: A review*. Tumour Biology. 2017 Mar;39(3):1010428317694550.

CHING, T.; HIMMELSTEIN, D.S.; BEAULIEU-JONES, B.K.; KALININ, A.A.; DO, B.T.; WAY, G.P.; FERRERO, E.; AGAPOW, P.M.; ZIETZ, M.; HOFFMAN, M.M.; XIE, W.; ROSEN, G.L.; LINGERICH, B.J.; ISRAELI, J.; LANCHANTIN, J.; WOLOSZYNEK, S.; CARPENTER, A.E.; SHRIKUMAR, A.; XU, J.; COFER, E.M.; LAVENDER, C.A.; TURAGA, S.C.; ALEXANDARI, A.M.; LU, Z.; HARRIS, D.J.; DECAPRIO, D.; QI, Y.; KUNDAJE, A.; PENG, Y.; WILEY, L.K.; SEGLER, M.H.S.; BOCA, S.M.; S. SWAMIDASS, S.J.; HUANG, A.; GITTER, A.; GREENE, C.S. *Opportunities and obstacles for deep learning in biology and medicine.* Journal of the Royal Society Interface. 2018 Apr; 15(141): 20170387

DORDEA, A.C.; BRAY, M.A.; ALLEN, K.; LOGAN, D.J.; FEI, F.; MALHOTRA, R.; GREGORY, M.S.; CARPENTER, A.E.; BUYS, E.S. *An open-source computational tool to automatically quantify immunolabeled retinal ganglion cells.* Experimental Eye Research. 2016 Jun; 147:50-56.

EULENBERG, P.; KÖHLER, N.; BLASI, T.; FILBY, A.; CARPENTER, A.E.; REES, P.; THEIS, F.J.; WOLF, F.A. *Reconstructing cell cycle and disease progression using deep learning.* Nature Communications. 2017; 8: 463.

HAN, Z.; WEI, B.; ZHENG, Y.; YIN, Y.; LI, K.; LI, S. *Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model.* Scientific Reports. 2017 Jun 23;7(1):4172.

HENNIG, H.; REES, P.; BLASI, T.; KAMENSKY, L.; HUNG, J.; DAO, D.; CARPENTER, A.E.; FILBY, A. *An open-source solution for advanced imaging flow cytometry data analysis using machine learning.* Methods. 2017 Jan 1; 112:201-210.

HITCHCOCK, L. *The future of telepathology for the developing world.* Archives of Pathology & Laboratory Medicine. 2011 Feb;135(2):211-4.

HOLTEN-ROSSING, H.; TALMAN, M.M.; JYLLING, A.M.B.; LAENKHOLM, A.V.; KRISTENSSON, M.; VAINER, B. *Application of automated image analysis reduces the workload of manual screening of sentinel lymph node biopsies in breast cancer.* Histopathology. 2017 Dec;71(6):866-873.

IBGE - Instituto Brasileiro de Geografia e Estatística. 2019. www.ibge.gov.br acessado em 20/11/2019.

INSTITUTO NACIONAL DE CÂNCER (INCA, BRASIL). Estimativa 2018. Incidência do Câncer no Brasil. Rio de Janeiro: INCA, 2017.

INSTITUTO NACIONAL DE CÂNCER (INCA, BRASIL). Estimativa 2020. Incidência do Câncer no Brasil. Rio de Janeiro: INCA, 2020.

JIMSON, S.; MALATHI, L.; KUMAR, G.M.K.; BALACHANDER, N. *Artifact in Histological Section.* Biomedical and Pharmacology Journal. 2016; 9(2).

JONES, T.R.; CARPENTER, A.E.; LAMPRECHT, M.R.; MOFFAT, J.; SILVER, S.J.; GRENIER, J.K.; CASTORENO, A.B.; EGGERT, U.S.; ROOT, D.E.; GOLLAND, P.; SABATINI, D.M. *Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning*. Proceedings of the National Academics of Sciences of the United States of America. 2009 Feb 10;106(6):1826-31.

