## UNIVERSIDADE DE SÃO PAULO ESCOLA DE ENGENHARIA DE SÃO CARLOS DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Metodologia de estimação de idade óssea baseada em características métricas utilizando mineradores de dados e classificador neural.

# **EVANDRA MARIA RAYMUNDO**

Orientador: Prof. Dr. Evandro Luís Linhari Rodrigues

## UNIVERSIDADE DE SÃO PAULO ESCOLA DE ENGENHARIA DE SÃO CARLOS DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Metodologia de estimação de idade óssea baseada em características métricas utilizando mineradores de dados e classificador neural.

# **EVANDRA MARIA RAYMUNDO**

Dissertação apresentada à Escola de Engenharia de São Carlos, da Universidade de São Paulo, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Evandro Luís Linhari Rodrigues

Busquei ao SENHOR, e ele me respondeu; livrou-me de todos os meus temores.(Sl. 34,4) Às mulheres de minha vida, minha mãe Wilma que me ensinou a sonhar e me deu forças para que esse sonho se concretizasse, à minha irmã Eliane que esteve presente em todos os momentos da concretização desse sonho. Aos meus tios Waldomiro minha tia Maria Inês, que foram os alicerces desse sonho, ajudaram em etapas importantes e hoje vencem comigo. A toda Família Ferreira Salles pela força e pela presença em todos os momentos. A cada um dos que cruzaram meu caminho e com sorriso ou palavra de apoio me ajudaram a chegar até aqui.

Agradeço primeiramente a Deus por ter chegado aqui realizando este trabalho, pela inspiração nos momentos de dúvida e pela coragem nas horas difíceis. Por todas as vezes que Ele me carregou no colo, quando as forças me faltaram. Por todas as "mãos" que ele me fez entender e compreender, por ter me dado o dom de AMAR esse trabalho.

À minha mãe pela ajuda financeira e por toda ajuda incondicional e incentivo que me deu, à minha irmã pela ajuda incondicional em todos os momentos.

A toda a minha família pelo incentivo para a realização do meu mestrado por que se tornaram alicerces do meu sonho que hoje está construído

A meu orientador e xará, Prof. Dr. Evandro Luís Linhari Rodrigues pela confiança em meu trabalho, por ter acreditado que daria certo e por me orientar.

À banca examinadora pelas dicas e correções desse trabalho.

À Universidade de São Paulo por colocar à minha disposição computadores e todo material necessário para a realização deste trabalho.

Ao Prof. Dr. Marcelo Andrade Vieira e ao Prof. Dr. Ivan Nunes da Silva, pelo auxílio em várias fases deste trabalho.

A Prof.Dra. Marilde Terezinha Prado Santos e a seu aluno e meu amigo Rodrigo Bella, por toda colaboração, para a realização deste trabalho.

Aos meus amigos de Laboratório Ednilson, por toda ajuda na elaboração do banco de métricas, pela paciência e dedicação em ajudar no sucesso do meu trabalho, ao Celso Olivete Júnior que me forneceu toda base de conhecimento para realização deste trabalho, e por toda a ajuda prestada para a sua finalização. Ao André Delorme, pelas dicas, Frank, Rodrigo e Eduardo Llappa que mesmo de longe torceram pelo meu sucesso.

Aos meus amigos do Laboratório de Visão Computacional (LAVI), Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas (LAPIMO), Laboratório de Análise Computacional em Sistemas Elétricos de Potência (LACO), Laboratório de Automação Inteligente de Processos e Sistemas (LAIPS) pelo companheirismo durante período em que estava cursando as disciplinas, por toda ajuda e pelas horas descontraídas que passamos.

Ao Lúcio Jorge pelo auxílio com os mineradores.

Quero agradecer principalmente aos amigos que fiz durante todo esse período e que foram de grande importância, Claúdia e Augusto, Antonio Louro, Gustavo, Ricardo Casado, Michele, Juliana e Eduardo que ajudaram em tudo. Leandro Sebastian, pelo grupo de estudo e por tudo que compartilhamos.

Agradeço especialmente a Luciana de Toro e ao Cristiano Minotti, estiveram presentes em momentos cruciais, tornaram-se meus grandes amigos.

Agradeço ao Paulo Fonseca Filho e Paulo Cereda, pela idéia e ajuda fundamental na criação do banco de imagens. Fábio por todas as caronas e pela ajuda nos trabalhos.

À amiga Elaine, pela paciência e generosidade.

À Ivani Albers, uma querida amiga e minha paciente mestra de inglês.

Aos todos os amigos do trabalho que torceram a cada instante para que esse dia chegasse, Alecsandro, Casagrande, Daniela, Edvaldo, João Paes, José Carlos, Patrícia, Stevanatto, Wagner, in memória ao querido Néris que tanto quis estar aqui e assistir esse momento.

Sou grata aos funcionários que se tornaram meus amigos do Departamento de Engenharia Elétrica da EESC - USP, José, Jussara, Marisa, Rosane, Pedro e aos professores da pós-graduação, principalmente ao Prof. Dr. Adilson Gonzaga, que foi a primeira pessoa a possibilitar a realização desse sonho.

Às faxineiras que sempre foram compreensivas pelo cuidado conosco e principalmente, preocupação com nosso ambiente de trabalho.

Ao Prof. MSc. Gilmar Cação, que acreditou em mim e me forneceu bases para realizar melhor esse trabalho.

Agradeço a cada um que diariamente fez parte desse sonho concretizado, desta etapa vencida.

Agradeço ao CNPq por todo apoio financeiro.

"Mãos que se fizeram amigas, companheiras de todas as horas, plantaram conhecimento, semearam entre lágrimas e muitos risos, transformando convivência em sentimento saboreiam hoje a glória da vitória.

Muito Obrigada!!!!"

Os olhos do SENHOR estão sobre os justos, e os seus ouvidos atentos ao seu clamor.(Sl. 34, 4.15.)

Oferecer-te-ei sacrifícios de louvor, e invocarei o nome do SENHOR. Pagarei os meus votos ao SENHOR, na presença de todo o seu povo, Nos átrios da casa do SENHOR, no meio de ti, ó Jerusalém. Louvai ao SENHOR. (Sl. 116, 17-19)

O SENHOR guardará a tua entrada e a tua saída, desde agora e para sempre.(Sl. 121,8)

#### Resumo.

RAYMUNDO, E. M., Metodologia de estimação de idade óssea baseada em características métricas utilizando mineradores de dados e classificador neural. Dissertação de Mestrado - Escola de Engenharia de São Carlos Universidade de São Carlos, 2009.

Este trabalho apresenta uma proposta de metodologia de estimação de idade óssea baseada em características métricas, utilizando o banco de imagens carpais da Escola de Engenharia de São Carlos (EESC). As imagens foram devidamente segmentadas para obtenção da área, perímetro e comprimento de cada osso, gerando, assim, um banco de dados métricos o *CarpEven*. As informações da base métrica *CarpEven* foram submetidas a dois mineradores de dados: ao *StARMiner*, (*Statistical Association Rules*) uma metodologia de mineração de dados criada por um grupo de pesquisadores do ICMC-USP, e ao *Weka* (*Waikato Environment for Knowledge Analysis*), desenvolvido pela Universidade Waikato da Nova Zelândia. As informações foram submetidas a classificadores neurais, contribuindo, assim, para a criação de uma nova metodologia de estimação de idade óssea. Finalmente, é feita uma comparação entre os resultados obtidos e os resultados já alcançados por outras pesquisas.

**Palavras-chave:** Mineração de dados, estimação de idade óssea, imagem carpal, segmentação de imagens, classificadores neurais.

#### Abstract

RAYMUNDO, E. M., Methodology for bone age estimation based on metric characteristics using data mining and neural classifier. Mater Degree - School of Engenier de São Carlos University of San Carlos, 2009.

This work presents a methodology for bone age estimation based on metric characteristics using the carpal images database from Engineering School of São Carlos (EESC-USP). The images were properly segmented to obtain the area, perimeter and length of each bone, thus generating a metric database named CarpEven. The database information were submitted to two data miners: the StarMiner (Statistical Association Rules Miner) a methodology for data mining created by a group of researchers from ICMC-USP, and the Weka (Waikato Environment for Knowledge Analysis), developed by the University of Waikato in New Zealand. The information was submitted to the neural classifiers contributing to the creation of a new methodology for bone age estimation. The results are compared with those obtained by others research.

Keywords: Mining, bone age estimation, carpal image, image segmentation, neural classifiers.

# Lista de Figuras.

Figura 1 – Estrutura de um osso longo. Fonte: Avancini & Favaretto (1997)6
Figura 2- Esquema anatômico dos ossos cárpicos e da mão. Fonte: Tavano (2001)8
Figura 3 - Centros de ossificação da mão e punho. Fonte: Greulich & Pyle (1992)11
Figura 4 - Estágios de desenvolvimento do centro de ossificação da falange distal. Fonte:12
Figura 5 - Índice de Eklof & Ringertz, centros de ossificação utilizados. Fonte:
Figura 6 – (a) Marcações estimadas dos centros de ossificação. (b): localização15
Figura 7 - Etapas a serem desenvolvidas do algoritmo de classificação da idade óssea17
Figura 8 - Presença da sequência Fibonacci entre as falanges e metacarpo. Fonte:20
Figura 9 - Radiografia carpal ilustrando o efeito <i>Heel</i> . Fonte: Tavano (2001)23
Figura 10 - Fluxograma da metodologia aplicada. Fonte: Raymundo et al. (2007)24
Figura 11 – Resultados: a) imagem original, b) imagem otimizada, c) imagem após realce26
Figura 12 - Resultados de acordo com Land & Mccann (1971) (a) imagem original, (b) 28
Figura 13 - Imagem das falanges e metacarpos segmentados, mão feminina. Fonte:33
Figura 14 - Imagem das falanges e metacárpos com o traçado de comprimento, mão33
Figura 15 - Nomenclatura de atributos da base de métricas "CarpEven", Fonte:34
Figura 16- Etapas do processo de extração de conhecimento de dados. Fonte: Fayyad et al37
Figura 17- Visão hierárquica do processo de KDD. Fonte: Souza &Rocha (2005)38
Figura 18– Papel das atividades preditiva e descritiva na mineração de dados. Fonte: Garcia 39
Figura 19 – Ilustração das regiões de rejeição para teste de uma hipótese. Fonte: Ribeiro 45
Figura 20 - Neurônio Como Unidade Limiar. Fonte: Fernandes & Portugal (1995)50
Figura 21 – Exemplos de função de ativação. Fonte: Fernandes & Portugal (1995)51
Figura 22 – Representação das camadas de uma rede neural artificial. Fonte: Ortega (2002).53
Figura 23 – Diagrama esquemático da rede <i>Perceptron</i> . Fonte: Goedtel (2007)55
Figura 24 - Centros ósseos utilizados para compor o conjunto de métricas selecionadas para
cada faixa etária da Base Fem, utilizando algoritmos CSF, RepTree e M5P, destacados
em cores - métrica Área
Figura 25 - Centros ósseos utilizados para compor o conjunto de métricas selecionadas para
cada faixa etária da Base Masc, utilizando algoritmos CSF, RepTree e M5P, destacados
em cores - métrica Área74

Figura 26 - Desenvolvimento médio dos centros ósseos selecionados com a métrica Área para
(a) Base Fem e (b) Base Masc
Figura 27 - Desenvolvimento médio dos centros ósseos selecionados com a métrica Perímetro
para (a) Base Fem e (b) Base Masc
Figura 28 - Desenvolvimento médio dos centros ósseos selecionados com a métrica
Comprimento para (a) Base Fem e (b) Base Masc
Figura 29 – Gráfico taxa média de concordância dos algoritmos com a média dos laudos nos
testes de validação ( <i>LM/rB</i> ) Base Geral
Figura 30 – Gráfico taxa de concordância com a média dos laudos (Área Weka) para cada
faixa etária Base Geral82
Figura 31 - Gráfico taxa de concordância com a média dos laudos (Área <i>StARminer</i> ) para cada
faixa etária Base Geral83
Figura 32 - Gráfico taxa de concordância com a média dos laudos (Área Weka) para cada
faixa etária Base Fem84
Figura 33 - Taxa de concordância com a média dos laudos para cada faixa etária, utilizando
centros ósseos selecionados para 6 anos (Base Masc - Área)
Figura 34 - Taxa de concordância com a média dos laudos para cada faixa etária, quando
aplicadas aos centros ósseos mais vezes selecionados (Base Fem – Área)87
Figura 35 – Taxa de concordância com a média dos laudos para cada faixa etária, quando
aplicadas aos centros ósseos mais vezes selecionados (Base Masc $-$ Área)
Figura 36 – Centros ósseos utilizados para compor o conjunto da Análise I90
Figura 37 - Taxa de concordância com a média dos laudos para cada faixa etária, quando
aplicadas aos centros ósseos utilizados na Análise I (Base Fem)93
Figura 38 - Taxa de concordância com a média dos laudos para cada faixa etária, quando
aplicadas aos centros ósseos utilizados na Análise I (Base Masc)
Figura 39 - Centros ósseos utilizados para compor o conjunto da Análise II95
Figura 40 - Taxa de concordância com a média dos laudos para cada faixa etária, metacarpo
do dedo indicador (mind) e as falanges distais dos dedos polegar e médio (fdpol e fdmed)
- Análise II (Base Fem)97
Figura 41 - Taxa de concordância com a média dos laudos para cada faixa etária, metacarpo
do dedo indicador (mind) e as falanges distais dos dedos polegar e médio (fdpol e fdmed)
- Análise II (Base Masc)98
Figura 42 - Artigos e periódicos utilizados na pesquisa de acordo com ano de publicação103

# Lista de Tabelas.

Tabela 1 - Valores críticos de Z em relação aos valores de $\gamma_{\min}$ . Fonte: Ribeiro (2008)45
Tabela 2 – Padrão de entrada dos limiares no algoritmo StARminer e padrão estabelecido para
62
Tabela 3 – Atributos selecionados pelo software Weka; em destaque, os atributos utilizados 67
Tabela 4 - Conjunto final de atributos de cada uma das amostras, de acordo com Weka; em
destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a)68
Tabela 5 - Atributos selecionados com algoritmo StARminer; em destaque, os atributos
utilizados pela metodologia de Olivete Jr. (2005a)
Tabela 6 - Atributos da Base Geral selecionados com os algoritmos do Weka para Área; em
destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a)69
Tabela 7 - Atributos da Base Geral selecionados com os algoritmos do Weka para Área e
Perímetro; em destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a).
70
Tabela 8 - Resultado da aplicação do teste T-Student, para os atributos encontrados com
algoritmo RepTree para todas as faixas etárias71
Tabela 9 - Atributos da Base Geral selecionados com o algoritmo StARminer e parâmetros
utilizados na seleção para Área; em destaque, os atributos utilizados pela metodologia de
Olivete Jr. (2005a)
Tabela 10 - Atributos selecionados para Base Fem para faixa etária de 6 anos métrica-Área;
em destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a)72
Tabela 11 - Atributos selecionados para Base Masc para faixa etária de 6 anos - métrica Área;
não foram selecionados atributos utilizados pela metodologia de Olivete Jr.(2005a)72
Tabela 12 - Parâmetros utilizados nas redes neurais com os algoritmos Levenberg-Marquardt
e resilient-Backpropagation80
Tabela 13 - Resultado da aplicação do teste <i>T-Student</i> , para os resultados obtidos com a rede
neural <i>LM</i> para todas as faixas etárias (Base Geral)81
Tabela 14 – Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos encontrados
com algoritmo CSF para todas as faixas etárias Base Geral83
Tabela 15 - Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos encontrados
com algoritmo <i>StARminer</i> para todas as faixas etárias Base Geral84

Tabela 16 - Resultado da aplicação do teste T-Student, para os centros osseos encontrados
com algoritmo (Weka -CSF) para todas as faixas etárias Base Fem85
Tabela 17 - Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos encontrados
com algoritmo (StARminer) para todas as faixas etárias Base Fem
Tabela 18 - Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos encontrados
para a faixa etária de 6 anos com (Weka -CSF) para todas as faixas etárias Base Masc. 86
Tabela 19 – Centros ósseos mais vezes selecionados em todas as faixas etárias para Base Fem.
87
Tabela 20 - Centros ósseos mais vezes selecionados em todas as faixas etárias para Base
Masc
Tabela 21 – Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos mais vezes
selecionados com algoritmo (Weka ) para todas as faixas etárias Base Fem
Tabela 22 - Resultado da aplicação do teste T-Student, para os centros ósseos mais vezes
selecionados com algoritmo (Weka ) para todas as faixas etárias Base Masc89
Tabela 23 – Valores dimensionais (pixels) <sup>2</sup> para cada idade óssea (Base Fem) – Análise I,
conforme equação 1192
Tabela 24 – Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos utilizados na
Análise I para todas as faixas etárias Base Fem
Tabela 25 - Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos utilizados na
Análise I para todas as faixas etárias Base Masc94
Tabela 26 - Valores dimensionais (pixels) para cada centro óssseo (Base Masc) - Análise II.
96
Tabela 27 - Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos utilizados na
Análise II para todas as faixas etárias Base Fem
Tabela 28 - Resultado da aplicação do teste <i>T-Student</i> , para os centros ósseos utilizados na
Análise II para todas as faixas etárias Base Masc98

#### Lista de Abreviaturas.

Anacarp - Estimador da Idade Óssea via Análise Carpal.

ANSI – American National Standards Institute.

ARFF - Attribute Relationship File Format.

CBIR - Content-based Image Retrieval.

CFS – Correlation-based Feature Selection.

EESC – Escola de Engenharia de São Carlos.

E&R - designação para metodologia criada por Eklof e Ringertz.

G&P – designação para metodologia criada por Greulich e Pyle.

JDBC – Java data base connectivity.

KDD – Knowledge discovery in database.

LVQ- Learning Vector Quantization.

Mipav – Medical image processing analysis and visualization.

PMC - Perceptron Multicamadas

ODBC – Open data base connectivity.

RNA - Redes Neurais Artificiais.

RUS – Radio, Ulna e Short Bones.

SGBD – Sistema de gerenciamento de banco de dados.

SQL - Structured query language.

SSR – Single-scale retinex.

StARMiner – Statistical association rules.

TW - designação para metodologia criada por Tanner e Whitehouse.

TW2 – designação para segunda metodologia criada por Tanner e Whitehouse.

WEKA – Waikato environment for knowledge analysis.

# Sumário.

Res	umo		i	
Abs	tract		ii	
List	a de Figu	ras	iii	
List	a de Tabe	elas	v	
List	a de Abre	eviaturas	vii	
Sun	nário		ix	
1	Introd	lução	1	
1.1		ivo da pesquisa		
1.2	J	ância da pesquisa		
1.3				
	1 3			
2		ação de idade óssea.		
2.1		mação de idade óssea - formulação do problema de pesquisa		
2.2	A org	anização dos ossos da mão	8	
2.3	Méto	dos clássicos de avaliação da idade óssea	10	
	2.3.1	GREULICH & PYLE.	10	
	2.3.2	TANNER & WHITEHOUSE	12	
	2.3.3	EKLOF & RINGERTZ.	13	
	2.3.4	Estimativa de idade óssea através da análise carpal baseada na si	mplificação do	
	método	de Eklof & Ringertz	14	
	2.3.5	Metodologia de extração automática de características da mão pa	ra a estimação	
	da idade	óssea utilizando redes neurais artificiais no processo de decisão	16	
	2.3.6	Novas metodologias automáticas para estimativa da idade óssea	a, baseadas no	
	processa	mento de imagens radiográficas da mão	18	
2.4	Consi	derações a respeito da idade óssea.	20	
3	Considerações sobre as imagens de raio X			
3.1	O pré	-processamento como ferramenta para gerar informação	23	
	3.1.1	Metodologia adaptativa de processamento de imagem radiográfi	ca carpal com	
	realce de	e bordas.	24	

	3.1.2	Metodologia de pré-processamento utilizando o algoritmo retinex em	imagens		
	radiográ	ficas carpais.	26		
3.2	Avali	ação do processo de pré-processamento	28		
3.3	Conclusões a respeito da aplicação do pré-processamento em imagens carpais.29				
4	O arm	O armazenamento de dados			
4.1	A aplicação de sistemas de gerenciamento de dados				
4.2	Justificativa da aplicação do Banco de dados				
4.3	O processo de formação do banco de dados				
4.4	Consi	derações a respeito da base de dados métrica.	34		
5	Proce	sso de descoberta de conhecimento em bases de dados	36		
5.1	O pap	pel dos processos de KDD quando aplicados à base de métricas	39		
	5.1.1	O Weka	40		
	5.1.2	O StARMiner	42		
5.2	Impoi	rtância da aplicação de processos de mineração nos dados	46		
5.3	A con	stribuição da descoberta de conhecimento para base métrica	47		
6	Redes	S Neurais Artificiais	48		
6.1	Conceitos iniciais				
6.2	Componentes das RNAs				
6.3					
	6.3.1	Rede LVQ (Learning Vector Quantization) e Counterpropagation	54		
	6.3.2	O Perceptron multicamadas	54		
6.4	Consi	derações finais sobre redes neurais artificiais.	56		
7	Desenvolvimento da Metodologia58				
7.1	Introdução		58		
7.2	Etapa	s utilizadas no desenvolvimento da metodologia	58		
	7.2.1	Geração dos dados e constituição da base CarpEven	58		
	7.2.2	Preparação dos dados.	59		
	7.2.3	O processo de mineração de acordo com Weka.	59		
	7.2.4	O processo de mineração de acordo com StARrminer	61		
	7.2.5	Redes neurais artificiais aplicadas como classificadores	62		

7.3	Considerações finais			
8	Discussões e resultados.			
8.1	Introdução	65		
8.2	O processo de geração de informação	65		
8.3	O processo de mineração de dados.	65		
	8.3.1 Características utilizadas pelo Weka	66		
	8.3.2 Características selecionadas pelo StARminer	68		
8.4	A aplicação das redes neurais artificiais.	79		
8.5	Combinações dos centros ósseos			
8.6	Considerações a respeito dos resultados			
9	Conclusão.			
9.1	Trabalhos futuros			
10	Referências.	103		

## 1 - Introdução.

A estimação de idade óssea utilizando ossos da mão e punho tem sido empregada devido à grande quantidade de ossos e epífises localizados em uma área não muito extensa, que sofrem mudanças em diferentes tempos e velocidades. O grau de desenvolvimento é, frequentemente, avaliado através do uso de indicadores que refletem as mudanças físicas ocorridas no indivíduo durante o processo de maturação (CASANOVA et al., 2006).

