INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES

Autarquia associada à Universidade de São Paulo

**ANÁLISE MORFOLÓGICA DE IMAGENS E CLASSIFICAÇÃO DE
ABERRAÇÕES CROMOSSOMICAS POR MEIO DE LÓGICA FUZZY**

LEONARDO PERES SOUZA

Dissertação apresentada como parte dos
requisitos para obtenção do Grau de Mestre
em Ciências na Área de Tecnologia
Nuclear- Reatores.

Orientador:

Prof. Dr. Roberto Navarro de Mesquita

São Paulo

2011

*Computers are to biology what
mathematics is to physics.
- Harold Morowitz*

*Dedico este trabalho a minha mãe e aos
meus eternos irmãos, são a minha vida, o
que eu sou!*

AGRADECIMENTOS

Ao prof. Dr. Roberto Navarro de Mesquita

Ao IPEN (Instituto de Pesquisas Energéticas Nucleares)

Ao CEN-IPEN (Centro de Engenharia Nuclear)

Ao CB-IPEN (Centro de Biotecnologia)

Às Dr^a Kayo Okazaki e Dr^a Márcia Augusta da Silva

Ao Dr. Marco Antônio Rodrigues Fernandes, Dr. Helio Yoriyaz e Dr. Paulo de Tarso Dalledone Siqueira

À minha Mãe e meus eternos Irmãos Ricardo e Eduardo

Aos meus avós.

Aos meus tios Ney, Bia, e primos Fernanda, Neyto e Marcelo

Aos meus tios Osmar, Célia, e primos Mateus e Lucas

À galera da Toca do Leitão

Aos meus grandes amigos Rafael, Bocão e Paulinho

Aos meus novos/velhos amigos, Gregório, Gabriel, Paulinha, Juliana, Jéssica, Jú, Carlos, Chris, Bruno, Arthur, Henrique, Luciano, Paula, Gustavo, Luiz, Rafael Borba, Jaime, Osmar, Mariana, Lídia, Aline e Cia.

Aos bolsistas do CEN

Ao CNPQ pelo financiamento dessa pesquisa.

ANÁLISE MORFOLÓGICA DE IMAGENS E CLASSIFICAÇÃO DE ABERRAÇÕES CROMOSSÔMICAS POR MEIO DE LÓGICA *FUZZY*

Leonardo Peres Souza

RESUMO

Este trabalho desenvolve uma metodologia para a automação da análise morfológica de imagens de cromossomos humanos irradiados no reator nuclear IEA-R1 (localizado no Instituto de Pesquisas Energéticas e Nucleares, IPEN, em São Paulo, Brasil) e, portanto, sujeitos a aberrações morfológicas. Esta metodologia se propõe a auxiliar na identificação, caracterização e classificação de cromossomos pelo profissional citogeneticista. O desenvolvimento da metodologia inclui a elaboração de um aplicativo baseado em técnicas de inteligência artificial utilizando Lógica *Fuzzy* e técnicas de processamento de imagens. O aplicativo desenvolvido foi denominado de CHRIMAN e é composto de módulos que contêm etapas metodológicas que suprem aspectos importantes para a obtenção de uma análise automatizada. A primeira etapa é a padronização dos procedimentos de aquisição das imagens digitais bidimensionais de metáfases através do acoplamento de uma câmera fotográfica digital comercial comum à ocular do microscópio utilizado na análise metafásica convencional. A segunda etapa é relativa ao tratamento das imagens obtidas através da aplicação de filtros digitais, armazenamento e organização das informações tanto do conteúdo da imagem em si, como das características extraídas e selecionadas, para posterior utilização nos algoritmos de reconhecimento de padrões. A terceira etapa consiste na utilização do banco de imagens digitalizadas e informações extraídas e armazenadas para a identificação dos cromossomos, sua caracterização, contagem e posterior classificação. O acerto no reconhecimento das imagens cromossômicas é de 93,9%. Esta classificação é baseada nos padrões encontrados classicamente em Buckton [1973], e possibilita o auxílio ao geneticista no procedimento de análise dos cromossomos com diminuição do tempo de análise e criando condições para a inclusão deste método num sistema mais amplo de avaliação de danos causados às células pela exposição à radiação ionizante.

MORPHOLOGICAL IMAGES ANALYSIS AND CHROMOSOMIC ABERRATIONS CLASSIFICATION BASED ON *FUZZY* LOGIC

Leonardo Peres Souza

ABSTRACT

This work has implemented a methodology for automation of images analysis of chromosomes of human cells irradiated at IEA-R1 nuclear reactor (located at IPEN, São Paulo, Brazil), and therefore subject to morphological aberrations. This methodology intends to be a tool for helping cytogeneticists on identification, characterization and classification of chromosomal metaphasic analysis. The methodology development has included the creation of a software application based on artificial intelligence techniques using *Fuzzy* Logic combined with image processing techniques. The developed application was named CHRIMAN and is composed of modules that contain the methodological steps which are important requirements in order to achieve an automated analysis. The first step is the standardization of the bi-dimensional digital image acquisition procedure through coupling a simple digital camera to the ocular of the conventional metaphasic analysis microscope. Second step is related to the image treatment achieved through digital filters application; storing and organization of information obtained both from image content itself, and from selected extracted features, for further use on pattern recognition algorithms. The third step consists on characterizing, counting and classification of stored digital images and extracted features information. The accuracy in the recognition of chromosome images is 93.9%. This classification is based on classical standards obtained at Buckton [1973], and enables support to geneticist on chromosomic analysis procedure, decreasing analysis time, and creating conditions to include this method on a broader evaluation system on human cell damage due to ionizing radiation exposure.

SUMÁRIO

DEDICATÓRIA	iii
AGRADECIMENTOS	iv
RESUMO	v
ABSTRACT	vi
LISTA DE FIGURAS	ix
SUMÁRIO	xii
I INTRODUÇÃO	1
I-1. Análise Cromossômica	2
I-2. Processamento de Imagens	2
I-3. Reconhecimento de Padrões em Imagens	3
I-4. Objetivo e Justificativa	4
I-4.1. Objetivos Específicos	5
II REVISÃO BIBLIOGRAFICA	7
II-1. Aspectos Biológicos	7
II-1.1. Tipos de Aberrações Cromossômicas	9
II-2. Aspectos Computacionais em Análises de Imagens	10
III FUNDAMENTOS TEÓRICOS	13
III-1. Processamento de Imagens Digitais	13
III-2. Segmentação de Imagens	15
III-3. Registro em Imagens	18
III-4. Reconhecimento de Padrões	19
III-5. Inteligência Artificial	20
III-5.1. Modelo <i>Fuzzy</i>	21
IV - MATERIAIS E MÉTODOS	26
IV-1. Infra-Estrutura	26
IV-1.1. <i>Facility</i> do reator de pesquisa (IEA-R1)	26
IV-1.2. Laboratório de Cultura Celular do Centro de Biotecnologia (CB)	26
IV-1.3. Equipamento e sistema operacional para o desenvolvimento do aplicativo	27
IV-2. Descrições das Etapas para o Desenvolvimento da Metodologia	27
IV-3. Apresentação da Metodologia	29
IV-3.1. Aquisições de Imagens de Células Irradiadas	29
IV-3.2. Desenvolvimento do Aplicativo	34
IV-3.2.1 Transformação das Imagens Cromossômicas para Tons de Cinza	35
IV-3.2.2 Ajuste de Histograma das Imagens Cromossômicas	36
IV-3.2.2 Eliminação do Background das Imagens Cromossômicas	38
IV-3.2.4 Segmentação das Imagens Cromossômicas	41

IV.3.2.5 Registro das Imagens Cromossômicas _____	42
IV.3.2.6 Caracterização das Imagens Cromossômicas _____	46
IV.3.2.6.1 Análise dos perfis de uma imagem cromossômica _____	46
IV.3.2.6.2 Sistema de Lógica <i>Fuzzy</i> para Caracterização dos Perfis _____	50
IV.3.2.6.3 Sistema de Lógica Fuzzy para Caracterização das Imagens Cromossômicas _____	55
IV.3.2.6.4 Localização e Identificação de Centrômero(s) na Imagem Cromossômica _____	58
V RESULTADOS E DISCUSSÕES _____	67
V-1. Apresentação da Interface do Usuário - CHRIMAN _____	73
VI CONCLUSÃO _____	76
REFERÊNCIAS BIBLIOGRÁFICAS _____	78
ANEXO I - Resultados Obtidos – Segmentação e Registro das Imagens Cromossômicas. _____	1
ANEXO II - Resultados Obtidos – Reconhecimento das Imagens Cromossômicas – <i>Threshold</i> = 70. _____	2
ANEXO III – Regras <i>Fuzzy</i> PerfilClassif. _____	3
ANEXO VI – Regras <i>Fuzzy</i> ChromoClassif. _____	4

LISTA DE FIGURAS

<i>Figura 1 – Exemplo de um núcleo de célula humano durante a metáfase [IAEA, 2001].</i>	8
<i>Figura 2 – A) O pixel que está sendo analisado; B), A vizinhança N-4 de um pixel; C) A vizinhança N-8 de um pixel; D) vizinhança N-26 de um voxel.</i>	16
<i>Figura 3 – Representação das coordenadas da vizinhança do pixel analisado.</i>	17
<i>Figura 4 – Sistema de Lógica Fuzzy [Cecconello, 2006].</i>	23
<i>Figura 5 - Diagrama de etapas de um sistema de reconhecimento de padrões em imagens digitais.</i>	28
<i>Figura 6 – Imagens cedidas pela equipe da Dr^a Kayo Okazaki e Dr^a Márcia Augusta da Silva.</i>	31
<i>Figura 7 – Metáfase com a presença de aberrações cromossômicas [Dr^a. Kayo Okazaki e Dr^a. Márcia Augusta da Silva (IPEN)].</i>	32
<i>Figura 8 – Diagrama de sequência de execução do aplicativo.</i>	34
<i>Figura 9 – Aplicação do método de transformação em tons de cinza. Input Image e Gray Image.</i>	36
<i>Figura 10 – Aplicação do método de transformação de ajuste de contraste. Input Image e Contrast Enhanced Image.</i>	37
<i>Figura 11 – A) Histograma de intensidades da imagem em tons de cinza; B) Histograma de intensidades da imagem após a aplicação do ajuste de contraste.</i>	37
<i>Figura 12 – Aplicação do método de transformação de ajuste de contraste. Input Image e No background Image. Remoção do background pelo limiar 70.</i>	39
<i>Figura 13 - Aplicação do método de eliminação de background pelo valor do limiar. Da esquerda para a direita, são aplicados os métodos Otsu, e os valores de corte (limiar) por intensidade de 80 (31,37%), 70 (27,45%) e 60 (23,52%) respectivamente. A seta aponta para a imagem do mesmo conjunto de cromossomos exemplificando uma correta segmentação.</i>	40
<i>Figura 14 – Imagem cromossômica com acumulo de Giemsa.</i>	41
<i>Figura 15 – Imagem onde os cromossomos estão sobrepostos e não foram segmentados corretamente, devido ao fato de que o mesmo rótulo foi atribuído a todo o subconjunto de pixels desta imagem.</i>	42
<i>Figura 16 – Aplicação da transposta na imagem cromossômica segmentada. Primeira transformação morfológica em que as imagens cromossômicas são submetidas.</i>	43

<i>Figura 17 – Pixels de referência e identificação do eixo central do cromossomo utilizado para o registro da imagem cromossômica.</i>	<i>44</i>
<i>Figura 18 – Aplicação da operação de rotação utilizando o eixo central do cromossomo disposto na imagem.</i>	<i>44</i>
<i>Figura 19 – Imagem onde o cromossomo está curvado de forma que dificulte o resultado da classificação.....</i>	<i>45</i>
<i>Figura 20 – Imagem onde os cromossomos foram segmentados e registrados corretamente.</i>	<i>46</i>
<i>Figura 21 – Esquema representativo perfis traçados nas imagens cromossômicas com a finalidade do reconhecimento da disposição dos cromossomos.</i>	<i>47</i>
<i>Figura 22 – Banda com um pixel isolado. Aplica-se a regra de adição de pixels isolados à banda válida mais próxima.....</i>	<i>47</i>
<i>Figura 23 – A) Perfil com banda longa e única; B) Perfil com banda média e única; C) Perfil com banda curta e única; D) Perfil com duas bandas de contagens médias; E) Perfil com duas bandas de contagens distintas; F) Perfil com o número de quantidade de bandas superior a 2 (dois) e perfil com a presença de banda não validas.....</i>	<i>49</i>
<i>Figura 24 – Demonstração dos nove perfis traçados ao longo do eixo y, cada perfil é traçado a cada 10% da altura da imagem cromossômica gerada.</i>	<i>50</i>
<i>Figura 25 – FLS perfilClassif: A) Variável de entrada quantidadeBandas, funções de relacionamento: um, dois, três, quatro e cinco. B) Variável de entrada contagemBanda, funções de relacionamento: pouco, meioPouco, meio, meioMuito e muito.</i>	<i>51</i>
<i>Figura 26 – FLS perfilClassif: Funções de relacionamento de saída grupoPerfil: longoUnico [0-1], medioUnico [1-2], curtoUnico [2-3], curtoDuplo]3-4], medioLongoDuplo [4-5] e naoValido]5-6].</i>	<i>52</i>
<i>Figura 27 – Mapeamento de 9 (nove) perfis na imagem e a caracterização destes pelo FLS perfilClassif A) Mapeamento e caracterização dos perfis de uma imagem cromossômica normal. B) Mapeamento e caracterização dos perfis de uma imagem cromossômica com a aberração cromossômica: dicêntrico.</i>	<i>54</i>
<i>Figura 28 – A) Funções de relacionamento das variáveis de entrada do FLS: tamanhoMatriz: pequena, medioPequena, meioMediaPequena, meioMediaGrande, mediaGrande e grande; B) Funções de relacionamento das variáveis de entrada do FLS: preenchimentoImagem: pouco, medioPouco, meioMedioPouco, meioMedioMuito, medioMuito, muito; C) Funções de relacionamento das variáveis de entrada do FLS: situacaoPerfil: contagemInsuficiente e contagemSuficiente.</i>	<i>56</i>
<i>Figura 29 – Função de saída escalar do FLS, chamada sitImagem: naoPadronizado e padronizado.</i>	<i>58</i>

<i>Figura 30 – Utilização do mapeamento de 9 (nove) perfis para identificar as coordenadas de início e fim da busca pelos centrômeros.....</i>	<i>61</i>
<i>Figura 31 – Linhas identificadas como início e fim do(s) centrômero(s) nas imagens cromossômicas geradas.....</i>	<i>62</i>
<i>Figura 32 – Perfis caracterizados como perfis do mesmo grupo, porém com diferentes características.....</i>	<i>62</i>
<i>Figura 33 – Histograma de contagens de pixels em cada perfil da imagem analisada.....</i>	<i>64</i>
<i>Figura 34 – Perfil identificado como a região onde se localiza o centrômero na imagem cromossômica gerada, utilizando a segunda técnica de reconhecimento de padrões.</i>	<i>64</i>
<i>Figura 35 – Total de reconhecimentos nos cromossomos com as três intensidades (60, 70 e 80) em cada imagem metafásica analisadas. Número de cromossomos analisados e porcentagem do total de cromossomos em uma célula x Imagens (sem fundo) escolhidas para análise.</i>	<i>69</i>
<i>Figura 36 – Resultados obtidos pela primeira técnica de reconhecimento de padrões de identificação dos perfis de localização dos centrômeros em uma imagem cromossômica. Número de centrômeros analisados e porcentagem de acerto da primeira técnica x Imagens analisadas com limiar igual a 70.....</i>	<i>70</i>
<i>Figura 37 – Resultados obtidos pela segunda técnica de reconhecimento de padrões de identificação dos perfis de localização dos centrômeros em uma imagem cromossômica. Número de centrômeros analisados e porcentagem de acerto da segunda técnica x Imagens analisadas com limiar igual a 70.....</i>	<i>71</i>
<i>Figura 38 – Resultados obtidos pelas duas técnicas de reconhecimento de padrões de identificação dos perfis de localização dos centrômeros em uma imagem cromossômica. Número de centrômeros analisados e porcentagem de acerto somando as duas técnicas x Imagens analisadas com limiar igual a 70.....</i>	<i>72</i>
<i>Figura 39 – Interface do usuário desenvolvida pelo para o aplicativo CHRIMAN.....</i>	<i>74</i>

SUMÁRIO

IPEN - Instituto de Pesquisas Energéticas e Nucleares

CNEN - Conselho Nacional de Energia Nuclear

CEN - Centro de Engenharia Nuclear

CB - Centro de Biotecnologia

TC - Tomografia Computadorizada

IEA R1 - Reator de Pesquisas IEA R1

GUI - Graphic User Interface

RGB - Red, Green, Blue (sistema de coloração)

IA - Inteligência Artificial

EM - Espectro Eletromagnético

HD - Hard Disk

I INTRODUÇÃO

O presente trabalho faz parte do grupo de pesquisas de Inteligência Artificial e Aplicações do Centro de Engenharia Nuclear – IPEN / USP (CEN). O trabalho consiste no desenvolvimento de uma metodologia para a automação da caracterização da morfologia de cromossomos humanos em imagens bidimensionais digitais com a identificação e classificação das possíveis anormalidades causadas em virtude dos danos biológicos causados pela radiação ionizante.

Diferentes áreas do conhecimento fazem parte desta da pesquisa que contou com a colaboração do Centro de Biotecnologia – IPEN / USP (CB), na tentativa de reproduzir o conhecimento do especialista citogeneticista em analisar e caracterizar os cromossomos e as chamadas aberrações cromossômicas através de algoritmos que utilizam inteligência artificial, em especial um sistema de inferências com Lógica *Fuzzy*.

A primeira seção aborda uma introdução sobre as tecnologias e rotinas científicas empregadas neste trabalho e descreve os objetivos e as justificativas para o desenvolvimento do aplicativo. O segundo capítulo apresenta uma breve cronologia dos aspectos biológicos e computacionais para a melhor compreensão da evolução da tecnologia envolvida na análise cromossômica. O terceiro item apresenta a base de fundamentação deste projeto. O quarto capítulo apresenta toda a metodologia empregada para a aquisição das imagens e desenvolvimento do aplicativo, seguido pelo item de resultados obtidos e discussões sobre os problemas encontrados, como também a apresentação da *interface com o usuário* (*Graphic User Interface - GUI*) desenvolvida no trabalho. O sexto capítulo trata das conclusões e sugestões para trabalhos futuros. E, por fim, as referências bibliográficas e os anexos são apresentados.

I-1. Análise Cromossômica

O conhecimento e a tecnologia do comportamento e da estrutura cromossômica é um tema amplamente discutido, como por exemplo, a busca por informações sobre a replicação e, a localização, tanto das células quanto a de seus genes, como também a identificação das chamadas aberrações cromossômicas, causadas por agentes genotóxicos endógenos ou exógenos, e fazem parte da área de estudo para inúmeros pesquisadores geneticistas e citogeneticista.

A análise cromossômica é uma importante e difícil tarefa no diagnóstico clínico, na dosimetria mutagênica e na pesquisa biológica. Trata-se de uma prática cara, que consome muito tempo e, de certa forma, imprecisa quando realizada manualmente [J. Piper, 1980].

Os avanços nas técnicas citogenéticas estudadas durante décadas passadas, encorajaram os estudos clínicos em diagnósticos pré-natais e as pesquisas em patologias de câncer, em dosimetria de mutações induzidas, entre outras [Q. Wu, 1987].

O objetivo da análise cromossômica é determinar, mediante a morfologia dos cromossomos, as anormalidades na estrutura biológica e revelar os danos causados por agentes genotóxicos.

Dentre algumas técnicas de análise e diagnóstico por meio de imagens cromossômicas, escopo deste projeto, destaca-se a análise computacional da imagem celular com o objetivo de detectar essas anormalidades na morfologia cromossômica na imagem, onde, inúmeras técnicas computacionais foram desenvolvidas para auxiliar este diagnóstico. Algumas delas são apresentadas neste documento.

I-2. Processamento de Imagens

As primeiras aplicações em processamento de imagens apareceram no final da década de 60 e começo da década de 70, com o surgimento de computadores capazes de desempenhar tal tarefa. As principais preocupações eram de melhorar e realçar imagens obtidas da lua. Em paralelo às aplicações espaciais, as técnicas de processamento digital de imagens foram aplicadas às imagens médicas, a imagens obtidas de lugares remotos da Terra e no campo da astronomia [Gonzalez et al, 2008]. As técnicas pioneiramente

implementadas foram a utilização da escala de tons de cinza, a aplicações de *zooms* e a binarização da imagem.

Com o desenvolvimento da tecnologia, tanto relativa a “hardwares” mais robustos, com maior capacidade de armazenamento, velocidade de execução de equações matemáticas, incluindo a sofisticação e complexidade dos algoritmos, consagrou a década de 80 como um período significativo do crescimento e amadurecimento do campo da pesquisa computacional, principalmente envolvendo a utilização de imagens digitais.