LAMPRECHT, M.R.; SABATINI, D.M.; CARPENTER, A.E. *CellProfiler: free, versatile software for automated biological image analysis*. Biotechniques. 2007 Jan;42(1):71-5.

LENZ, D.; GASPARINI, L.S.; MACEDO, N.D.; PIMENTEL, E.F.; FRONZA, M.; JUNIOR, V.L.; BORGES, W.S.; COLE, E.R.; ANDRADE, T.U.; ENDRINGER, D.C. *In vitro cell viability by CellProfiler® software as equivalent to MTT assay*. Pharmacognosy Magazine, v. 13, p. 365, 2017.

LOUKAS, C.; KOSTOPOULOS, S.; TANOGLIDI, A.; GLOTSOS, D.; SFIKAS, C.; CAVOURAS, D. *Breast cancer characterization based on image classification of tissue sections visualized under low magnification*. Computational and Mathematical Methods in Medicine. 2013; 2013:829461.

LU, C.; ROMO-BUCHELI, D.; WANG, X.; JANOWCZYK, A.; GANESAN, S.; GILMORE, H.; RIMM, D.; MADABHUSHI, A. *Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers*. Laboratory Investigation. 2018 Nov; 98(11): 1438–1448.

MCQUIN, C.; GOODMAN, A.; CHERNYSHEV, V.; KAMENTSKY, L.; CIMINI, B.A.; KARHOHS, K.W.; DOAN, M.; DING, L.; RAFELSKI, S.M.; THIRSTRUP, D.; WIEGRAEBE, W.; SINGH, S.; BECKER, T.; CAICEDO, J.C.; CARPENTER, A.E. *CellProfiler 3.0: Next-generation image processing for biology*. PLoS Biology. 2018 Jul 3;16(7):e2005970.

MINISTÉRIO DA SAÚDE. Secretaria de Vigilância da Saúde. Departamento de Análise de Situação de Saúde. Coordenação Geral de Informações e Análises Epidemiológicas. Sistema de Informação sobre Mortalidade, 2015.

MISSELWITZ, B.; STRITTMATTER, G.; PERIASWAMY, B.; SCHLUMBERGER, M.C.; ROUT, S.; HORVATH, P.; KOZAK, K.; HARDT, W.D. *Enhanced CellClassifier: a multi-class classification tool for microscopy images*. BMC Bioinformatics. 2010 Jan 14; 11:30.

MONTAZERI, M.; MONTAZERI, M.; MONTAZERI, M.; BEIGZADEH, A. *Machine learning models in breast cancer survival prediction*. Technology and Health Care. 2016;24(1):31-42.

MULRANE, L.; REXHEPAJ, E.; PENNEY, S.; CALLANAN, J.J.; GALLAGHER, W.M. *Automated image analysis in histopathology: a valuable tool in medical diagnostics*. Expert Review of Molecular Diagnostics. 2008 Nov;8(6):707-25.

PEIKARI, M.; SALAMA, S.; NOFECH-MOZES, S.; MARTEL, A.L. *Automatic Cellularity Assessment from Post-Treated Breast Surgical Specimens*. Cytometry Analytical. 2017 Nov; 91(11): 1078–1087.

PESAPANE, F.; CODARI, M.; SARDANELLI, F. *Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine*. European Radiology Experimental. 2018 Oct 24;2(1):35.

RADIYA-DIXIT, E.; ZHU, D.; BECK, A.H. *Automated Classification of Benign and Malignant Proliferative Breast Lesions*. Scientific Reports. 2017; 7: 9900.

RAKOTOMALALA, R. *Tanagra: un logiciel gratuit pour l'enseignement et la recherche*. In: 5^{es} Journées d'Extraction et Gestion des Connaissances (EGC 05), Paris. 2005; 2:697–702.