Segundo Olivete Jr (2005a), o maior desafio na utilização de métodos de visão computacional para alcançar resultados satisfatórios na estimação da idade óssea é a utilização de uma metodologia que seja capaz de obter diagnósticos computadorizados e análises quantitativas para estudos em grandes populações, com diferentes faixas etárias e contrastes.

Incorporar técnicas computacionais para análise de imagens médicas, e proporcionar meios que auxiliem o diagnóstico médico é a função desta pesquisa. Características como dimensão, podem ser medidas com relativa facilidade por meio do amparo computacional, servindo como uma segunda opinião na análise de um especialista. Determinar a idade óssea através da imagem digitalizada de uma radiografia de mão e punho (radiografia carpal) é um procedimento frequentemente empregado em pacientes pediátricos que apresentam desordens no crescimento e, mais recentemente na área odontológica, nos procedimentos ortodônticos (GREULICH & PYLE, 1992). A forma clássica mais antiga de avaliação é por meio de comparações do desenvolvimento de ossos de determinadas regiões da radiografia carpal do paciente com um modelo de referência. Existem dois métodos muito utilizados neste tipo de procedimento: Greulich, (GREULICH & PYLE 1992) e Tanner, (TANNER & WHITEHOUSE, 1969); ambos necessitam de uma comparação visual entre a radiografia do paciente e modelos padrão fornecidos por Atlas. Outro método utilizado é o de Eklof, (EKLOF & RINGERTZ, 1967), que se baseia na análise do comprimento e largura de 10 centros de ossificação.

Mesmo este tipo de análise, baseado em dimensões, ainda é subjetivo, pois varia de acordo com a experiência dos radiologistas e consome tempo para a avaliação de todos os ossos ou todas as regiões de interesse dependendo do método. O auxílio computacional, representa um avanço para essas análises, já que as estruturas ósseas de interesse podem ser

realçadas, e as medidas dimensionais podem ser obtidas, facilitando, assim, sua análise (OLIVETE JR., 2005 b, c, d, e 2006 a, b).

#### 1.1 - Objetivo da pesquisa.

Este trabalho tem como objetivo principal a criação de uma nova metodologia para a estimação de idade óssea, a partir da investigação exaustiva das relações entre as métricas obtidas, através da segmentação das imagens de raio X e o uso de mineração de dados associada à aplicação de classificação neural. A pesquisa utiliza métricas baseadas na metodologia de Eklof & Ringertz - que foi simplificada por Olivete Jr (2005a) -, analisando em conjunto os resultados obtidos por Queiroz (2006) que, em sua pesquisa, não só endossa a eficiência do método de Olivete Jr (2005a), como também avalia que o uso de redes neurais, para o processo de classificação, auxilia, de forma importante, o procedimento como um todo.

### 1.2 - Relevância da pesquisa

Sabe-se que a ciência, em todas as suas vertentes, busca o bem-estar do indivíduo; portanto, a importância de diagnósticos precoces é fundamental para o restabelecimento da saúde ou para correção de disfunções em tempo hábil.

Para embasar a relevância da pesquisa são apresentados, neste trabalho, os métodos clássicos para estimação de idade óssea, nos quais é possível notar que as novas metodologias, oriundas desses métodos, estão não somente, baseadas na redução dos centros analisados, mas também buscam fornecer, ao radiologista, meios para o diagnóstico prévio de anomalias ou mesmo disfunções de qualquer natureza.

Assim, as métricas analisadas por este trabalho dão continuidade ao ciclo de pesquisas que vêm sendo realizadas, e abrangem os vários aspectos utilizados pelos métodos clássicos e outros métodos que podem ser associados a estes, tais como área e perímetro.

Reunir as informações a respeito de estimação de idade óssea cria possibilidades para que haja correções de desordens de crescimento, já que as avaliações podem ser feitas de forma prática, permitindo, aos especialistas da área médica, a escolha de uma metodologia que se adapte às suas necessidades ou mesmo utilizá-la associada a outras.

Um breve relato apresenta os resultados prévios das últimas pesquisas de Olivete Jr (2007), com relação à presença do número de ouro e da sequência *Fibonacci* nas falanges dos

dedos. Tais pesquisas impulsionam investigações que visam encontrar novas formas de estimação de idade óssea que culminem, não só com metodologias precisas, mas também com a redução da quantidade de centros de ossificação analisados, o que diretamente indica uma redução do tempo envolvido no processo e, supressão da subjetividade existente em métodos cuja decisão da análise é feita somente pelo ser humano.

### 1.3 - Estrutura de apresentação do trabalho

O presente trabalho encontra-se dividido em 9 capítulos que são descritos sucintamente a seguir.

O Capítulo 2, Revisão Bibliográfica, apresenta uma revisão literária sobre a Estimação de idade óssea, motivo de toda a pesquisa. Relaciona as metodologias clássicas e as pesquisas atuais realizadas na área de Estimação de idade óssea. As metodologias estão relacionadas em ordem cronológica, permitindo, assim, a prévia avaliação da contribuição de cada uma para o objetivo do trabalho.

O Capítulo 3, A problemática das imagens de raio X, apresenta um breve relato a respeito das imagens que são utilizadas nesta pesquisa, os problemas que as mesmas podem apresentar, e as técnicas utilizadas para amenizá-los. Apresenta a solução encontrada por esta pesquisa para auxílio ao processo de segmentação.

O Capítulo 4, O armazenamento de dados, relata as principais características do banco de dados *CarpEven*, sua concepção e forma. É um capítulo descritivo, ilustra o funcionamento e as características dos dados utilizados nesta pesquisa, como os mesmos são armazenados e sua nomenclatura.

O Capítulo 5, Processo de descoberta de conhecimento em banco de dados, traz uma abordagem a respeito do processo de descoberta do conhecimento, relatando as vantagens desse método que tem, como principal objetivo, a mineração de dados, ou seja, a descoberta de informações relevantes em grandes volumes de dados. Descreve o software e o algoritmo que são aplicados à base de dados *CarpEven*, suas particularidades e modo de funcionamento.

O Capítulo 6, Redes Neurais Artificiais, faz uma breve revisão literária sobre redes neurais artificiais. Descreve o funcionamento das mesmas, suas utilidades e aplicações, abordando os tipos utilizados nessa pesquisas e suas particularidades.

O Capítulo 7, Metodologia, aborda aspectos da metodologia utilizada, desde a formação da base de dados até o processo de validação. Descreve o funcionamento dos mineradores e redes neurais, suas características e desempenho perante as informações da base *CarpEven*.

O Capítulo 8, Discussão e resultados, apresenta os resultados obtidos e uma discussão a respeito dos mesmos, por meio de gráficos e tabelas; o desempenho tanto dos mineradores quanto das redes neurais é analisado e discutido. São observadas as particularidades e similaridades dos conjuntos de dados obtidos e a eficiência de sua aplicação na classificação dos dados.

O Capítulo 9, Conclusões e sua seção sobre trabalhos futuros, apresenta, conclusões a respeito dos resultados da aplicação dos mineradores como metodologia de extração de características, e do classificador neural utilizado. Analisa as avaliações feitas e aponta para o conjunto eleito para compor a nova metodologia, baseada na utilização da base *CarpEven* e da associação dessas novas informações.

## 2 - Estimação de idade óssea.

De acordo com Enlow (1993), o crescimento não é um mero processo de aumento de tamanho. Ele envolve, também, uma sucessão desconcertante de mudanças regionais nas proporções e requer incontáveis ajustes localizados para alcançar a devida função e estabilidade entre todas as partes.

O indicador comumente utilizado em estudos a respeito do crescimento e desenvolvimento é a idade óssea (TAVANO, 2004), sendo que, a radiografia de mão e punho representa um auxílio na avaliação da idade óssea individual, detectando, por meio dos eventos de ossificação, o período de surto de crescimento puberal (MERCADANTE, 1996).

Neste capítulo, é feita uma revisão da literatura sobre a estimação de idade óssea, métodos conhecidos e utilizados pela comunidade médica, bem como todos os avanços ocorridos nas pesquisas a este respeito.

## 2.1 - A estimação de idade óssea - formulação do problema de pesquisa.

A análise de radiografias carpais tem sido muito utilizada para estimar a idade óssea, obtendo o quanto o crescimento evolui em relação à maturidade óssea, pois a idade cronológica não é um bom critério para avaliar o desenvolvimento e o crescimento de uma pessoa, devido às diferenças relacionadas ao sexo e outros fatores que contribuem para sua variação, tais como: genética, condição nutricional e ambiental, (OLIVETE JR, 2005a).

Durante a infância, conforme afirmação de Tanner & Whitehouse (1969), o crescimento está intimamente relacionado à idade cronológica, sofrendo um distúrbio no período da puberdade. O crescimento ósseo está relacionado à maturidade estrutural dos ossos do corpo. Analisando-se individualmente o osso, verifica-se que sua região periférica ou seus terminais, no caso de ossos longos (com estrutura cilíndrica), como pode ser visto na Figura 1, à medida que cada osso cresce, passa por alguns estágios (MICHAEL & NELSON, 1989). Há, segundo Queiroz (2006), um esqueleto cartilaginoso durante a vida embrionária, o qual será quase totalmente substituído por um esqueleto ósseo. É o que se denomina ossificação endocondral (do grego endos, dentro, e chondros, cartilagem).

A partir do segundo mês de vida intra-uterina, começam a se formar os ossos. Ao nascer, a criança já apresenta um esqueleto bastante ossificado, mas as extremidades de

diversos ossos ainda mantêm regiões cartilaginosas que permitem o crescimento (Queiroz 2006).

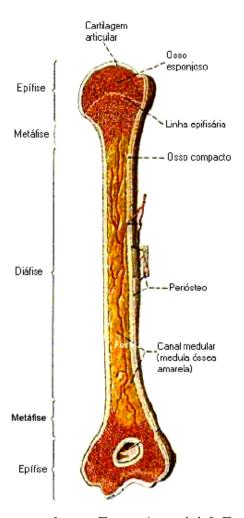


Figura 1 – Estrutura de um osso longo. Fonte: Avancini & Favaretto (1997).

Segundo Bosqueiro et al. (2001), pode-se observar o crescimento e o desenvolvimento de um indivíduo comparando a altura e o peso com tabelas de padrões existentes - método chamado de cálculo de idade morfológica; - contudo, não é um método eficaz, pois essas características variam de acordo com a raça e outros fatores.

A idade óssea, por sua vez, tem sido o fator mais utilizado nos estudos de estimação de crescimento e desenvolvimento, de acordo com Tavano (2001), devido à apresentação de resultados muito satisfatórios e, até certo ponto, seguros, quando comparados à idade biológica. Uma das maneiras de se determinar a idade óssea é através da análise do exame radiológico da mão e punho esquerdo. Os principais pontos de análise são os centros de

ossificação, forma e fusão dos ossos. Essa região tem sido muito utilizada devido à sequência cronológica que apresenta.

Devido a essas variações, os métodos clássicos de análise têm uma grande importância, pois os mesmos são divididos em dois tipos: baseado em padrões, no qual a idade é obtida por meio de uma comparação visual com os padrões existentes em um Atlas, e o método de escores, no qual cada centro de ossificação recebe um valor, e no final da análise, determina-se a idade comparando com uma escala de desenvolvimento.

Citar trabalhos como o de Taffarel et al. (2003) é muito apropriado, pois o mesmo desenvolveu uma metodologia de estimação de idade óssea automatizando o método de Tanner & Whitehouse; já Bosquieiro et al. (2001), realizaram um estudo cujo objetivo foi verificar se os métodos de estimativa de idade óssea de Greulich & Pyle e Tanner & Whitehouse poderiam ser aplicados à população brasileira, e qual dos dois métodos seria o mais confiável, se comparado com a idade cronológica. Tavano (2001) apresentou um importante estudo, em sua tese de doutorado (1976), determinando a idade óssea de 590 crianças brasileiras, sendo 295 de cada sexo, com idades entre 3 a 17 anos, utilizando, para tal, os índices de Greulich & Pyle(GP), Tanner & Whitehouse (TW) e de Eklof & Ringertz (ER). Tavano (2001) verificou a possibilidade da utilização destes índices em território nacional; o método de TW foi simplificado, utilizando-se apenas os escores dos ossos cárpicos - os outros centros foram desprezados. Para o método ER, as variações dimensionais de largura dos ossos utilizados foram modificadas, e o mesmo provou que essas modificações não alteram os resultados finais.

Partindo da análise das metodologias clássicas em conjuntos com as metodologias desenvolvidas por pesquisadores, essa pesquisa, baseada nas métricas extraídas das imagens carpais, visa encontrar relações que possam ser utilizadas, para estimar a idade óssea, de forma simples, sem a necessidade de exaustivas análises, com auxilio de ferramentas específicas para descoberta de conhecimento em grande volume de dados.

Aborda-se a organização dos ossos da mão para elucidar a dificuldade encontrada, quando é necessária a avaliação de regiões da mão que apresentam sobreposição de estruturas, justificando, assim, a busca por métodos que não necessitem da avaliação dessas regiões e utilizem conjuntos reduzidos para a estimação.

## 2.2 - A organização dos ossos da mão.

A mão é composta por 27 ossos, conforme mostra a Figura 2. Destes, 19 são alongados e localizam-se na palma da mão; e, nos dedos, os oito ossos restantes são curtos e formam o carpo, articulados entre si e, juntamente com os ossos do metacarpo, a ulna e o rádio correspondem ao punho.

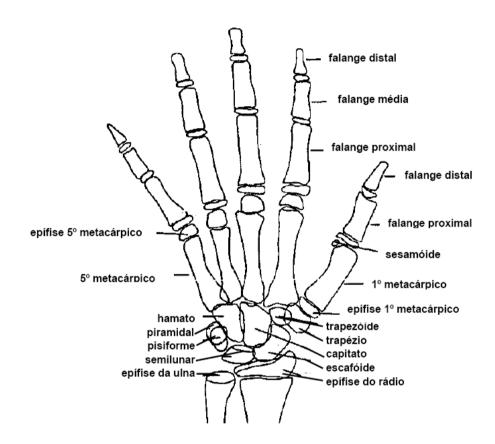


Figura 2- Esquema anatômico dos ossos cárpicos e da mão. Fonte: Tavano (2001).

Os ossos metacárpicos correspondem a um conjunto de cinco ossos, contados a partir do polegar, que formam a palma da mão e estão ligados às falanges e aos ossos do carpo. Cada um dos ossos do metacarpo é formado por um corpo e duas extremidades que se articulam com as falanges proximais, formando os nós do punho. O início da ossificação dos ossos metacarpianos se dá no terceiro mês antes do nascimento da criança (NETLER, 1990).

Por outro lado, as falanges são os dedos propriamente ditos, sendo que o polegar, primeiro dedo, é formado por apenas duas falanges a proximal e a distal; os outros quatro dedos são formado por três falanges: distal, média e proximal. A articulação com o metacárpico é feita pela falange proximal; a falange distal é livre em sua extremidade distal, e a média está posicionada entre as duas. A ossificação das falanges ocorre entre o terceiro e o quarto mês antes do nascimento da criança, segundo Tavano (2001).

O punho é formado pelos ossos do carpo, que são descritos abaixo, e observados do sentido do rádio para a ulna, conforme apresentado na Figura 2.

- Escafoide: é o osso de maior volume da fileira proximal e tem o formato de um barquinho;
- Semilunar: apresenta uma concavidade profunda na sua superfície inferior e tem o formato de uma meia lua;
- Trapézio: situa-se entre o Escafoide e a epífise do 1º Metacárpico;
- Trapezoide: é o mais largo dorsal que ventralmente articula-se com o Escafoide, o
   2º Metacárpico, o Trapézio e o Capitato;
- Capitato: é o maior dos ossos do carpo e o primeiro a se ossificar;
- Hamato: tem um formato piramidal com a sua base fixada nos dois últimos metacárpicos;
- Pisiforme: é o menor osso do carpo, tem formato arredondado é o último a se ossificar;
- Piramidal: apresenta um formato de pirâmide com a sua base voltada para cima e para fora.

O início da ossificação do carpo ocorre após o nascimento da criança; entretanto, há registros de casos de nascimento de crianças com alguns ossos do carpo já ossificados.

No esqueleto da mão existem ossos ditos fixos, que são os da série distal do carpo e os metacarpianos centrais (segundo e terceiro metacarpos); e os móveis, que são os demais e estão divididos principalmente em duas partes: as falanges que dão forma aos dedos e os metacarpianos periféricos - o primeiro, o quarto e o quinto metacarpiano (BOSQUIEIRO et al., 2001).

## 2.3 - Métodos clássicos de avaliação da idade óssea.

Existem diversos métodos para a estimação da idade óssea, que utilizam vários pontos e estruturas do corpo para tal avaliação. De acordo com (DUARTE & NEVES,2008), há 800 centros de ossificação no corpo humano que podem ser utilizados para estimação de idade óssea. Metade desses centros desenvolve-se durante a vida uterina e são chamados de primários; e os outros, chamados secundários, surgem mais tarde e se desenvolvem ao longo de linhas definidas até a sua fusão quando a maturação óssea se encerra.

Os métodos clássicos de estimação de idade óssea mais difundidos e utilizados no Brasil são os métodos Greulich & Pyle, Tanner & Whitehouse e Eklof & Ringertz, baseados na análise dos ossos da mão e punho.

O primeiro utiliza um Atlas de padrões para fazer a avaliação por meio de inspeção dos ossos e estimar a idade óssea (OLIVETE JR, 2005a). O segundo utiliza escores, e o método de Eklof & Ringertz utiliza medidas de comprimento e/ou largura de determinados centros de ossificação, com distinção entre sexos.

A grande vantagem do Atlas, segundo Haiter et al., (2000) é que ele faz a distinção entre os sexos na apresentação dos padrões.

#### 2.3.1 - GREULICH & PYLE

Em 1950, Greulich & Pyle publicaram um Atlas radiográfico de desenvolvimento ósseo da mão e punho. Destacaram a necessidade da criação de um método que promovesse informações mais precisas sobre o desenvolvimento de um indivíduo e não apenas informações, como estatura, peso e idade cronológica (QUEIROZ, 2006).

A comparação da radiografia carpal do indivíduo e as radiografias contidas em um Atlas é a essência da metodologia de Greulich & Pyle. O Atlas é baseado em dados coletados durante estudos e apresenta diferentes padrões para cada sexo.

O método de Greulich & Pyle envolve um grande número de comparações de todos os ossos da mão com as radiografias de diferentes idades disponibilizadas no Atlas; e, de acordo com o sexo, recebem uma idade óssea igual àquela do padrão ao qual se assemelham. No total, 28 pontos de ossificação precisam ser examinados durante a comparação (NIEMEIJER, 2002).

O Atlas é composto por 58 pranchas radiográficas, com padrões que variam de acordo com sexo, e estão dispostos em intervalos de três meses de idade, nos primeiros 18 meses, e, a

partir daí, até 5 anos, semianualmente; e, depois, anualmente, exceto por alguns intervalos semestrais na puberdade. Cobre até a idade de 18 anos para o sexo feminino e 19 anos para o sexo masculino (OLIVETE JR, 2005a). O Atlas apresenta a idade esqueletal dos ossos, individualmente, e uma parte descritiva em quadros que se encontram junto a cada padrão para cada sexo (TAVANO, 2001).

A estimativa da idade óssea é obtida por meio de uma inspeção visual de 28 centros de ossificação, mostrados na Figura 3.