O advento da produção em massa de mecanismos de aquisição digital reduziu os custos e gerou a expansão da pesquisa, utilizando imagens obtidas por câmeras fotográficas e de vídeo de alta resolução digitais. Desde então, a aplicação de técnicas em processamento de imagens tem se firmado como uma grande área da engenharia [Wiley, 2007].

Na área médica, as preocupações iniciais das pesquisas em processamento de imagens, eram voltadas para a melhoria na forma da aquisição e da digitalização destas imagens. A primeira grande aplicação na medicina foi a sua utilização no desenvolvimento da Tomografia Computadorizada (TC). Durante este desenvolvimento, ficou evidente que a imagem resultante do processo tomográfico não possuía a qualidade necessária para se realizar um possível diagnóstico, justificando, a necessidade de se processar e/ou transformar digitalmente as imagens adquiridas.

Para isso, tornou-se indispensável o desenvolvimento de algoritmos capazes de realçar cores e formas e trabalhar sobre a qualidade das imagens resposta, dentre eles: o controle e/ou equalização de histogramas, o realce de contraste a partir de um limiar (*threshold*), a eliminação de informações desnecessárias de fundo (*background*), a utilização e remoção de ruídos oriundos da forma de aquisição, por exemplo, o estudo de reconhecimento de padrões, para que, a partir de imagens aprimoradas digitalmente, fosse possível adquirir dados e informações relevantes a um determinado estudo clínico, antes não visível ao especialista, entre diversas outras aplicações.

I-3. Reconhecimento de Padrões em Imagens

O reconhecimento de padrões em imagens digitais é um das grandes áreas de desenvolvimento de reconhecimento por imagens, aplicados à inteligência artificial. Basicamente, consiste em identificar semelhanças para classificar e caracterizar objetos

presentes em uma imagem ou mais imagens. Esta classificação é realizada baseada em grande conhecimento do problema proposto (experiência do profissional), ou por meio de métodos computacionais que reproduzem a análise visual por meio de informações de características inerentes às imagens. Estabelecem-se implicitamente ou estatisticamente regras, correlações e relações de comparação entre as imagens tratadas.

Um grande espectro de aplicações de reconhecimento de padrões é encontrado na literatura, com diferentes formas de aplicação utilizando-se de técnicas de inteligência artificial na melhora da eficiência e velocidade da obtenção da informação de interesse.

O recente avanço, e a consequente redução de custo da tecnologia envolvida na utilização de imagens digitais, desencadearam um grande impulso no desenvolvimento desta área de conhecimento [Jain et.al., 2000].

Especificamente na área de imagens cromossômicas, trabalhos foram publicados relatando desde os desenvolvimentos da aplicação de técnicas paramétricas e não paramétricas de classificação cromossômica supervisionada, como em [Sampat et.al., 2005], até a descrição do desenvolvimento de um aplicativo, semelhante ao objetivo deste projeto, que analisa imagens cromossômicas com a finalidade de identifica-las [Detcheva, 1996].

A análise cromossômica utilizando imagens tratadas e analisadas por um algoritmo com recursos de inteligência artificial (como a lógica *Fuzzy*), pode funcionar como um apoio à tomada de decisões e auxiliar na caracterização e classificação cromossômica.

I-4. Objetivo e Justificativa

O objetivo deste trabalho é desenvolver um aplicativo baseado em técnicas de reconhecimento de padrões em imagens digitais, utilizando-se de recursos de inteligência artificial, com aplicação de regras *Fuzzy*, melhor detalhadas na seção (IV.3.2.6.2 Sistema de Lógica Fuzzy para Caracterização dos Perfis) e (IV.3.2.6.3 Sistema de Lógica Fuzzy para Caracterização das Imagens Cromossômicas), que trabalhe como uma ferramenta de apoio ao profissional citogeneticista na análise e caracterização de imagens cromossômicas, assim como suas anormalidades morfológicas, identificadas como aberrações cromossômicas.

O aplicativo projetado tem como funções um módulo que realiza a contagem do número de cromossomos presentes numa imagem cromossômica, outro que identifica e

classifica dentre os cromossomos presentes em uma imagem, os aqueles normais e os com aberração. Um módulo que obtenha a classificação de alguns dos tipos de aberração foi desenvolvido usando como base, as referências de classificação padronizadas em Buckton [1973].

A proposta de desenvolvimento de uma metodologia de auxílio ao diagnóstico clínico cromossômico por meio de imagens digitais é de grande valia para o profissional citogeneticista. Incluem-se nesta metodologia: a padronização na aquisição da imagem digital de uma metáfase; o tratamento destas imagens por algoritmos de processamento e realce das imagens obtidas, preparando-as para as etapas seguintes de caracterização individual das imagens cromossômicas; a elaboração e desenvolvimento de uma interface gráfica para usuários - *GUI* que permita a análise, o armazenamento e a organização das informações referente a cada célula analisada.

Sendo assim, a automação na contabilidade e caracterização dos cromossomos pode auxiliar o estudo citogenético, pela análise classificatória sobre um conjunto de cromossomos presentes na imagem que apresentem características que sejam classificadas como cromossomos não normais, ou indicar ao profissional quando não foi possível o reconhecimento da imagem por meio das técnicas propostas.

Este aplicativo pode reduzir os custos operacionais necessários para o estudo dos cromossomos e suas aberrações cromossômicas e representar um passo importante para a validação de uma metodologia de classificação de aberrações cromossômicas a partir de imagens cromossômicas digitais

I-4.1. Objetivos Específicos

- Aplicar sobre as imagens metafásicas técnicas de processamento de imagens para realçar a informação de interesse, identificação cromossômica;
- Elaboração de algoritmos de reconhecimento de padrões para a identificação de perfis traçados na imagem cromossômica estudada;
- Identificação da morfologia dos cromossomos nas imagens como as anormalidades na sua disposição causadas por danos provocados pela radiação ionizante;
- Apresentar uma metodologia de identificação de imagens cromossômicas e suas aberrações por meio de técnicas de reconhecimento de padrões de perfis utilizando-se de técnicas de Inteligência Artificial por meio de Lógicas *Fuzzy*.

- Desenvolvimento de uma interface gráfica para usuários - *GUI* que permita a análise, o armazenamento e a organização das informações referente a cada célula analisada.

II REVISÃO BIBLIOGRAFICA

A revisão bibliográfica está dividida em dois aspectos envolvidos neste estudo: os aspectos biológicos e os aspectos computacionais.

II-1. Aspectos Biológicos

Os cromossomos estão presentes em todas as células nucleadas de todos os organismos vivos. Um cromossomo é uma longa sequência de DNA (ácido desoxirribonucleico) que contém genes e outras sequências de nucleótidos ou nucleotídeos com funções específicas nas células dos seres vivos.

Os processos de mitose e meiose das células indicam que os cromossomos desempenham um papel de extrema importância nas transmissões de informações genéticas. Essa indicação ganha um maior reforço pela existência de casos em que os processos de divisão nuclear apresentam irregularidades, gerando as chamadas anomalias cromossômicas, também denominadas aberrações ou mutações cromossômicas [Kavalco, 1999].

A visualização dos cromossomos somente é possível durante a metáfase, fase da divisão celular onde os cromossomos estão dispersos de forma que possibilite a sua análise. Como mostrado na Figura 1.

Figura 1 – Exemplo de um núcleo de célula humano durante a metáfase [IAEA, 2001].

O total de cromossomos em células humanas é de 46 (quarenta e seis) sendo 44 (quarenta e quatro) destes, cromossomos homólogos (22 pares autossômicos), agrupados em classes, e 2 (dois), ou um par, de cromossomos sexuais X e Y (um par de cromossomos X no caso feminino e um cromossomo X e um Y para masculino).

Os cromossomos normais humanos são classificados de acordo com o seu comprimento e a posição em que se encontra o centrômero, ao longo dos filamentos de cromátide. Dividem-se em quatro grupos: os metacêntricos, cromossomos cujo centrômero fica posicionado no meio, separando cada cromátide em dois braços de mesmo tamanho; os submetacêntricos possuem o centrômero afastado do meio e, em cada cromátide, os dois braços têm tamanhos um pouco diferentes; os acrocêntricos, cromossomos cujos centrômeros localizam-se próximos a uma das extremidades, com um braço bem menor que outro; e por fim o telocêntrico, onde o centrômero localiza-se em uma das extremidades do cromossomo [Kasahara, 2001].

Caso um ou mais cromossomos estejam alterados morfológicamente, indica que houve algum dano no material genético da célula ou DNA, por motivos naturais ou induzidos, que podem ser visualizados como aberrações nos cromossomos. As aberrações cromossômicas são alterações cromossômicas, seja em sua integridade estrutural ou em sua quantidade presente no núcleo de uma célula. Tais alterações na contagem podem ser fatores que definem diversos tipos de doenças genéticas como a síndrome de Down, síndrome de Turner, síndrome de Klinefelter, entre outras.

Ao longo de muitos anos, as imagens cromossômicas reais foram objeto de estudo de inúmeros cientistas, biólogos, pesquisadores, entre outros membros do estudo citogenético, desde a descoberta da existência dos cromossomos, até a descoberta de suas divisões durante o processo de proliferação das células. Podemos encontrar todo este desenvolvimento bem representado no clássico *Methods for the Analysis of Human Chromosome Aberrations*, WHO (World Health Organization), onde estudos, análises, monitoramento dos efeitos da radiação nas células e outros fatores que contribuam para a formação de aberrações cromossômicas, são estudados e catalogados no manual de [Buckton, 1973]. As classificações a partir das imagens cromossômicas deste estudo serão baseadas nesta referência.

II-1.1. Tipos de Aberrações Cromossômicas

Os primeiros trabalhos em efeitos biológicos da radiação ionizante foram publicados no início do século XX e tratavam de anormalidades cromossômicas nos tecidos de mamíferos em testes, por exemplo, com camundongos [Bergonie, et al., 1904], testes em ovos de *Ascaris* (larvas) [Perthes, 1904], testes em gafanhotos [Amato, 1911], testes em células de larvas de *Amphibia* [Grasnick, 1918], e também em células somáticas e meióticas de plantas [Koernicke, 1905]. Entretanto, em razão do baixo nível das técnicas citogenéticas disponíveis na época, principalmente seções histológicas, a natureza e a sequência da origem dos vários tipos de anormalidades não puderam ser corretamente interpretadas.

As aberrações cromossômicas ocorrem basicamente de duas formas distintas, espontâneas ou estimuladas (induzidas) por radiação, ou outros agentes genotóxicos. Ambas as aberrações cromossômicas (estrutural ou numérica) ocorrem espontaneamente em virtude de diversos fatores intrínsecos ou extrínsecos das células. A frequência de aberrações espontâneas é de cerca de 0.6% em células vivas [Natarajan, 2002].

A indução das aberrações por radiação baseia-se em duas formas básicas de classificação, que são dependentes da fase do ciclo celular onde se dá a ação dos agentes genotóxicos, sendo elas: as do tipo cromossômico e as do tipo cromatídico. As classificações a partir das imagens cromossômicas propostas por este trabalho serão baseadas nos critérios descritos em [Buckton, 1973].

As aberrações do tipo-cromossômico ocorrem antes da duplicação do DNA: fragmentos acêntricos (deleção terminal), minutos (deleção intersticial), anéis acêntricos e cêntricos e inversões pericêntricas, dicêntricos, tricêntricos e translocações. As aberrações do tipo-cromatídico ocorrem após a duplicação do DNA e envolvem apenas uma das cromátides do cromossomo.

Existem duas principais teorias para a formação de aberrações cromossômicas, chamadas de teoria de “*breakage-and-reunion*” (quebra-e-arranjo) de [Sax, 1940-1941], e teoria “*exchange*” (trocas) de [Revell, 1953]. São adaptações das teorias de “*breakage first*” (primeira quebra) de [Stadler, 1931] e “*contact first*” (primeiro contato) de [Serebrovsky, 1929] respectivamente.

De acordo com a teoria *breakage-and-reunion*, o principal evento causado por radiação é a quebra cromatídica ou cromossômica, e os tipos de aberrações observados na mitose dependem do resultado das quebras, por exemplo, as cromátides podem se restituírem ou rejuntarem (fusão não legítima) com outra quebra que está próxima o suficiente em espaço e tempo. Da mesma forma, de acordo com a teoria de *exchange* a principal lesão causada pela radiação não é a quebra (cromossômica ou cromatídica), mas a menor lesão que decai com o tempo.

Baseados nestas teorias de trocas e translocações, muitos pesquisadores projetam que o futuro da análise cromossômica e das pesquisas em aberrações cromossômicas deverá estar focado no estudo dos mecanismos envolvidos, para que haja uma melhor compreensão neste campo [Natarajan, 2002].

Desde os eventos iniciais que ocorrem no núcleo durante a interfase, esforços devem ser prestados para desvendar esta estrutura cromossômica, e responder a questionamentos, como por exemplo, como os cromossomos são organizados nos núcleos, se ocupam domínios específicos, como também se os movimentos dos cromossomos seguem ações específicas.

II-2. Aspectos Computacionais em Análises de Imagens

Atualmente, um grande espectro de diferentes aplicações em técnicas de reconhecimento de padrões e inteligência artificial é encontrado na literatura principalmente envolvendo temas como a melhor forma de obtenção da informação de interesse como também a melhora da eficiência em tempo de resposta. Este avanço e a

consequente redução de custo de produção da tecnologia envolvida desencadearam um grande impulso no desenvolvimento do auxílio computacional, principalmente n auxílio à medicina [Jain et.al., 2000].

Os aplicativos chamados de sistemas especialistas são projetados para que o trabalho de um especialista do qual se exige um esforço excessivo da atenção como, por exemplo, a identificação e o reconhecimento de características e classificação de objetos conhecidos em determinada imagem, cada vez mais vêm se difundindo entre cientistas, pesquisadores especialistas e empresas em todo o mundo. Dentre eles, sistemas responsáveis pela automatização da análise cromossômica se desenvolveram a partir dessas aplicações computacionais, e gera esforços de pesquisadores desde o início das aplicações em processamento de imagens e reconhecimento de padrões (início da década de 70).

Uma grande dificuldade do estudo surge com a multidisciplinaridade do conhecimento, que de certa forma, apresenta resultados restritos a determinadas aplicações, onde parcerias se tornam indispensáveis para a obtenção de resultados satisfatórios.

As rotinas computacionais em análises cromossômicas, que incluem processamento de imagens, detecção de bordas, segmentação, registro, extração de características e classificação são apresentadas por um espectro amplo de pesquisadores em diferentes áreas do conhecimento. São demonstrados e discutidos desde as técnicas de filtros em pré-processamento [Nickolls, 1981], limiar (*threshold*) e binarização para eliminação de fundo [Moradi, 2006], técnicas de análise de histogramas, para realizar a detecção do centrômero do cromossomo analisado [Detcheva, 1998], [Moradi, 2006] e correções angulares das cromátides nas imagens cromossômicas [Roshtkhari, 2008].

Destacam-se em Roshtkhari [2008], as técnicas de rotação e concatenação de partes da imagem para se realizar registros, cálculos matemáticos na análise horizontal de histogramas propostos, extração de características. Esta referência serviu de base para o segundo método de reconhecimento da localização dos centrômeros nas imagens cromossômicas propostos por este trabalho. Esta funcionalidade será melhor detalhada na seção (IV.3.2.6.4 Localização e Identificação de Centrômero(s) na Imagem Cromossômica).

Outro exemplo, com o desenvolvimento de um bom controle de escalas, é possível determinar o comprimento do cromossomo, de suas cromátides. Uma enorme variedade de ferramentas ainda está por ser desenvolvidas e disponibilizadas para a pesquisa

computacional, como para a pesquisa citogenética propriamente dita, e consequentemente para a pesquisa dos efeitos biológicos da radiação ionizante.

Um aplicativo de destaque chama-se *Chromosome Analysis Intelligent System (CAIS)* [Detcheva, 1998], *software* que apresenta características semelhantes ao aplicativo deste projeto, porém para outros fins específicos. A preocupação principal deste aplicativo é a detecção das bandas que contém o material genético através de imagens bidimensionais em tons de cinza, mas também, realizar a cariotipagem (separação dos cromossomos por características), localização do centrômero, entre outras ferramentas.

Porém, a área de atuação de diagnósticos por imagem não está restrita ao meio acadêmico. Inúmeras empresas se utilizam desta ferramenta, para disponibilizar instrumentos, equipamentos e *softwares* de altíssima qualidade e tecnologia para hospitais e clínicas em todo o mundo, empresas como Ge®, Lg®, Samsung®, etc. Na área de pesquisa deste projeto, destaca-se a empresa MetaSystems® com *softwares* de busca por metáfases (*Metaphase Finder*®), de cariotipagem (*Ikaros*®), entre outros.

Dentre os aplicativos desenvolvidos mundialmente para auxiliar o diagnóstico clínico de aberrações em cromossomos humanos desenvolvidos mundialmente, um exemplo é o *software* Metafer2 [Huber, 2001], cujas pesquisas da empresa MetaSystems® envolvem estudos na área de aquisição, processamento, tratamento, extração de características de imagens cromossômicas. Este aplicativo destaca-se por analisar aberrações cromossômicas e consiste em analisar imagens de cromossomos, porém trata-se de outro tipo de técnica citogenética, a *FISH* (*fluorescence in situ hybridization*), que consiste em uma técnica utilizada para detectar e localizar a presença ou ausência de sequências específicas de DNA nos cromossomos.

Porém, para a utilização desta ferramenta é necessária a compra de todo o equipamento de aquisição de imagens específico associado (microscópio, câmeras, *software*), elevando o custo associado a esta utilização.

III FUNDAMENTOS TEÓRICOS

Esta seção descreve alguns fundamentos teóricos das técnicas computacionais utilizadas para o desenvolvimento do aplicativo. Cada etapa apresentada neste capítulo, como os processamentos e as transformações em que a imagem cromossômica é submetida, foi incluída no aplicativo.

III-1. Processamento de Imagens Digitais

Uma imagem pode ser definida como uma função bidimensional $f(x, y)$, onde x e y são coordenadas espaciais (planas) e possuem a amplitude com o valor de f para qualquer par (x, y) , chamada de intensidade ou *gray level* de uma imagem para qualquer ponto. Quando x , y e o valor da amplitude f são finitos e discretos, chamamos de uma imagem digital. O campo de pesquisas *Processamento de Imagens Digitais* refere-se a processar imagens digitais por meio da computação digital [Gonzalez e Woods, 2000].

As aplicações de técnicas de pré-processamento de imagens digitais, são utilizadas quando o problema estudado necessita da melhoria na qualidade inicial de uma imagem, antes mesmo de as imagens serem analisadas para a extração de características e aquisição de parâmetros, ou apenas o termo processamento de imagens digitais, para diversas aplicações de transformações espaciais e reconhecimento em imagens.

Estas técnicas são utilizadas quando é necessária a obtenção de informações presentes nas imagens e/ou em parte delas e abrangem as mais diversas áreas do conhecimento, como expressões faciais, imagens de satélites, e principalmente em imagens médicas.

Após a aquisição de uma imagem, para qualquer valor de energia no espectro EM, estas imagens podem e normalmente devem ser tratadas de forma que possibilite ou apenas facilite o entendimento do problema analisado, ou seja, para que uma imagem possa ser

analisada ou estudada ela deve estar adequada à extração do maior número de informações de interesse ao pesquisador.

O pré-processamento de imagens envolve as técnicas computacionais para o realce de contrastes, histogramas de representação das intensidades dos *pixels*, aplicação de filtros para remoção de ruído, a preparação para a separabilidade de regiões de interesse, o realce para a extração de características, entre outras, possibilitando a obtenção de entendimentos e informações sobre a imagem estudada.

O estudo em questão trata as imagens em tons de cinza, sendo assim, dispensará a utilização de filtros digitais associados à cor. A função do pré-processamento é adequar as imagens de forma a aumentar as chances para o sucesso dos processos seguintes.

Na literatura encontram-se diversos tipos de filtros de pré-processamento em imagens digitais, destacam-se: os métodos baseados em convolução, transformada de Fourier, filtros não lineares de ajuste e manipulação do contraste da imagem, entre outros. Para este projeto o método computacional utilizado para o realce da imagem foi a manipulação do contraste da imagem cromossômica. A utilização e aplicação serão mais bem detalhadas no capítulo (IV - MATERIAIS E MÉTODOS).

A manipulação do contraste é uma técnica de pré-processamento de imagem comumente utilizada em sistemas especialistas envolvendo imagens digitais, entre elas: o mapeamento de valores de cinza, modificações em histogramas, na qual as intensidades de cinza são dimensionadas e seu histograma segue a uma determinada distribuição de intensidades no intervalo de 0 (zero) a 255 (duzentos e cinquenta e cinco), a limiarização (*thresholding*) de intensidades, na qual, através de um limiar se determina certo intervalo de intensidades de tons cinza, de acordo com o interesse e preservação dos dados. A manipulação do contraste e a modificação do histograma de intensidades permitem que as intensidades em tons de cinza de um intervalo de intensidades, sejam utilizadas de acordo a uma tarefa específica [Nickolls, 1980].