RAWAT, R.R.; RUDERMAN, D.; MACKLIN, P.; RIMM, D.L.; AGUS, D.B. *Correlating nuclear morphometric patterns with estrogen receptor status in breast cancer pathologic specimens*. Npj Breast Cancer volume 4, Article number: 32 (2018)

ROMO-BUCHELI, D.; JANOWCZYK, A.; GILMORE, H.; ROMERO, E.; MADABHUSHI, A. *A Deep Learning Based Strategy for Identifying and Associating Mitotic Activity with Gene Expression Derived Risk Categories in Estrogen Receptor Positive Breast Cancers*. Cytometry Analytical. 2017 Jun; 91(6): 566–573.

SAMAR, K.; MANISHA, T.; MEGHA, J.; AMI, D. *Artifacts in Histopathology: A Potential Cause of Misinterpretation*. Research and Reviews: Journal of Dental Sciences. 2014; 2(2). RRJDS | Volume 2 | Issue 2 | April - June, 2014

SINTHIA, P.; MALATHI, M. *An Effective Two Way Classification of Breast Cancer Images: A Detailed Review*. Asian Pacific Journal of Cancer Prevention. 2018 Dec 25;19(12):3335-3339. Review.

SHI, B.; GRIMM, L.J.; MAZUROWSKI, M.A.; BAKER, J.A.; MARKS, J.R.; KING, L.M.; MALEY, C.C.; HWANG E. S.; LO, J.Y. *Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features*. Journal of the American College Radiology. 2018 Mar; 15(3 Pt B): 527–534.

SINGH, S.; CARPENTER, A.E.; GENOVESIO, A. *Increasing the Content of High-Content Screening An Overview*. Journal of Biomolecular Screening. 2014 Jun; 19(5): 640–650.

SOMMER, C.; GERLICH, D.W. *MACHINE LEARNING IN CELL BIOLOGY - TEACHING COMPUTERS TO RECOGNIZE PHENOTYPES*. Journal of Cell Science. 2013 Dec 15;126(Pt 24):5529-39.

TOPRAK, A. *Extreme Learning Machine (ELM)-Based Classification of Benign and Malignant Cells in Breast Cancer*. Medical Science Monitor. 2018; 24: 6537–6543.

TOZETTI, P.B.; LIMA, E.M.; NASCIMENTO, A.M.; ENDRINGER, D.C.; PINTO, F.E.; ANDRADE, T.U.; MITTAG, A.; TARNOK, A.; LENZ, D. *Morphometry to identify subtypes of leukocytes*. Hematology Oncology Stem Cell Therapy 2014; 7(2): 69–7.

- VU, T.H.; MOUSAVI, H.S.; MONGA, V.; RAO, G.; RAO, U.K.A. *Histopathological Image Classification using Discriminative Feature-oriented Dictionary Learning*. IEEE Transactions on Medical Imaging. 2016 Mar; 35(3): 738–751.
- WERNICK, M.N. ; YANG, Y. ; BRANKOV, J.G. ; YOURGANOV, G. ; STROTHER, S.C. *Machine Learning in Medical Imaging*. IEEE Signal Processing Magazine, 2010 Jul; 27(4): 25–38.
- WHITNEY, J.; CORREDOR, G.; JANOWCZYK, A.; GANESAN, S.; DOYLE, S.; TOMASZEWSKI, J.; FELDMAN, M.; GILMORE, H.; ANANT, MADABHUSHI. *Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer*. BMC Cancer. 2018 May 30;18(1):610.
- XU, J.; LEI, X.Q.; GILMORE, H.; WU, J.; TANG, J.; MADABHUSHI, A. *Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images*. IEEE Transactions on Medical Imaging. 2016 Jan; 35(1): 119–130.
- YU, K.H.; ZHANG, C.; BERRY, G.J.; ALTMAN, R.B.; RÉ, C.; RUBIN, D.L.; SNYDER, M. *Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features*. Nature Communications. 2016 Aug 16; 7:12474.
- ZHONG, F.; BI, R.; YU, B.; YANG, F.; YANG, W.; SHUI, R. *A Comparison of Visual Assessment and Automated Digital Image Analysis of Ki67 Labeling Index in Breast Cancer*. PLoS One. 2016; 11(2): e0150505.