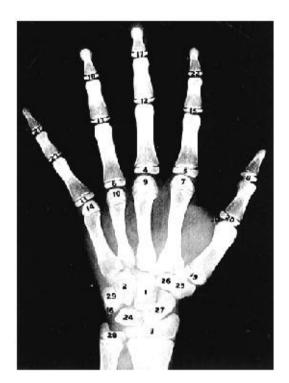


Figura 3 - Centros de ossificação da mão e punho. Fonte: Greulich & Pyle (1992).

Segundo Greulich & Pyle, (1992), estes centros devem ser comparados com todos os padrões dispostos no Atlas, até que se encontre um semelhante à radiografia que está sendo analisada. Estes padrões são formados por informações extraídas de radiografias carpais, obedecendo aos intervalos de idade. Na maioria dos casos analisados, nem sempre a fase em que a radiografia está, coincide exatamente com algum padrão do Atlas. Quando isso ocorre, seleciona-se aquele que mais se aproxima da radiografia em análise (OLIVETE JR, 2005a).

#### 2.3.2 - TANNER & WHITEHOUSE

O primeiro método de Tanner & Whitehouse foi criado em 1959; o chamado TW. Nesse método, cada um dos vinte ossos selecionados, da mão e punho, são classificados separadamente em estágios existentes de acordo com letras que vão de **a** até **j** (QUEIROZ, 2006). Em 1969, baseado nos escores de Acheson et al.(1966), este sistema foi revisado, culminando no método denominado TW2, que contém alterações no sistema de escores, e, ainda, estabelece escores diferentes para cada sexo. Este novo sistema também prevê maturações ósseas separadas para o rádio, ulna e ossos curtos, metacarpos e falanges, (método RUS) e para os ossos carpais (método CARPAIS), segundo Queiroz (2006).

A avaliação dos vinte centros de ossificação entre os ossos da mão e do punho ainda é realizada. O segundo e o quarto dedos não são considerados, porque a maturação dos metacárpicos e das falanges, geralmente, está intimamente ligada, prejudicando o isolamento preciso dessas áreas. A Figura 4 ilustra os estágios da falange distal.



**Figura 4 -** Estágios de desenvolvimento do centro de ossificação da falange distal. Fonte:Haiter et al (2006).

Como observado na Figura 4, foram escolhidos apenas os padrões dos centros de ossificação da falange distal para exemplificar suas fases de crescimento, excluindo-se o restante, que pode ser encontrado em Tanner et al. (1969).

Baseado na atribuição de escores para cada um dos 20 ossos, o método segue a seguinte ordem: rádio, ulna, metacárpicos I, II e V; falanges proximais I, II e V; falanges médias III e V; falanges distais I, III e V; capitato, hamato, piramidal, semilunar, escafoide, trapézio e trapezoide.

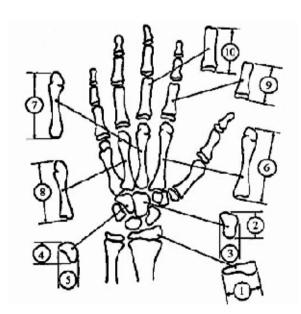
Os ossos são avaliados em uma escala de A até I, exceto o rádio, que é avaliado de A – J. Caso não haja sinal da presença de osso é dada a avaliação A. As avaliações são registradas comparando-se o osso em questão com os diagramas que existem no Atlas para cada ponto de

ossificação. Tal informação encontra-se em uma página do Atlas com todos os estágios de cada osso, bem como as especificações de idades (OLIVETE JR, 2005a).

#### 2.3.3 - EKLOF & RINGERTZ.

O método Eklof & Ringertz é baseado nas medidas de comprimento e/ou largura de determinados centros de ossificação. Inicialmente, foram realizados todos os testes possíveis com todos os ossos para determinar quais proporcionavam resultados estatisticamente mais precisos na determinação da idade óssea em relação a laudos médicos. Com a realização de testes, verificou-se que o crescimento de alguns ossos da mão e do pulso satisfaziam esta condição no intervalo de 1 a 15 anos de idade. Verificou-se, também, que as crianças com idade inferior a 1 ano e meio apresentavam uma regressão não linear entre idade cronológica e idade óssea e, com isso, concluiu-se que elas não poderiam ser incluídas nesta análise, pois apresentavam índices discrepantes de idade óssea (TAVANO, 2001).

Para determinar os padrões normais, foram escolhidos 10 parâmetros (ossos) que possuíam os mais altos coeficientes de correlação (0,922 – 0,964) entre idades e medidas em questão (OLIVETE JR, 2005a). Os 10 parâmetros são apresentados a seguir, na Figura 5.



**Figura 5** - Índice de Eklof & Ringertz, centros de ossificação utilizados. Fonte:Tavano(2001).

A estimativa da idade óssea é obtida através da análise dos centros de ossificação, formados pelas regiões mostradas na Figura 5: 1- largura da epífise distal do rádio; 2-comprimento do capitato; 3- largura do capitato; 4- comprimento do hamato; 5- largura do hamato; 6- comprimento do metacárpico II; 7- comprimento do metacárpico III; 8-comprimento do metacárpico IV; 9- comprimento da falange proximal II; e 10- comprimento da falange proximal III. Uma série de 1013 crianças foi analisada pelos autores, para estabelecer as variações normais. Usando cálculos estatísticos, eles apresentaram 2 tabelas, uma para cada sexo, em que cada um dos 10 centros de ossificação tem um padrão mínimo e um máximo de medida (em milímetros). Estes parâmetros são apresentados para todas as idades de 1 até 15 anos, com intervalos de 3 meses entre eles, representando um total de 56 parâmetros para cada centro de ossificação.

Existem vários trabalhos que utilizam estes métodos para estimar a idade óssea. Um exemplo é Marques et al. (2001), que utiliza imagens carpais para estimar a idade óssea; outro trabalho importante foi o de França et al. (1998), que buscou estabelecer cientificamente as idades ósseas e dentárias utilizando radiografias da face, bacia, joelho, cotovelo, entre outros, que apresentam centros de ossificação.

No trabalho de França et al. (1998), as imagens foram inicialmente segmentadas, utilizando técnicas de processamento de imagens e, posteriormente, foram analisadas utilizando o método de Tanner & Whitehouse.

Moraes et al. (2003) também realizou um estudo comparativo entre os métodos de Greulich & Pyle, e Eklof & Ringertz para verificar parâmetros como: a variabilidade, fidelidade, confiabilidade e a praticidade entre os dois (OLIVETE JR, 2005a).

A busca pelo aprimoramento dessas metodologias clássicas e a necessidade de adaptálas às novas tecnologias motivam pesquisas que objetivam melhorar, ainda mais, a estimação. A seguir, discorrer-se-á a respeito de alguns trabalhos relacionados à estimativa de idade óssea, produzidos pelo grupo de pesquisa da EESC.

# 2.3.4 - Estimativa de idade óssea através da análise carpal baseada na simplificação do método de Eklof & Ringertz.

Olivete Jr (2005a), em sua pesquisa, obtém todas as medidas objetivas dos ossos, permitindo, assim, uma diferenciação da idade óssea nas diferentes fases de crescimento.

Para tanto, 450 imagens de radiografia da mão esquerda do banco de dados da EESC, de pacientes de ambos os sexos, foram submetidas a um pré-processamento com intuito de

amenizar efeitos indesejáveis que porventura levassem a conclusões dúbias. Em seguida, foram submetidas ao processo de segmentação, que possibilitou a construção de um Atlas digital com todas as imagens e dados refletindo o desenvolvimento ósseo.

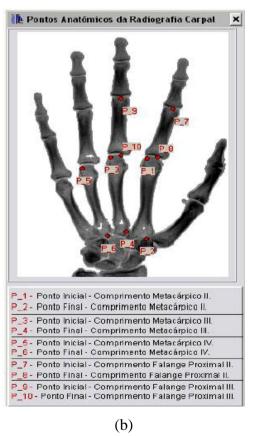
O banco de dados de imagens carpais foi utilizado para armazenar todas as imagens préprocessadas (segmentadas), com as suas respectivas informações, facilitando, desta forma, o acesso às mesmas. As medidas são armazenadas de acordo com o proposto pelo método de Eklof & Ringertz (OLIVETE JR, 2005a).

A metodologia utiliza apenas 5 centros de ossificação (localizados apenas nos dedos), sendo necessária a inserção de apenas 10 marcadores. Para facilitar a estimativa, os marcadores são inseridos automaticamente.

Olivete Jr (2005a) criou um software que realiza marcações automáticas dos centros de ossificação Figura 6 (a) e, em seguida, disponibiliza um mapa, Figura 6 (b), com as posições corretas dos 10 centros de ossificação. Após essa etapa, é verificado se os pontos que foram automaticamente inseridos estão nas posições corretas; caso contrário, é necessário ajustá-los.



(a)



**Figura 6** – (a) Marcações estimadas dos centros de ossificação. (b): localização correta dos pontos utilizados por Eklof & Ringertz. Fonte: Olivete Jr (2005a).

A idade óssea final é obtida através da média das idades encontradas para cada um dos cinco ossos.

Com os resultados obtidos pela simplificação, (5 centros de ossificação), Olivete Jr, (2005a), buscou uma nova simplificação, baseando-se em um número inferior a 5 centros. Para isso, foram feitas todas as combinações possíveis entre os 5 ossos selecionados no estudo, comparando-as com o laudo médico e também com os valores obtidos através do uso de todos os centros de ossificação, (num total de 10 ossos) propostos pelo método de Eklof & Ringertz. O experimento constatou que esta simplificação poderia ser ainda maior, utilizando apenas 3 ossos para estimar a idade, fazendo com que o tempo de processamento fosse ainda mais reduzido. Assim, são necessários apenas 6 pontos contra 20 pontos do método original.

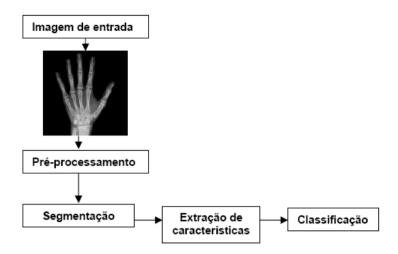
Com base nos estudos de Olivete Jr (2005a), conclui-se que é possível estimar, com confiança, a idade óssea, baseando-se apenas nos ossos da mão.

# 2.3.5 - Metodologia de extração automática de características da mão para a estimação da idade óssea utilizando redes neurais artificiais no processo de decisão.

Com base nas informações citadas anteriormente, Queiroz (2006) desenvolveu uma metodologia capaz de automatizar a estimação da idade óssea e auxiliar o radiologista em sua análise, servindo como uma segunda opinião na decisão final, que diminui o tempo necessário para análise de todos os centros de ossificação. Com o emprego de redes neurais artificiais e de técnicas de processamento de imagens, foi possível investigar as relações entre os centros de ossificação utilizados pelo método de Eklof & Ringertz – tradicional, e o método simplificado proposto por Olivete Jr, et al. (2005e) - buscando, também, investigar a possibilidade de se reduzir, ainda mais, o número de centros de ossificação necessários para estimação da idade óssea.

O objetivo dessa metodologia foi desenvolver um algoritmo para estimar a idade óssea, baseado no método de Eklof & Ringertz, utilizando redes neurais.

O algoritmo desenvolvido é composto por quatro partes, como mostra a Figura 7. A primeira consiste no pré-processamento da imagem; a segunda parte consiste na segmentação, cujas estruturas de interesse, ou seja, os centros de ossificação, são destacados; na terceira parte são extraídas as características que servirão como dados de entrada para a rede neural, que irá estimar a idade óssea automaticamente, compreendendo, assim, a quarta parte do processo (QUEIROZ, 2006).



**Figura 7** - Etapas a serem desenvolvidas do algoritmo de classificação da idade óssea. Fonte: Queiroz, (2006).

Queiroz (2006) afirma que a extração de características, de imagens radiográficas da mão, de forma automática, é bastante prejudicada quando se trata da região do punho, devido à sobreposição de ossos. Porém, utilizando-se o método de Eklof & Ringertz, simplificado, proposto por Olivete Jr et al. (2005e), em que essa região é desprezada, este procedimento pode ser implementado de forma mais consistente.

O método de assinatura horizontal foi utilizado para encontrar características, e procedimentos de varredura foram aplicados para localizar os dedos na imagem que é binarizada. Para que ocorra essa localização, as posições iniciais são armazenadas e aplicadas na imagem em nível de cinza, após a etapa de suavização (QUEIROZ 2006).

Feitas as demarcações, o passo seguinte é o processo de classificação, realizado com auxílio de redes neurais.

No processo de classificação, foram utilizadas as redes *Perceptron* multicamadas, devido ao seu processo de aprendizado supervisionado e a maneira com que os clusters são formados, (QUEIROZ,2006). Para que a rede possa classificar através do método tradicional de Eklof & Ringertz, o conjunto de treinamento é composto pelos dados dimensionais do Atlas, que apresenta dimensões de idades entre 1 e 15 anos.

Os resultados e testes feitos no trabalho de Queiroz (2006) mostraram-se satisfatórios quando comparados aos métodos clássicos na etapa de classificação. A idade estimada corresponde à idade óssea estimada de acordo com o laudo médico.

A metodologia proposta para a extração automática de características da mão mostrouse eficiente, devido à capacidade de localizar regiões de interesse e de extração da dimensão de centros de ossificação em imagens de baixa qualidade. Foi possível, ainda, observar que, mesmo em imagens que possuem inclinação nos dedos, o algoritmo localiza e extrai as informações dimensionais de maneira correta.

A aplicação de redes neurais artificiais no processo de classificação possibilitou a estimação da idade óssea de vários pacientes simultaneamente, auxiliando o trabalho do radiologista especialmente em grandes centros radiológicos, onde vários padrões de entrada podem ser testados, e o resultado na saída da rede é a idade óssea estimada (QUEIROZ, 2006).

# 2.3.6 - Novas metodologias automáticas para estimativa da idade, óssea baseadas no processamento de imagens radiográficas da mão.

A proposta de Olivete Jr. (2007), contribui para aperfeiçoar a metodologia desenvolvida em sua dissertação de mestrado, com a finalidade de criar novos métodos de pré-processamento para extração automática de informações dos ossos da mão (centros de ossificação), via abordagem de contornos ativos e apresentar uma evolução do modelo tradicional, que seja capaz de se adaptar às formas ósseas encontradas nas radiografias da mão.

Em síntese, a pretensão foi comprovar a eficiência e a agilidade na extração de características, verificando a existência da proporção áurea e da sequência de *Fibonacci* em medidas dos centros de ossificação, a fim de analisar se o crescimento ósseo do paciente está de acordo com a sua idade cronológica.

Para a realização desta pesquisa, foram selecionadas 640 imagens radiografias digitalizadas da mão de indivíduos dos sexos masculino e feminino de uma mesma região brasileira; as radiografias foram obtidas junto ao Prof. Dr. Francisco Haiter Neto<sup>1</sup>.

Após o processo de digitalização, as imagens foram separadas de acordo com a idade cronológica do indivíduo e através do laudo médico obtido das metodologias de E&R, G&P e T&W. Os laudos foram feitos pelo Prof. Dr. Francisco Haiter Neto. Parte dos laudos baseados no método de E&R, também foi obtida através do software *Anacarp* (OLIVETE & RODRIGUES, 2006).

A metodologia foi dividida em 3 etapas, iniciando-se pela segmentação das imagens, seguida pela utilização dos modelos de contornos ativos. A última etapa utiliza a sequência *Fibonacci* e o número de ouro para estimar a idade óssea.

\_

<sup>&</sup>lt;sup>1</sup> Departamento de Odontologia da Faculdade de Odontologia de Piracicaba, UNICAMP.

### Segmentação - o modelo de contornos ativos.

O método de contornos ativos foi proposto para resolver o problema de localização de bordas de objetos em imagem. De acordo com Olivete Jr. (2007), estes métodos iniciam-se em uma configuração mais ou menos arbitrária, com um contorno inicial que evolui até contornar o objeto de interesse. É devido ao seu comportamento dinâmico que se tem a definição de "modelos deformáveis" (RODRIGO et al., 2003).

As estimativas foram feitas tomando informações de dimensões (área) de determinados centros de ossificação, por meio de proporções áureas encontradas nos ossos da mão, mais especificamente nos ossos das falanges e do metacarpo, e através da presença da sequência de *Fibonacci* destes ossos.

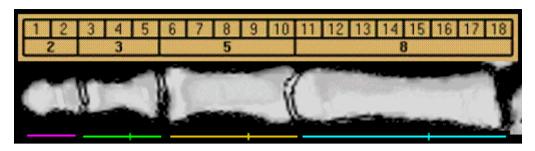
O número de ouro e a sequência de *Fibonacci*, presentes entre as proporções analisadas, serviram para indicar o crescimento correto.

# A estimativa da idade óssea baseada na proporção áurea e na sequência de Fibonacci dos ossos da mão.

As propostas de estimativa da idade óssea baseando-se na proporção áurea envolvem algumas análises que são apresentadas por Olivete Jr (2007), com a utilização dos centros de ossificação formados pelas falanges e metacarpos.

Nesta estimativa analisa-se se o crescimento do paciente está de acordo com a idade cronológica, levando em consideração a presença da proporção áurea entre a soma das falanges.

Para embasar sua pesquisa, Olivete Jr. (2007) cita Torres (1970), que realizou um estudo no qual faz uma descrição da presença da proporção áurea no ser humano, como pode ser visto na Figura 8, que analisa a proporção entre as falanges e metacarpo.



**Figura 8 -** Presença da sequência *Fibonacci* entre as falanges e metacarpo. Fonte: Olivete Jr. (2007).

Olivete Jr. (2007) verificou que os resultados obtidos por todos os métodos de estimativa apresentaram um elevado coeficiente de correlação (próximo de 1.0), entre as idades cronológicas e as idades ósseas estimadas. Estes resultados indicam que todos os métodos utilizados podem ser aplicados no acompanhamento do crescimento ósseo.

A metodologia baseada em área foi desenvolvida devido ao método de E&R, baseado em medidas de comprimento e/ou largura de determinados ossos. As medidas de comprimento de alguns centros de ossificação foram utilizadas para verificar a existência do número de ouro entre eles. Foram geradas três metodologias que proporcionaram várias razões, formadas por comprimentos dos centros de ossificação.

Olivete Jr. (2007), conclui que estas metodologias podem ser utilizadas como um novo padrão para a análise do crescimento ósseo, em pacientes na faixa etária de 6 a 16 anos, com a vantagem de fácil utilização e rápida execução.

Olivete Jr. (2007) avalia, também, que, metodologia baseada na sequência de *Fibonacci* trazer resultados satisfatórios, embora inferiores aos obtidos com as metodologias baseadas em número de ouro e em medidas de área, baseada em contornos ativos, sua principal vantagem é o fato de apresentar resultados precisos e que não se alteram após várias estimativas, bem como uma elevada velocidade de processamento, operando de forma automática e simplificada.

#### 2.4 Considerações a respeito da idade óssea

Este capítulo fez uma revisão da literatura dos métodos tradicionais para estimação de idade óssea, abrangendo as metodologias que sucederam esta pesquisa, a fim de reunir informações que agreguem valor não somente à investigação, mas que possibilitem mais qualidade e velocidade da estimação agindo como veículo propulsor para pesquisas que procuram viabilizar novas metodologias de estimação de idade óssea, de forma simples.

A automação ou semi-automação desses procedimentos, além de evitar estimações que necessitam de exaustivas análises, permite um aumento no nível de precisão e diminui o desgaste dos profissionais da área médica, revertendo o quadro de diagnósticos tardios.

Para que a automação seja uma ferramenta eficiente nesses processos, é necessário que as imagens utilizadas sejam submetidas a um pré-processamento que uniformize contraste e melhore as condições das mesmas para processos de segmentação e análises. No próximo capítulo, serão abordadas algumas técnicas que foram utilizadas como ferramentas de pré-processamento com intuito de amenizar os problemas apresentados pelas imagens carpais.

# 3- Considerações sobre as imagens de raio X.

Uma radiografia é resultado da reprodução da imagem de um objeto em um material fotossensível através do uso dos raios X. Essa reprodução, muitas vezes, apresenta algumas alterações de intensidade dos *pixels*, dependendo dos aparelhos utilizados, ou ainda, devido à pele, ou, mesmo, sobreposição de estruturas. Existem vários problemas que dificultam ou tornam equivocada a análise das imagens reproduzidas muitas vezes; eles estão relacionados à qualidade do filme, processo de formação da imagem, ruídos, ou a problemas inerentes ao processo de radiografia denominado efeito *Heel* (OLIVETE JR, 2005a).

O efeito *Heel* é uma das maiores dificuldades encontradas quando se realiza o processamento de imagens de raios X. Dá-se em virtude da variação não uniforme do contraste de fundo, produzindo uma intensidade de radiação variável, em função da localização espacial, provocando mudança de contraste por toda a imagem (OLIVETE JR et al., 2005b).