Imagens em formato digital também permitem o armazenamento de informações para diversas finalidades, tais como: registro do problema (possibilitando análises posteriores), a reanálise do cromossomo, objeto estudado, amostras ou mesmo de informações referentes a um paciente, entre outras possibilidades. Estas imagens apresentam-se em uma variedade de formatos de armazenamento digital, dos quais se destacam: *Joint Pictures Expert Group (JPEG)*, *Graphics Interchange Format (GIF)*,

Tagged Image File Format (TIFF), *Portable Network Graphics (PNG)*, e para imagens médicas o DICOM (imagens médicas com formato padronizado – contêm a imagem e informações técnicas sobre a imagem adquirida), tem a finalidade de que máquinas hospitalares de aquisição de imagens (CT, PET, etc.) possam reconhecer informações de outras máquinas em diferentes hospitais, entre outras possibilidades de informações registradas. Neste trabalho todas as imagens que foram fornecidas pela equipe da Dr^a Kayo estão em formato JPG, obtidos diretamente de uma câmera fotográfica comum acoplada à ocular do microscópio, como também as imagens resultantes após a análise mantêm o mesmo formato.

III-2. Segmentação de Imagens

A segmentação é utilizada na maioria dos tratamentos de imagens digitais, cuja aplicação é um tema de extrema relevância na área de imageamento. A técnica computacional de segmentar as imagens refere-se à possibilidade de subdividir imagens, em “porções” de *pixels* diferenciados por regiões, retirando o fundo (*background*) com informações desnecessárias ao estudo e destacando os objetos de interesse presentes na imagem, no entanto, com o compromisso de preservar ao máximo as características iniciais das imagens analisadas.

Inúmeras técnicas de segmentação são objeto de estudo em diferentes áreas. Algumas de maior interesse para este trabalho são: a segmentação por método de *thresholding* (limiar) [O’Gorman, 1994], algoritmos de detecção de borda (como o Sobel, o Canny, o Snakes [Ballerini, 2003]), entre outras. No caso de reconhecimento dos cromossomos presentes na imagem, o papel básico da segmentação está ligado a *pixels* conectados após a eliminação do *background* e a partir disto, auxiliar na eliminação do fundo da imagem e tornar possível a extração características individuais dos cromossomos.

A segmentação de uma imagem subdivide uma imagem em suas partes ou objetos constituintes, cujo nível até o qual essa subdivisão deve ser realizada, assim como a técnica utilizada, depende do problema que está sendo resolvido. A segmentação de imagens com resolução e compreensão do objeto estudado é uma das tarefas mais difíceis do processamento em imagens digitais. A acurácia da segmentação determina o eventual sucesso ou falha do procedimento de análise computadorizada [Gonzalez et al., 2004].

Algoritmos de segmentação normalmente são baseados em uma das duas propriedades em valores de intensidades: descontinuidade e similaridade. Na primeira, a técnica consiste em particionar a imagem baseada em mudanças brutas em intensidades, como as bordas em uma imagem. E a principal, a segunda categoria, que consiste em particionar a imagem de acordo com um critério pré-definido.

A segmentação de um objeto em uma imagem é o processo de definição da região/limites deste, numa determinada cena. Consiste basicamente em duas tarefas: reconhecimento e delimitação. O reconhecimento é uma tarefa difícil e qualitativa para a determinação mais próxima da área do objeto; já a delimitação é uma tarefa mais fácil e quantitativa, que consiste em especificar precisamente a localização e as extensões do objeto, a menos que eles estejam sobrepostos. Algoritmos de delimitação são extremamente precisos, obtendo resultados mais próximos do real do que quando realizados manualmente, mesmo que por um especialista [Udupa, 2002].

Uma técnica de segmentação em imagens digitais, muito utilizada por pesquisadores da área de diagnóstico por imagens é a detecção de borda e a rotulação de objetos presentes na imagem. A rotulação de objetos tem como objetivo testar todas as conectividades de cada um dos *pixels* examinados na imagem. Cada *pixel* de uma imagem digital possui seus adjacentes, ou seja, cada *pixel* presente em uma imagem digital possui certo número de conexões com outros *pixels*, de intensidade (*gray scale*), dependente do intervalo de intensidade da imagem analisada.

Teoricamente, definem-se alguns tipos de conexões possíveis, como a “N-4”, que considera os quatro vizinhos do *pixel* analisado, a “N-8”, que analisa os 4 (quatro) vizinhos laterais e os de aresta, e a conexão “N-26” que analisa em 3 (três) dimensões o *voxel* (*volume pixel*), entre outras. A Figura 2 apresenta um esquema representativo das vizinhanças possíveis de um *pixel*.

Figura 2 – A) O pixel que está sendo analisado; B), A vizinhança N-4 de um pixel; C) A vizinhança N-8 de um pixel; D) vizinhança N-26 de um voxel.

A vizinhança N-4 de um *pixel* $p(x, y)$ é determinada pelos *pixels* $(x-1, y)$, $(x+1, y)$, $(x, y-1)$ e $(x, y+1)$, da mesma forma que a vizinhança N-8 é determinada por $(x-1, y+1)$, $(x-1, y)$, $(x-1, y-1)$, $(x, y+1)$, $(x, y-1)$, $(x+1, y+1)$, $(x+1, y)$ e $(x+1, y-1)$, e assim sucessivamente, para as outras vizinhanças. Essas coordenadas são muito utilizadas durante o processo de segmentação e registro da imagem estudada neste trabalho. Como já mencionado, este projeto trata de imagens em tons de cinza e as conectividades testadas são as de vizinhança N-8, como mostra a Figura 3.

	$(x-1, y+1)$	$(x, y+1)$	$(x+1, y+1)$	
	$(x-1, y)$		$(x+1, y)$	
	$(x-1, y-1)$	$(x, y-1)$	$(x+1, y-1)$	

Figura 3 – Representação das coordenadas da vizinhança do pixel analisado.

A determinação dos vizinhos de um *pixel* pertencente ao objeto que se deseja segmentar é essencial a todos os processos seguintes na análise da imagem. Se por um lado, um procedimento de segmentação robusto favorece substancialmente uma solução bem sucedida de um problema em imagens, de forma diferente, algoritmos de segmentação fracos ou errôneos geram falhas graves no processamento das informações de interesse.

O processo de segmentação envolve essencialmente a delineação das características de interesse no interior de uma imagem, normalmente é utilizada para as seguintes aplicações em imagens digitais:

- Cálculo de volumes;
- Quantificação de anormalidades;
- Visualização de estruturas;
- Identificação automática de lesões;
- Morfologia de estruturas, entre outras.

Um exemplo clássico de segmentação que é pertinente a este trabalho é o problema da identificação e extração dos objetos de interesse da imagem quanto aos chamados falsos positivos e falsos negativos. Os falsos positivos são caracterizados pela informação desnecessária relativa ao fundo, sendo considerada como válida e contada de forma

verdadeira, e os falsos negativos são caracterizados pela informação de parte do objeto segmentado que foi desconsiderada por se assemelhar ao fundo (*background*).

Quanto à resolução espacial de uma imagem considerada para uma classificação, a perda de informação real e/ou obtenção de informação não verdadeira, pode prejudicar a eficiência identificação dos padrões.

III-3. Registro em Imagens

O registro em imagens é utilizado no caso de uma imagem aberta (*input image*) ou no caso de alguma imagem não estar posicionada de forma ideal para que um determinado estudo/aplicação seja feita, ou de alguma forma, seja necessário o re-mapeamento da imagem para um domínio conhecido. Sendo assim, uma imagem armazenada com posicionamento inadequado pode ser registrada, deformada e/ou reposicionada novamente, de acordo com o estudo em questão.

O registro espacial em uma imagem pode ser realizado por diferentes métodos de transformações espaciais em uma imagem, destacando entre eles: a rotação, a translação, o esticamento, o reposicionamento de parte da imagem ou da imagem inteira, entre outras operações morfológicas. O objetivo desta técnica computacional é identificar correspondências entre diferentes imagens, utilizando a disposição dos objetos na imagem para encontrar características morfológicas semelhantes, ou pela busca de objetos (pontos) comuns na segunda imagem, conhecidos como *landmarks* ou pontos de controle. Estes pontos similares em mais de uma imagem são base para o estabelecimento de correspondências e para a escolha de qual transformação espacial determinada é mais apropriada para mapear uma imagem na outra. O passo final é sujeitar uma imagem a esta transformação [Hejlesen et al., 2004].

Outra aplicação do registro em imagens é o reposicionamento da imagem para a posterior análise, como por exemplo, o reposicionamento de uma imagem para ser analisada por um algoritmo de extração de características através do reconhecimento de padrões. Neste caso, procura-se otimizar o posicionamento adequado para a análise (escopo deste projeto), muitas vezes de forma empírica.

Nas imagens cromossômicas, o registro foi utilizado para alinhá-los de duas formas: na primeira, aplicando-se a transformação pela transposta da matriz (imagem cromossômica) para posicionar a imagem de forma vertical, ou seja, sua maior coordenada

de comprimento da matriz $M \times N$ será a coordenada N ; na segunda rotacional, aplica-se a transformação rotacional até uma posição tal, na qual as imagens individuais estejam posicionadas de modo que possam ser analisadas tanto pelo especialista como pelo próprio *software* desenvolvido neste trabalho.

A preocupação com a perda das informações relevantes é importante nesta etapa do processamento das imagens. As operações morfológicas realizam transformações matriciais a “base” da imagem alterando na disposição da imagem e afetando diretamente a possível identificação posterior do objeto estudado na imagem.

As imagens cromossômicas utilizadas neste projeto apresentam, em média, um conjunto dos 46 cromossomos, e a análise individual de cada cromossomo exige a segmentação e registro apropriado da imagem em cada cromossomo. Posteriormente serão discutidos estes desenvolvimentos.

III-4. Reconhecimento de Padrões

O reconhecimento de padrões em imagens digitais é um dos núcleos de desenvolvimento da inteligência artificial e está relacionado com a caracterização de um objeto de interesse presente na imagem em estudo. Este processo consiste em identificar e classificar objetos presentes nas imagens, baseado em conhecimentos prévios do problema proposto, ou em informações estatísticas extraídas a partir de características das imagens.

Um conjunto de medidas que denotam características das imagens cromossômicas é extraído dos dados de entrada e usado para associar cada vetor de características a uma classe de imagem. Muitas vezes pode-se presumir que as características obtidas estão associadas a um estado intrínseco da natureza da imagem, e, portanto, potencialmente existe um modelo subordinado a um estado, ou a um conjunto de probabilidades, e/ou funções de densidade de probabilidade correspondente, passível de ser usado para representá-lo [Schalkoff, 1992].

A partir deste conjunto de características presentes nas imagens, que podem ser relacionadas, mensuradas e comparadas, forma-se uma base de conhecimento sobre o padrão de interesse, que neste trabalho tem a finalidade de identificar cada cromossomo presente em uma imagem. Com esta tarefa realizada, técnicas de inteligência artificial, neste estudo utilizado um Sistema de Inferências *Fuzzy* podem ser aplicadas para a caracterização e identificação de cada objeto na imagem.

Sem estas características retiradas diretamente da imagem, não seria possível distinguir classes distintas de imagens. O reconhecimento de padrões característicos é a parte de fundamentação da classificação, e será baseado neste processo que o aplicativo poderá ser validado por especialistas.

III-5. Inteligência Artificial

A Inteligência Artificial (IA) tem se tornado um ramo do conhecimento computacional que têm incluído aplicações em diversas áreas do conhecimento e utilizada tanto no campo acadêmico, como no comercial ou industrial. Este ramo têm abrangido um conjunto amplo de técnicas que já foram implementadas em produtos diversos, desde jogos de última geração à máquina de lavar, elevadores, entre outros.

A discussão histórica da definição que envolve o termo Inteligência Artificial envolve avaliar o que pode ser considerado um conhecimento inteligente, ou a capacidade de tomar decisões por determinados algoritmos, ou mesmo, reproduzir numa máquina, indícios de consciência semelhantes ou não a de humanos.

Encontrar uma definição que satisfaça a todas as partes ainda não foi um fato concretizado. Do mesmo modo que a IA “forte” refere-se a uma possibilidade de reproduzir a inteligência, como a de humanos. Neste caso, as decisões são tomadas a partir da capacidade de reconhecer sensorialmente o ambiente que o cerca, e tomar “independentemente” uma ou mais atitudes, decisões, diante a determinadas situações, ou problemas.

Na área computacional a inteligência artificial, também chamada de inteligência computacional, surgiu por volta de 1950 com Alan Turing [1950], e seu modelo de avaliação de inteligência conhecido como teste de Turing, porém, somente em 1956 é que a inteligência artificial começou a ser reconhecida como ciência.

Cronologicamente, em 1956, o termo Inteligência Artificial foi usado pela primeira vez por John McCarthy em uma conferência em Dartmouth *College*, em Hanover, New Hampshire. Em 1957, Newell e Simon desenvolveram a ideia do *GPS (General Problem Solver)*, cujo núcleo do algoritmo utilizava uma metodologia conhecida como análise de meios extremos, baseada na ideia de determinar antecipadamente o que precisava ser feito e então processar uma maneira de executar a ação. Isso funciona bem para problemas com soluções simples, mas os pesquisadores logo perceberam que este tipo de método não

poderia ser aplicado de uma maneira geral, concluindo que o *GPS* somente poderia resolver alguns problemas bastante específicos. E em 1958, McCarthy desenvolveu a linguagem de programação LISP, que ainda é amplamente utilizada hoje em pesquisas em Inteligência Artificial [Coppin, 2004].

A partir daí, a evolução na área de IA, ocorreu lado a lado com a evolução da tecnologia de computadores, na qual, as máquinas passaram a ser programadas e “treinadas” para desenvolver e resolver tarefas. Atualmente, a IA computacional é uma área da ciência que estuda a teoria em aplicação de técnicas de algoritmos de IA com lógicas básicas como um conjunto de regras básicas de computação até os chamados neurônios lógicos utilizados no conceito de redes neurais. Dentre estas técnicas pode-se destacar: Redes Neurais, Sistemas de Lógica *Fuzzy*, Algoritmos Genéticos e Computação Evolucionária, entre outras.

Aplicações em chamados sistemas especialistas são algoritmos desenvolvidos com o intuito de reproduzir determinado conhecimento humano relacionado a algum conhecimento específico (especialista, médico, biólogo, entre outros). Estas aplicações têm como objetivo, por exemplo, controlar um sistema com algumas variáveis ou mesmo automatizar uma rotina exaustiva de trabalho. Para que seja possível o desenvolvimento deste tipo de algoritmo, é fundamental um alto detalhamento do problema estudado e a ser resolvido. Neste trabalho a intenção principal é reproduzir o conhecimento do profissional citogeneticista em análise de células irradiadas no reator de pesquisas IEA_R1.

Para que este objetivo fosse alcançado este estudo foi fundamental a utilização de técnicas avançadas de inteligência artificial no que diz respeito ao reconhecimento de padrões (características) extraídos de cada cromossomo. Sendo assim, um modelo de lógica *Fuzzy* será proposto para realizar computacionalmente esta caracterização.

III-5.1. Modelo *Fuzzy*

O *FLS - Fuzzy Logic System* ou *SLN - Sistema de Lógica Nebulosa* é um dos poucos métodos lógicos aptos a lidar simultaneamente com dados numéricos e conhecimentos linguísticos. Em última instância, faz um mapeamento não linear de um vetor de dados de entrada (características) em outro vetor de saída, mapeando números em números, “*Crisp Inputs*” em “*Crisp Outputs*”, através de um sistema de regras linguísticas, e este mapeamento pode ser expresso quantitativamente como $y = f(x)$. Os Conjuntos *Fuzzy*

(*Fuzzy Sets*) e a Lógica *Fuzzy* determinam as especificidades do mapeamento não linear [Mendel, 1995].

Um Conjunto *Fuzzy* é uma classe de objetos com uma grade de relacionamentos. Cada *conjunto* é caracterizado por uma função de relação (característica) que atribui para cada objeto, uma grade de relacionamento que pode ser associada a um valor verdade entre 0 (zero) e 1 (um). As operações de inclusão, união, intersecção, complemento, relação, convexidade, etc., estabelecem os vários contextos de possíveis aplicações dos conjuntos *Fuzzy* [Zadeh, 1965].

Existem vários tipos de problemas de representação do conhecimento na literatura: os baseados em modelos matemáticos conhecidos (conhecimento objetivo), baseados unicamente na base de dados (*data-driven*), e os chamados modelos-livres (conhecimento subjetivo) representados por afirmações linguísticas, que normalmente, não pode ser traduzido em modelos matemáticos e caracterizado em regras, e então quantificado. A Lógica *Fuzzy* e outras semelhantes técnicas de inteligência artificial permitem estes tipos de representação.

A Lógica *Fuzzy* pode ser vista como uma importante inovação e aperfeiçoamento da interface homem-máquina no desenvolvimento de sistemas computacionais inteligentes, sendo uma generalização da Lógica Clássica ou Lógica de Predicados, admitindo valores lógicos intermediários entre o falso e o verdadeiro (como o talvez, o muito, o médio, o pouco, entre outros).

A contribuição principal deste tipo de lógica está na possibilidade da representação do conhecimento não modelado utilizando uma representação matemática e saída lógica associada. A riqueza da Lógica *Fuzzy* requer um entendimento cuidadoso dos elementos que a compõem, onde, cada elemento pode validar ou inviabilizar a capacidade de representar o conhecimento do problema.

O pesquisador Lotfi Zadeh é considerado o pai fundador de todo o campo da Lógica *Fuzzy*. Em seus artigos encontram-se definições como: “... o fato permanece tal que... ‘classes’ imprecisamente definidas, desempenham um papel importante no pensamento humano, particularmente, no domínio do reconhecimento de características, comunicação da informação e abstração”, e também, “o quanto aumenta a complexidade do sistema, nossa habilidade de tornar preciso, como também de definir especificidades

sobre o ambiente que se trata, diminui até atingir um limiar no qual a precisão e a significância tornam-se quase características mutualmente exclusivas” [Mendel,1995].

Teoricamente, um sistema *Fuzzy* é uma estrutura constituída basicamente por um fuzzificador, uma base de regras, um método de inferência e um defuzzificador, que através de variáveis linguísticas (estados) podem quantitativamente expressar o conhecimento de um especialista, em determinada área do conhecimento, em termos comumente utilizados pela linguagem humana. A Figura 4 esquematiza as componentes de um Sistema de Lógica *Fuzzy*.

Figura 4 – Sistema de Lógica *Fuzzy* [Cecconello, 2006].

No fuzzificador cada entrada “*crisp*” (X) é transformada em um conjunto *Fuzzy*, ou seja, se $X \in \mathbb{R}^n$, uma variável de entrada do sistema, o fuzzificador associa a esta entrada, uma função de pertinência (relacionamento) a um conjunto *Fuzzy* $\mu_p(x)$ $[0-1]$, $x \in \mathbf{D}$, conjunto Universo de Discurso Linguístico. Assim, um chamado número “*crisp*” sempre estará adequado a uma variável linguística com valor dado por $\mu_p(x)$. A base de regras é um conjunto formado por regras *Fuzzy* que relacionam os termos linguísticos das variáveis de entrada $\mu_{pENTRADA}(x)$ e saída $\mu_{pSAÍDA}(x)$ (relações entre conjuntos *Fuzzy*), e que é construída baseada na tradução do conhecimento do especialista sobre a inferência a ser feita. E por fim, a “defuzzificação” das variáveis de saída, por meio da base de regras são obtidas as respostas “*crisps*” do sistema (Y) $Y \in \mathbf{V}$, avaliadas a partir do valor lógico assumido pelas variáveis linguísticas de saída do sistema de Inferências. Para cada estado definido pelos termos linguísticos das variáveis de entrada, é definida uma regra, e a partir deste conjunto de regras é definida uma variável de saída [Cecconello, 2006].

O método de inferência é o mecanismo pelo qual as diferentes relações lógicas, definidas pela base de regras, são avaliadas. As operações lógicas de composição entre as diferentes regras utilizam diferentes operadores matemáticos para calcular e atribuir pertinências aos consequentes destas regras. Em geral estas regras estão na forma de implicações do tipo *IF-THEN*, e têm o Valor Verdade atribuído a cada uma. Cada regra *Fuzzy* representa uma relação entre os Conjuntos *Fuzzy* de Entrada e Saída. Sendo assim, quanto mais termos linguísticos, mais informações são incorporadas na modelagem, e, quanto melhor detalhado é o problema, maiores são as chances de acerto do sistema Fuzzy.