Este fenômeno, segundo Fritz & Livingston (1985), é um dos grandes problemas encontrados em radiografias carpais, relacionado diretamente à distribuição de intensidade dos raios X, causando iluminação não uniforme em algumas partes da imagem. Tomando como base imagens radiográficas da mão, a parte inferior dessas imagens é pouco sensibilizada, enquanto a parte superior é sensibilizada em excesso. Essa variação de intensidade não homogênea é facilmente corrigida pelos sistemas de percepção da visão humana, mas dificulta o uso de técnicas de processamento automáticas, pela grande diferença de intensidade em toda imagem (OLIVETE JR, 2005a). Tal fenômeno pode ser observado na Figura 9.



Figura 9 - Radiografia carpal ilustrando o efeito *Heel*. Fonte: Tavano (2001).

## 3.1 - O pré-processamento como ferramenta para gerar informação

Nas diversas aplicações que utilizam imagens de raios X, alguns dos problemas encontrados são a presença de regiões que não correspondem à região de interesse e o efeito *Heel*. Detalhes na visualização da imagem são de grande importância em sistemas médicos, principalmente em algumas imagens de raios X, em que é possível observar pequenas estruturas ósseas que, muitas vezes, são suprimidas por problemas relacionados ao efeito *Heel*, acarretando diagnósticos imprecisos ou, mesmo, o não diagnóstico de problemas. Amenizar ou eliminar tais problemas, garantindo uma melhor qualidade dessas imagens, é a função das várias investigações de pré-processamento.

A uniformização de contraste apresenta-se como ferramenta de preparação dessas imagens para extração de informações como, por exemplo, a segmentação de estruturas anatômicas para auxiliar o diagnóstico computadorizado, tal como instrumentação para diagnóstico. A resolução de problemas como estes passa, sem dúvida, pelo caminho do préprocessamento. Raymundo et al.(2007), desenvolveu dois estudos que foram aplicados às imagens utilizadas nesse trabalho, descritos nas seções seguintes, baseados em duas técnicas que, certamente, agregam grandes contribuições neste sentido.

# 3.1.1 - Metodologia adaptativa de processamento de imagem radiográfica carpal com realce de bordas.

A metodologia adaptada por Raymundo et al. (2007), propõe-se a apresentar uma solução para o processamento de imagens radiográficas através do uso de um algoritmo, dividido em três etapas principais:

- A primeira etapa consiste na seleção automática da região que melhor enquadra toda a estrutura óssea presente na imagem.
- A segunda realiza a transformação de contraste com o objetivo de eliminar estruturas que correspondem a tecido não ósseo.
- Por último, é realizado o realce de bordas para evidenciar os contornos das falanges.

A Figura 10 ilustra o fluxograma do procedimento adotado neste trabalho, de forma mais detalhada.

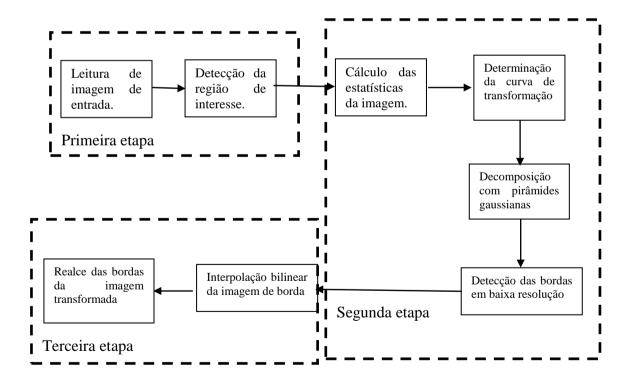


Figura 10 - Fluxograma da metodologia aplicada. Fonte: Raymundo et al. (2007).

# Detecção de região de interesse.

Segundo Raymundo et al. (2007), para evitar influências provocadas por regiões fora da área de interesse, um método de modulação de variância é aplicado para eliminar picos indesejáveis. É realizado um cálculo das projeções e, para delimitar a região de interesse, são encontrados os gradientes mais significantes e opostos no vetor de modulação.

# Transformação de contraste.

A curva de transformação de contraste é criada automaticamente, avaliando o histograma da região selecionada. Por se tratar de imagens de raios X, esse histograma apresenta basicamente dois picos distintos, que representam a região de fundo e a região de tecido e ossos. O vale que separa esses dois picos é calculado através do método *threshold* de Otsu (1979). Esse valor é empregado para definir a curva de transformação da imagem, que é formada utilizando a interpolação *spline* cúbica. Para obter uma curva mais suave, dois pontos adicionais são considerados na etapa de interpolação, sendo estes localizados nos pontos médios dos valores obtidos anteriormente Raymundo et al., 2007.

#### Realce de bordas.

O método de realce de bordas é aplicado na imagem transformada, melhorando a visualização das estruturas ósseas. Para amenizar os efeitos de ruído na imagem, o algoritmo utiliza-se de um filtro passa-baixa gaussiano 3X3. Uma pirâmide gaussiana é aplicada, como solução, para o elevado tempo de processamento na identificação das bordas, reduzindo a imagem em escala de 1:3 e, então executa- se a extração das bordas.

Neste nível de resolução, a detecção de bordas privilegia aquelas que correspondem aos contornos das estruturas ósseas maiores. A imagem resultante desta etapa e que contém as bordas é submetida a uma interpolação bilinear para alcançar novamente o tamanho da imagem original.

A metodologia adaptada mostrou-se eficiente, embora com a necessidade de ajustes nos parâmetros, em virtude de algumas imagens apresentarem transição suave entre tecido e ossos em termos de valores de *pixels*, o que poderia acarretar perda de informações de borda da região de interesse. Apesar das intensidades de *pixels* dos ossos serem semelhantes às de

*pixels* de tecido, o algoritmo é capaz de realizar uma correta distinção. Abaixo, a Figura 11 apresenta o resultado do pré-processamento adaptado por Raymundo et al. (2007).

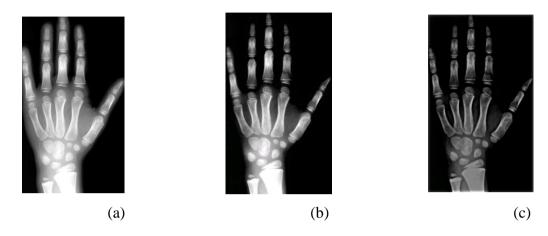


Figura 11 – Resultados: a) imagem original, b) imagem otimizada, c) imagem após realce de bordas e interpolação. Fonte: Raymundo et al. (2007).

# 3.1.2 - Metodologia de pré-processamento utilizando o algoritmo retinex em imagens radiográficas carpais.

A metodologia proposta por Land & Mccann (1971), como solução para os problemas de contraste presentes nas imagens, mostrou-se viável para as imagens em escala de cinza do banco de imagens da EESC. Esta metodologia é inspirada no mecanismo biológico de adaptação visual do olho humano e tem como propósito original a constância das cores; porém, apresentou eficiente desempenho no realce de contraste (Baliga, 2004). O algoritmo de processamento de imagens *SSR* (*Single-Scale Retinex*) tem por objetivo melhorar a qualidade das imagens tal como faz o olho humano, razão pela qual é analisado neste trabalho (Baliga, 2004).

Segundo Hines et al. (2004), o algoritmo foi aplicado com sucesso na melhoria da aparência de imagens em diversas áreas, tais como a radiografia médica, a fotografia subaquática e as investigações judiciais.

Segundo Olivete Jr. et al. (2006c), aplicando-se métodos de segmentação clássicos sobre as imagens carpais, verifica-se que as mesmas apresentam dificuldades marcantes na etapa de binarização, devido aos diferentes níveis de intensidade presentes. Estudos a respeito da viabilidade de algoritmos de segmentação para este tipo de imagem são importantes para que os processos se tornem abordagens simples e eficientes. A melhoria da qualidade dessas

imagens com metodologias como o *SSR* traz benefícios para o processo de identificação das regiões de interesse, auxiliando diretamente o procedimento de segmentação que é essencial para análise, identificação e extração de características marcantes em uma imagem.

## O algoritmo Single Scale Retinex.

A metodologia de Land & Mccann (1971), quando aplicada na fase de préprocessamento, apresenta uma solução para o problema de contraste das imagens. Neste caso específico de análise, utiliza-se o *Single-Scale Retinex (SSR)* que realiza operações relacionadas com a função neurofisiológica dos neurônios individuais da retina, núcleos de geniculação lateral e córtex cerebral (RAYMUNDO & RODRIGUES, 2008).

Após o pré-processamento com o algoritmo *Retinex*, a imagem foi submetida a uma melhoria de contraste para realçar os resultados.

#### Resultados.

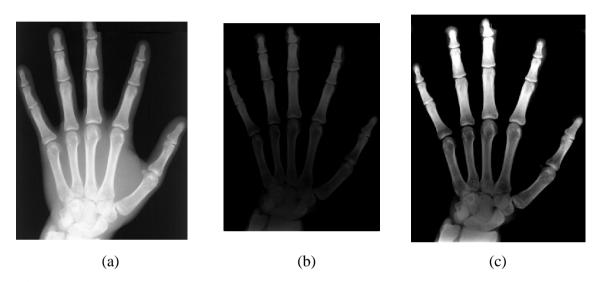
Os resultados da metodologia de Land & Mccann (1971), foram submetidos à função *strechlim*, do *toolbox* de processamento de imagens do Matlab®, que calcula o histograma da imagem e determina os limites de ajuste automaticamente. A função *strechlim* retorna estes valores como frações em um vetor, em que podem ser determinados os valores para a mais alta e mais baixa intensidade. Ao cortar os extremos em ambos os lados do intervalo da escala de intensidade, o *strechlim* cria mais espaço numa faixa dinâmica para ajustar as intensidades restantes. Outros valores podem ser especificados como limites de escala para esta função. Neste estudo foi utilizado o *strechlim* padrão (RAYMUNDO & RODRIGUES, 2008).

Adicionando esta função aos procedimentos da metodologia de Land & Mccann (1971), obteve-se resultados superiores à metodologia adaptada.

Os resultados apresentados com relação à metodologia adaptada, são melhores pois não há necessidade de ajustes durante o pré-processamento. Quanto à imagem final alcançada, os resultados também foram animadores, apresentaram uma superioridade em torno de 40% com relação à remoção de tecido (RAYMUNDO & RODRIGUES, 2008).

É importante observar que a excelência da metodologia de Land & Mccann (1971) pode ser, facilmente, comprovada por uma inspeção visual direta nas imagens, conforme aponta a Figura 12.

Pontos correspondentes às estruturas externas (tecido) são suprimidos, enquanto que os *pixels* pertencentes aos ossos são preservados. Há maior contraste em relação aos *pixels* dos ossos, tornando as falanges mais definidas. Apesar das intensidades de *pixels* dos ossos serem, muitas vezes, semelhantes às de *pixels* de tecido, uma distinção melhor entre as áreas é realizada.



**Figura 12** - Resultados de acordo com Land & Mccann (1971) (a) imagem original, (b) imagem após uso do *RSS*, (c) imagem após uso da função *strechlim*. Fonte: Raymundo & Rodrigues (2008).

Além da análise visual indicar claramente as diferenças entre as metodologias, foram realizadas operações de intersecções entre imagem original e imagens processadas em que se comprovou que a metodologia de Land & Mccann (1971) preserva melhor os centros de ossificação e elimina, quase por completo, o tecido em torno dos mesmos. Os resultados apontaram uma superioridade de desempenho em torno de 40% para a metodologia de Land & Mccann (1971), em termos de área de remoção de tecido (RAYMUNDO & RODRIGUES ,2008).

### 3.2 - Avaliação do processo de pré-processamento.

Como solução para problemas de iluminação em imagem de raios X de mão e punho, o pré-processamento tem-se mostrado uma etapa de grande importância, pois melhora a região de interesse da imagem e serve de base para o sucesso do processo de segmentação. Portanto, ações que tenham como objetivo suprimir as deficiências causadas pelo efeito *Heel* nas

imagens, são bem-vindas em qualquer pesquisa voltada ao uso de imagens radiográficas. É também o propósito, a investigação de procedimentos que atendam às necessidades de processos de segmentação e contribuam positivamente, como etapa importante em sistemas, que buscam melhorar a qualidade de imagens que apresentam problemas de iluminação.

O pré-processamento das imagens é crucial, pois resulta numa melhor segmentação da região de interesse, qualificando-se para ser utilizado em imagens de raios X da mão que se utilizem de processos de segmentação.

A segmentação é um processo essencial para análise e identificação de características marcantes em uma imagem (TRUCCO & VERRI, 1998); para tanto, metodologias de detecção de bordas apresentam-se como excelente opção para os processos de segmentação. Porém, essas metodologias, possuem a característica de localização por regiões, nas imagens em que os níveis de cinza dos *pixels* variam bruscamente. Tal procedimento torna-se impraticável, pois, num determinado *pixel*, a intensidade é uma, enquanto os *pixels* de sua vizinhança podem possuir intensidades bem diferentes, conhecidas como "descontinuidades"; elas fazem parte do contorno dos objetos (XU & PRICE, 1997).

# 3.3 - Conclusões a respeito da aplicação do pre-processamento em imagens carpais

A aplicação dessas metodologias em imagens carpais apontou resultados mais precisos e contribuiu de forma significativa para a estimação de idade óssea baseada na extração de bordas.

Os resultados apresentados favorecem a aplicação de algoritmos de segmentação de região por apresentarem maior contraste entre *pixels* de osso e de fundo. Essa metodologia amplia o leque de possibilidades não apenas no que diz respeito ao pré-processamento, mas também contribui para a área de extração de características; especialmente a identificação das "bordas dos ossos" tem grande importância.

O pré-processamento teve uma importante contribuição na criação da base de dados *CarpEven*, pois as imagens necessitaram de segmentação completa.

Mas todo esse trabalho seria inútil, se os dados não fossem armazenados de forma adequada e segura que permitisse acesso fácil e rápido; daí a necessidade de se utilizar um sistema gerenciador de banco de dados que pudesse suportar um volume robusto de informações e garantisse, além de confiabilidade, acessibilidade e um armazenamento seguro. No capítulo seguinte, é feita uma breve abordagem a respeito do servidor de banco de dados utilizado, as vantagens apresentadas pelo mesmo e as justificativas de sua escolha.

#### 4 -O armazenamento de dados.

Considerando que, na área de pesquisa, os dados são o item mais importante a ser tratado, os pesquisadores tendem a investigar as melhores formas de acessá-los e armazená-los e investigam, portanto, as melhores soluções disponíveis no mercado.

Segundo Santos (2005), os sistemas de dados informatizados têm como requisitos: atualidade, correção, relevância, disponibilidade e legibilidade, pois as bases de informações, atuais, permitem decisões acertadas. Eles devem fornecer informações efetivas, conter mecanismos de atualização que permitam correção dos dados, eles devem ser filtrados de modo a apresentar apenas o que é relevante.

Se a informação não está disponível quando é necessária, ela não tem utilidade. Além disso, a mesma deve ser apresentada num formato legível e de fácil interpretação para o utilizador.

Os *SGBDs* (Sistemas Gerenciadores de Banco de Dados) representam uma das melhores formas de armazenamento de dados, pois não trabalham com armazenamento em base de dados orientada a registros.

O amadurecimento dos sistemas legados e o crescimento da quantidade de informação gerada fazem com que a utilização de um *SGBD* seja necessária, aproveitando, assim, todas as suas vantagens, como, por exemplo: a acessibilidade por meio de várias ferramentas; tempo de resposta; segurança; sistema de tolerância a falhas; sistemas distribuídos; replicação de informações; estratégias consistentes de cópias de segurança (*backup*), restauração (*recovery*), entre outras (WILDEROM E FRANK, 2002).

#### 4.1 - A aplicação de sistemas de gerenciamento de dados

Investigar a existência de novas relações em imagens de raio X da mão esquerda para uma melhor estimação de idade óssea é uma necessidade constante no grupo de pesquisa de Visão Computacional da EESC.

Para atender essa necessidade, as imagens da base de dados da EESC foram todas préprocessadas e segmentadas, para que fossem extraídas as medidas de área, perímetro e comprimento de todas as falanges e metacarpos dos dedos. Como o volume de dados gerado resultou em algo robusto, a melhor alternativa para armazená-los foi utilizar um sistema de gerenciamento de banco de dados que fosse conhecido da comunidade científica e de fácil utilização para os profissionais da área médica. A partir dessa premissa, a base métrica *CarpEven* foi gerada, utilizando as vantagens oferecidas pelos *SGBDs*. Por se tratar de uma base de pesquisas com um grande volume de dados métricos, ela foi desenvolvida em *MySQL*, uma linguagem de uso gratuito, facilmente disponível na *WEB* para *downloads*.

## O Banco de Dados MySQL

Segundo Cox Jr. (2000), o *MySQL* é um servidor de banco de dados multiusuário, multitarefa que trabalha com uma das linguagens de manipulação de dados mais populares do mundo.

O SQL (Structured Query Language) é uma linguagem simples, em que se pode gravar, alterar e recuperar informações num web site com segurança e rapidez. Foi desenvolvido pelo Departamento de Pesquisas da IBM como forma de interface para o Sistema de Banco de Dados Relacionais **SYSTEM R**, no início dos anos 70. Em 1996, a ANSI (American National Standards Institute) publicou um padrão SQL. Desta forma, ela estabeleceu-se como linguagem padrão de Banco de Dados Relacional.

A linguagem tem como grande vantagem sua capacidade de gerenciar índices sem a necessidade de controle individualizado, algo muito comum nos Sistemas Gerenciadores de Arquivos.

A versão utilizada do *MySQL* neste trabalho é a *mysql-essential-5.0.51a-win32* e a *mysql-gui-tools-5.0-r12-win32*. Esta última, na realidade, é um complemento do software, pois o mesmo habilita a opção de interface gráfica e disponibiliza uma janela com todas as opções do *MySQL*, não exigindo do usuário grande habilidade para utilizá-la.

## 4.2 - Justificativa da aplicação do Banco de dados

A aplicação das informações ao banco de dados acrescenta possibilidades aos objetivos da pesquisa, que apresenta uma nova metodologia para estimação da idade óssea, baseada no processamento de imagens com uso de suas métricas. Esta nova metodologia explora a extração de características das imagens, utilizando informações dimensionais das regiões de interesse, além de concentrar e disponibilizar tais informações, via *Web*, para a comunidade científica, que realiza pesquisas na área de estimativa da idade óssea e também

para profissionais (médicos, dentistas) que utilizam esse tipo de informação no diagnóstico de patologias relacionadas ao crescimento humano. Ainda, amplia o contexto de informações sobre as metodologias clássicas e suas variantes, vislumbrando novos horizontes para as metodologias sobre estimação da idade óssea (carpal, cefalométrica, etc) e crescimento humano.

## 4.3 -O processo de formação do banco de dados.

O banco de dados com informações métricas, extraídas das imagens carpais da EESC"CarpEven", foi criado com uma estrutura baseada nas metodologias existentes e elaborado
com auxílio do software Mipav (Medical Image Processing, Analysis and Visualization), de
Mcauliffe (2007). Compreende medidas de comprimento, área e perímetro das falanges dos
dedos da mão esquerda, cujas imagens foram pré-processadas e segmentadas. Para obter o
comprimento horizontal, traça-se uma linha para calcular o tamanho dos ossos.

Na etapa de segmentação, cada uma das falanges e metacárpicos dos dedos foram segmentadas manualmente, com ajuda do software *Mipav*. Após a segmentação, o software calculou o valor da área e do perímetro das regiões de interesse. As figuras 13 e 14, a seguir, mostram como as imagens foram demarcadas.



**Figura 13 -** Imagem das falanges e metacarpos segmentados, mão feminina. Fonte : Raymundo (2008).

O comprimento dos ossos, também, foi calculado com ajuda do software *Mipav*. A Figura 14 ilustra como foram feitas as retas sobre a imagem.



**Figura 14** - Imagem das falanges e metacárpos com o traçado de comprimento, mão feminina. Fonte: Raymundo (2008).

Finalmente, todas as informações métricas foram transferidas para a base de métricas "CarpEven", na qual cada atributo significa uma medida, iniciando pelo dedo mínimo e terminando com dedo polegar. A identificação de cada uma das mãos é única, baseada numa simplificação do número de identificação das imagens utilizadas, chamada atributo id\_mao. Para facilitar consultas à base e não gerar dúvidas quanto à imagem geratriz das informações, criou-se um atributo id\_imagem, que traz a identificação exata da imagem no banco. O nome de cada atributo da base é formado pela letra que define a métrica utilizada: "a" foi designado para área, "p" para perímetro e "t" para o comprimento. A cada um dos dedos foram designados quatro atributos referentes às falanges e metacarpo; portanto, as quatro primeiras informações da base são do dedo mínimo, respectivamente os atributos, atr1, atr2, atr3 e atr4. Para identificar cada uma das falanges ou metacarpos foram utilizadas as definições: "fd"

para falange distal, "fm" para falange média, "fp" para falange proximal e "m" para metacarpo.