Dentre os Conjuntos Fuzzy e suas operações lógicas, pode se aplicar, de forma análoga, operações elementares da teoria de conjuntos, como por exemplo, as operações de união, interseção e complemento. Alguns exemplos destas operações são: Sendo A e B um subconjunto do universo **D**. A união de A e B, denotada por $A \cup B$, contendo todos os elementos em A e B, sendo, $\mu_{A \cup B}(x) = 1$, se $x \in A$ ou B, e $\mu_{A \cup B}(x) = 0$, se $x \notin A$ e $x \notin B$. A interseção de A e B, denotada por $A \cap B$, contendo todos os elementos que estão simultaneamente em A e B, sendo, $\mu_{A \cap B}(x) = 1$, se $x \in A$ e $x \in B$, e $\mu_{A \cap B}(x) = 0$, se $x \notin A$ ou $x \notin B$. Sendo $\bar{A}$ denotado pelo complemento de A; este conjunto contém os elementos que não estão em A, sendo, $\mu_{\bar{A}}(x) = 1$, se $x \notin A$, e $\mu_{\bar{A}}(x) = 0$, se $x \in A$ [Mendel, 1995]. Entre outras operações baseadas na teoria de conjuntos.

Uma representação de um simples modelo Fuzzy, é a associação de um valor $\mu_p(x)$ a uma proposição p, indicando o grau de veracidade dessa proposição (não probabilidade), sendo que $\mu_p(x)$ é uma função cujo conjunto imagem pertence ao intervalo [0-1]. Outra forma de se exemplificar um sistema *Fuzzy* é através da composição de diferentes modos de representar suas variáveis, desde que, utilizadas para compreensão e descrição do problema estudado. Essas variáveis possuem funções de pertinência dentro de um intervalo conhecido, com valores discretos. Esta forma de compreensão de um sistema *Fuzzy* (valores discretos) pode ser usada, como por exemplo, uma forma de representar valores de critérios específicos, se tornando uma variável de inclusão ou exclusão, ou mesmo, uma variável de classificação.

A metodologia *Fuzzy* na definição dos parâmetros de caracterização, de cada padrão de segmentação, e das aberrações cromossômicas que possam ser detectáveis numa imagem pelo aplicativo, foi realizada a partir de dados obtidos no reconhecimento de características: tamanho, preenchimento e identificação e reconhecimento de perfis para a

classificação final, com a adição da característica de presença e localização do(s) centrômero(s). Outras características intrínsecas de determinada imagem cromossômica podem ainda ser adicionadas a este conjunto de regras.

IV - MATERIAIS E MÉTODOS

Nesta seção são apresentados os recursos utilizados e as etapas de desenvolvimento da metodologia proposta por este trabalho.

IV-1. Infra-Estrutura

O trabalho engloba três ambientes de desenvolvimento: *facility* do reator de pesquisa (IEA-R1) onde as amostras foram irradiadas com nêutrons; laboratório de cultura celular do centro de biotecnologia e realizada a cultura de células e aquisição das imagens; e por fim, o ambiente de desenvolvimento do software a ser projetado.

IV-1.1. Facility do reator de pesquisa (IEA-R1)

A equipe do Centro de Engenharia Nuclear (CEN) cujo responsável é o Dr. Paulo Rogério Pinto Coelho projetou e mantém as aplicações do *facility* de irradiação no reator IEA-R1 para o desenvolvimento da técnica de *NTC - neutron capture therapy* (terapia por captura de nêutrons). Estas instalações possibilitam a utilização da radiação emitida pelo reator IEA-R1 nas amostras das células analisadas por este projeto. Conforme [Coelho et. al., 2008], as condições térmicas atuais da irradiação por nêutrons são de 32.2 ± 0.1 Gy/h de taxa de dose, com 25% de contaminação gama para 3.5MW de energia operacional do reator. Estas células são preparadas pela equipe da Dra. Kayo Okazaki e irradiadas no *facility* do reator pela equipe do Dr. Coelho.

IV-1.2. Laboratório de Cultura Celular do Centro de Biotecnologia (CB)

Esse laboratório, cuja responsável a Dra. Kayo Okazaki, conta com uma sala estéril, com os seguintes itens relacionados: fluxos laminares TROX e FORMA SCIENTIFIC, estufas de CO₂ FORMA SCIENTIFICA, estufa bacteriológica FANEM e microscópio

invertido CARL ZEISS; conta também com salas de preparação e lavagem, com os seguintes itens: centrífugas SORVALL e FANEM, banhos-maria FANEM, autoclave, balança METTLER, geladeiras e freezer, estufa FANEM para esterilização e purificador de água – sistema Milli-Q; e por fim, a sala de microscopia, com os seguintes itens relacionados: fotomicroscópio CARL ZEISS, microscópio de fluorescência CARL ZEISS e lupas CARL ZEISS.

IV-1.3. Equipamento e sistema operacional para o desenvolvimento do aplicativo

Para o desenvolvimento do aplicativo utilizou-se microcomputadores do Centro de Engenharia Nuclear, que detém licenças do software MATLAB®, e de seus acessórios (toolboxes) relativos à Lógica *Fuzzy* e ao Processamento de Imagens. Também será utilizado um computador pessoal do aluno com processador AMD® Athlon XP 64 3200+ @2.2GHz, 1GB de memória RAM e placa de vídeo GeForce® 7300GT com 256MB de memória RAM. O sistema operacional utilizado Microsoft® Windows 7.

IV-2. Descrições das Etapas para o Desenvolvimento da Metodologia

Inicialmente foi estudada a aplicação de filtros digitais com a finalidade de realçar os objetos presentes na imagem (cromossomos) preservando a informação de interesse e eliminando o que não se deseja estudar. Na segunda fase do estudo foram aplicadas técnicas de seleção, reconhecimento e extração de características e técnicas de processamento digital como, segmentação e registro na imagem bidimensional, seguidas pela quantização de variáveis de interesse (tamanhos, preenchimentos, etc.) e consequentemente o reconhecimento destes padrões. Numa etapa final, foi realizada a classificação das aberrações cromossômicas utilizando a lógica *Fuzzy* para inferir a que classe de aberração a imagem pertence, a partir dos valores das variáveis obtidas para cada característica selecionada na fase anterior.

Para a obtenção de uma classificação otimizada das aberrações cromossômicas, uma sequência de processos foram estudados: aquisição da imagem, pré-processamento, representação e reconhecimento das características em imagens e classificação cromossômica. A Figura 5 apresenta um modelo de um diagrama de etapas da metodologia

de reconhecimento de padrões em imagens digitais deste trabalho que inclui os módulos do aplicativo.

Figura 5 - Diagrama de etapas de um sistema de reconhecimento de padrões em imagens digitais.

A etapa de formação da base de dados é composta pelos processos de aquisição e armazenamento da imagem a ser estudada. Esta etapa do envolve desde a cultura de células até o posicionamento da câmera fotográfica.

A etapa de identificação é constituída por um conjunto de técnicas de pré-processamento de imagens que são aplicadas com o intuito de melhorar o acesso às características desejadas a serem extraídas das imagens, obtendo-se assim uma base de dados adequada à classificação dos cromossomos. Dentre as diversas opções, foram testados e utilizados filtros para o realce de contraste, segmentação para a detecção de bordas e extração de características, e registro, realizando um alinhamento e posicionamento mais adequado da imagem.

Na terceira etapa, valores e critérios de comparação foram estabelecidos para a futura classificação. Em termos práticos, uma característica importante considerada, por exemplo, o comprimento normal de cada cromossomo (reconhecidos pelo tamanho das imagens cromossômicas), já que eles normalmente são classificados de acordo com o seu tamanho.

Para uma classificação *Fuzzy* é necessária a utilização de várias características inerentes ao objeto a ser classificado. O conjunto destas características inseridas em termos linguísticos permite a utilização de um sistema de Inferências *Fuzzy*.

IV-3. Apresentação da Metodologia

O desenvolvimento da metodologia consistiu em duas fases principais: a aquisição das imagens e o desenvolvimento do aplicativo para a determinação do número de cromossomos em cada imagem e posterior a identificação da localização dos centrômeros para a classificação das aberrações cromossômicas. O intuito principal de classificação do aplicativo proposto é a localização e contagem dos centrômeros em imagens cromossômicas.

IV-3.1. Aquisições de Imagens de Células Irradiadas

As imagens analisadas neste estudo foram fornecidas pelas pesquisadoras Dr^a. Kayo Okazaki e Dr^a. Márcia Augusta da Silva – CB-IPEN/USP.

A aquisição das imagens é feita de acordo com o seguinte processo: as amostras de sangue humano são irradiadas no *beam-hole* do reator de pesquisas IEA-R1 (IPEN-CNEN), na qual, aberrações cromossômicas são estimuladas propositadamente. Não há uma relação direta entre dose e tipo de aberração, ou seja, a irradiação de células humanas pode resultar em qualquer tipo de aberrações cromossômicas. Após o processo de irradiação as amostras são analisadas visualmente pelo profissional citogeneticista.

Com o desenvolvimento deste projeto, as imagens de células destas amostras são aquiradas e submetidas ao processamento, reconhecimento e caracterização das imagens digitais. Estas foram obtidas pela equipe da Dr^a Okazaki (obtidas acoplado uma câmera fotográfica ao ocular¹ de um microscópio). As aquisições foram realizadas através de uma câmera fotográfica comum Sony® modelo: Cyber-Shot DSC-P73, acoplada à ocular do microscópio CARL ZEISS. Foram selecionadas algumas das imagens com resolução e nitidez consideradas adequadas para análise e desenvolvimento dos algoritmos, criando-se assim uma etapa de estabelecimento de critérios mínimos de qualidade para as imagens. Foram descartadas as imagens em que os níveis de intensidade de cinza dos cromossomos, ou mesmo parte de um mesmo cromossomo, se assemelhassem ao valor de intensidades do *background*.

¹ A lente ocular é a parte de qualquer equipamento de observação óptica direta (telescópio, microscópio, binóculo, lupa binocular) que se destina à aplicação direta (com contato ou proximidade extrema) de um ou de ambos os olhos do observador. A lente ocular tem um diâmetro menor que o da lente objetiva, ela forma uma imagem virtual, maior e invertida com relação ao objeto.

As imagens apresentadas na Figura 6 demonstram as 11 imagens utilizadas neste trabalho.

Metáfase 002r

Metáfase 003r

metáfases 008b (1)r

Metáfase 011r

Metáfase 016r

Metáfase 019

Metáfase 070

Metáfase 0121

Metáfase 0214 Gy r

Metáfase 0224 Gy r

Figura 6 – Imagens cedidas pela equipe da Dr^a Kayo Okazaki e Dr^a Márcia Augusta da Silva.

A Figura 7 (imagem da Metáfase 0501) apresenta uma metáfase com a presença de um cromossomo com aberração do tipo dicêntrico e a presença de deleções terminais ou fragmentos (identificados com uma seta).

Figura 7 – Metáfase com a presença de aberrações cromossômicas [Dr^a. Kayo Okazaki e Dr^a. Márcia Augusta da Silva (IPEN)].

Alguns testes foram realizados com câmeras de diferentes resoluções e arranjos ópticos para a formação de um banco de imagens para o desenvolvimento desta metodologia de análise cromossômica automatizada. Um acoplador óptico foi testado para adaptar uma câmera Fuji® modelo: Finepix S9600 (8.0 megapixels) à ocular do microscópio CARL ZEISS. Este acoplamento foi testado com diferentes limites de configurações, entretanto o ganho em resolução em relação à máquina Sony® modelo: Cyber-Shot DSC-P73® (3.9 megapixels) previamente utilizada pela equipe da Dr^a Kayo, e não apresentou melhorias na eficácia dos algoritmos desenvolvidos para o reconhecimento dos padrões.

Optou-se pela configuração que oferecia uma resolução e acoplamentos simplificados, porém capazes de fornecer as informações necessárias e suficientes para a identificação e classificação das imagens. Desta forma, optou-se pela criação de uma rotina de aquisição e obtenção dos resultados com maior praticidade e utilizando-se de recursos básicos de aquisição de imagens. E assim, estabeleceu-se um padrão mínimo de qualidade na aquisição destas imagens para que a análise seja eficiente. Este padrão pode ser objeto de futuros estudos e o desenvolvimento de uma nova metodologia.

O aumento da resolução das imagens de aquisição não afeta significativamente os resultados de classificação pelos algoritmos deste projeto e acarreta o aumento significativo do tempo de processamento do aplicativo.

A opção pela simplificação de aquisição permite que o desenvolvimento do aplicativo possa ser direcionado a pesquisadores que não possam arcar com sistemas completos de análise de imagens de grande porte. Como já citado, estes “kits” subentendem a aquisição pelo pesquisador de todo o sistema (*hardware + software*).

Atualmente no CB-IPEN, a contagem de cromossomos como também sua classificação e diagnóstico é realizada manualmente, lâmina a lâmina, tomando um tempo excessivo do profissional, tornando-se uma prática minuciosa e exaustiva.

Para os testes realizados neste projeto, foram fornecidas 33 imagens de metáfases celulares, sendo que, foram selecionadas 11, de onde foram geradas (obtidas) 1.465 imagens cromossômicas individuais (considerando as três intensidades analisadas), utilizadas para a análise de dados e o desenvolvimento dos algoritmos de identificação e classificação de imagens cromossômicas, conforme mostrado na Tabela 1.

Metáfase	Dimensões	Nº imagens Limiar – 60 (23,52%)	Nº imagens Limiar – 70 (27,45%)	Nº imagens Limiar – 80 (31,37%)
002r	845 x 500	43	43	40
003r	640 x 511	45	45	45
008b (1)r	574 x 795	43	41	41
011r	885 x 810	44	44	40
016r	637 x 656	45	45	45
019	505 x 481	42	41	40
070	680 x 815	47	48	49
0121	604 x 517	43	43	43
0214 Gy r	779 x 807	49	49	49
0224 Gy r	387 x 784	47	47	47
0501	644 x 649	45	44	43
Nº de imagens cromossômicas	1.465	493	490	482

A Tabela 1 - Imagens metafásicas e a quantidade de imagens cromossômicas geradas para os estudos realizados neste trabalho.

O critério para a escolha destas 11 imagens cromossômicas se deve ao fato destas imagens apresentarem a melhor relação entre as intensidades dos *pixels* nos cromossomos e às intensidades dos *backgrounds*, implicando principalmente na otimização da segmentação (separabilidade) dos cromossomos e consequentemente o reconhecimento.

IV.3.2. Desenvolvimento do Aplicativo

O desenvolvimento do aplicativo elaborado é descrito por um diagrama de sequência demonstrando desde os filtros de pré-processamento da imagem, até a classificação do objeto (cromossomo). Foram estudadas e algumas implementadas no aplicativo diversas técnicas de processamento, segmentação, registro e classificação de imagens por meio de técnicas de reconhecimento de padrões, objetivando o desenvolvimento da metodologia proposta neste trabalho. A Figura 8 apresenta estas etapas do sistema de classificação das imagens cromossômicas num diagrama de sequência de execução.

Figura 8 – Diagrama de sequência de execução do aplicativo.

IV.3.2.1 Transformação das Imagens Cromossômicas para Tons de Cinza

O primeiro processamento a que a imagem cromossômica é submetida pelo *software* é o processo de “transformar” a imagem, inicialmente em tons *RGB* (*red-green-blue*), para a escala em tons de cinza (*gray scale*), intervalo de intensidades de pixels é igual a [0-255]. A aplicação desta transformação se baseia em formar uma imagem resultante, a partir da soma do mesmo índice de um pixel em cada uma das três “fatias” (diferente de três dimensões) da imagem. A determinação das tonalidades de cinza da nova imagem é obtida pela soma ponderada de três valores para todos os pixels que compõem a imagem metafásica de acordo com a Equação 1.0.

$$\text{Gray Level} = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (\text{Eq 1.0})$$

Onde, R, G e B correspondem às cores dos pixels em cada matriz de composição de uma imagem em RGB, respectivamente.

A equação 1.0 apresenta a aplicação da função do Matlab® ‘*rgb2gray*’. A utilização desta função nas imagens cromossômicas está ligada ao fato de o objeto ser contrastado do fundo da imagem, ou seja, o cromossomo tende a preto (*pixels* mais intensos e com menos valor de índice) e o fundo tende ao branco (baixo nível de coloração e alto valor de índice). Como a segmentação de interesse está baseada na morfologia dos cromossomos dispostos nas imagens, as cores não apresentam informações necessárias a este estudo.

Nas imagens em tons de cinza é possível aplicar inúmeros filtros já existentes e processamentos específicos para este tipo de imagem e o custo computacional é menor, havendo a necessidade de menos espaço tanto em memória para o processamento quanto em disco para o armazenamento. A Figura 9 ilustra a imagem submetida a uma transformação de RGB para tons de cinza.

Figura 9 – Aplicação do método de transformação em tons de cinza. Input Image e Gray Image.

IV.3.2.2 Ajuste de Histograma das Imagens Cromossômicas

A segunda transformação a que a imagem é submetida, é o ajuste de contraste. Esta etapa aplica sobre a imagem uma saturação das intensidades de cinza no histograma gerado a partir da imagem analisada, fazendo com que os tons claros tendam a 0 (preto) e os altos níveis tendam ao 255 (branco). Esta operação realiza uma melhor distribuição das intensidades ao longo de todo o eixo x do histograma (intervalo [0-255]) e preservando suas características iniciais. A Figura 10 apresenta o método de ajuste de contraste aplicado na imagem em tons de cinza.

Figura 10 – Aplicação do método de transformação de ajuste de contraste. Input Image e Contrast Enhanced Image.

Este processamento da imagem é realizado pela função do Matlab® ‘*imadjust*’. Esta técnica foi escolhida empiricamente com a finalidade de auxiliar na separabilidade e segmentação dos cromossomos. Existem diversas técnicas que envolvem operações em histogramas e “borramentos” nas imagens, entretanto, o ajuste de contraste se mostrou eficaz a este tipo de problema, uma vez que, após a transformação, preserva as características iniciais do histograma de intensidades. Uma equalização histogramica, por exemplo, faria com que a diferença de intensidade dos *pixels* entre objeto e fundo sendo menos aparente. A Figura 11 demonstra o histograma inicial e o resultado da manipulação do contraste aplicado sobre a imagem cromossômica.

Figura 13-A

Figura 13-B

Figura 11 – A) Histograma de intensidades da imagem em tons de cinza; B) Histograma de intensidades da imagem após a aplicação do ajuste de contraste.

A Figura 11-A apresenta o histograma de intensidades da imagem em tons de cinza, sem qualquer tipo de processamento de realce de contraste, e na Figura 11-B, pode-se notar que as contagens para cada intensidade diminuíram, porém as características da dispersão da função distribuição do histograma (curva de intensidades) foram mantidas.

IV.3.2.2 Eliminação do Background das Imagens Cromossômicas

A imagem em tons de cinza com contraste adequado possibilita a etapa de escolha do limiar (*threshold*) de segmentação dos cromossomos do *background*. Etapa crucial para a análise cromossômica automatizada, pois é a partir deste limiar que os cromossomos são segmentados para posteriormente serem analisados.

A escolha do limiar é uma difícil tarefa devido ao processo de segmentação o qual os cromossomos devem ser segmentados da imagem de entrada como subconjuntos de *pixels* dentro de um intervalo de intensidades, neste caso em tons de cinza. O estabelecimento de um critério para a qualidade mínima do contraste nas imagens originais é uma importante decisão, pois em alguns casos as intensidades dos pixels do cromossomo se assemelham aos valores encontrados no *background*.

A seleção destas imagens se deu pela avaliação dos limites de qualidade de aquisição contra a obtenção dos melhores resultados de segmentação e consequente de classificação e caracterização das imagens cromossômicas pelo *aplicativo*. Um conjunto de 11 metáfases, representando um total de 1.465 imagens cromossômicas geradas, foram escolhidas e consideradas suficientes para avaliar o aplicativo, considerando a soma das 3 intensidades avaliadas (*threshold* com valor de 60 (23,52%), 70 (27,45%) e 80 (31,37%)) neste trabalho (ANEXO-I).

O valor do limiar é variável por estas imagens não terem sido adquiridas de modo padronizado, havendo diferenças de nitidez e foco decorrentes das diferenças no modo de aquisição. Assim, a decisão da escolha do limiar foi desenvolvida como sendo dependente da interação e decisão do especialista e, do resultado final da análise (segmentação – identificação). Este aspecto é o único no “software” criado CHRIMAN em que há a interferência do usuário para ajuste deste valor. Entretanto, caso necessitem realizar uma sequência de análise em imagens semelhantes e com aquisição padronizada, basta escolher previamente o limiar que melhor se aplica a elas.

A faixa de intensidade de corte que resultou em melhores análises corresponde a 25% a 30% da intensidade máxima das imagens em tons de cinza - 255. A eliminação do fundo otimizada consiste em eliminar *pixels* com valores de intensidades próximas de 70 (27,45%).

Assim, todos os *pixels* pertencentes ao intervalo [limiar - 255] são eliminados da imagem, ou seja, preenchidos com o valor de 255 (branco). A Figura 12 apresenta a imagem de uma metáfase com o *background* removido. O *threshold* está configurado para a intensidade de 70 (27,45%).

Figura 12 – Aplicação do método de transformação de ajuste de contraste. Input Image e No background Image. Remoção do background pelo limiar 70.

Um modo teórico obtido na literatura para se calcular o valor do limiar é por meio do método *Otsu* [1979] de separabilidade entre histogramas, no qual, a partir da análise histogramica na imagem de estudo, este método computacional calcula a máxima separabilidade entre os histogramas de intensidades (claro e escuro). Porém, este método se revelou ineficaz para a maioria das imagens deste estudo, pois não elimina os *pixels* com em torno dos cromossomos, *pixels* estes com altos valores de intensidades, preservando informações desnecessárias prejudicando a efetiva segmentação e posteriormente de caracterização da imagem do cromossomo.

A Figura 13 apresenta os quatro valores utilizados no aplicativo desenvolvido neste estudo. O software permite que o usuário verifique o melhor valor de limiar e o aplique sobre a imagem estudada com o intuito da remoção do *background*.

Figura 13 - Aplicação do método de eliminação de background pelo valor do limiar. Da esquerda para a direita, são aplicados os métodos Otsu, e os valores de corte (limiar) por intensidade de 80 (31,37%), 70 (27,45%) e 60 (23,52%) respectivamente. A seta aponta para a imagem do mesmo conjunto de cromossomos exemplificando uma correta segmentação.

Na Figura 13 pode-se observar pela indicação da seta nas quatro imagens, a remoção dos pixels no espaço entre os cromossomos, cuja informação é diretamente relacionada a análise do cromossomo segmentado. A tarefa de encontrar o valor deste limiar “ótimo”, significa a melhor segmentação do objeto e a preservação das características iniciais, essenciais a sua classificação final.

Após a eliminação do fundo da imagem inicia-se o processo de contagem de objetos na cena. O algoritmo utilizado realizar esta tarefa é o *labeling* (rotulação), função nativa do Matlab® - “*bwlabeln*”. Este algoritmo, basicamente, testa as conectividades entre o pixel analisado pelo *raster* (conjunto de regras “for” que tem como função varrer a imagem analisada *pixel a pixel*) e os *pixels* adjacentes a ele. já apresentado e explicado na seção (III-2. Segmentação de Imagens).

IV.3.2.4 Segmentação das Imagens Cromossômicas

A segmentação é realizada pelo isolamento dos subconjuntos de *pixels* de mesmo rótulo. A correta segmentação dos cromossomos das imagens metafásicas está diretamente ligada às etapas iniciais do aplicativo, que consistem em processar a imagem, de forma que o a relação de contraste entre o fundo (*background*) e os objetos de interesse (cromossomos) permitam que sejam eliminados os *pixels* indesejáveis na imagem.

Dentre os problemas encontrados durante a etapa de segmentação, pode-se citar o acúmulo de coloração nas células das imagens aquiritadas, decorrentes do processo de gotejamento das células na lâmina (acumulo de Giemsa permanece em determinadas áreas das células), como também os problemas resultantes da metodologia utilizada na aquisição da imagem, como foco, nitidez, etc. A Figura 14 apresenta uma imagem com difícil interpretação, na parte inferior da imagem, devido ao acúmulo de Giemsa entre as cromátides irmãs.

Figura 14 – Imagem cromossômica com acúmulo de Giemsa.

Outra situação em que a técnica de segmentação proposta apresenta certas limitações é no caso em que os cromossomos sobrepostos ou muito próximos (conectados) e, são selecionados na mesma imagem segmentada (atribuídos com o mesmo rótulo), interligados por conexões de *pixels*, gerando erros, não apenas na contagem dos cromossomos, como também na identificação e classificação do mesmo.

As chances de acerto do *software* estão diretamente ligadas ao sucesso da segmentação de cada cromossomo. A Figura 15 apresenta uma imagem em que não foi possível a segmentação dos cromossomos utilizando esta técnica.

Figura 15 – Imagem onde os cromossomos estão sobrepostos e não foram segmentados corretamente, devido ao fato de que o mesmo rótulo foi atribuído a todo o subconjunto de pixels desta imagem.

As imagens cromossômicas de uma imagem metafásica são segmentadas corretamente na proporção de 97,1% de imagens cromossômicas quando utilizado o valor de limiar igual a 70 (27,45%) (ANEXO-I). Este valor se aproxima da proporção da análise feita por profissionais. A técnica reproduz essa dificuldade.

Com os cromossomos segmentados, foi desenvolvido um algoritmo que gere uma nova imagem, com as dimensões da imagem cromossômicas, ou seja, as dimensões de um subconjunto de *pixels* identificados com o mesmo rótulo, onde para cada subconjunto é realizada uma contagem do número de *pixels*.

A rotulação é utilizada para ordenar as imagens cromossômicas, de acordo com a contagem dos *pixels* preenchidos de cada rótulo atribuído. Este valor se apresentou mais eficaz em relação a ordenação do que os valores da dimensão total da matriz ($x * y$), devido ao fato cromossomos longos e/ou curvados na sua disposição aumentam as dimensões da imagem cromossômica segmentada gerada, identificados também na análise visual.

Durante esta etapa acontece o único corte de “objetos”, que elimina subconjuntos de pixels de mesmo rótulo com contagens muito baixas. Estes subconjuntos são caracterizados como *pixels* isolados: objetos identificados como uma matriz abaixo de 20 (vinte) contagens. Etapa esta ainda em estudo pela equipe do projeto.

IV.3.2.5 Registro das Imagens Cromossômicas

A primeira alteração morfológica aplicada durante a rotina de classificação do aplicativo consiste em posicionar o cromossomo, no mesmo plano, de forma que o seu comprimento alinhe-se com a coordenada y (eixo vertical do plano cartesiano), e, conseqüentemente a largura das cromátides se alinhem com a coordenada de x . Sendo

assim, as imagens são rotacionadas de modo que facilite a sua análise. Uma exceção nesta regra acontece quando a própria análise da segmentação indica (a partir das referências obtidas) que a imagem cromossômica está na posição “adequada” para a o início das análises.

O total registro aplicado sobre as imagens cromossômicas segmentadas segue a 2 passos, detalhados nos parágrafos seguintes.

O primeiro passo compara as diferenças entre as dimensões x e y da matriz:

Se $x > y$; Então: $A(x, y) = A(x, y)$; A matriz permanece igual.

Senão, $x \leq y$; Então: $A(x, y) = A(x, y)'$, ou $A(x, y) = A(y, x)$; Obtêm-se a transposta da matriz da imagem.

A Figura 16 mostra a primeira transformação morfológica aplicada sobre as imagens cromossômicas.

Figura 16 – Aplicação da transposta na imagem cromossômica segmentada. Primeira transformação morfológica em que as imagens cromossômicas são submetidas.

Com a imagem cromossômica reposicionada inicia-se o segundo passo, que tem como objetivo, a identificação dos pontos de referência, essenciais para se efetuar o reposicionamento; a pela busca do eixo central do cromossomo disposto na imagem; e ainda, identificar em quantos graus e para qual sentido, a imagem deverá ser rotacionada.

Para isso, dois perfis são traçados: o primeiro na linha $y*0.25$ e na linha $y*0.75$, (25%) e (75%) da extensão total da coordenada y do plano cartesiano, respectivamente. Os perfis são traçados e comparados quanto ao alinhamento dos eixos citados, onde 3 pontos de referência são identificados durante o *raster* (Figura 17), em ambos os perfis traçados, o primeiro e o último *pixel* preenchido ao longo da linha selecionada no cromossomo, e o *pixel* médio entre os dois pontos.

Se $f(x, y) \neq 255$, $P_{\text{inicial}} = x$; break;

Raster em todo o perfil traçado, se $f(x, y) \neq 255$, $P_{\text{final}} = x$;

$P_{\text{médio}} = (P_{\text{final}} - P_{\text{inicial}}) / 2$;

A Figura 17 apresenta os 3 pixels identificados como pontos de controle, utilizados para a rotação do cromossomo no alinhamento ao eixo y.

Figura 17 – Pixels de referência e identificação do eixo central do cromossomo utilizado para o registro da imagem cromossômica.

Caso o *pixel* central (*pixel* de localização médio entre o primeiro e o último *pixel* preenchido nos perfis selecionados) do primeiro perfil esteja com o valor menor ao *pixel* central no segundo perfil, a imagem deverá ser rotacionada em sentido horário, caso contrário, sentido anti-horário. O cálculo de quantos graus a imagem deverá ser rotacionada, é obtido pelo valor de α : $\alpha = \arctangente(P_{\text{médio}2} / P_{\text{médio}1})$, como mostra a Figura 18-a e 18-b.

Figura 18 – Aplicação da operação de rotação utilizando o eixo central do cromossomo disposto na imagem.

O efetivo posicionamento das imagens cromossômicas é realizado pela função do MATLAB® “imrotate” utilizando o método de interpolação *Nearest neighbor*. Esta técnica encontra limitações com alguns cromossomos que estão dispostos de forma curvada, imagens estas em que o cromossomo está “dobrado” em razão do processo de gotejamento das células na lâmina. A Figura 19 apresenta uma imagem onde a aplicação do algoritmo proposto para a rotação dificulta o resultado do registro, implicando, em possíveis erros de classificação ou mesmo, em alguns casos, inviabilizando a caracterização do mesmo.

Figura 19 – Imagem onde o cromossomo está curvado de forma que dificulte o resultado da classificação.

A partir dos cromossomos segmentados, os rótulos são atribuídos e aplica-se o módulo de ordenação das imagens cromossômicas, utilizando-se das contagens de *pixels* preenchidos, para subconjuntos com um mesmo rótulo. Imagens com a segmentação incorreta ou sobrepostas, consideradas “inseparáveis” para o aplicativo, em geral, tornam-se subconjuntos com contagens maiores em tamanho e preenchimento, apresentando um forte indicador de que estas imagens devem ser melhor analisadas, tanto pelo *software* quanto pelo especialista citogeneticista. A Figura 20 apresenta imagens cromossômicas apropriadas, possíveis de serem segmentadas individualmente, e que foram registradas corretamente.

Figura 20 – Imagem onde os cromossomos foram segmentados e registrados corretamente.

IV.3.2.6 Caracterização das Imagens Cromossômicas

Nesta seção serão apresentadas as técnicas de reconhecimento de padrões em imagens cromossômicas utilizadas para a caracterização e classificação dos cromossomos pela quantidade de centrômeros encontrados.

IV.3.2.6.1 Análise dos perfis de uma imagem cromossômica

Com o intuito identificar objetos em uma imagem, uma técnica utilizada em reconhecimento de padrões, consiste no estudo de perfis verticais e/ou horizontais traçados na imagem estudada. Esta técnica de reconhecimento foi aplicada nas imagens cromossômicas para a obtenção de um indicador referente à disposição do cromossomo na imagem gerada, como também, auxiliar na avaliação do processo de segmentação e registro destas imagens. A Figura 21 apresenta a ideia de se traçar perfis nas imagens cromossômicas com a finalidade principal de reconhecer a disposição e a morfologia do cromossomo segmentado da imagem metafásica.

Figura 21 – Esquema representativo perfis traçados nas imagens cromossômicas com a finalidade do reconhecimento da disposição dos cromossomos.

Os padrões de perfil são baseados em duas características obtidas nos “vetores” traçados nas imagens cromossômicas: a contagem, ou comprimento, de cada banda e a quantidade de bandas presentes em um perfil.

As contagens são obtidas por valores de *pixels* diferentes a 255 no perfil analisado, já que o fundo foi eliminado durante o processo de segmentação do cromossomo, ressaltando que, estes valores pertencem ao intervalo $[0(\text{zero}) - \text{threshold}]$. A quantidade de bandas em um perfil é obtida pela continuidade de cada uma das bandas encontradas (número de contagens de bandas válidas), e, um intervalo na contagem acontece quando a busca encontra no perfil valores igual a 255.

Para a obtenção destas características nos perfis (contagem de banda e quantidade de banda), alguns ajustes são necessários para a correta interpretação das informações. Um exemplo ocorre quando um *pixel* está isolado e próximo a uma banda válida, obviamente 1 *pixel* isolado não pode ser considerado como sendo uma banda válida. E sendo assim, caso haja a interrupção entre bandas com valores mínimos de contagens (de 1 até 3 *pixels*) e uma banda válida (contagens maiores ou iguais a 3), e este intervalo seja menor ou igual a 3 contagens de *pixels*, este(s) é(são) adicionado(s) à banda válida mais próxima a ele(s). A Figura 22 apresenta uma perfil com um *pixel* isolado que será incorporado à contagem da banda 2.

Figura 22 – Banda com um *pixel* isolado. Aplica-se a regra de adição de *pixels* isolados à banda válida mais próxima.

A partir destes estudos nos perfis verticais foram identificados 6 (seis) diferentes grupos de perfis. O primeiro grupo de perfil se caracteriza por uma banda única e suas contagens apresentam valores acima de 70% da largura total da imagem (Figura 23-A); o segundo grupo é representado por uma banda com valores médios em relação à coordenada x total (Figura 23-B); o terceiro grupo é representado por uma banda única e com baixas contagens de preenchimento do perfil (Figura 23-C); de modo geral, estes dois últimos grupos citados representam a(s) constrição(ões) na morfologia dos cromossomos, região onde provavelmente está localizada a região de um centrômero; o quarto e quinto grupos possuem duas bandas e cada uma delas tem um reconhecimento, de certa forma individual, pois o resultado final de um perfil como estes necessariamente tem duas respostas, sendo uma para cada perfil (Figura 23-D e 23-E); e por fim, o sexto grupo representa pontos isolados na imagem ou quantidades de banda superior a 2 (dois), ou seja, perfis não válidos que demonstram que o cromossomo, ou não foi segmentado e/ou não foi registrado corretamente da imagem metafásica (Figura 23-F).

Figura 23-E

Figura 23-F

Figura 23 – A) Perfil com banda longa e única; B) Perfil com banda média e única; C) Perfil com banda curta e única; D) Perfil com duas bandas de contagens médias; E) Perfil com duas bandas de contagens distintas; F) Perfil com o número de quantidade de bandas superior a 2 (dois) e perfil com a presença de banda não válidas.

O resultado do perfil na Figura 23-F apresenta um perfil com mais de uma banda e um perfil com duas bandas, onde cada valor de contagem é igual a 1 e 4, respectivamente, sendo que na segunda banda se aplicou a regra de adição de *pixel* à banda mais próxima. Neste caso, este perfil é relacionado ao grupo de perfis que indicam possíveis anormalidades na imagem gerada, pois, apesar de possuir 2 (duas) bandas, a primeira contagem caracteriza-se como uma banda não válida presente no perfil analisado (contagens muito baixas e isoladas).

Sendo assim, o algoritmo elaborado para identificar os perfis dedica-se a avaliar as contagens das bandas válidas e a quantidade das mesmas presentes no perfil, onde, de forma resumida, os cromossomos possuem 4 (quatro) grupos de bandas classificados como “corretos” e 2 (dois) grupos referentes a alguma anormalidade na imagem.

Este procedimento é aplicado por nove vezes, mapeando a imagem analisada, sendo que, a localização de cada perfil traçado equivale à “altura” da imagem variando a cada 10% da imagem, ou seja, um perfil para cada valor de comprimento como: $\text{altura} \times 0.10$, $\text{altura} \times 0.20$, até $\text{altura} \times 0.90$.

Cada perfil é analisado pelo algoritmo de reconhecimento, e as respostas (saídas) deste, servem de valor de ativação de uma variável de entrada do sistema de lógica *Fuzzy* proposto, melhor detalhado na seção (IV.3.2.6.2 Sistema de Lógica Fuzzy para

Caracterização dos Perfis). A Figura 24 representa o mapeamento ao longo do eixo y, realizado durante a etapa de obtenção das características da imagem cromossômica.

Figura 24 – Demonstração dos nove perfis traçados ao longo do eixo y, cada perfil é traçado a cada 10% da altura da imagem cromossômica gerada.

IV.3.2.6.2 Sistema de Lógica *Fuzzy* para Caracterização dos Perfis

Um Sistema de Lógica *Fuzzy* foi implementado para suprir a necessidade de se interpretar as características deste sistema de caracterização de imagens cromossômicas, e visando combiná-las obtendo assim um resultado escalar possível de ser interpretado. Este sistema *Fuzzy* foi criado para auxiliar na avaliação os algoritmos propostos para a segmentação e registro, e identificar as variações de contagens presentes nos perfis analisados, com o intuito principal de classificar principalmente a variável referente ao comprimento das bandas. Este sistema desenvolvido foi chamado de “perfilClassif”.

As respostas obtidas por este *FLS* de classificação dos perfis auxiliam na identificação da disposição dos cromossomos nas imagens (localização das cromátides e identificação do centrômero). Esta análise está melhor detalhada na seção (IV-2.5. Localização e Identificação de Centrômero).

As características para o *FLS* “perfilClassif” (contagem e quantidade de bandas) são obtidas pelo algoritmo “analísPerfil” e posteriormente utilizadas para ativar as funções de relacionamento das variáveis de entrada: “quantidadeBandas” e “tamanhoBanda”, respectivamente. As Figuras 25-A e 25-B apresentam a configuração das variáveis de entrada do sistema *Fuzzy* “perfilClassif”.

Figura 25-A

Figura 25-B

Figura 25 – FLS perfilClassif: A) Variável de entrada *quantidadeBandas*, funções de relacionamento: *um*, *dois*, *três*, *quatro* e *cinco*. B) Variável de entrada *contagemBanda*, funções de relacionamento: *pouco*, *meioPouco*, *medio*, *medioMuito* e *muito*.

A variável de entrada deste Sistema de Lógica *Fuzzy* correspondente a quantidade de bandas em cada perfil analisado possui 5 funções de relacionamento representando os possíveis valores obtidos pela função “*analisePerfil*”. Caso o perfil contenha um número de bandas igual ou superior a 5, ativa a função de relacionamento “cinco” na variável de entrada “*quantidadeBandas*”. Um importante detalhe referente a esta variável de entrada refere-se ao fato de que somente valores inteiros são aceitos (*singletons*), representando na prática, não ser possível a identificação da quantidade de $\frac{1}{2}$ (meia) banda. Este detalhe apresenta a robustez das possibilidades de representação dos problemas através da Lógica *Fuzzy*.

A variável de entrada referente às contagens (comprimento) de cada banda em um perfil, já apresentadas as definições, apresentam também 5 possibilidades de funções de relacionamento distintas dentro do intervalo de valores [0-1], “pouco”, “meioPouco”, “medio”, “medioMuito” e “muito”. Estas funções de relacionamento foram definidas à partir da análise visual (método empírico) e por meio de uma análise que comparou os dados referentes às larguras das bandas encontradas em cada perfil traçado nas imagens cromossômicas e as agrupou de acordo com o número de ocorrências dentro de um

intervalo de pré-definido de tamanhos, e determinou o método no qual, estas funções de pertinência foram ajustadas. Através deste *FLS* é possível realizar a classificação dos perfis com resultados um pouco mais refinados.

Um conjunto de inferências foi criado, como um grande sistema *IF-THEN*, onde as características dos perfis são fuzzificadas e interpretadas por um conjunto de regras (ANEXO-III), gerando uma resposta escalar para avaliação do perfil, e indica uma resposta que será utilizada como uma característica da imagem cromossômica, utilizadas durante processos seguintes de classificação.

A saída (resposta) deste *FLS* apresenta um valor que indica se o perfil está coerente com as regras de padronização na disposição após os processos de segmentação e registro: perfis que contenham até duas contagens de bandas e, se são bandas proporcionais; e banda única, avaliando o seu comprimento, e indicando a possível localização do(s) centrômero(s). Procedimento melhor detalhado na seção (IV-2.5. Localização e Identificação de Centrômero(s) na Imagem Cromossômica).

Os padrões destes perfis são classificados em 6 (seis) grupos, e a análise realizada consiste em avaliar por meio das variáveis de resposta do *FLS* “perfilClassif”, as variações no preenchimento de cada banda válida identificada em um perfil.

A Figura 26 apresenta as funções de relacionamento das variáveis de saída do *FLS* “perfilClassif” para a identificação das bandas, apresentando as 6 (seis) possíveis variáveis de saída (grupos): “longoÚnico”; “médioÚnico”; “curtoÚnico”; “curtoDuplo”; “médioMuitoDuplo”; e “naoValido”. O método de inferências aplicado neste *SLF* é o Método de Inferência de Mamdani e o método de defuzzificação é o de centro de massa.

Figura 26 – *FLS* perfilClassif: Funções de relacionamento de saída grupoPerfil: longoUnico [0-1], medioUnico [1-2], curtoUnico [2-3], curtoDuplo]3-4], medioLongoDuplo [4-5] e naoValido]5-6].