Para que os atributos não ficassem com a nomenclatura muito longa, os nomes dos dedos foram substituídos por suas abreviações: "min" para mínimo, "anu" para anular, "med" para médio, "ind" para indicador e "pol" para polegar.

O conjunto de todas essas definições gerou a nomenclatura dos atributos da base; sendo assim, têm-se:

Exemplo do atributo área da falange distal do dedo mínimo, "a\_atr1\_fdmin", para o perímetro do mesmo dedo, apenas a letra "a" é substituída por "p", "p\_atr1\_fdmin" e no caso do tamanho, a letra "t" é a utilizada "t\_atr1\_fdmin".

Conforme apresentado na Figura 15, os atributos estão dispostos na base de dados da seguinte forma:



**Figura 15 -** Nomenclatura de atributos da base de métricas "*CarpEven*", Fonte: Raymundo (2008).

## 4.4 - Considerações a respeito da base de dados métrica.

A criação da base de dados com informações métricas, "CarpEven", representa a possibilidade de agregar novas formas de avaliação às metodologias existentes, pois acrescentam às pesquisas da área de estimação de idade óssea, informações da população brasileira de região determinada, a respeito do crescimento e de suas características em cada uma das faixas etárias. A base de informações carpais atual vem sendo complementada graças ao surgimento de novas pesquisas, o que torna a mesma mais versátil e permite que a quantidade de informações seja sempre crescente. Essas informações poderão ser amplamente utilizadas por estudos que venham, não só enriquecer a principal função desta pesquisa, mas desbravar novos caminhos e auxiliar o trabalho da comunidade médica que busca por procedimentos que colaborem com o diagnóstico, de forma precisa e tempo reduzido na análise.

A base métrica armazena dados que foram coletados das imagens. O desafio dessa nova metodologia é não só contribuir com melhorias no processo de estimação de idade óssea, mas ser capaz de avaliar se alguns ou vários atributos associados podem estimar a idade óssea. Para que tais informações possam ser extraídas dos dados coletados e contribuam valiosamente com a pesquisa, os mesmos foram aplicados a processos de descoberta de conhecimento, existentes na literatura, que serão abordados no próximo capítulo.

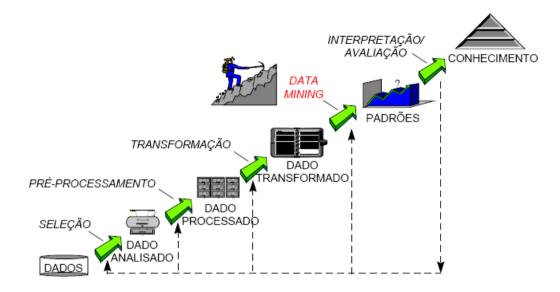
#### 5 - Processo de descoberta de conhecimento em bases de dados.

Antigamente, tanto em relação ao hardware, como em relação ao software, a tecnologia era limitada e permitia armazenamento de um pequeno volume de dados. Consultas simples eram suficientes para a análise dos dados, mas os avanços tecnológicos na forma de armazenamento, processamento de grandes volumes de dados e a necessidade de conhecer e entender a base de dados tornou-se cada vez maior (FERRO & LEE, 2001).

Análises manuais tornaram-se impraticáveis, métodos eficientes para a análise computacional dos dados tornaram-se indispensáveis, bem como a extração das informações e padrões que existem nesses dados. A lacuna que existe entre a geração de grandes volumes de dados, a compreensão e a extração de novos conhecimentos sobre estes vem crescendo em todas as áreas da atividade humana, e suprimir essa lacuna é crucial, especialmente, nas áreas médicas. Desse modo, é de fundamental importância a aplicação de métodos que colaborem no processo de tomada de decisão e descoberta de conhecimento médico específico, (diagnósticos, prognósticos, monitoramento, etc), sobre um único paciente ou de um conjunto de casos de pacientes distintos (FERRO&LEE, 2001).

Com o objetivo de suprir essas necessidades, surgiram os processos de *KDD*, (*Knowledge Discovery in Database*) (FERRO & LEE, 2001), que envolvem a preparação dos dados, seleção, limpeza, *data mining*, incorporação de conhecimento anterior apropriado e interpretação formal dos resultados de mineração. Isso permite que o conhecimento derivado dos dados possa ser usado para auxiliar nas tomadas de decisões (FAYYAD et al.,1996).

A Figura 16, a seguir, ilustra as diversas fases do *KDD*:



**Figura 16**- Etapas do processo de extração de conhecimento de dados. Fonte: Fayyad et al. (2002).

A finalidade da *KDD* é extrair padrões dos dados, sendo a fase de extração considerada central no processo, pois se preocupa com a busca de padrões nos dados e com o ajuste de modelos ou determinação de padrões a partir dos dados observados (SOUZA & ROCHA, 2005).

Segundo Lima (2008), o KDD possui alguns elementos essenciais; aqui elencados:

- Todas as soluções lidam com grande número de dados.
- É necessário eficiência frente ao volume de dados.
- A acurácia é um elemento essencial.
- Requerem uso de linguagens de alto-nível.
- Utilizam alguma forma de aprendizado automático.
- Produzem algum resultado interessante.

O termo *KDD* refere-se ao processo completo de descoberta de conhecimento em dados e envolve diversas fases e tarefas. A Figura 17, a seguir, apresenta uma visão hierárquica do processo de *KDD*, suas fases e tarefas.

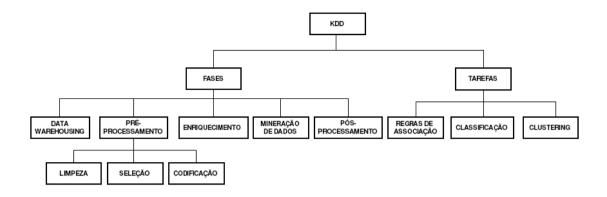


Figura 17- Visão hierárquica do processo de KDD. Fonte: Souza &Rocha (2005).

A mineração de dados é um dos principais passos no processo *KDD* e pode ser definida como um processo de descoberta de padrões nos dados. Pode ser automático ou, geralmente, semi-automático (Ferro & Lee, 2001).

Segundo Garcia (2008), as tarefas de mineração representam uma especificação do que se deseja buscar nos dados. Elas podem extrair diferentes tipos de conhecimento, sendo necessário decidir, no início do processo de mineração, qual tipo de conhecimento o algoritmo deve extrair.

Tais tarefas classificam-se em:

- Atividade Preditiva: busca identificar a classe de uma nova amostra de dados a partir do conhecimento adquirido de um conjunto de amostras com classes conhecidas.
- Atividade Descritiva: trabalha com um conjunto de dados que não possuem uma classe determinada, buscando identificar padrões de comportamento comuns nestes dados (GARCIA, 2008).

Cada uma das atividades desempenha tarefas relacionadas a seus processos de resolução de problemas. Atividades preditivas trabalham com tarefas de regressão e classificação, cuja classificação consiste no processo de encontrar um conjunto de modelos (funções) que descrevam ou diferenciem classes de dados ou conceitos. A regressão é utilizada para predizer valores de dados numéricos não disponíveis nos bancos de dados e se assemelha à tarefa de classificação; no entanto, trabalha com atributos contínuos, ou seja, atributos que possuem uma infinidade de valores (GARCIA, 2008).

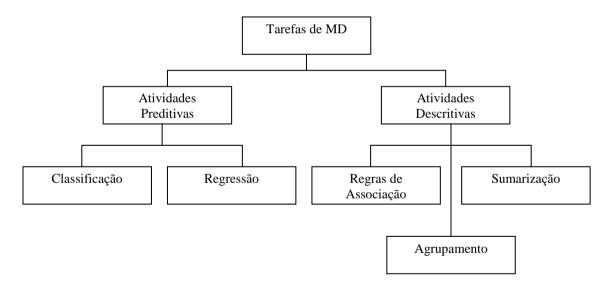
Por outro lado, as atividades descritivas utilizam-se de tarefas de regras de associação, agrupamento e sumarização para resolver os problemas. As tarefa de regras de associação estabelecem uma relação estatística entre os dados, em que uma regra de associação é da forma:  $X \rightarrow Y$ ,  $X \in Y \subset I$ ,  $X \cap Y = \emptyset$ .

- Suporte: porcentagem que  $X \cup Y$  ocorre em D.
- Confiança: porcentagem de transações em D contendo X que também contém Y.

Onde I é o conjunto de itens e D é o conjunto de transações do banco de dados, na qual cada transação  $T \subseteq I$ .

Para a tarefa de agrupamento, os registros são agrupados de acordo com uma semelhança, não havendo classes anteriormente definidas; e esta característica o diferencia do agrupamento da tarefa de classificação. Registros dentro de um determinado grupo possuem um certo grau de semelhança e, comparando-se um registro de um grupo com o de outro, a semelhança é mínima ou nula.

A tarefa de sumarização envolve métodos para encontrar uma descrição compacta de um subconjunto de dados. A Figura 18 apresenta um esquema do papel de cada uma das atividades no processo de mineração e o tipo de tarefa desempenhado por elas, individualmente (GARCIA, 2008).



**Figura 18**– Papel das atividades preditiva e descritiva na mineração de dados. Fonte: Garcia (2008).

## 5.1 - O papel dos processos de KDD quando aplicados à base de métricas.

Dados compõem o conhecimento; sendo assim, são essenciais para qualquer organização, e seu acúmulo se tornou relativamente fácil com a tecnologia. O dado é algo

bruto, é a matéria-prima da qual se pode extrair informação, que é o dado processado com significado e contexto bem definidos. O computador, em essência, transforma dados em informações. O uso das informações de forma inteligente gera o conhecimento. Dessa forma, a qualidade da informação sustenta o conhecimento (SOUZA & ROCHA, 2005).

O processo não trivial de identificação de novos padrões válidos, potencialmente, úteis e compreensíveis em conjuntos de dados refere-se ao *KDD*.

A descoberta de conhecimento em banco de dados tem ampla aplicação prática em diversos domínios da ciência, principalmente aplicações que envolvem a tecnologia da informação.

Descobrir informações presentes no banco de métricas, criado a partir das imagens da EESC, e classificá-las, é o intuito desta pesquisa.

Baseando-se nos processos *KDD* e utilizando ferramentas de mineração de dados, a fim de encontrar informações relevantes que possam ser empregadas para determinar a idade óssea, utilizou-se a base de dados *CarpEven*, aplicada ao software *Weka*( *Waikato Environment Knowledge Analysis*) e ao algoritmo *StARMiner* (*Statistical Association Rule Miner*).

Os resultados obtidos em cada uma das aplicações foram testados e avaliados, tomando como base pesquisas anteriores. A comparação avalia o número de atributos encontrados, ressalta a eficácia dos testes, reconhece qual conjunto é o mais adequado às necessidades da pesquisa e analisa se ambos conseguiram desempenho semelhante no processo de mineração. Por fim, foi feita uma análise de quão precisa a estimação de idade óssea pode se tornar com o uso das características encontradas.

## 5.1.1 - O Weka

O Weka, (Waikato Environment Knowledge Analysis) é uma ferramenta de mineração de dados com crescente utilização. Encontra-se disponível na web², além de sua facilidade de instalação. A implementação feita em Java torna o sistema portável (ALMEIDA et al. 2003).

Segundo Witten & Frank (2005), o *Weka* fornece implementações de algoritmos de aprendizagem que podem facilmente ser aplicáveis a um conjunto de dados, incluindo uma variedade de ferramentas para transformá-los, tais como os algoritmos para discretização.

<sup>&</sup>lt;sup>2</sup> (http://www.cs.waikato.ac.nz/ml/weka/)

Pode pré-processar um conjunto de dados em um regime de aprendizagem, analisar o resultado de determinado classificador e seu desempenho sem qualquer restrição ao tipo, programa ou código escrito. Inclui métodos padrão para todos os problemas de mineração de dados; porém, conhecer os tipos de dados é parte integrante do trabalho. Fornece vários tipos de visualização de dados e prévias das ferramentas. Todos os algoritmos são exibidos em forma de tabela única, relacionada no formato *ARFF* (*Attribute Relationship File Format*), e podem ser lidos ou gerados por consulta em forma de banco de dados.

# Formas de utilização.

Uma das formas de utilizar o *Weka* é aplicar um método de aprendizagem e analisar o conjunto de dados resultante para saber mais sobre os dados. Outra, é a utilização de modelos de aprendizagem; a terceira forma é a aplicação de vários algoritmos e comparação de seu desempenho, a fim de escolher um para predição (WITTEN & FRANK, 2005).

Os métodos de aprendizagem são chamados classificadores. As ferramentas de previsão para os dados são chamadas filtros. Tais como os classificadores, os filtros são selecionados a partir de um menu e sempre devem ser adequados aos requisitos. Existem diferentes tipos de filtros que podem ser utilizados. O *Weka* inclui, também, implementações de algoritmos de regras de associação que agregam dados, quando nenhum valor de classe for especificado, além de selecionar os atributos relevantes (WITTEN & FRANK, 2005).

Ele também dispõe de métodos de meta aprendizagem, utilizados para a construção de conjuntos de classificadores. Dentre os métodos disponíveis destacam-se: *Bagging* e *Boosting*. O método *Bagging* constrói os classificadores a partir de sucessivos e independentes conjuntos de amostras de dados. Estas amostras são geradas a partir de um conjunto de dados de treinamento, em que, randomicamente, são extraídas "n" instâncias com substituição a partir do conjunto original, ou seja, ocorrendo a repetição de instâncias nas amostras. Já no método *Boosting*, cada instância de treinamento tem um certo peso associado, ao induzir o primeiro classificador, e todas as instâncias são equiprováveis, isto é, têm o mesmo peso (Almeida et al., 2003). Após a redução do primeiro classificador, os pesos das instâncias de treinamento, classificadas incorretamente, são alterados com base nos classificadores que foram construídos anteriormente (WITTEN & FRANK, 1999).

## Interfaces gráficas.

A forma mais fácil de utilização do software é por meio de uma interface gráfica chamada *Explorer*, que permite acesso a todas as suas instalações usando o menu de seleção. A fundamental desvantagem do Explorer é que, quando se abre um conjunto de dados, ele o detém na memória principal; isso significa que ele só pode ser aplicado a dados de pequena e média dimensão (WITTEN & FRANK, 2005).

Existem, ainda, outros dois tipos de interfaces gráficas; elas fazem parte do *Weka* incremental e contêm alguns algoritmos que podem ser usados para processar grandes volumes de dados.

A interface *Knowedge Flow* permite que caixas representem os algoritmos e sejam adicionadas a fontes de dados ao redor da tela; elas podem ser unificadas para a configuração desejada. A interface *Experimenter*, projetada para auxiliar em questões básicas na aplicação de técnicas de classificação e regressão, indica quais métodos e valores funcionam melhor para determinado problema.

O *Weka* pode ser, ainda, empregado para realizar processamento com distribuição de carga em várias máquinas usando o *RMI* (*Remote Method Invocation*).

O Weka é, portanto, uma ferramenta para mineração de dados de interface amigável, que agrega um conjunto de algoritmos, todos implementados em JAVA (WITTEN & FRANK 2005; BASU & MELVILLE, 2005), cuja aplicação pode acessar dados oriundos de bancos de dados via JDBC (Java Database Connectivity) ou através da chamada de arquivos de dados próprios.

Sendo assim, todas estas características fazem da ferramenta desenvolvida por pesquisadores da Universidade de Waikato, na Nova Zelândia, uma das ferramentas mais populares entre a comunidade de mineração de dados (VALENTIM et al., 2005).

#### 5.1.2 - O StARMiner

Segundo Ribeiro, (2008), o *StARMiner* (*Statistical Association Rule Miner*) é um algoritmo estatístico de regra de associação para mineração. Ele executa a seleção de características através da *Rule Minning* (associação estatística). O algoritmo foi desenvolvido para associar características de baixo nível das imagens com o seu significado semântico, sendo também utilizado para realizar seleção de características em bases de imagens médicas, melhorando a

precisão dos sistemas *CBIR* (*Content-Based Image Retrieval*). Atua na tarefa de redução de dimensionalidade dos vetores de características, lidando com a alta dimensionalidade que degrada a recuperação de imagens por conteúdo.

O algoritmo usa medidas estatísticas para descrever o comportamento das características, considerando as categorias (classes) das imagens para encontrar regras de associação representativas. As regras encontradas pelo algoritmo *StARMiner* também são aplicadas para ponderar características dos vetores e para melhorar a precisão das consultas (RIBEIRO, 2008).

A regra de associação estatística tem por função mostrar relações interessantes entre subconjuntos de dados, com base na distribuição dos valores quantitativos. O motivo de se trabalhar com regra de associação estatística é que não exige discretização de dados; apenas das classes, quando essas forem numéricas e contínuas.

De acordo com Ribeiro (2008), um processo de discretização, muitas vezes, leva a uma perda de informação e pode distorcer os resultados de um algoritmo de mineração. As regras de associação representam uma das tarefas mais importantes na mineração de dados; elas têm sido amplamente estudadas e aplicadas.

O algoritmo visa encontrar regras de associação que selecionem o conjunto mínimo de características e preservem a habilidade de diferenciar imagens de acordo com suas categorias.

Seja  $x_j$  uma categoria (classe) e  $f_i$  uma característica (atributo) de uma imagem. As regras retornadas pelo algoritmo StARMiner têm o formato:

$$x_i \rightarrow f_i$$

na qual o antecessor da regra indica um subconjunto de imagens que pertencem à categoria  $x_j$ , e o sucessor da regra  $f_i$  é uma característica que tem um comportamento diferente em imagens da categoria  $x_j$  em relação às demais imagens da base de dados. O algoritmo *StARMiner* somente retorna regras que satisfazem as duas condições a seguir:

- A característica  $f_i$  deve ter um comportamento em imagens da categoria  $x_j$  diferente do seu comportamento em imagens das demais categorias da base.
- A característica  $f_i$  deve apresentar um comportamento uniforme nas imagens da categoria  $x_i$ .

As condições acima são implementadas no algoritmo *StARMiner* através da incorporação de restrições de interesse para a mineração das regras. Essas restrições de interesse são descritas a seguir, por Ribeiro (2008):

- Seja T uma base de imagens médicas,  $x_j$  uma categoria de uma imagem,  $Tx_j \in T$  o subconjunto de imagens da categoria  $x_j$ ,  $f_i$  a i-ésima característica do vetor de características F, e  $fi_k$  o valor da característica  $f_i$  na imagem k.
- Sejam  $\mu f_i(V)$  e  $\sigma f_i(V)$ , respectivamente, a média e o desvio padrão dos valores da característica  $f_i$  no subconjunto de imagens V.

O algoritmo utiliza ainda três limiares definidos pelo usuário:

- $\Delta\mu_{min}$  a mínima diferença das médias da característica fi entre as imagens da categoria  $x_i$  e as demais imagens da base;
- σ<sub>max</sub> o máximo desvio padrão permitido da característica f<sub>i</sub> em imagens da categoria
   x<sub>i</sub>;
- $\gamma_{min}$  a mínima confiança para rejeitar a hipótese H0, de que são iguais, estatisticamente, às médias dos valores de  $f_i$  nos conjuntos  $Tx_j$  (imagens da categoria  $x_i$ ) e  $T Tx_i$  (imagens das demais categorias).

O StARMiner minera regras da forma  $x_j \rightarrow f_i$ , se as condições fornecidas nas equações 1, 2, 3, 4, a seguir, forem satisfeitas:

$$\mu F_i(V) = \frac{\sum_{k \in V} (F_{i_k})}{|V|} \tag{1}$$

$$\sigma F_{i}(V) = \sqrt{\frac{\sum_{k \in V} ((F_{ik}) - \mu F_{i}(V))^{2}}{|V|}}$$
(2)

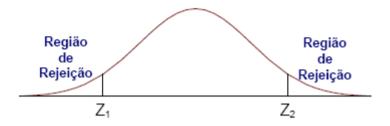
$$\left| \mu F_i(T_{xj}) - \mu F_i(T - T_{xj}) \right| \ge \Delta \mu_{\min} \tag{3}$$

$$\sigma F_i(T_{x_i}) \le \sigma_{\max} \tag{4}$$

Rejeição da Hipótese  $H0: \mu F_i(T_{x_j}) = \mu F_i(T - T_{x_j})$ : a hipótese H0 deve ser rejeitada com a confiança igual ou maior do que  $\gamma_{min}$ , em favor da hipótese de que as médias  $\mu f_i(Tx_j)$  e  $\mu f_i(T - Tx_i)$  são estatisticamente diferentes.