Para as respostas com quantidades de banda superiores a 2 há a indicação de que a imagem está incoerente, e neste caso, este perfil é classificado como um forte candidato a pertencer ao grupo 6, acionando a variável de saída “naoValido” do FLS “perfilClassif”, e ativando um “flag” durante a rotina de identificação de imagens cromossômicas. Este “flag” indica uma anormalidade no perfil, e representa um indicador de que a imagem não foi corretamente segmentada e/ou que provavelmente existam mais do que 1 cromossomo nesta imagem gerada.

Contando com caracterização dos perfis, proposta pelo LFS “perfilClassif” foi idealizado um mapeamento das imagens cromossômicas traçando, a cada intervalo de $(0.1 * \text{altura})$ para a obtenção de uma característica de caráter crucial para a classificação de aberrações cromossômicas: “situacaoPerfil”. Melhor detalhado na seção (IV.3.2.6.3 Sistema de Lógica Fuzzy para Caracterização das Imagens Cromossômicas).

A Figura 27 apresenta a aplicação do algoritmo de identificação de perfis proposto, demonstrando o mapeamento de 9 perfis na imagem e a caracterização destes pelo FLS “perfilClassif” em exemplos obtidos após a classificação dos perfis traçados nas imagens cromossômicas.

Figura 27-A

Figura 27-B

Figura 27 – Mapeamento de 9 (nove) perfis na imagem e a caracterização destes pelo FLS perfilClassif A) Mapeamento e caracterização dos perfis de uma imagem cromossômica normal. B) Mapeamento e caracterização dos perfis de uma imagem cromossômica com a aberração cromossômica: dicêntrico.

Além de identificar o conteúdo da imagem cromossômica gerada, este sistema auxilia a avaliação do processo de segmentação e registro. Assim, a etapa de identificação das imagens cromossômicas conta uma característica inerente ao objeto de estudo (cromossomo), provendo uma das variáveis CRISP de entrada para o segundo Sistema de Lógica *Fuzzy* referente à caracterização da padronização na disposição dos cromossomos.

Outra funcionalidade obtida por meio da análise dos perfis é referente à segunda etapa da interpretação de características presentes na imagem cromossômica, no que se refere à contagem e localização dos centrômeros.

De acordo com os especialistas citogeneticistas, o centrômero é reconhecido visualmente como sendo uma constrição na morfologia do corpo do cromossomo. Este critério é utilizado para cromossomos metacêntricos, submetacêntricos, e acrocêntricos, pois a constrição está disposta no corpo do cromossomo analisado, e obviamente, para os cromossomos telocêntricos (centrômero localizado em uma das extremidades das

cromátides), uma segunda técnica de reconhecimento da localização dos centrômeros foi proposta a fim de localizar o menor valor de perfil e com quantidade de banda igual a 1.

Esta informação é obtida através da análise dos perfis superiores e inferiores (busca vertical) no mapeamento, buscando perfis representativos destas constrições na morfologia cromossômica, identificadas como bandas curtas e únicas, tornando a busca mais eficiente devido a menor quantidade de informação processada.

A quantidade e a localização dos centrômeros em uma imagem representa a resolução da busca pela identificação dos cromossomos normais e algumas das aberrações cromossômicas do tipo cromossômico (acontecem antes da duplicação do material genético), representando no caso de duas contagens como dicêntrico, e no caso de três contagens como tricêntricos, principais aberrações cromossômicas investigadas pelo aplicativo desenvolvido por este projeto.

IV.3.2.6.3 Sistema de Lógica Fuzzy para Caracterização das Imagens Cromossômicas

O primeiro passo para a caracterização das imagens cromossômicas “ChromoClassif” é identificar as características que serão utilizadas como as variáveis de entrada do sistema de lógica *Fuzzy* para a caracterização das imagens. O segundo sistema *Fuzzy* implementado, trata da identificação da disposição dos cromossomos nas imagens obtidas após as etapas de transformação, ajustes de contraste, limiarização, rotulação, ordenação, segmentação, registro, e, a análise dos perfis, utilizando as características obtidas individualmente e também considerando todo o conjunto de imagens geradas a partir da imagem metafásica.

As características analisadas são: o tamanho total da imagem, considerando as dimensões da matriz total da imagem, variável de entrada: “tamanhoMatriz”; o preenchimento da imagem, número de *pixels* com intensidades diferentes de 255 (branco), variável de entrada: “preenchimentoImagem”; e o resultado da análise de perfis, variável de entrada: “situacaoPerfil”. Esta ultima, utiliza um contador que é incrementado no caso em que não seja indicada anormalidades nos perfis, e no caso de o valor deste contador seja diferente de 9, há um forte indicador de alguma anormalidade na morfologia do cromossomo na imagem gerada.

As Figuras 28 (A, B e C) apresentam as funções de relacionamento das 3 (três) variáveis de entrada para a avaliação da segmentação e identificação da disposição do cromossomo na imagem gerada, o FLS “ChromoClassif”.

Figura 28-A

Figura 28-B

Figura 28-C

Figura 28 – A) Funções de relacionamento das variáveis de entrada do FLS: *tamanhoMatriz*: *pequena*, *medioPequena*, *meioMediaPequena*, *meioMediaGrande*, *mediaGrande* e *grande*; B) Funções de relacionamento das variáveis de entrada do FLS: *preenchimentoImagem*: *pouco*, *medioPouco*, *meioMedioPouco*, *meioMedioMuito*, *medioMuito*, *muito*; C) Funções de relacionamento das variáveis de entrada do FLS: *situacaoPerfil*: *contagemInsuficiente* e *contagemSuficiente*.

A variável de entrada “*tamanhoMatriz*” compara as dimensões da maior matriz obtida pelo método de ordenação das imagens (a matriz de rótulo com valor 1),

representando o valor de referência a ser comparado com as demais imagens. Os valores para as funções de relacionamento das variáveis de entrada “tamanhoMatriz” determinam o quanto essa imagem é menor proporcionalmente ao valor de referência.

A segunda característica utilizada como variável de entrada representa a característica de preenchimento da imagem que está sendo analisada: “preenchimentoImagem”. Esta variável é obtida de forma que, considerando as dimensões da imagem gerada, um contador é incrementado quando o *raster* detecta um *pixel* com intensidade diferente de 255. Sendo assim, obtemos o valor representado no intervalo [0-1], de o quanto a imagem é preenchida pela morfologia do cromossomo. Esta variável é um importante indicador sobre a disposição do cromossomo na imagem gerada, pois preenchimentos baixos indicam uma possível falha na segmentação dos cromossomos da imagem metafásica.

E por fim, como detalhado na seção (IV.3.2.6.1 Análise dos perfis de uma imagem cromossômica), a resultante da análise de cada um dos 9 perfis de referência, resultando no número de perfis válidos encontrados, considerado pela função de entrada do sistema *Fuzzy* “situacaoPerfil”, atribuídos à variável de entrada “contagemSuficiente”.

A combinação destas 3 entradas e o conjunto de regras *Fuzzy* associadas a elas resultam na indicação de coerência com as regras de segmentação e classificação propostas (ANEXO-IV).

A saída resposta do Sistema de Lógica *Fuzzy* apresenta valores bastante simples, indicando uma classificação de “padronizado” ou “nãoPadronizado” de acordo com as características inferidas. O método de inferências aplicado neste *SLF* é o Método de Inferência de Mamdani e o método de defuzzificação é o de centro de massa. A Figura 29 apresenta as funções de relacionamento da variável de saída “sitImagem” para classificação da segmentação e classificação da padronização da morfologia do cromossomo disposto na imagem gerada.

Figura 29 – Função de saída escalar do FLS, chamada *sitImagem*: *naoPadronizado* e *padronizado*.

As imagens classificadas como “padronizado” são então submetidas à identificação da localização dos centrômeros (processo seguinte) e as imagens identificadas como “naoPadronizado” são indicadas pelo aplicativo para que haja uma pós análise visual pelo profissional especialista.

IV.3.2.6.4 Localização e Identificação de Centrômero(s) na Imagem Cromossômica

As imagens cromossômicas válidas pelos critérios já mencionados na seção anterior são investigadas para a identificação dos dados referentes à localização do centrômero. A identificação dos centrômeros é uma etapa essencial para a final classificação da morfologia do cromossomo na imagem cromossômica digital. Esta análise é uma tentativa de reproduzir a forma humana de identificação visual de uma constrição na morfologia do cromossomo, e reforçar a classificação final das imagens geradas.

Duas técnicas de reconhecimento de padrões foram utilizadas para a identificação dos centrômeros na imagem cromossômica: a primeira utiliza-se da lógica *Fuzzy* para a determinação e identificação do centrômero, baseando-se nas técnicas de reconhecimento de padrões nos perfis das imagens; e a segunda técnica é uma adaptação da metodologia usada por Roshtkhari [2008]. Este método utiliza o histograma de contagens de *pixels*, contagens horizontais em cada perfil, e identifica o menor valor de contagem no histograma como sendo a localização do intervalo de perfis referente ao centrômero, para, posteriormente, registrar a imagem cromossômica.

O autor Roshtkhari [2008] faz a caracterização por meio de uma técnica de reconhecimento de histogramas identificando o ângulo de rotação ideal para que, cromossomos muito curvados possam ser morfologicamente alterados, gerando uma nova imagem. A disposição do cromossomo na imagem gerada se difere da disposição do mesmo na imagem metafásica, porém, auxilia na identificação, tanto pelo profissional citogeneticista como pelos algoritmos de reconhecimento de padrões.

Esta segunda técnica foi utilizada somente quando a primeira técnica, baseada em buscas e identificação por Sistemas de Lógicas *Fuzzy* para localização dos centrômeros, não foi robusto o suficiente para encontrar a constrição na morfologia do cromossomo disposto na imagem. Esta técnica se mostrou eficiente principalmente para os casos onde os centrômeros estão localizados na extremidade das cromátides. O número de centrômeros encontrados por esta segunda técnica é de 83 (20,2%) do total de 410 identificados em 396 imagens cromossômicas analisadas com o *threshold* configurado para o valor de 70 (27,45%) (ANEXO II).

A primeira técnica, maior escopo deste trabalho, realiza a caracterização das imagens cromossômicas utilizando Sistemas de Lógica *Fuzzy* para a identificação de perfis com a finalidade da localização dos centrômeros.

O primeiro passo para a busca pelo(s) centrômero(s) é identificar, dentre os 9 perfis de referência traçados na imagem, quais deles possuem os perfis de quantidade de banda igual a 1 e contagens menores em relação aos demais perfis da imagem. Como já mencionado, somente os perfis pertencentes aos grupos “medioUnico” e “curtoUnico” podem ser identificados como pertencentes ao intervalo onde está localizado o(s) centrômero(s). Da mesma forma que, para o caso de um perfil caracterizado com quantidade de banda igual ou superior a 2, este perfil é necessariamente desconsiderado.

O algoritmo desenvolvido para esta tarefa realiza uma busca vertical na imagem cromossômica aproveitando-se dos 9 (nove) perfis horizontais traçados como referência ao mapeamento da imagem, e selecionando os que satisfazem as características com que os classifiquem como possível localização de um centrômero.

Nesta etapa do aplicativo há uma primeira busca por perfis classificados como “curtoUnico”, perfis caracterizados por quantidade de banda igual a 1 e contagens proporcionalmente baixas, ou seja, ativando as funções de relacionamento “pouco” e “medioPouco” no *FLS* “perfilClassif”. Quando este grupo de perfil é identificado em uma

imagem cromossômica, há um forte indício de que este perfil está contido em um intervalo de perfis referente a localização de um centrômero. Esta regra não se aplicou quando houve alguma dificuldade, por algum dos motivos já mencionados, na segmentação e/ou no registro da imagem cromossômica gerada. Maiores esforços na análise deste tipo de perfil poderiam sem escopo de trabalhos futuros a este projeto.

Identificados os perfis de referência pertencentes ao grupo “curtoUnico”, inicia-se o processo de busca pelos perfis pertencentes ao segundo grupo de interesse, o grupo “medioUnico”, perfis caracterizados por quantidade de banda igual a 1 e contagens proporcionalmente medianas, ou seja, ativando as funções de relacionamento “medio” e “medioMuito” no *FLS* “perfilClassif”.

Após a identificação de ambos estes casos de interesse, estas coordenadas são armazenadas em um vetor para a posterior otimização na busca pelo centrômero, utilizando-se deste(s) valores para identificar o(s) perfil(s) de início e o(s) perfil(s) de fim da busca.

Identificada a coordenada do primeiro perfil pertencente a um destes grupos, ele é referenciado (pertencente ao intervalo de um centrômero), e o antecessor a ele é identificado como a coordenada de início da busca. Caso o perfil 1 seja a referência, a busca se inicia na terceira linha da imagem (devido ao fato de as primeiras linhas normalmente não possuírem informações confiáveis, oriundos da remoção do *background*).

O segundo passo trata da identificação da coordenada que representa a linha de fim de busca. Para isto, o vetor seguinte ao perfil referenciado é analisado, e caso seja um pertencente a um dos grupos inerentes aos grupos de centrômeros, indica que a morfologia do centrômero se entende por mais um dos perfis traçados, e, assim, um incrementador no valor referente ao vetor do fim da busca é ativado. Este processo se repete até o caso de a coordenada analisada apresentar características diferentes às dos grupos de interesse. Este perfil é então, identificado como a coordenada de fim da busca.

Caso o perfil identificado como a coordenada de fim da busca 1 seja diferente do último traçado na imagem, perfil 9, ou igual a última coordenada do eixo y da imagem, casos em que os perfis pertencentes a estes grupos se estendem além do perfil 9 (nove), a busca por perfis pertencente aos grupos “medioUnico” e “curtoUnico” se repete identificando o caso de 2 regiões de localização do centrômero. E assim, sucessivamente.

A Figura 30 apresenta os perfis selecionados como coordenada inicial e final da busca por centrômeros nas imagens cromossômicas geradas por este trabalho.

Figura 30 – Utilização do mapeamento de 9 (nove) perfis para identificar as coordenadas de início e fim da busca pelos centrômeros.

Após identificados os perfis de início e fim da busca por centrômero(s), se inicia o processo de análise para a caracterização e consequentemente sua localização.

A primeira checagem realizada verifica se a linha selecionada como início da busca, pertence aos grupos caracterizados como perfis de centrômeros, e analisa se os próximos a ele pertencem ao mesmo grupo. Este procedimento é realizado para identificar as coordenadas das possíveis localizações dos centrômeros, onde, para representar um centrômero, a primeira técnica de identificação de centrômero busca uma faixa de perfis únicos e curtos e/ou médios.

A Figura 31 apresenta nas imagens as coordenadas obtidas pelo primeiro método de identificação de centrômero, identificadas como coordenadas inicial e final dos centrômeros nas imagens geradas.

Figura 31 – Linhas identificadas como início e fim do(s) centrômero(s) nas imagens cromossômicas geradas.

Durante a rotina de localização dos centrômeros, foram desenvolvidas algumas verificações para evitar “falsos positivos” na classificação. Algumas formas de refinamento na busca foram identificadas e implementadas no aplicativo .

A primeira técnica de refinamento utilizada analisa os perfis iniciais das imagens cromossômicas, onde criou-se um critério para que estes perfis, não fossem identificados como pertencentes ao grupo dos perfis de centrômero. A análise estabeleceu que para ser um perfil pertencente a uma faixa de centrômero, as contagens dos perfis, no caso de únicos, devem começar a diminuir, caracterizando uma contração na morfologia do cromossomo na imagem.

A segunda verificação consiste em analisar os perfis identificados como possível perfil pertencente a um centrômero, se estes possuem características que representam os perfis de fim das cromátides (perfis únicos próximos a perfis com quantidades de banda igual a 2). Apesar do fato de possuírem apenas 1 banda, as intensidades das coordenadas centrais possuem intensidades maiores comparadas às intensidades periféricas. A Figura 32 apresenta perfis caracterizados como perfis “unicoMedio” com diferentes características.

Figura 32 – Perfis caracterizados como perfis do mesmo grupo, porém com diferentes características.

Visualmente, é possível diferenciar as características entre os perfis 1 e 2 na Figura 32, entretanto, estes perfis são caracterizados como pertencentes ao grupo dos perfis que pertencem à possível localização dos centrômeros, devido a algumas de suas características morfológicas. Uma forma de testar a veracidade desta informação é, por meio de uma comparação entre as intensidades centrais do perfil, com os valores das intensidades dos pixels “periféricos”.

Esta verificação é realizada da seguinte forma: o perfil selecionado como pertencente ao grupo de perfis que podem representar um centrômero. Deste perfil, são eliminados os *pixels* com valor igual a 255, como este perfil possui apenas uma banda, o vetor resultante desta eliminação, é um vetor com as intensidades válidas consecutivas no mesmo. Deste vetor, o ponto médio ($x/2$), o ponto de 25% da coordenada de x ($x/4$), e o ponto de 75% da coordenada x ($3x/4$) são as referências, e os três pixels anteriores e sucessores a eles são utilizados para calcular a intensidade média destas localizações no perfil. Caso o valor da média dos pontos centrais seja maior à média referente aos pontos de $1/4$ e $3/4$ do perfil, este, é necessariamente desclassificado como pertencente a um centrômero.

Após a eliminação dos perfis não válidos, obtém-se a localização e a extensão de um centrômero na morfologia cromossômica. Caso o número de centrômeros identificados por esta técnica seja igual a 0 (zero), a imagem cromossômica é submetida à segunda técnica para a identificação da localização do centrômero utilizando histogramas de contagens.

Para a elaboração do histograma de contagens de pixels preenchidos um incrementador é aplicado sobre cada perfil da imagem cromossômica, para o caso da intensidade do pixel ser diferente de 255. Com isso, temos um vetor com a contagem destes *pixels* em vetores indicativos de cada coordenada da imagem. A Figura 33 apresenta o histograma de contagens de pixels gerado para o reconhecimento da localização do centrômero pela segunda técnica proposta no trabalho.

Figura 33 – Histograma de contagens de pixels em cada perfil da imagem analisada.

O primeiro métodos de refinamento, envolvendo a eliminação dos perfis iniciais das imagens, foi tido como uma etapa obrigatória desta segunda técnica de localização de centrômeros, sendo que o perfil de início da busca pelo centrômero é identificado como sendo o primeiro perfil com contagem inferior ao perfil anterior a ele. De forma similar o perfil de fim da busca é identificado considerando o intervalo entre eles, como o corpo do cromossomo que deve ser analisado.

Com a eliminação destes, o menor valor de contagens é localizado e analisado pelo Sistema *Fuzzy* “*analisPerfil*”. No caso deste perfil possuir apenas uma banda de contagens, este é identificado como a possível região do centrômero na imagem cromossômica. A Figura 34 demonstra imagens em que a localização do centrômero foi identificada com a segunda metodologia de localização de centrômeros proposta.

Figura 34 – Perfil identificado como a região onde se localiza o centrômero na imagem cromossômica gerada, utilizando a segunda técnica de reconhecimento de padrões.

Caso o menor valor encontrado seja classificado como um perfil de duas bandas, ou mais, o segundo valor do vetor ordenado é selecionado e classificado, este procedimento se repete até o menor valor pertencente a um dos grupos de classificação, com apenas uma banda, seja identificado. Uma justificativa para considerar todos os perfis com contagem única é o fato de que através da primeira técnica, somente perfis pertencentes aos grupos “medioUnico” e “curtoUnico” poderiam ser classificados como centrômero.

A limitação dessa metodologia se deve ao fato de que consegue identificar apenas uma constrição na morfologia do cromossomo, justificando a sua utilização apenas em casos bastante complexos. Outra utilização seria em imagens onde o centrômero está localizado nas extremidades das cromátides, desclassificando-os da primeira técnica proposta, que busca avaliar os perfis e identificá-los como a extensão da morfologia do centrômero.

Uma decisão tomada para a melhor avaliação dos resultados foi a de não analisar com as técnicas de reconhecimento propostas, as imagens cromossômicas geradas com rótulo superior a 36, a análise realizada nestas imagens é apenas a tentativa de identificação pelos fragmentos cromossômicos (aberração cromossômica).

Os fragmentos são identificados como as imagens cromossômicas com o preenchimento menor do que 80% da média dos preenchimentos da imagem.

Se, $\text{PreenchimentoImagemAtual} / \text{MédiaPreenchimento} > -80\%$ $\text{Image}_{\text{frag}}$;

Senão, $\text{Image}_{\text{normal}}$; e $\text{quantidadeCentromero} = 1$.