Para rejeitar H0 com confiança  $\gamma_{min}$ , o valor de Z, calculado usando a equação 5 abaixo, deve estar na região de rejeição ilustrada na Figura 19. Os valores críticos de Z, que são Z1 e Z2, dependem do valor de  $\gamma_{min}$  como é apresentado na Tabela 1, (RIBEIRO,2008).

$$Z = \frac{\mu F_i(T_x) - \mu F_i(T - T_x)}{\frac{\sigma F_i(T_x)}{\sqrt{|T_x|}}}$$
 (5)



**Figura 19** – Ilustração das regiões de rejeição para teste de uma hipótese. Fonte: Ribeiro (2008).

**Tabela 1** - Valores críticos de Z em relação aos valores de  $\gamma_{\min}$ . Fonte: Ribeiro (2008).

${\gamma}_{ m min}$	0.9	0.95	0.99
$Z_1$	-1.64	-1.96	-2.58
$Z_2$	1.64	1.96	2.58

O algoritmo *StARMiner*, também, fornece informações sobre o comportamento em função das regras mineradas. Uma regra mineirada pelo *StARMiner*, em sua forma completa, é:

$$x \to F_i, \mu F_i(T_x), \mu F_i(T - T_x), \sigma F_i(T_x), \sigma F_i(T - T_x),$$

na qual,

 $\mu F_i(T_x)$  e o  $\sigma F_i(T_x)$  são, respectivamente, a média e o desvio-padrão dos valores de  $F_i$  nas imagens da categoria x;  $\mu F_i(T-T_x)$  e o  $\sigma F_i(T-T_x)$  são, respectivamente, a média e o desvio-padrão dos valores " $F_i$ " nas imagens que não são da categoria "x".

O algoritmo *StARMiner* torna possível encontrar atributos que categorizam as imagens, ou seja, o algoritmo assinala as funcionalidades com alto poder de diferenciação das categorias da imagem, uma vez que têm um comportamento uniforme e particular em determinada categoria. Isso é importante, porque as características (ou atributos) que

apresentam um comportamento uniforme para cada imagem do conjunto de dados, independentemente da categoria da imagem, não contribuem para categorizá-la e devem ser eliminadas (Ribeiro, 2008).

A complexidade do algoritmo, segundo Ribeiro(2008), é dada pela função  $\Theta$  (ckN), na qual "N" é o número de instâncias do conjunto de dados, "k" é o número de funcionalidades e "c" é o número de categorias.

O *StARMiner*, em resumo, é um algoritmo de extração de característica da imagem que pode ser dividido em treinamento do conjunto, definição de conjunto e teste. No treinamento do conjunto, o algoritmo executa o processo de seleção, selecionando o conjunto. O teste é usado para definir o método de avaliação do conteúdo com base em consultas (RIBEIRO, 2008).

# 5.2 - Importância da aplicação de processos de mineração nos dados.

Várias mudanças surgiram com a tecnologia e uma delas se refere ao aumento da importância da informação no contexto científico. A informação tornou-se um grande diferencial, auxiliando desde a tomada de decisão até a descoberta de fraudes ou perfis de consumidores (FAYYAD et al., 1996). No entanto, para que estas informações fossem extraídas corretamente, fez-se necessária a utilização de técnicas e ferramentas que propiciem a descoberta ou mineração de padrões (MITTCHEL, 1997).

A experiência mostra que nenhum algoritmo de aprendizagem de máquina é único, pois existe a necessidade de adequá-lo aos problemas de mineração. A mineração de dados é uma ciência experimental, de modo que permite a comparação e/ou utilização de diferentes métodos, visando à identificação dos mais adequados para o problema (WITTEN & FRANK, 2005).

A máquina, na maioria das aplicações de mineração de dados, é apenas uma componente da aprendizagem, pequena parte de um sistema de software; esse, sim, de suma importância, pois é quem define o resultado final a ser aplicado aos dados.

Encontrar características que apontem para uma estimação de idade óssea acurada, baseada na premissa da experimentação, ou seja, da investigação de qual tarefa de mineração resultará em características relevantes, é o objetivo. Para tanto, a extração de dados métricos das imagens carpais e a posterior criação da base de dados "CarpEven" foi de fundamental importância, pois os dados pertencentes a essa base são avaliados.O uso das métricas, área

perímetro e tamanho das falanges e metacarpos dos dedos, que até o presente eram analisados separadamente e agora formam um único conjunto, pretende reduzir a quantidade de informações analisadas para avaliação de idade óssea e facilitar o diagnóstico, além de permitir geração de conhecimento.

## 5.3 - A contribuição da descoberta de conhecimento para base métrica.

A utilização de informações métricas reunidas em uma única base de dados representa um fator determinante do êxito desta pesquisa. As informações a respeito da estimação de idade óssea, empregadas na atualidade, dependem dos Atlas convencionais disponíveis, alguns com informações a respeito de tamanho dos ossos, ou sobre as várias alterações de determinados ossos ao longo do estágio de crescimento humano; outras, ainda, consideram aspectos de vários ossos, como já mencionado.

A reunião de novas informações na base já existente permite que novos conhecimentos sejam associados a antigos e acrescentem, a estes, novas formas de se analisar, ou mesmo combinar metodologias, o que certamente contribui para um rápido e fácil diagnóstico. Baseando-se nos resultados das pesquisas existentes, associados à análise das novas informações, obtidas no processo de descoberta de conhecimento, a atual pesquisa será direcionada para geração de um robusto volume de conhecimento para a tomada de decisão. Esse volume de conhecimento possibilita a estimação de idade óssea graças à aplicação feita com apoio de redes neurais. Isso permite que os especialistas possam combinar essa metodologia a outras, ou utilizar a que lhe seja mais exequível, no momento do diagnóstico, e acompanhar, de várias maneiras, a maturação óssea.

#### 6 - Redes Neurais Artificiais

Neste capítulo é apresentada uma breve revisão sobre redes neurais artificiais, suas particularidades e vantagens. Explicar-se-ão as redes utilizadas neste trabalho, a função e contribuição de cada uma delas para o êxito do mesmo.

As Redes Neurais Artificiais (*RNA*s) têm sido usadas com grande sucesso em diferentes áreas do conhecimento. O cérebro humano é extremamente eficaz no reconhecimento de padrões e regularidades, sendo, ainda, capaz de generalizar baseado no conhecimento acumulado. As RNA incorporam estas qualidades do cérebro, sendo, portanto, capazes de realizar previsões (FERNANDES & PORTUGAL, 1995).

#### 6.1 - Conceitos iniciais

As Redes Neurais Artificiais fundamentam-se nos estudos sobre a estrutura do cérebro humano, para tentar emular sua forma inteligente de processar informação. São modelos computacionais inspirados no cérebro humano, os quais possuem capacidade de aquisição e manutenção de conhecimento. Alguns estudos da neurofisiologia consideram que a riqueza computacional do cérebro humano está associada ao grande número de neurônios interconectados por uma rede complexa de "sinapses" (FERNANDES & PORTUGAL, 1995).

Em outras palavras, as *RNA*s podem ser definidas como um conjunto de unidades de processamento (neurônios), interligadas por um grande números de interconexões (sinapses).

Segundo Gaspar (2006), atualmente as principais aplicações de redes neurais artificiais são:

**Reconhecimento de padrões**: atribui um padrão de entrada a uma das várias classes predefinidas. Exemplo :reconhecimento de imagens, retina, análise de voz, etc.

**Clustering** / **Categorização**: explora a semelhança entre padrões e agrupa padrões parecidos em classes. Exemplos: *data minning*, compressão de dados.

**Aproximação de função**: encontra uma estimativa y de uma função desconhecida f. Exemplo: previsão de tempo, mercado financeiro, séries temporais, etc.

**Otimização**: maximizar e otimizar uma função-objetivo, sujeita a um conjunto de restrições. Exemplo: programação linear e não-linear, otimização combinatorial, etc. **Memórias associativas**: recuperar itens corretos, mesmo que a entrada seja parcial, ou distorcida. Exemplo: recuperação de imagens em base de dados, identificação de curvas, etc.

Controle: controlar sistemas de engenharia. Exemplo: controle de processos, robótica, etc.

## 6.2 - Componentes das RNAs

De acordo com Rumelhart (1986), um modelo conexionista pode ser descrito por oito elementos principais:

- Um conjunto de unidades de processamento.
- Um estado de ativação.
- Uma função saída.
- Um padrão de interconexão.
- Uma regra de propagação.
- Uma regra de ativação.
- Uma regra de aprendizado.
- Um ambiente onde o sistema deve funcionar.

## Unidade de Processamento.

O neurônio é a célula fundamental do sistema nervoso cerebral. Uma rede neural recebe continuamente informações, percebe-as e toma decisões apropriadas (HAIKYN, 2001).

De acordo com Fernandes & Portugal (1995), a unidade de processamento (neurônio) é o componente básico das RNAs e corresponde ao neurônio humano. A Figura 20 apresenta a ilustração de um neurônio como unidade que representa um limite a ser ultrapassado. Suas entradas representam os dendritos; cada dendrito possui um sinal (x), que é adicionado (S). Depois da adição, este sinal é processado através da função, que representa o limite ou função de ativação  $f(\cdot)$ , a qual produz um sinal de saída. Na Figura 20, o neurônio pode ser considerado uma representação simplificada dos neurônios biológicos.

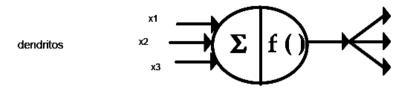


Figura 20 - Neurônio Como Unidade Limiar. Fonte: Fernandes & Portugal (1995).

**Estado de Ativação**: o estado de ativação de todos os neurônios da rede ou do sistema específica o que está sendo representado em um determinado instante *t*.

Os valores das ativações existentes na rede podem ser discretos, por exemplo, assumindo valores {0,1} ou {-1,0,1}, como também podem ser contínuos, assumindo valores no intervalo [0,1] ou [-1,1], que são calculados pela regra de ativação (LIPPMAN, 1987);

**Função de Saída**: os neurônios interagem por meio de um valor que é transmitido pelas conexões. Este valor é determinado pela ativação do neurônio estimulador. A "função de saída" é o mapeamento do estado de ativação num sinal de saída. A "função de saída" pode ser a função identidade, isto é, o sinal de saída é igual ao estado de ativação. Em muitos casos, a função de saída é uma função limiar, que só emite sinal quando o seu estado ultrapassa um certo limite.

**Padrão de Interconexão**: pode-se representar o padrão de interconexão da rede por uma matriz de pesos w, na qual um elemento  $w_{ij}$  corresponde à influência do neurônio  $x_i$  sobre o neurônio  $x_j$ . Conexões com pesos positivos, chamadas de excitatórias, indicam o reforço na ativação do neurônio  $x_j$ . Conexões com pesos negativos, chamadas de inibitórias, indicam inibição na ativação do neurônio  $x_j$ . O conjunto das ligações excitatórias e inibitórias existentes na rede determina o comportamento da mesma (FERNANDES & PORTUGAL, 1995).

**Regra de Propagação**: cada neurônio  $x_i$  calcula sua nova ativação por meio de uma regra de propagação. Em geral, ela é definida como sendo uma função somatório da entrada dos pesos dos neurônios  $x_j$  que estão diretamente conectados a  $x_i$ . Isso é feito através da multiplicação do estado do j-ésimo neurônio ( $x_j$ ) pelo peso da conexão  $w_{ij}$ , do i-ésimo para o j-ésimo neurônio, para cada um dos j neurônios que estão conectados à entrada do neurônio i. A regra de propagação se completa através do limite que deve ser ultrapassado pelo somatório. Este limiar, que pode ser nulo, inclusive, deve ser superado para que ocorra a ativação da célula. Matematicamente, o neurônio pode ser representado por:

$$u_k = \sum_{j=1}^m w_{kj} x_j \tag{6}$$

 $y_k = \varphi(u_k + b_k)$ , na qual  $u_k$  é a saída linear e  $y_k$  é o sinal de saída do neurônio.

**Regra de Ativação**: é necessária para calcular o valor de ativação de um neurônio no instante t. É preciso uma função f que calcule a nova ativação a(t), utilizando as entradas. Geralmente, esta função possui a forma  $a_i(t+1) = f(ai(t), net_i(t))$ , na qual f é a função de ativação. Esta função mapeia os neurônios de entrada para um intervalo pré-especificado de saída. As quatro funções de ativação mais utilizadas são: linear, rampa, degrau e sigmóide, representadas na Figura 21.

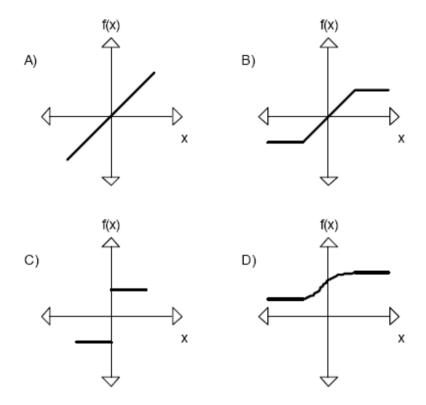


Figura 21 – Exemplos de função de ativação. Fonte: Fernandes & Portugal (1995).

A equação abaixo apresenta a função sigmóide ou função logística, cujos pontos de saturação são 0 e 1 (SIMPSON, 1990).

$$S(x) = \frac{1}{(1 - e^{-x})} \tag{7}$$

**Regra de Aprendizado**: a modificação do processamento ou da estrutura de conhecimento de uma rede neural envolve a alteração do seu padrão de interconexão. Em princípio, isso pode ser feito de três maneiras:

- Desenvolvimento de novas conexões.
- Perda de conexões existentes na rede.
- Modificação dos pesos das conexões já existentes.

Em geral, as regras de aprendizado podem ser consideradas como uma variante da Regra de Hebb (1949), que estabeleceu o princípio da alteração da eficiência das conexões como base do aprendizado. Segundo o postulado, "quando o neurônio A se encontra próximo do neurônio B, de forma a poder excitá-lo, e o faz repetidas vezes, algum processo desconhecido provoca o crescimento de conexões entre as células A e B, facilitando, assim, a excitação de B por A".

## 6.3 - Topologias de redes neurais utilizadas.

Um fator muito importante na utilização de redes neurais é a escolha da topologia de rede. O modelo da topologia está diretamente ligado à configuração das interligações entre os neurônios e ao modelo do algoritmo de treinamento. O que faz com que a topologia de rede escolhida tenha reflexos diretos nas características e propriedades apresentadas pelo modelo (ORTEGA,2002).

As redes neurais possuem uma estrutura que pode ser dividida em três partes básicas:

A primeira; chamada camada de entrada, na qual os padrões são apresentados à rede.

A segunda; chamada camada intermediária ou oculta na qual o processamento da rede é realizado.

A terceira; chamada camada de saída, na qual são apresentados os resultados alcançados pela rede.

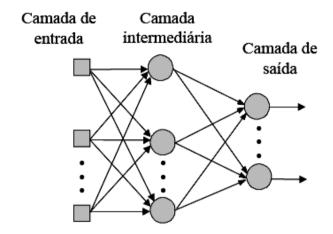


Figura 22 – Representação das camadas de uma rede neural artificial. Fonte: Ortega (2002).

As RNAs possuem diferentes arquiteturas que, na realidade, são a maneira como seus neurônios estão organizados, fortemente relacionada com o algoritmo utilizado para treinar a rede. As principais arquiteturas são:

**Redes** *Feed Forward* (camada única): neste tipo de rede tem-se uma camada de entrada e uma única camada de neurônios, que é a própria camada de saída; não existe realimentação;

**Redes** *Feed Forward* (multicamadas): este tipo de rede se distingue da anterior pela presença de uma ou mais camadas escondidas;

Redes recorrentes: contêm retroalimentação entre neurônios de camadas diferentes;

**Estrutura** *Lattice*: rede *Feed Forward* consiste em uma *array* de neurônios de uma ou mais dimensões, na qual os sinais de entrada são os mesmos para todos os neurônios.

# Tipos de treinamento.

O treinamento ou processo de aprendizagem de uma *RNA* consiste em ajustar os pesos sinápticos (matriz de pesos *w*), de forma que a aplicação de um conjunto de entradas produza um conjunto de saídas desejadas. O treinamento das redes pode ser classificado em dois tipos: **Treinamento Supervisionado**: a rede é treinada para fornecer uma saída desejada em relação a um estímulo de entrada específico;

**Treinamento Não-Supervisionado**: não há uma saída específica em relação aos estímulos de entrada. A rede se auto-organiza em relação a algum subconjunto de entradas, cujos elementos possuem características similares.

Neste trabalho utilizar-se-á o treinamento supervisionado, devido à existência do conjunto de saídas desejadas.

## 6.3.1 - Rede LVQ (Learning Vector Quantization) e Counterpropagation

Segundo Oliani(2005), o objetivo principal, envolvido com um processo de quantização vetorial por meio de redes neurais, consiste em dividir o espaço amostral de entrada em diversos subespaços disjuntos, sendo que cada um dos vetores de entrada (padrões) deve pertencer somente a um desses subespaços, os quais estarão representando a classe associada ao problema considerado.

As redes LVQ (*Learning Vector Quantization*) são redes cujo treinamento é supervisionado, sendo que utilizam um processo competitivo para o ajuste de seus pesos, os quais estarão representando os respectivos vetores quantizadores.

Nesta abordagem, todas as *n*-classes associadas aos padrões de entrada do sistema são conhecidas. O objetivo da rede é a quantização ótima do espaço de entrada em *n*-espaços.

Após o treinamento, a rede pode ser utilizada para classificar outros padrões de entrada entre as várias classes do sistema (OLIANI, 2005).

## Rede LVQ-1

Neste caso, considera-se que cada vetor de entrada x(i), utilizado no treinamento, pertença a uma classe "j". O algoritmo de aprendizagem é idêntico ao de *Kohonen*, modificando apenas o processo de ajuste dos pesos, o qual é aplicado ao vencedor (OLIANI, 2005).

#### 6.3.2 - O Perceptron multicamadas

O *Perceptron* é a forma mais simples de uma rede neural e é utilizado para a classificação de padrões que devem ser linearmente separáveis.

As redes *Perceptron* multicamadas (MLP) são consideradas um aproximador universal de funções. A base teórica deste modelo foi apresentada originalmente em 1974 por Paul WERBOS, David PARKER e David RUMELHART. O modelo do *Perceptron* multicamadas utiliza uma topologia de três ou mais camadas, cujas conexões entre as unidades são do tipo intercamadas e são direcionadas da camada de entrada para a camada de saída (FERNANDES & PORTUGAL,1995).

O modelo do *Perceptron* é composto por uma camada de entrada, uma ou mais camadas escondidas e uma camada neural de saída, como apresentado abaixo.

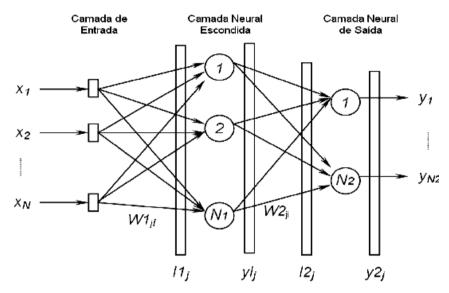


Figura 23 – Diagrama esquemático da rede *Perceptron*. Fonte: Goedtel (2007).

Os algoritmos de treinamento para a rede *Perceptron*, testados nesta pesquisa, foram o algoritmo de *Levenberg - Marquardt*, (HAGAN & MENHAJ,1994) e o algoritmo *Resilient-Backpropagation*.

# Levenberg-Marquardt.

O algoritmo de *Levenberg-Marquardt* é uma técnica baseada no método dos mínimos quadrados para modelos não-lineares, que pode ser incorporada ao algoritmo *Backpropagation* a fim de aumentar a eficiência do processo de treinamento. Neste algoritmo as funções erro quadrático e erro quadrático médio são fornecidas (GOEDTEL, 2007). Esse algoritmo consiste em ajustar os pesos sinápticos e limiares, de forma que a aplicação de um conjunto de entradas produza um conjunto de saídas desejadas. Possui um tempo reduzido de convergência e um baixo esforço computacional, pois é baseado em uma aproximação do método de Newton; sua principal característica é a computação da matriz jacobiana, como mostra a equação abaixo (QUEIROZ, 2006).

$$\Delta z = [J^{T}(w).J(w) + \mu I]^{-1}.J^{T}(w).e(w)$$
 (8)

na qual: J é a matriz jacobiana, w são os pesos sinápticos,  $\mu$  é um parâmetro que ajusta a taxa de convergência, I é a matriz identidade, e e(w) é o vetor erro ( em relação à matriz w) referente aos p-padrões de treinamento.

### Resilient-Backpropagation.

Segundo Borucki et al.(2006), o algoritmo *resilient-Backpropagation* apresenta-se como uma rede neural multicamadas unidirecional, aplicada na realização de tarefas relacionadas ao reconhecimento. Utiliza uma rede tipo Feed-Forward Backpropagation Network (*FF BP*), em que cada neurônio tem uma função sigmóide de ativação. Seu treinamento é de forma supervisionada, a entrada dos pesos é baseada no algoritmo de retropropagação - *resilientes Backpropagation (RPROP)*, descrito pela equação 9.

$$w_{ij}^{(k)}(n+1) = w_{ij}^{k}(n) - \eta_{ij}^{k}(n)\operatorname{sgn}(\nabla_{ij}^{(k)}(n))$$
(9)

na qual:

 $\eta_{ij}^{(k)}$  - é o coeficiente de aprendizagem individual para cada peso,  $\nabla_{ij}^{(k)}$  - erro em função do gradiente (BORUCKI et al.,2006).

O coeficiente de aprendizagem é selecionado individualmente a cada iteração, para cada peso  $w_{ij}$  baseado no valor de mudança do gradiente. Se nas iterações consecutivas o sinal do gradiente permanecer o mesmo, o coeficiente  $\eta$  de aprendizagem é aumentado; caso contrário, é reduzido, (BORUCKI et al.,2006).

### 6.4 - Considerações finais sobre redes neurais artificiais.

Os sistemas baseados em inteligência artificial têm se desenvolvido como uma abordagem promissora e flexível, que pode ser aplicada a diversos tipos de problemas. Entre as muitas ferramentas que constituem a inteligência artificial, destacam-se as redes neurais

artificiais que são utilizadas neste trabalho, no processo de classificação dos dados que será abordado no próximo capítulo.

As redes neurais têm hoje um papel de destaque junto à comunidade científica, graças ao seu grande poder de aquisição e manutenção de conhecimento, que permitem que as mesmas atuem na solução de problemas complexos e variados, baseando-se no funcionamento do cérebro humano. Esses modelos computacionais com diversas particularidades têm permitido que problemas robustos sejam solucionados, de forma mais rápida, a partir de sua aplicação.

Com grande poder de resolução de problemas complexos e não-lineares, com aplicações em várias áreas, as redes *Perceptron* Multicamadas (MLP) são computacionalmente mais poderosas do que as redes sem camadas escondidas. Podem tratar dados que não são linearmente separáveis. A precisão obtida e a implementação da função dependem do número de neurônios utilizados nas camadas escondidas. O processamento realizado por cada neurônio é definido pela combinação dos processamentos realizados pelos neurônios da camada anterior que estão conectados a ele. A partir da primeira camada escondida até a camada de saída, as funções implementadas tornam-se cada vez mais complexas. Essas funções definem como é formada a divisão do espaço de decisão (HAYKIN, 2001).

Devido a essa característica e por sua implementação simples, essas redes foram aplicadas aos dados da base *CarpEven*.

### 7 - Desenvolvimento da Metodologia.

Neste capítulo, aborda-se a metodologia utilizada na pesquisa para a criação de uma nova forma de estimação de idade óssea baseada em extração de características, empregando redes neurais artificiais no processo de classificação.

### 7.1 - Introdução.

O objetivo desta pesquisa é contribuir com as metodologias de estimação de idade óssea existentes, baseando-se em extração de características, com o propósito de reduzir o número de centros analisados e possibilitar mais precisão de resultados. Essa metodologia é descrita em quatro etapas, apresentadas nas seções seguintes.

### 7.2 - Etapas utilizadas no desenvolvimento da metodologia

### 7.2.1 - Geração dos dados de constituição da base CarpEven

Nesta primeira etapa, as imagens carpais foram segmentadas manualmente, com auxílio do software *Mipav* (*Medical Image Processing, Analysis and Visualization*), (MCAULIFFE, 2007), apresentado nas Figuras 13 e 14 deste trabalho, no capítulo que trata do processo de formação dos dados. Com essa segmentação obteve-se um conjunto de métricas (área, perímetro e comprimento) que foram calculadas pelo software, retornando valores em pixels. Esses valores foram armazenados em planilhas do Microsoft Excel® e, quando transferidos para a base de dados *CarpEven*, são convertidos em milímetros, através da seguinte Equação (10), presente no algoritmo de criação do banco:

$$dimen = \left(v_{mipav} * \left(\frac{2,54}{res_{img}}\right)\right) * 10$$
 (10)

na qual: *dimen* é o resultado do cálculo, *v\_mipav* é o valor resultado dos cálculos feitos pelo software, e *res\_img* é a resolução da imagem.

O resultado foi, então, agrupado em um banco de dados, que permite acesso *on line* a todas as suas informações, bem como identifica as imagens no banco de imagens e suas métricas na base *CarpEven*. O banco é composto por 666 imagens de indivíduos na faixa etária que vai de 6 a 16 anos, sendo que 320 são de mãos femininas e 346 são de mãos masculinas.

### 7.2.2 - Preparação dos dados.

A segunda etapa foi a de preparação dos dados e aplicação de mineradores. Os dados da base *CarpEven* foram separados em três conjuntos. Um conjunto com os dados da base completa, chamado de "Base Geral", outro só com informações de mãos femininas, "Base Fem", e outro conjunto com informações de mãos masculinas, "Base Masc". Estes conjuntos foram separados para que se avaliasse o comportamento de cada um dos mineradores diante das informações.

Assim, a mineração de dados deve contribuir para o processo de descoberta de padrões válidos, novos, potencialmente úteis e compreensíveis, embutidos no conjunto de dados (FAYYAD et al, 1996).

### 7.2.3 - O processo de mineração de acordo com Weka.

Na terceira destaca-se o papel dos mineradores. A mineração de imagens utiliza os princípios da mineração de dados, porém de uma maneira mais complexa, envolvendo um conjunto de técnicas de processamento de imagens, extração de características, indexação de objetos complexos e visualização (RIBEIRO, 2008).

A possibilidade de utilizar a performance do *Weka* no trabalho de mineração das informações da base *CarpEven* foi o que motivou seu emprego nesta pesquisa. Convém ressaltar que os atributos foram normalizados antes da aplicação no software, e o atributo idade necessitou de discretização (consideram-se apenas as faixas etárias, como, por exemplo, 6 anos, 7 anos, etc...) para um melhor desempenho.

As amostras da base *CarpEven* foram aplicadas ao selecionador de característica do *Weka* e aos algoritmos de árvore de decisão, utilizando o padrão proposto sem alterações nos mesmos. Foi empregado mais de um tipo de algoritmo, (como descrito a seguir) para verificar o comportamento dos dados. O conjunto final de características relevantes e extraídas pelo *Weka* é composto pelo conjunto de características presentes nos três testes. Os dados

necessitaram de normalização e o atributo idade foi também discretizado. A seguir, segue uma breve explicação a respeito dos algoritmos que foram aplicadas aos dados.

### CFS (Correlation-based Feature Selection).

Avalia o valor de um subconjunto de características individuais, considerando a capacidade preditiva de cada característica, juntamente com o grau de redundância entre eles; subconjuntos de funcionalidades que estão altamente correlacionados com a classe são preferidos (WITTEN & FRANK, 2005).

O método de avaliação do *CFS* é uma heurística baseada em correlação com base em redução de atributos. Pode avaliar o "mérito" do subconjunto de atributos. O objetivo é identificar um único atributo com capacidade de predição e suas correlações com a classe. O algoritmo heurístico é baseado na hipótese abaixo:

Os atributos que pertencem a um subconjunto  $F_m$  estão altamente correlacionados à classe  $C_i$ , enquanto os outros atributos são irrelevantes.

Os atributos irrelevantes no subconjunto dificilmente são relacionados com a classificação; podem ser ignorados. Os atributos redundantes, também, podem ser eliminados para que ocorra uma alta correlação. A aceitação de um atributo como relevante deve-se à sua capacidade de discriminação quando comparado a outros (NI et al., 2007).

### RepTree.

Método rápido de aprendizagem baseado em árvores de decisão (*Fast decision tree learner*). Constrói, por meio de divisão com critério, uma árvore de decisão / regressão, utilizando informações ganho / redução, variância e podas; utiliza-se do erro (com *backfitting*), para fazer cortes. Faz uso apenas de atributos com valores numéricos. Valores em falta são tratados através da divisão do correspondente, isto é, como no algoritmo *C4.5*<sup>3</sup>( O algoritmo *C4.5* gera um classificador na forma de uma árvore de decisão, com uma estrutura composta por uma folha, indicando uma classe e o nó de decisão que especifica um teste a ser realizado no valor de um atributo, com um galho, para cada resposta possível do teste, que levará para um sub-árvore ou uma folha.) (BASU & MELVILLE, 2005).

\_

<sup>&</sup>lt;sup>3</sup> http://www.inf.ufpr.br/aurora/tutoriais/arvoresdecisao/IAcap01.html

#### **M5P.**

O *M5P* é um procedimento que gera rotinas no modelo de árvores e regras. O algoritmo original *M5* foi inventado por Quinlan (1992) e melhorado nesta versão de nome *M5P*. O processo de aprendizagem de árvore de regressão começa com o treinamento de todos os dados; em seguida, o *M5P* seleciona o atributo mais discriminante, como um nó raiz, e divide a formação em subgrupos, de acordo com o valor do atributo selecionado: nó raiz. O mais discriminante é o atributo que mais reduz a variância dos valores da modelagem (ŠTRAVS & BRILLY, 2007).

A entrada de dados no algoritmo *M5* trabalha com método de aprendizagem, cujas entradas devem ser suficientemente relevantes, e a saída de dados modelada nesse processo. Ele pode aprender automaticamente os padrões subjacentes do processo de modelagem, e por si só dividir os atributos em subespaços onde existam semelhanças entre certas características ou padrões. A divisão em subespaços inclui sucessivas divisões dos dados e segue com o objetivo de selecionar e determinar o critério de separação em cada iteração, na qual maximiza o nó distância entre grupos de dados (ŠTRAVS & BRILLY, 2007).

Os atributos selecionados de cada uma das bases, de acordo com os mineradores, são listados no capítulo seguinte. Alguns destes atributos selecionados coincidem com os mesmos atributos utilizados para a criação da metodologia de Olivete Jr. (2005a). Os atributos comuns, encontrados pelos três algoritmos, formaram o conjunto que foi aplicado ao classificador neural.

### 7.2.4 - O processo de mineração de acordo com StARrminer

Assim como ocorreu com o *Weka* para avaliação da aplicação ao algoritmo *StARMiner*, amostras foram separadas em grupos. Houve, também, a necessidade de discretização do atributo idade, já que o mesmo representava a classe e, nesse algoritmo, as classes devem ser sempre discretas; os dados, porém, não necessitaram de um tratamento.

O algoritmo utiliza ainda três limiares que são definidos pelo usuário antes de seu início. Ele possui um padrão, mas para cada um dos conjuntos de amostra foi necessário estabelecerem-se novos limiares para que o algoritmo extraísse as características.

Na Tabela 2, abaixo, são definidos os valores padrão, e também os valores que foram utilizados para cada um dos conjuntos de acordo com Ribeiro (2008).

**Tabela 2** – Padrão de entrada dos limiares no algoritmo *StARminer* e padrão estabelecido para cada um dos conjuntos.

Conjunto  $\Delta \mu_{min}$  - $\sigma_{max}$  $\gamma_{min}$ Padrão algoritmo 0.99 0.15 0.10 **Base Geral** 0.95 0.22 0.15 0.96 0.5 **Base Fem** 0.20 **Base Masc** 0.96 0.20 0.25

Dentre os atributos selecionados pelo *StARminer*, destacam-se, no próximo capítulo, os atributos que coincidem com os utilizados por Olivete Jr. (2005a) em sua metodologia. Tal fator de concordância pode ser aplicado como avaliador desse novo conjunto de características, já que a mineração de imagens usa os princípios da mineração de dados, envolvendo um conjunto de técnicas de processamento de imagens, extração de características, indexação de objetos complexos e visualização (RIBEIRO, 2008).

### 7.2.5 - Redes neurais artificiais aplicadas como classificadores.

Na quarta e última etapa, os resultados obtidos com os mineradores foram aplicados a um classificador neural. Devido ao desempenho apresentado pelas redes neurais nos diversos processos de classificação, elas foram aplicadas às características extraídas com o algoritmo *StARminer* e o software *Weka*.

Após a formação de cada um dos conjuntos de atributos minerados, foi realizado um estudo a respeito da melhor forma de classificação das informações. Com base na literatura existente e nas características apresentadas para resolução de problemas complexos e eficiência em processos de classificação, optou-se por redes neurais artificiais. A primeira idéia foi a de utilizar uma rede simples com características de um classificador. Sucedendo o processo de avaliação, a *LVQ* foi aplicada a cada um dos conjuntos.

## LVQ.

Este tipo de rede surgiu da derivação das redes SOM (*Self Organization Maps*) com a seguinte diferença: possui treinamento supervisionado, ou seja, a sua saída (desejada) é conhecida. O principal objetivo dessa rede é dividir o conjunto de entradas em diversos subconjuntos disjuntos, sendo que cada vetor de entrada (padrões) deve pertencer somente a

um desses subconjuntos, os quais estarão representando as classes associadas ao problema (OLIANI, 2005).

Estas redes foram aplicadas ao trabalho por sua facilidade de implementação e por possuirem um algoritmo de rápida implementação; mas não houve sucesso, no que diz respeito à convergência da mesma.

O fato de alguns conjuntos aplicados a esse tipo de rede neural apresentarem um grau de complexidade elevado, e a formação dos clusters ser muito próxima, dificulta a separação linear das amostras, tornando a sua utilização inadequada.

# Perceptron.

As redes *Perceptron* multicamadas apresentam melhor aplicabilidade no processo de classificação, devido ao seu processo de aprendizado supervisionado e à maneira com que os clusters são formados (QUEIROZ, 2006).

Essas características possibilitaram que tal rede neural fosse aplicada nesta pesquisa. A rede *Perceptron* foi utilizada com duas variações: o algoritmo de *Levenberg-Marquardt e o* algoritmo *resilient-Backpropagation*, para que se pudesse observar o comportamento dos conjuntos e, de acordo com a forma de classificação de cada um dos algoritmos, escolher o que fosse mais adequado.

### 7.3 - Considerações finais.

Neste capítulo foi apresentada a metodologia utilizada na pesquisa, que se baseia na extração de características utilizando mineradores reconhecidos na literatura e submissão desses resultados a redes neurais que apresentam características adequadas às necessidades do trabalho.

Como pode ser observado com base nos métodos existentes e nas pesquisas realizadas, novos parâmetros foram analisados, culminando com a metodologia apresentada. Com auxílio de pré-processamento descrito no capítulo 3; as imagens puderam ser segmentadas manualmente e suas métricas armazenadas.

Finalizado o processo de extração, as informações foram armazenadas e posteriormente aplicadas ao processo de mineração, as informações foram avaliadas em conjunto e também separadas, por sexo, faixa etária e tipo de métrica, resultando características que, com auxílio

de redes neurais, provêem a estimação de idade óssea. Os resultados desta metodologia podem ser observados no capítulo a seguir.

#### 8 - Discussões e resultados.

### 8.1 - Introdução

Neste capítulo são apresentados os resultados finais e as discussões cabíveis. É descrita a forma como os atributos encontrados com a aplicação dos mineradores e classificados pela rede neural, foram selecionados para os testes e como resultaram na nova metodologia. Além disso, o capítulo descreve a constituição dos conjuntos de atributos, como foi feita a combinação entre eles e a avaliação dos resultados. Finalmente são mostrados os processos de classificação com a aplicação de redes neurais artificiais e o desempenho de cada uma das análises que foram utilizadas.

### 8.2 - O processo de geração de informação.

A base de dados *CarpEven* (descrita no capítulo 4), teve suas informações separadas em três conjuntos (Base Geral, Base Masc e Base Fem). Tais informações foram primeiramente aplicadas a mineradores de dados e, posteriormente, a redes neurais para classificação. Para a realização dos testes de validação da metodologia, foram utilizadas 20% das amostras de cada faixa etária, de 6 a 16 anos, conforme orientado na literatura. Como o conjunto possui número de amostras distinto para cada faixa de idade, todas tiveram 20% do seu número de amostras reservado para os testes de validação.

As amostras utilizadas foram laudadas por especialistas e pelo software *Anarcarp* criado por Olivete Jr. & Rodrigues, (2006).

Os resultados obtidos com a utilização dos mineradores e a aplicação da rede neural foram separados em conjuntos. Estes foram combinados 3 a 3 e avaliados a fim de encontrar a combinação ou combinações que obtivessem maior taxa de concordância com a média dos laudos. Por fim a nova metodologia foi composta.

### 8.3 - O processo de mineração de dados.

Como observado na seção anterior, a aplicação do conjunto de parâmetros ao processo de mineração resultou em um novo conjunto reduzido de características. Ressalta-se que alguns deles também fazem parte das características já selecionadas por Olivete Jr. (2005a)

em sua proposta de redução do conjunto utilizado pelo método Eklof & Ringertz (1959). Isso endossa, mais uma vez, a metodologia de Olivete Jr. (2005a) e comprova que tais centros de ossificação devem estar presentes em conjuntos que avaliem o desenvolvimento ósseo. Apresenta-se a seguir, o conjunto de características selecionadas pelo *Weka* e, em destaque, as características que coincidem com as utilizadas pela metodologia de Olivete Jr. (2005a).

## 8.3.1 - Características utilizadas pelo Weka.

Com o uso do *Weka*, os conjuntos gerados foram aplicados a três algoritmos diferentes para se observar o comportamento das informações. Os resultados, mostrados na Tabela 3, foram, a princípio, muito similares. Graças a essas similaridades, pode-se gerar um conjunto final com os atributos comuns selecionados, para utilização no processo de classificação.

**Tabela 3** – Atributos selecionados pelo software Weka; em destaque, os atributos utilizados pela metodologia de Olivete Jr. (2005a).

Identificação	Subset Evaluator (CFS).	RepTree.	M5P.
Base Geral	a_atr1_fdmin	a_atr1_fdmin	a_atr1_fdmin
	a_atr2_fmmin	a_atr2_fmmin	a_atr2_fmmin
	a_atr8_manu	a_atr8_manu	a_atr8_manu
	a_atr17_fdpol	a_atr17_fdpol	a_atr17_fdpol
	a_atr19_fppol	a_atr19_fppol	a_atr19_fppol
	p_atr5_fdanu	p_atr2_fmmin	p_atr5_fdanu
	p_atr7_fpanu	p_atr5_fdanu	p_atr7_fpanu
	p_atr10_fmmed	p_atr7_fpanu	p_atr10_fmmed
	p_atr14_fmind	p_atr10_fmmed	p_atr14_fmind
	p_atr15_fpind	p_atr11_fpmed	p_atr15_fpind
	p_atr17_fdpol	p_atr14_fmind	p_atr17_fdpol
	t_atr3_fpmin	p_atr15_fpind	t_atr4_mmin
	t_atr4_mmin	p_atr17_fdpol	t_atr16_mind
	t_atr16_mind	t_atr4_mmin	t_atr19_fppol
		t_atr16_mind	
	a_atr9_fdmed	a_atr9_fdmed	a_atr9_fdmed
Base Fem	a_atr20_mpol	a_atr17_fdpol	a_atr20_mpol
	p_atr2_fmmin	a_atr20_mpol	p_atr2_fmmin
	p_atr11_fpmed	p_atr2_fmmin	p_atr8_manu
	p_atr19_fppol	p_atr11_fpmed	p_atr11_fpmed
	p_atr20_mpol	p_atr16_mind	p_atr19_fppol
	t_atr3_fpmin	p_atr19_fppol	p_atr20_mpol
	t_atr8_manu	p_atr20_mpol	t_atr2_fmmin
	t_atr9_fdmed	t_atr3_fpmin	t_atr3_fpmin
	t_atr12_mmed	t_atr8_manu	t_atr8_manu
	t_atr15_fpind	t_atr9_fdmed	t_atr9_fdmed
	t_atr16_mind	t_atr12_mmed	t_atr12_mmed
	t_atr19_fppol	t_atr15_fpind	t_atr15_fpind
		t_atr16_mind	t_atr16_mind
		t_atr19_fppol	t_atr19_fppol
Base Masc	a_atr20_mpol	a_atr20_mpol	a_atr20_mpol
	p_atr1_fdmin	p_atr1_fdmin	p_atr1_fdmin
	p_atr11_fpmed	p_atr11_fpmed	p_atr11_fpmed
	p_atr19_fppol	p_atr19_fppol	p_atr19_fppol
	p_atr20_mpol	p_atr20_mpol	p_atr20_mpol
	t_atr3_fpmin	t_atr9_fdmed	p_atr19_fppol
	t_atr4_mmin	t_atr12_mmed	p_atr20_mpol
	t_atr8_manu	t_atr3_fpmin	t_atr3_fpmin
	t_atr9_fdmed	t_atr4_mmin	t_atr4_mmin
	t_atr15_fpind	t_atr15_fpind	t_atr9_fdmed
	t_atr16_mind	t_atr16_mind	t_atr12_mmed
	t_atr19_fppol	t_atr19_fppol	t_atr15_fpind
			t_atr16_mind
			t_atr19_fppol

Com base nos atributos selecionados pelo algoritmo do *Weka*, formou-se um conjunto final dos atributos selecionados - como pode ser visto na Tabela 4 - que foram comuns a todos os algoritmos, para cada uma das bases utilizadas. Destaca-se nesse conjunto os atributos que coincidem com os utilizados por Olivete Jr. (2005a).