A decisão de não analisar as imagens com rótulo maior que 36 está baseada principalmente ao erro associado à dificuldade de registrar estas imagens a uma posição considerada boa, para a então análise. Como estes cromossomos possuem uma morfologia de dimensões reduzidas, por muitas vezes, algumas das características essenciais para a classificação, ou foram removidas durante a eliminação do *background* (remoção dos falsos negativos – pixels pertencentes ao objeto de interesse na imagem e foram eliminados erroneamente), ou foram incorretamente registradas. Entretanto, dificilmente essas imagens representam alguma aberração cromossômica como cromossomos dicêntricos ou tricêntricos; e sendo assim, considerar que estas imagens possuem apenas um centrômero, ou nenhum no caso de fragmento, foi considerado bastante razoável.

Desta forma, o total de 396 imagens cromossômicas foram analisadas para avaliar os algoritmos de reconhecimento de padrões propostos para a identificação da localização dos centrômeros. Cada uma dessas imagens foi submetida a todas as etapas de reconhecimento do aplicativo: transformação de intensidades, segmentação, registro e, como mencionado, algoritmos de identificação utilizados para a classificação final da imagem cromossômica.

V RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados obtidos pelo aplicativo criado no desenvolvimento deste trabalho. Apresenta também comentários e discussões referente a estes resultados, bem como, as dificuldades encontradas durante as etapas de caracterização das imagens cromossômicas.

As imagens cromossômicas foram submetidas a diversas análises por técnicas de processamento e reconhecimento de padrões, e obtenção das características morfológicas quantificadas para a identificação e a classificação da imagem estudada.

A primeira tarefa dedica-se a segmentar, quando possível, individualmente os cromossomos de uma imagem metafásica. Para realizar esta tarefa, as primeiras transformações impostas às imagens são: a conversão das imagens de entrada, em novas imagens com tons de cinza, o ajuste do histograma de intensidades, utilizado com o intuito de realçar os objetos de interesse (cromossomos), e efetivar a segmentação dos cromossomos, por meio de um valor de limiar de intensidades para este histograma de tons de cinza previamente ajustado.

Para a segmentação, a escolha do limiar (*threshold*) para este conjunto de 11 imagens metafásicas selecionado, como mencionado anteriormente, a recomendação é de se utilizar valores de corte em torno de 25% a 30% da intensidade máxima de tons de cinza da imagem. Dentre os limiares fixados estudados neste trabalho, o limiar de 70 (27,45%) foi o *threshold* que apresentou os melhores resultados na classificação.

O acerto médio do *software* neste limiar está na taxa de 93.9% das imagens cromossômicas individuais segmentadas corretamente (ANEXO-I), e assim, consideradas como padronizadas de acordo com as regras propostas. Pôde-se também concluir que o acerto na segmentação da imagem não se deve levar em consideração apenas a separabilidade dos objetos (cromossomos), mas também se deve considerar a preservação das informações iniciais das imagens adquiridas, características estas, essenciais para a classificação e caracterização final.

Após a etapa de segmentação das imagens cromossômicas à partir das imagens de células metafásicas, uma técnica de registro foi elaborada e aplicada sobre as mesmas, objetivando o reposicionamento da disposição do cromossomo e as adequando para a tentativa de interpretação das características pelos algoritmos propostos. O erro de 12% na identificação dos centrômeros e o erro de 0.5% de centrômeros não encontrados estão diretamente ligados a esta transformação, pois os algoritmos de identificação são baseados sobre a disposição dos cromossomos.

Como forma de comparação, as mesmas transformações foram aplicadas sobre as imagens cromossômicas utilizando os valores de limiar em 60 (sessenta) e 80 (oitenta), obtendo 90.9% e 91.7% de acerto no reconhecimento da padronização da disposição dos cromossomos nas imagens, respectivamente.

Entretanto, apesar dessas imagens apresentarem valores pouco inferiores ao acerto na identificação da disposição dos cromossomos nas imagens geradas pelo limiar de 70, o limiar de 60 apresentou um valor de corte não recomendado, devido ao fato da alta taxa de falsos negativos serem eliminados durante a remoção do *background*, e, de forma inversa, o limiar de 80 apresentou uma alta taxa de falsos positivos, preservando um alto nível de ruído como, já citado, o acúmulo de Giemsa e inviabilizando segmentações relativamente simples.

O mapeamento de identificação das imagens cromossômicas, traçando os 9 perfis nas imagens, demonstrou resultados satisfatórios ao auxílio no reconhecimento da segmentação e classificação da disposição do cromossomo na imagem cromossômica. Para cada imagem cromossômica, este mapeamento foi realizado. Posteriormente, este método auxilia na determinação dos intervalos de perfis onde estão localizados os possíveis centrômeros, tornando busca pelos centrômeros de forma mais eficiente.

A Figura 35 apresenta os resultados de segmentação e interpretação das imagens cromossômicas geradas pelo aplicativo com o limiar configurado para cada uma das intensidades estudadas.

Figura 35 – Total de reconhecimentos nos cromossomos com as três intensidades (60, 70 e 80) em cada imagem metafásica analisadas. Número de cromossomos analisados e porcentagem do total de cromossomos em uma célula x Imagens (sem fundo) escolhidas para análise.

Diversas outras características, de certa forma, elementares aos especialistas citologistas podem (com absoluta certeza) ser identificadas e tratadas de acordo com regras de classificação adequadas, sugerindo uma proposta de continuidade deste trabalho. Fato este, reforçado devido à complexidade da interpretação das informações referente aos cromossomos sem qualquer tipo de supervisão (interferência do usuário). Algumas ideias estão sugeridas no capítulo (VI CONCLUSÃO).

A identificação final de cada uma das 36 (trinta e seis) imagens cromossômicas foi baseada na segmentação, registro e interpretação dos dados referentes a cada imagem. Esta identificação foi feita por meio de características intrínsecas das imagens analisadas e compreendeu a comparação entre suas dimensões, o preenchimento da disposição do cromossomo e a identificação do número de centrômeros. Esta última baseia-se em dois Sistemas de Lógica *Fuzzy* propostos para reconhecimento de padrões, definidos para busca a localização das constrições na morfologia dos cromossomos.

Para as imagens reconhecidas como não padronizadas a busca pelos centrômeros não é realizada, ou seja, as características referentes à morfologia do cromossomo são

iguais a 0 (zero), e um indicador é acionado para que o usuário analise estas imagens visualmente.

A primeira técnica proposta para a identificação de perfis pertencentes à localização do(s) centrômero(s) nas imagens cromossômicas identificou corretamente 79,8% do total de 410 centrômeros encontrados nas imagens cromossômicas, como o valor do limiar configurado em 70 (27,45%).

A Figura 36 apresenta os resultados obtidos pela primeira técnica de identificação de localização do centrômero na imagem cromossômica gerada pelo aplicativo.

Figura 36 – Resultados obtidos pela primeira técnica de reconhecimento de padrões de identificação dos perfis de localização dos centrômeros em uma imagem cromossômica. Número de centrômeros analisados e porcentagem de acerto da primeira técnica x Imagens analisadas com limiar igual a 70.

O acerto da primeira técnica de reconhecimento da localização dos centrômeros em imagens cromossômicas é de 296 (90,5%) do total de 327 centrômeros encontrados por esta técnica, nas 396 imagens cromossômicas analisadas pelos algoritmos de reconhecimento da localização dos centrômeros (ANEXO-II).

A segunda técnica proposta para a identificação de perfis pertencentes à localização do(s) centrômero(s) nas imagens cromossômicas identificou corretamente 20,2% do total de centrômeros encontrados nas imagens cromossômicas, como o valor do limiar

configurado em 70. A Figura 37 apresenta os resultados obtidos pela segunda técnica de busca pela localização dos centrômeros na imagem cromossômica gerada pelo aplicativo.

Figura 37 – Resultados obtidos pela segunda técnica de reconhecimento de padrões de identificação dos perfis de localização dos centrômeros em uma imagem cromossômica. Número de centrômeros analisados e porcentagem de acerto da segunda técnica x Imagens analisadas com limiar igual a 70.

O acerto da segunda técnica de reconhecimento da localização do centrômero em imagens cromossômicas é de 65 (78,3%) do total de 83 centrômeros encontrados por esta técnica, nas 396 imagens cromossômicas analisadas pelos algoritmos de reconhecimento da localização dos centrômeros (ANEXO-II).

O acerto total da localização dos centrômeros pelo aplicativo somando os resultados de ambas as técnicas de reconhecimento são apresentados na Figura 38.

Figura 38 – Resultados obtidos pelas duas técnicas de reconhecimento de padrões de identificação dos perfis de localização dos centrômeros em uma imagem cromossômica. Número de centrômeros analisados e porcentagem de acerto somando as duas técnicas x Imagens analisadas com limiar igual a 70.

A Figura 38 demonstra o número de centrômeros analisados por ambas as técnicas propostas para identificação a localização de centrômeros nas imagens cromossômica. O valor máximo individual de erro está na imagem 070 e 008b(1)r com 21,1% dos centrômeros foram reconhecidos erroneamente. Estas imagens foram difíceis de serem identificadas devido à morfologia bastante irregular destes cromossomos, onde, o método de identificação dos perfis mostrou-se rigoroso com os perfis candidatos a pertencer a um centrômero, e os desclassificando de forma incorreta.

O acerto de ambas as técnica de reconhecimento da localização dos centrômeros em imagens cromossômicas é de 361 (88%) do total de 410 centrômeros encontrados nas 396 imagens cromossômicas analisadas pelos algoritmos de reconhecimento da localização dos centrômeros (ANEXO-II).

O gráfico da Figura 38 apresenta ainda os erros na identificação da localização dos centrômeros, onde apenas 2 centrômeros visualmente identificados não foram considerados, 1 na imagem 070 e 1 na imagem 003r.

Uma forma de avaliar todo o aplicativo, considerando cada etapa de processamento em que as imagens foram submetidas, e apresentar os resultados finais para a identificação e classificação das imagens cromossômicas, possíveis de serem identificadas, desenvolvida por este trabalho, tem o acerto médio de 93,9% das imagens cromossômicas geradas identificadas corretamente, com o limiar de 70.

A seção a seguir apresenta a *GUI* desenvolvida para o usuário. Uma ferramenta com algumas funcionalidades está implementada e atualmente está sendo utilizada uma versão beta pela equipe da Dr^a Okazaki. Esta ferramenta recebeu o nome de CHRIMAN.

V-1. Apresentação da Interface do Usuário - CHRIMAN

Foi desenvolvida uma interface amigável (*user-friendly*) para o profissional citogeneticista. A equipe de desenvolvimento escolheu o nome de CHRIMAN (*CHROMosome IMAGE ANalyser*). No aplicativo, estão disponíveis para a análise todas as etapas citadas na metodologia, e também um rápido e fácil manuseamento dessas informações, com funcionalidades de obtenção e análise dos resultados. O CHRIMAN realiza reconhecimento de características de imagens cromossômicas e apresenta essas informações em um painel de fácil visualização das imagens resultantes. Outra funcionalidade é a possibilidade de salvar essas informações obtidas em disco rígido.

A funcionalidade de visualização das imagens resultantes, onde os cromossomos presentes na imagem metafásica estejam dispostos e apresentados lado a lado facilitando a análise visual do profissional citogeneticista, e possibilita o reconhecimento de padrões para a localização dos centrômeros em imagens cromossômicas. A Figura 39 apresenta a interface gráfica do CHRIMAN.

Figura 39 – Interface do usuário desenvolvida pelo para o aplicativo CHRIMAN.

A interface é dividida em *headers* (painéis) de funcionalidades, onde cada painel representa uma funcionalidade implementada no CHRIMAN. Um painel de inicialização e inserção das imagens e o arquivamento dessas imagens cromossômicas (salvar as imagens de entrada e suas resultantes), um painel de análise dessas imagens, onde é possível aplicar *zooms*, manusear o posicionamento, rotacionar, assim como, identificar o valor de cada *pixel* presente em cada uma das imagens, outro painel com duas telas de visualização de imagens em seus diferentes processos, até a apresentação dos cromossomos no painel final de visualização dos cromossomos, e um ultimo painel com as informações obtidas após a análise celular.

O primeiro painel é o *Metaphase* – que consiste na apresentação da imagem a ser tratada pelo aplicativo, chamadas de *input image*, essas imagens estão no formato *RGB* e ficam posicionadas no lado esquerdo do painel. No lado direito deste mesmo painel fica posicionada a imagem após cada etapa do processamento e transformação a que a imagem é submetida. Este módulo apresenta ainda as dimensões da imagem.

O segundo módulo é o módulo de processamento *Image Processing* – este módulo representa a parte lógica de todo o sistema. Através de botões, o usuário consegue visualizar as etapas de processamento a que a imagem é submetida. Neste painel está

localizada a única interferência do usuário na decisão do aplicativo, a seleção do limiar (*threshold*) de intensidade que será utilizado para a segmentação.

O terceiro painel *Metaphase View* – apresenta a imagem resultante da união das imagens cromossômicas individuais geradas pelo *software* após todas as etapas de transformação, de forma que facilite a visualização do especialista. Apesar deste painel não apresentar fielmente uma cariotipagem, como mencionado, este módulo visa facilitar a visualização de forma semelhante a cariotipagem padrão.

E, por fim, no quarto painel *Chromosome Analysis* – que apresenta os resultados obtidos para cada imagem cromossômica individualmente. Neste painel são apresentados diversas informações referentes à célula que esta sendo analisada, como também dados referentes à imagem cromossômica individual gerada à partir da célula metafásica. Neste painel, foi criado também, um visualizador individual incremental, onde é possível analisar cada imagem cromossômica gerada e identificar o seu rótulo atribuído no painel *Metaphase*.

VI CONCLUSÃO

Está proposta uma metodologia que inclui: a aquisição da imagem de metáfases humanas irradiadas no reator nuclear IEA-R1 e a utilização do CHRIMAN como um auxílio ao estudo citogenético. O tratamento, análise e interpretação destas imagens, permitem dentre as funcionalidades do aplicativo, o armazenamento das informações referentes a cada uma das imagens de células e cromossômicas em disco rígido, a automatização da contagem e caracterização dos cromossomos, e, por meio de técnicas de reconhecimento de padrões, realizar uma análise classificatória sobre cromossomos presentes na imagem metafásica que os caracterize como cromossomos normais, ou com algum tipo de aberração cromossômica (dicêntrico, tricêntrico e fragmentos).

O aplicativo desenvolvido pode representar uma forma de reduzir os custos operacionais necessários para o estudo dos cromossomos e identificação das aberrações cromossômicas, e protagoniza um desenvolvimento inicial para a validação de uma metodologia de classificação de aberrações cromossômicas a partir de imagens digitais.

Alguns tipos de aberrações cromossômicas não são possíveis de serem identificadas por esta metodologia de auxílio ao diagnóstico citogenético (técnica convencional utilizando o corante Giemsa). Aberrações como as translocações e inversões dentro da mesma estrutura cromatídica (rearranjos) das cromátides dos cromossomos, contendo o conteúdo genético, por se tratarem de trocas entre as cromátides nas cromátides, não podem ser identificadas através do conjunto de aquisição de imagens: câmera fotográfica comum – ocular do microscópio. Sendo assim, o aplicativo se concentra na capacidade de simular a visão do profissional citogeneticista, cuja identificação está baseada na morfologia dos cromossomos na imagem metafásica.

Cerca de 93,9% de imagens cromossômicas com o limiar de segmentação sugerido foram classificadas corretamente. Este resultado foi satisfatório para etapa de segmentação e identificação da disposição dos cromossomos nas imagens. O desenvolvimento deste trabalho, que já está em grande parte implementado, propôs a elaboração de métodos

(regras) de identificação de centrômeros e, a identificação de cromossomos com aberrações do tipo dicêntrico e tricêntrico, como também os fragmentos cromossômicos.

Com a detecção dessas aberrações, torna-se possível o auxílio ao diagnóstico de um dos efeitos provocados pela radiação ionizante sobre as células humanas. Dentre as 36 imagens cromossômicas analisadas do conjunto de 11 imagens de células metafásicas, acredita-se que o valor de 88% para a proposta de apresentação desta metodologia, representa resultados satisfatórios para o reconhecimento das características de caracterização da localização dos centrômeros nas imagens cromossômicas geradas. O erro de 12% na identificação dos centrômeros, e o erro de 0.5% de centrômeros não encontrados está diretamente relacionados ao erro no registro e segmentação das imagens à partir da imagem metafásica.

Como mencionado ao longo de todo este documento, ainda há a possibilidade de se incorporar ao núcleo de reconhecimento deste aplicativo, um maior número de regras de classificação para cobrir todas as possibilidades de reconhecimento.

Como indicação para trabalhos futuros, acredita-se que possam ser estabelecidos critérios mais estritos e definidos para as etapas desta metodologia de análise, incluindo-se desde os critérios iniciais de irradiação das células, os critérios de aquisição de imagem, os critérios para validação dos resultados e padronização da apresentação das imagens processadas e classificadas. Assim, pode-se vir a estabelecer ao redor do aplicativo um conjunto de procedimentos padrões que funcionem como base para a potencial utilização do citogeneticista de forma validada cientificamente.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 Amato, A., **Ueber die Wirkung der Roentgenstrahlen auf in karyokinese begriffene Zellen**, Z. Roentgenk 13, 1–14, 1911.
- 2 Ballerini, L., **Genetic snakes for medical image segmentation**. Mathematical modeling and estimation techniques in computer vision, 284 – 295, 1998.
- 3 Ballerini, L., **Multiple genetic snakes for bone segmentation**. Applications of evolutionary computing, 346 – 356, 2003.
- 4 Bellman, R.E., Zadeh, L.A., **Decision-Making in a Fuzzy environment**. Management Science, v.17, n.4, p.B141-B164, 1970.
- 5 Bergonie, J., Tribondeau, L., **Action des rayons X sur le testicule du rat blanc**, Comp. Rend. Soc. Biol. 57, 400–402, 1904.
- 6 Buckton, K.E., Evans, H.J., **Methods for the analysis of human chromosome aberrations**. World Health Organization, Geneva, 1-66, 1973.
- 7 Castleman, K.R., Schulze, M., Wu, Q., **Automated biodosimetry using digital image analysis of fluorescence in situ hybridization specimens**. Radiat. Res. 148, 71–75, 1997.
- 8 Cecconello, M.S., Bassanezi, R.C., **Mudança na natureza da interação entre duas espécies**. Biomatemática, 16, 107-116, 2006.
- 9 Cecconello, M.S., **Modelagem alternativa para dinâmica populacional: Sistemas dinâmicos Fuzzy**. Dissertação de Mestrado, IMECC - UNICAMP, Campinas/SP, 2006.
- 10 Coelho, P.R., Muniz, R., Nascimento, J., Silva, G.S.A., Siqueira P.T., Yoriyaz, H., Carneiro, V., **Radiation field characterization of the NCT research facility at IEA-R1. In: 13th International Congress on Neutron Capture Therapy - A new option against cancer**. Florença. 13th International Congress on Neutron Capture Therapy - A new option against cancer, p. 553-555, 2008.
- 11 Coppin, B., **Artificial Intelligence Illuminated**. Jones & Bartlett Publishers, 2004.

- 12 Decheva E. V.Valev. **Knowledge based chromosome analysis system - a conception.** Proc. of Third International Conference on Automatic Image Processing CAIP'89, Leipzig, (99-101), 1989.
- 13 Decheva E. **Image processing in CAIS - an intelligent system for chromosome analysis and classification.** In: R.Klette (ed.), "Computer Analysis of Images and Patterns", Proc. of the Ivth International Conference CAIP'91, Akademie Verlag, Berlin, (260-264), 1991.
- 14 Decheva E. **Object-oriented approach to the design of CAIS - an intelligent system for chromosome analysis and classification.** In: Mathematics and Education in Mathematics - Proc. of 21 Spring Conference of UBM, Sofia, April 3-6, (226-230), 1992.
- 15 Detcheva E. **Heuristic knowledge based chromosome classification.** Proc. of First National Conference INFORMATICS'94, Sofia, (83-89), 1994.
- 16 Detcheva E. **CAIS - an intelligent system for chromosome analysis and classification.** Int. Workshop "Artificial Intelligence and the Humanities", Sozopol, (49-54), September 1996.
- 17 Detcheva E. **Processing and analysis of methaphase and chromosome images.** Mathematics and Education in Mathematics - Proc. of 27th Spring Conference of UBM, Sofia, (216-222), 1998.
- 18 Grasnick, W., **die Wirkung der Radiumstrahlen auf tierisches Gewebe. Experimentell-histologische Untersuchung as Geweben von Amphibianlarven,** Arch. Mikroskop. Anat. 90, 1–38, 1918.
- 19 Guimaraes, L.V., Schuck, A., Elbern, A., **Chromosome classification for karyotype composing applying shape representation on Wavelet packet transform,** In: Proc. 25th Annual Internat. Conf. of the IEEE EMBS, pp. 941–943, 2003.
- 20 Gonzalez, R.C., Woods, R. E., Eddins, S.L., **Digital Image Processing Using MATLAB.** Prentice Hall, 2004.
- 21 Gonzalez, R.C., Woods, R. E., **Digital Image Processing. 3rd Edition.** Prentice Hall, 2008.
- 22 Hejlesen, O., Ege, B., Englmeier, K.H. , Aldington, S., McCanna, L., and Bek, T., **TOSCA-Imaging – Developing Internet Based Image Processing Software for**

- Screening and Diagnosis of Diabetic Retinopathy.** Medinfo, Proceedings of the 11th World Congress on Medical Informatics, IOS Press (eds. M. Fieschi et al.) pp 222-226, 2004.
- 23 Huber, R., Kulka, U., Lörch, Th., Braselmann, H., Engert, D., Figel, M., Bauchinger, M., **Technical report: application of the Metafer2 fluorescence scanning system for the analysis of radiation-induced chromosome aberrations measured by FISH-chromosome painting.** Mutat. Res. 492, 51–57, 2001.
 - 24 Huber, R., Kulka, U., Lörch, Th., Braselmann, H., Bauchinger, M., **Automated metaphase finding: an assessment of the efficiency of the Metafer2 system in a routine mutagenicity assay.** Mutat. Res. 334, 97–102, 1995.
 - 25 Huber, R., Kulka, U., Lörch, Th., Braselmann, H., Bauchinger, M., **Technical report: automated classification of first and second cycle metaphases.** Mutat. Res. 419, 27–32, 1998.
 - 26 Hummel, R. A., **Image Enhancement by Histogram Transformation,** Computer Graphics and Image Processing, 6, 2, 184–195, 1977.
 - 27 IAEA; **Cytogenetic analysis for radiation dose assessment.** A Manual. Technical reports series no 405, International Atomic Energy Agency, Vienna, 2001.
 - 28 Jain, A.K., Duin, R.P.W., and Mao J., **Statistical Pattern Recognition: A Review.** IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 22, n.1, 4-37, 2000.
 - 29 Jiang, J., **Image compression with neural networks – a survey.** Signal Processing: Image Communication, 14, 737–760, 1999.
 - 30 Josien, P.W., Pluim et al., **Mutual Information Based Registration of Medical Images: A Survey.** IEEE Transactions on Medical Imaging, vol. XX, No. Y, 2003.
 - 31 Kavalco, K., (portal) **biociencia.org**, 1999 – acessado em 2010.
 - 32 Kasahara, S., **Práticas de Citogenética.** Rio Claro, SP: autor-editor. v. 1. 69 2001.
 - 33 Koernicke, M., **Ueber die Wirkung der Roentgen und Radiumstrahlen auf pflanzliche Gewebe und Zellen,** Ber.Deut. Bot. Ges. 23, 404–415, 1905.

- 34 Luomahaara, S., Lindholm, C., Mustonen, R., Salomaa, S., **Distribution of radiation-induced exchange aberrations in human chromosomes 1, 2, and 4.** Int. J. Radiat. Biol. 75, 1551–1556, 1999.
- 35 Matlab®: **The Language of Technical Computing - Creating Graphical User Interfaces.** Version 1, The MathWorks, 2000.
- 36 MATLAB® version R2009b. Natick, Massachusetts: The Mathwork Inc., 2010.
- 37 Matz, S.C., Figueiredo, R.J.P. de, **A localized nonlinear method for the contrast enhancement of images**, in: IEEE International Conference on Image Processing (ICIP '99), vol. 3, Los Alamitos, CA, pp. 484–488, 1999.
- 38 Mehmet, S., Bulent S., **Survey over Image Thresholding Techniques and Quantitative Performance Evaluation.** Journal of Electronic Imaging. 13(1):146-165, 2004.
- 39 Mendel, J.M., *Fuzzy Logic Systems for Engineering: a Tutorial.* Proc. IEEE, Vol. 83: 345-377, 1995.
- 40 Milios, E. E., **Shape matching using curvature processes.** Comput. Vision Graph. Image Process. 47, 203-226, 1989.
- 41 Moradi, M., Setarehdan, S.K. **New features for automatic classification of human chromosomes: a feasibility study**, Pattern Recogn. Lett. 27 (1), 19–28, 2006.
- 42 A.T. Natarajan, **Chromosome aberrations: past, present and future.** Mutat. Re. 504, 3-16, 2002.
- 43 Natarajan, A.T., Obe, G., **Molecular mechanisms involved in the production of chromosomal aberrations. Part I. Utilization of Neurospora endonuclease for the study of aberration production in G2 stage of the cell cycle**, Mutat. Res. 52, 137–149, 1978.
- 44 Nickolls, P., Piper, J., Rutovitz, D., Chisholm, A., Johnstone, I., Robertson, M. **Pre-processing of images in an automated chromosome analysis system.** Pattern Recognition 14(1-6): 219-229, 1981.

- 45 O’Gorman, L., **Binarization and multithresholding of document images using connectivity. Graphical models and image processing.** Vol.56, No.6, 494 – 506, 1994.
- 46 Odawara, K., Yamamoto, K., Kato, H., Hara, M., Kogiso, S., Kishida, F., Yoshitake, A., Nakatsuka, I., **A new semi-automated chromosome analysis system for in vitro chromosomal aberration tests.** Mutat. Res. 389, 207–212, 1997.
- 47 Otsu, N., **A Threshold Selection Method from Gray-Level Histograms.** IEEE Transactions on Systems, Man, and Cybernetics., 9(1): 62-66, 1979.
- 48 Perthes, G., **Versuche ueber den Einfluss der Roentgenstrahlen und der Radium strahlen auf die Zellteilung,** Deut. Med. Wochenschr. 30, 632–634, 1904.
- 49 Pizer, S.M. et al., **Adaptive Histogram Equalization and Its Variations,** Computer Vision, Graphics and Image Processing. 39, 3, 355–368, September 1987.
- 50 Revell, S.H., **Chromosome breakage by X-rays and radiomimetic substances in Vicia,** Heredity 6, 107–123, 1953.
- 51 Reed, T. R., & Wechsler, H., **Segmentation of texture images and gestalt organization** using spatial/spatial-frequency representations. IEEE Transactions on PAMI, 12, 1–12, 1990.
- 52 Ritter, G., Gaggermeier, K. **Automatic classification of chromosomes by means of quadratically asymmetric statistical distribution,** Pattern Recogn. 32 (6), 997–1008, 1999.
- 53 Roshtkhari, M.J., Setarehdan, S.M. **A novel algorithm for straightening highly curved images of human chromosome.** Pattern Recognition Letters, 1208~1217, 2008.
- 54 Russell, S.J., Norvig, P., **Artificial Intelligence: A Modern Approach.** 2nd edition, Prentice-Hall, 2002.

- 55 Sampat, M.P., Bovik, A.C., Aggarwal, J.K., Castlemand, K.R., **Supervised parametric and non-parametric classification of chromosome images**. Pattern Recognition 38, 1209 – 1223, 2005.
- 56 Sax, K., **An analysis of X-ray induced chromosomal aberrations in Tradescantia**, Genetics 25, 41–68, 1940.
- 57 Sax, K., **Types and frequencies chromosomal aberrations induced by X-rays**, Cold Spring Harb. Symp. Quant. Biol. 9, 93–101, 1941.
- 58 Schalkoff, R.J., **Pattern Recognition: Statistical, Structural and Neural Approaches**. John Wiley Sons, Inc., 1992.
- 59 Serebrovsky, A.S., **A general scheme for the origin of mutations**, Am. Nat. 63, 374–378, 1929.
- 60 Souza, L.P, Mesquita, R.N., **Classificação de dois tipos de aberrações cromossômicas por meio de lógica Fuzzy**, CONFIAM, Botucatu-SP, Brasil (2010).
- 61 Stadler, L.J., **The experimental modification of heredity in crop plants. Part I. Induced chromosomal irregularities**, Sci. Agric. 557–572, 1931.
- 62 Suetens, P., Fua, P., Hanson, A. J., **Computational strategies for object recognition**. ACM Comput. Suru. 24, 1(Mar.), 5-62, 1992.
- 63 Tsekouras, G.E., **A Fuzzy vector quantization approach to image compression**. Applied Mathematics and Computation, 167, 539–560, 2005.
- 64 Turing, A., **“Computing machinery and intelligence”**. Mind 59, 433–460, 1950.
- 65 Udupa, J.K. et al., **A methodology for evaluating image segmentation algorithms**. In Proceedings of SPIE: Medical Imaging, San Diego, CA, 266—277, 2002.
- 66 Zadeh, L.A., **Fuzzy sets**, Information and Control. Vol 8, pp. 338-353, 1965.
- 67 Zimmemann, H. J., **Fuzzy set theory and applications**. 2.ed. Boston: Dordrecht; London : Kluwer Academy, pp. 241-82, 1995.

ANEXO I - Resultados Obtidos – Segmentação e Registro das Imagens Cromossômicas.

Metáfase	Threshold	Count.	Padronizadas		NãoPadronizada		Acertos		Erros	
501	60	45	42	93.3%	3	6.7%	41	88.9%	4	8.9%
	70	44	42	95.5%	2	4.5%	42	94.4%	2	4.5%
	80	43	40	93%	3	7%	40	91.7%	3	7%
121	60	43	40	93%	3	7%	33	72.2%	10	23.3%
	70	43	41	95.3%	2	4.7%	38	86.1%	5	11.6%
	80	43	41	95.3%	2	4.7%	34	75%	9	20.9%
070	60	47	47	100%	0	0%	46	97.2%	1	2.1%
	70	48	48	100%	0	0%	46	94.4%	2	4.2%
	80	49	49	100%	0	0%	47	94.4%	2	4.1%
019	60	42	38	90.5%	4	9.5%	38	88.9%	4	9.5%
	70	41	38	92.7%	3	7.3%	38	91.7%	3	7.3%
	80	40	38	95%	2	5%	36	88.9%	4	10%
016r	60	45	43	95.6%	2	4.4%	42	91.7%	3	6.7%
	70	45	43	95.6%	2	4.4%	42	91.7%	3	6.7%
	80	45	44	97.8%	1	2.2%	41	88.9%	4	8.9%
011r	60	44	41	93.2%	3	6.8%	36	77.8%	8	18.2%
	70	44	42	95.5%	2	4.5%	40	88.9%	4	9.1%
	80	40	38	95%	2	5%	34	83.3%	6	15%
003r	60	45	43	95.6%	2	4.4%	43	94.4%	2	4.4%
	70	45	44	97.8%	1	2.2%	44	97.2%	1	2.2%
	80	45	44	97.8%	1	2.2%	44	97.2%	1	2.2%
002r	60	43	42	97.7%	1	2.3%	39	88.9%	4	9.3%
	70	43	42	97.7%	1	2.3%	39	88.9%	4	9.3%
	80	40	39	97.5%	1	2.5%	36	88.9%	4	10%
008b(1)r	60	43	42	97.7%	1	2.3%	37	83.3%	6	14%
	70	41	40	97.6%	1	2.4%	37	88.9%	4	9.8%
	80	41	40	97.6%	1	2.4%	36	86.1%	5	12.2%
0214Gyr	60	49	49	100%	0	0%	48	97.2%	1	2.0%
	70	49	49	100%	0	0%	49	100%	0	0%
	80	49	49	100%	0	0%	48	97.2%	1	2%
0224Gyr	60	47	46	97.9%	1	2.1%	45	94.4%	2	4.3%
	70	47	47	100%	0	0%	45	94.4%	2	4.3%
	80	47	47	100%	0	0%	46	97.2%	1	2.1%
Totais	-	1465	1418	96.8%	47	3.2%	1350	92.2%	115	7.8%
Totais	60	493	473	95.9%	20	4.1%	448	90.9%	45	9.1%
Totais	70	490	476	97.1%	14	2.9%	460	93.9%	30	6.1%
Totais	80	482	469	97.3%	13	2.7%	442	91.7%	40	8.3%

ANEXO II - Resultados Obtidos – Reconhecimento das Imagens Cromossômicas – *Threshold* = 70.

Metáfase	501		0121		019		070		016r		011r		003r		002r		008b(1)r		0124Gyr		0224Gyr		TOTAL	
Segmented Chromosomes Images	44		43		41		48		45		44		45		43		41		49		47		490	
Imagens Identificadas	42	95.5%	41	95.3%	38	92.7%	48	100.0%	43	95.6%	42	95.5%	44	97.8%	42	97.7%	40	97.6%	49	100.0%	47	100.0%	476	97.1%
Imagens Não Identificadas	2	4.5%	2	4.7%	3	7.3%	0	0.0%	2	4.4%	2	4.5%	1	2.2%	1	2.3%	1	2.4%	0	0.0%	0	0.0%	14	2.9%
Fragmentos Cromossômicas	2	4.5%	0	0.0%	0	0.0%	4	6.3%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	3	6.1%	1	2.1%	10	2.0%
Imagens com Rótulo > 36	6	13.6%	7	16.3%	5	12.2%	8	16.7%	9	20.0%	8	18.2%	9	20.0%	7	16.3%	5	12.2%	10	26.5%	10	21.3%	84	17.1%
Nº Imagens Cromossômicas Analisadas	36		36		36		36		36		36		36		36		36		36		36		396	
Erro - Segmentação	3	6.8%	2	4.7%	4	9.8%	1	2.1%	1	2.2%	1	2.3%	1	2.2%	5	11.6%	0	0.0%	0	0.0%	0	0.0%	18	4.5%
Erro - Registro	1	2.3%	0	0.0%	0	0.0%	7	14.6%	0	0.0%	0	0.0%	1	2.2%	3	7.0%	0	0.0%	1	2.0%	1	2.1%	14	3.5%
Nº Centrômeros Encontrados	41		46		40		46		46		46		46		42		43		48		50		494	
Nº Centrômeros Analisados	35		39		35		38		37		38		37		35		38		38		40		410	
Centrômeros Identificados	33	94.3%	33	84.6%	30	85.7%	30	78.9%	35	94.6%	34	89.5%	36	97.3%	30	85.7%	30	78.9%	35	92.1%	35	87.5%	361	88.0%
Centrômeros Não Identificados	2	5.7%	6	15.4%	5	14.3%	8	21.1%	2	5.4%	4	10.5%	1	2.7%	5	14.3%	8	21.1%	3	7.9%	5	12.5%	49	12.0%
Centrômeros Não Encontrados	0	0.0%	0	0.0%	0	0.0%	1	2.6%	0	0.0%	0	0.0%	1	2.7%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	2	0.5%
Classificação do Centrômero Técnica 1	28	80.0%	36	92.3%	25	71.4%	26	68.4%	33	89.2%	35	92.1%	30	81.1%	26	74.3%	26	68.4%	27	71.1%	35	87.5%	327	79.8%
Centrômeros Encontrados Corretamente	28	100.0%	30	83.3%	21	84.0%	24	92.3%	31	93.9%	31	88.6%	30	100.0%	24	92.3%	20	76.9%	26	96.3%	31	88.6%	296	90.5%
Centrômeros Encontrados Erroneamente	0	0.0%	6	16.7%	4	16.0%	2	7.7%	2	6.1%	4	11.4%	0	0.0%	2	7.7%	6	23.1%	1	3.7%	4	11.4%	31	10.5%
Classificação do Centrômero Técnica 2	7	20.0%	3	7.7%	10	28.6%	12	31.6%	4	10.8%	3	7.9%	7	18.9%	9	25.7%	12	31.6%	11	28.9%	5	12.5%	83	20.2%
Centrômeros Encontrados Corretamente	5	71.4%	3	100.0%	9	90.0%	6	50.0%	4	100.0%	3	100.0%	6	85.7%	6	66.7%	10	83.3%	9	81.8%	4	80.0%	65	78.3%
Centrômeros Encontrados Erroneamente	2	28.6%	0	0.0%	1	10.0%	6	50.0%	0	0.0%	0	0.0%	1	14.3%	3	33.3%	2	16.7%	2	18.2%	1	20.0%	18	21.7%
Cromossomos Classificados Corretamente	42	95.5%	38	88.4%	38	92.7%	46	95.8%	42	93.3%	40	90.9%	44	97.8%	39	90.7%	37	90.2%	49	100.0%	45	95.7%	460	93.9%
Cromossomos Classificados Erroneamente	2	4.5%	5	11.6%	3	7.3%	2	4.2%	3	6.7%	4	9.1%	1	2.2%	4	9.3%	4	9.8%	0	0.0%	2	4.3%	30	6.1%

ANEXO III – Regras *Fuzzy* PerfilClassif.

1. If (contagemBanda is pouco) and (quantidadeBanda is tres) then (padraoPerfil is naoValido) (1)
2. If (contagemBanda is medioPouco) and (quantidadeBanda is tres) then (padraoPerfil is naoValido) (1)
3. If (contagemBanda is medio) and (quantidadeBanda is tres) then (padraoPerfil is naoValido) (1)
4. If (contagemBanda is medioMuito) and (quantidadeBanda is tres) then (padraoPerfil is naoValido) (1)
5. If (contagemBanda is muito) and (quantidadeBanda is tres) then (padraoPerfil is naoValido) (1)
6. If (contagemBanda is pouco) and (quantidadeBanda is quatro) then (padraoPerfil is naoValido) (1)
7. If (contagemBanda is medioPouco) and (quantidadeBanda is quatro) then (padraoPerfil is naoValido) (1)
8. If (contagemBanda is medio) and (quantidadeBanda is quatro) then (padraoPerfil is naoValido) (1)
9. If (contagemBanda is medioMuito) and (quantidadeBanda is quatro) then (padraoPerfil is naoValido) (1)
10. If (contagemBanda is muito) and (quantidadeBanda is quatro) then (padraoPerfil is naoValido) (1)
11. If (contagemBanda is pouco) and (quantidadeBanda is cinco) then (padraoPerfil is naoValido) (1)
12. If (contagemBanda is medioPouco) and (quantidadeBanda is cinco) then (padraoPerfil is naoValido) (1)
13. If (contagemBanda is medio) and (quantidadeBanda is cinco) then (padraoPerfil is naoValido) (1)
14. If (contagemBanda is medioMuito) and (quantidadeBanda is cinco) then (padraoPerfil is naoValido) (1)
15. If (contagemBanda is muito) and (quantidadeBanda is cinco) then (padraoPerfil is naoValido) (1)
16. If (contagemBanda is pouco) and (quantidadeBanda is dois) then (padraoPerfil is curtoDuplo) (1)
17. If (contagemBanda is medioPouco) and (quantidadeBanda is dois) then (padraoPerfil is mediolongoDuplo) (1)
18. If (contagemBanda is medio) and (quantidadeBanda is dois) then (padraoPerfil is mediolongoDuplo) (1)
19. If (contagemBanda is medioMuito) and (quantidadeBanda is dois) then (padraoPerfil is mediolongoDuplo) (1)
20. If (contagemBanda is muito) and (quantidadeBanda is dois) then (padraoPerfil is mediolongoDuplo) (1)
21. If (contagemBanda is pouco) and (quantidadeBanda is um) then (padraoPerfil is curtoUnico) (1)
22. If (contagemBanda is medioPouco) and (quantidadeBanda is um) then (padraoPerfil is curtoUnico) (1)
23. If (contagemBanda is medio) and (quantidadeBanda is um) then (padraoPerfil is medioUnico) (1)
24. If (contagemBanda is medioMuito) and (quantidadeBanda is um) then (padraoPerfil is medioUnico) (1)
25. If (contagemBanda is muito) and (quantidadeBanda is um) then (padraoPerfil is longoUnico) (1)

ANEXO VI – Regras *Fuzzy* ChromoClassif.

- [illegible]

- [illegible]

63. If (sitPerfil is contSuficiente) and (tamMatriz is medGrande) and (preenchImagem is medioMediaPouco) then (sitImagem is padronizado) (1)
64. If (sitPerfil is contSuficiente) and (tamMatriz is medGrande) and (preenchImagem is medioMediaMuito) then (sitImagem is naoPadronizado) (1)
65. If (sitPerfil is contSuficiente) and (tamMatriz is medGrande) and (preenchImagem is medioMuito) then (sitImagem is naoPadronizado) (1)
66. If (sitPerfil is contSuficiente) and (tamMatriz is medGrande) and (preenchImagem is muito) then (sitImagem is naoPadronizado) (1)
67. If (sitPerfil is contSuficiente) and (tamMatriz is grande) and (preenchImagem is pouco) then (sitImagem is padronizado) (1)
68. If (sitPerfil is contSuficiente) and (tamMatriz is grande) and (preenchImagem is medioPouco) then (sitImagem is padronizado) (1)
69. If (sitPerfil is contSuficiente) and (tamMatriz is grande) and (preenchImagem is medioMediaPouco) then (sitImagem is padronizado) (1)
70. If (sitPerfil is contSuficiente) and (tamMatriz is grande) and (preenchImagem is medioMediaMuito) then (sitImagem is naoPadronizado) (1)
71. If (sitPerfil is contSuficiente) and (tamMatriz is grande) and (preenchImagem is medioMuito) then (sitImagem is naoPadronizado) (1)
72. If (sitPerfil is contSuficiente) and (tamMatriz is grande) and (preenchImagem is muito) then (sitImagem is naoPadronizado) (1)