Nota-se que na Base Geral, o número de atributos coincidentes com a metodologia de Olivete Jr. (2005a) é reduzido, enquanto têm-se 5 e 4 atributos coincidentes nas Bases Feminina e Masculina, respectivamente. Isto se deve ao fato das bases serem analisadas de forma separada, permitindo ao algoritmo fazer uma melhor distinção entre as amostras, pois, como se sabe, homens e mulheres possuem desenvolvimento ósseo diferente.

**Tabela 4** – Conjunto final de atributos de cada uma das amostras, de acordo com *Weka*; em destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a).

Conjunto	Atributos
Base Geral	a_atr1_fdmin, a_atr2_fmmin, a_atr8_manu, a_atr17_fdpol, a_atr19_fppol, p_atr5_fdanu,
	p_atr7_fpanu, p_atr10_fmmed, p_atr14_fmind, p_atr15_fpind,
	p_atr17_fdpol, t_atr4_mmin, t_atr16_mind
Base Fem	a_atr9_fdmed, a_atr20_mpol, p_atr2_fmmin, <b>p_atr11_fpmed</b> , p_atr19_fppol,
	p_atr20_mpol, t_atr3_fpmin, t_atr8_manu, t_atr9_fdmed, t_atr12_mmed, t_atr15_fpind,
	t_atr16_mind, t_atr19_fppol
Base Masc	a_atr20_mpol, p_atr1_fdmin, <b>p_atr11_fpmed</b> ,p_atr19_fppol, p_atr20_mpol, t_atr3_fpmin,
	t_atr4_mmin, t_atr15_fpind, t_atr16_mind, t_atr19_fppol

### 8.3.2 - Características selecionadas pelo StARminer.

Na Tabela 5, é apresentado o conjunto de atributos que foi selecionado pelo *StARminer*. Estão em destaque os atributos que coincidem com os utilizados pela metodologia de Oliverte Jr. (2005a), assim como para o conjunto selecionado pelo *Weka*.

**Tabela 5** – Atributos selecionados com algoritmo *StARminer*; em destaque, os atributos utilizados pela metodologia de Olivete Jr. (2005a).

Conjunto	Atributos	
Base	t_atr1_fdmin, t_atr2_fmmin, t_atr5_fdanu, t_atr6_fmanu, t_atr9_fdmed, t_atr13_fdind,	
Geral	t_atr14_fmind, <b>t_atr16_mind</b> , t_atr17_fdpol	
Base Fem	t_atr1_fdmin,t_atr2_fmmin, t_atr5_fdanu, t_atr6_fmanu, t_atr9_fdmed, t_atr11_fpmed,	
	t_atr13_fdind, t_atr14_fmind, t_atr17_fdpol	
Base	t_atr1_fdmin,t_atr2_fmmin, t_atr4_mmin, t_atr5_fdanu, t_atr6_fmanu, t_atr9_fdmed,	
Masc	t_atr10_fmmed,t_atr13_fdind,t_atr14_fmind,t_atr17_fdpol	

Como o algoritmo do *StARminer* trabalha com regras de associação, as variações em sua forma de funcionamento podem ocorrer apenas quando seus valores padrão são alterados. A alteração exagerada desses valores padrão acarreta alteração da confiança e desvio padrão. Isso pode causar a falha da análise do conjunto selecionado. Para obter máximo desempenho do algoritmo associado a um bom trabalho de mineração, procurou-se utilizar um conjunto de valores limiar o mais próximo possível do padrão estabelecido pelo algoritmo. Os limiares utilizados (que resultaram cada conjunto) foram apresentados anteriormente, no capítulo 7 de metodologia.

Após o processo de seleção de características utilizando a base completa, as métricas foram analisadas separadamente.

As informações foram aplicadas a cada um dos algoritmos para se observar o comportamento dos conjuntos. Um dos conjuntos de características gerados é apresentado a seguir, na Tabela 6.

Como as métricas foram aplicadas a três diferentes algoritmos, na Tabela 6 são listados os atributos selecionados e, em destaque, os que coincidem com os avaliados por Olivete Jr. (2005a).

**Tabela 6** - Atributos da Base Geral selecionados com os algoritmos do *Weka* para Área; em destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a).

Área				
CSF	RepTree	M5P		
a_atr3_fpmin	a_atr15_fpind	a_atr16_mind		
a_atr15_fpind	a_atr16_mind			
a_atr16_mind	a_atr19_fppol			
a_atr19_fppol	a_atr4_mmin			
	a_atr3_fpmin			
	a_atr1_fdmin			

Todas as combinações possíveis entre as métricas foram analisadas para que se chegasse a uma análise final. A seguir, na Tabela 7, apresenta-se o resultado da seleção utilizando as métricas combinadas duas a duas. Esse procedimento foi realizado para a Base Geral, Base Fem e Base Masc. As métricas também foram avaliadas separadamente, para que a análise das informações fosse completa e auxiliasse na criação da metodologia.

Para avaliação das métricas duas a duas, os dois mineradores foram utilizados; porém, no caso do *StARminer*, dependendo do valor utilizado e respeitando seus padrões de seleção, o

algoritmo seleciona atributos de um conjunto apenas, excluindo o outro, ou seja, seus limiares se tornaram tendenciosos quando aplicados a dois conjuntos ao mesmo tempo. É importante ressaltar que as combinações citadas no decorrer do texto foram realizadas para todas as bases, porém nem todos os resultados não são mostrados.

**Tabela 7** - Atributos da Base Geral selecionados com os algoritmos do *Weka* para Área e Perímetro; em destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a).

Área e Perímetro					
CSF	RepTree	M5P			
a_atr3_fpmin	p_atr3_fpmin	p_atr3_fpmin			
a_atr15_fpind	a_atr19_fppol	a_atr19_fppol			
a_atr16_mind	a_atr16_mind	p_atr12_mmed			
a_atr19_fppol	p_atr17_fdpol	p_atr8_manu			
p_atr3_fpmin	p_atr16_mind	p_atr5_fdanu			
p_atr4_mmin	p_atr19_fppol	p_atr19_fppol			
p_atr6_fmanu	p_atr2_fmmin	a_atr11_fpmed			
p_atr7_fpanu	p_atr6_fmanu				
p_atr10_fmmed	a_atr17_fdpol				
p_atr11_fpmed	p_atr1_fdmin				
p_atr12_mmed	p_atr19_fppol				
p_atr14_fmind	a_atr7_fpanu				
p_atr15_fpind	p_atr11_fpmed				
p_atr16_mind	p_atr7_fpanu				
p_atr19_fppol	a_atr15_fpind				
p_atr20_mpol	a_atr5_fdanu				
	p_atr12_ mmed				

O conjunto de atributos selecionado pelo algoritmo *CSF* apresentou um percentual médio de concordância com a média dos laudos de 91,3%, enquanto os algoritmos *M5P* e *RepTree* obtiveram um percentual de 88,0% e 85,0%, respectivamente. Os valores percentuais foram apenas utilizados como "guia", os resultados adotados como base para a tomada de decisão foram os do teste *T-Student*. Os resultados da utilização dos mineradores foram aplicados ao teste *T-Student* a fim de verificar se os mesmos estavam dentro da faixa de significância adotada, neste caso específico 0,05. Os valores obtidos com o algoritmo *CSF* serão apresentados posteriormente, pois o algoritmo apresentou taxa média de significância de 0,16, sendo escolhido para aplicação em outros testes. Como pode ser observado na Tabela 8,

os valores obtidos pela avaliação do algoritmo *RepTree*, estão acima da taxa de significância, com média de 0,20 sendo que a faixa etária de 16 anos foi a faixa que mais se aproximou da taxa de significância adotada, o algoritmo M5P apresentou na avaliação taxa de significância média de 0,19, também acima do adotado. No caso do algoritmo *M5P*, nenhuma das faixas etárias aproximou-se do adotado.

**Tabela 8** – Resultado da aplicação do teste *T-Student*, para os atributos encontrados com algoritmo *RepTree* para todas as faixas etárias.

Avaliação	RepTree
Idade	
6 anos	0,20
7 anos	0,18
8 anos	0,18
9 anos	0,17
10 anos	0,30
11 anos	0,19
12 anos	0,18
13 anos	0,18
14 anos	0,18
15 anos	0,41
16 anos	0,09
Média	0,20

As métricas também foram separadas e aplicadas ao algoritmo *StARminer*. Na Tabela 9, a seguir, são vistos os parâmetros utilizados e as características selecionadas para a métrica Área.

**Tabela 9 -** Atributos da Base Geral selecionados com o algoritmo *StARminer* e parâmetros utilizados na seleção para Área; em destaque, os atributos utilizados pela metodologia de Olivete Jr. (2005a).

Valores de limiares definidos pelo usuário:			Características selecionadas	
	$\Delta\mu_{min,}\sigma_{max,}\gamma_{min,}$	Qtdaa a mo		
Área	0.95 0.15 0.30	início	fim	a_atr1_fdmin
		0	666	a_atr5_fdanu
				a_atr8_manu
				a_atr9_fdmed
				a_atr13_fdind

Para resultar uma análise mais elaborada, as métricas foram separadas por faixa etária. Essas informações puderam apenas ser aplicadas ao *Weka*, pois o *StARminer* necessita que a classe dos dados (neste caso a idade) seja discretizada. Nesta análise específica, utilizaram-se as amostras das Bases Fem e Masc separadamente. A fim de ilustrar tal análise, a Tabela 10 apresenta os atributos selecionados pelo *Weka*, para as amostras femininas da faixa etária 6 anos, da métrica Área.

**Tabela 10** – Atributos selecionados para Base Fem para faixa etária de 6 anos métrica-Área; em destaque, os atributos utilizados pela metodologia de Olivete Jr.(2005a).

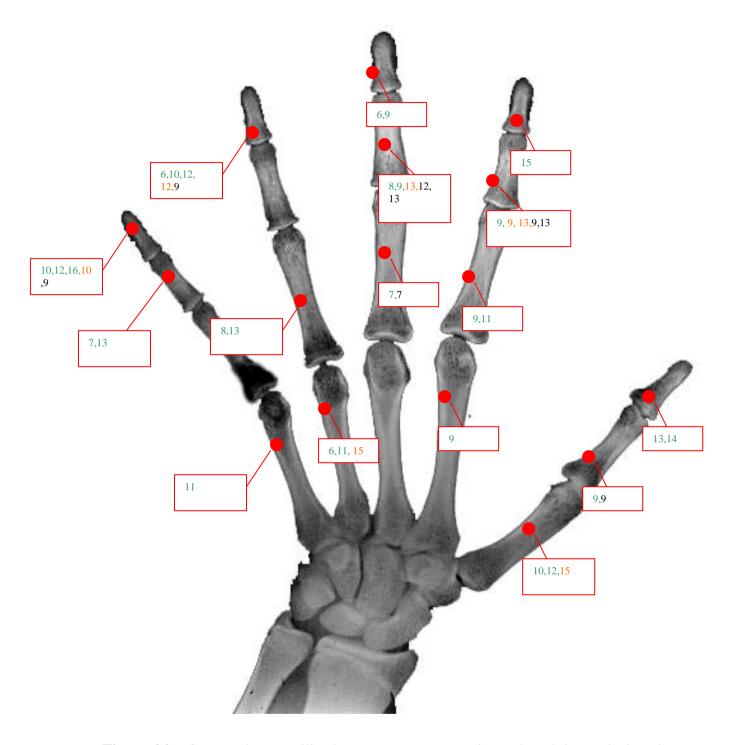
6 anos					
CSF	RepTree	M5P			
a_atr5_fdanu	a_atr16_mind	Sem atributos			
a_atr8_manu					
a_atr9_fdmed					
a_atr16_mind					

O mesmo procedimento foi realizado com a Base Masc, assim, além de avaliar o uso dos mineradores de forma global, pode-se também fazer uma análise local dos atributos selecionados. A lista de atributos selecionados para a faixa etária de 6 anos para métrica Área da Base Masc, pode ser vista na Tabela 11.

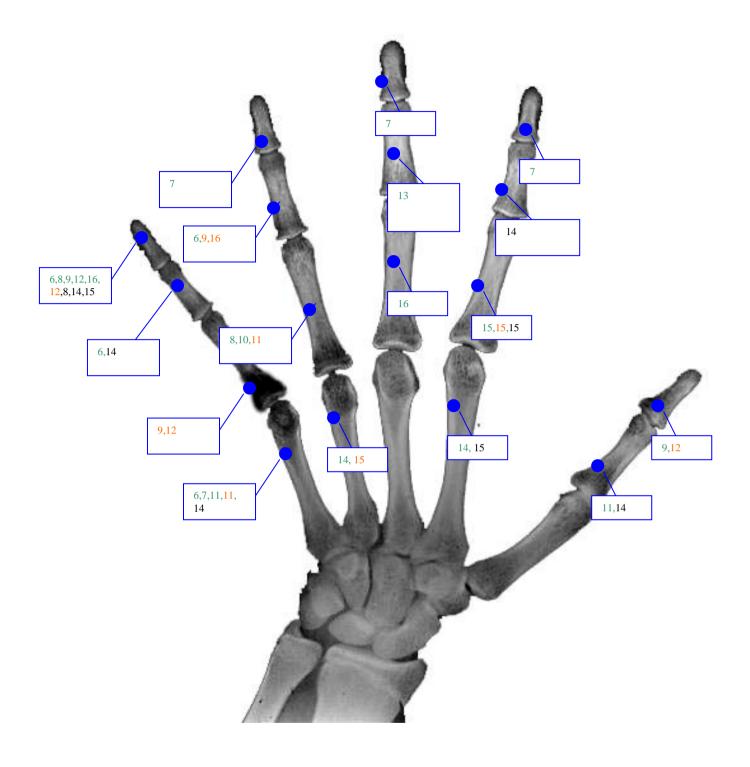
**Tabela 11** - Atributos selecionados para Base Masc para faixa etária de 6 anos - métrica Área; não foram selecionados atributos utilizados pela metodologia de Olivete Jr. (2005a).

6 anos					
CSF	RepTree	M5P			
a_atr1_fdmin	Sem atributos	Sem atributos			
a_atr2_fmmin					
a_atr4_mmin					
a_atr6_fmanu					

As figuras a seguir ilustram a contribuição de cada um dos atributos, que a partir de agora serão chamados de centros ósseos para composição dos conjuntos selecionados. Como visto anteriormente, as métricas separadas por faixa etária foram aplicadas apenas aos algoritmos do *Weka*. Nas Figuras 24 e 25, são identificados pelas cores de fonte verde (algoritmo *CSF*), laranja (algoritmo *RepTree*) e preto (algoritmo *M5P*) os centros ósseos selecionados pelos algoritmos do *Weka*, para a métrica Área, das bases Fem e Masc respectivamente. Os números presentes nos quadros representam a faixa etária em que o centro ósseo é selecionado como relevante.



**Figura 24** – Centros ósseos utilizados para compor o conjunto de métricas selecionadas para cada faixa etária da Base Fem, utilizando algoritmos CSF, RepTree e M5P, destacados em cores - métrica Área.

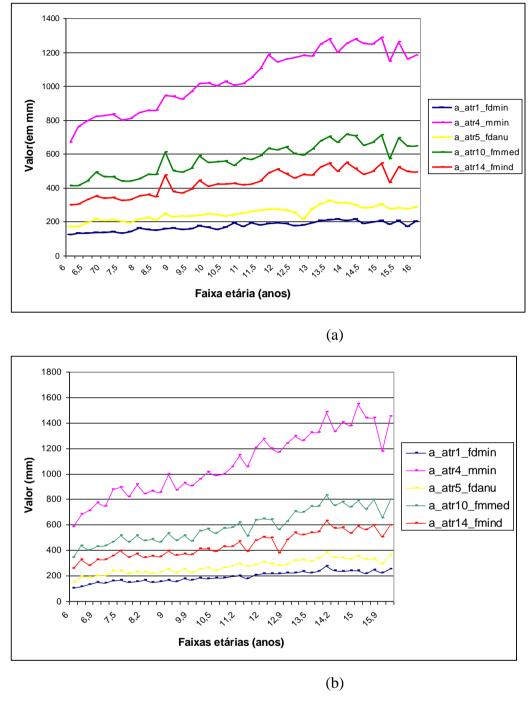


**Figura 25** – Centros ósseos utilizados para compor o conjunto de métricas selecionadas para cada faixa etária da Base Masc, utilizando algoritmos CSF, RepTree e M5P, destacados em cores - métrica Área.

Além dos centros ósseos vistos nas figuras acima, selecionados para a métrica Área, os centros ósseos selecionados para as outras duas métricas, foram também avaliados. Todos os centros ósseos que se destacaram podem ser vistos nos gráficos das Figuras 26, 27 e

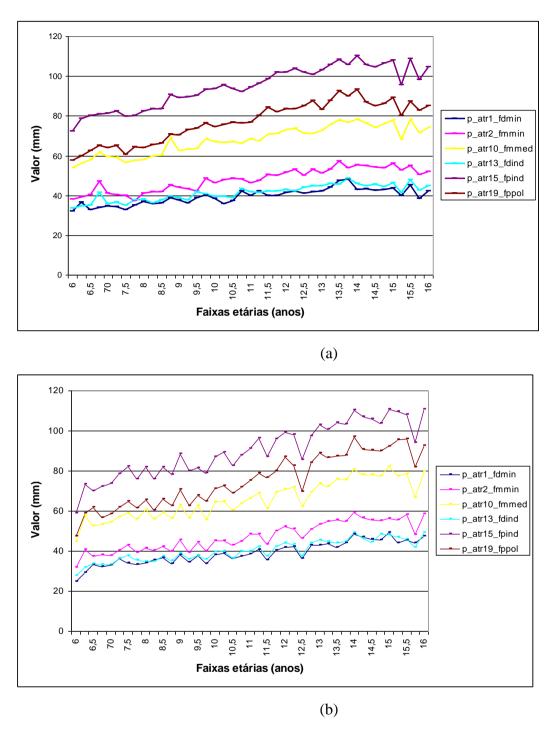
28. No caso da métrica Perímetro para a Base Fem, os centros ósseos que se destacaram foram: falange distal do dedo indicador (*fdmin*) e falange proximal do dedo anular (*fpanu*). Quanto à Base Masc os centros ósseos que se destacaram foram: falange distal do dedo mínimo (*fdmin*), falange média do dedo médio (*fmmed*) e falange proximal do dedo indicador (*fpind*). Finalmente para a métrica Comprimento os centros ósseos que se destacaram para a Base Fem foram: falange distal dos dedos indicador (*fdind*), e mínimo (*fdmin*). E os seguintes centros ósseos para a Base Masc: falange distal do dedo mínimo (*fdmin*), falange média dos dedos mínimo (*fmmin*) e médio (*fmmed*) e falange proximal do dedo indicador (*fpind*).

Na Figura 26 são mostrados os gráficos de desenvolvimento médio dos centros ósseos selecionados com a métrica Área para as bases Fem e Masc respectivamente. Em cores, estão representados os centros, falange média do dedo médio (verde), falange média do indicador (vermelho), falange distal do anular (amarelo), falange distal do mínimo (azul) e metacarpo do mínimo (rosa).



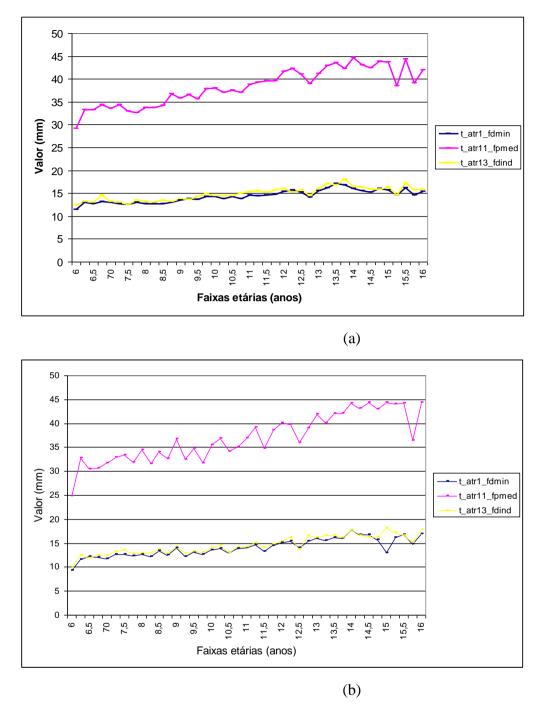
**Figura 26** - Desenvolvimento médio dos centros ósseos selecionados com a métrica Área para (a) Base Fem e (b) Base Masc.

Na Figura 27 é mostrado o gráfico de desenvolvimento para os centros ósseos selecionados com a métrica Perímetro, para as duas bases. Em cores, estão representados os centros, falange distal do dedo mínimo (azul escuro), falange média do dedo mínimo (rosa), falange média do dedo médio (amarelo), falange distal do dedo indicador (azul claro), falange proximal do indicador (ameixa escuro) e falange proximal do polegar (marrom).



**Figura 27** - Desenvolvimento médio dos centros ósseos selecionados com a métrica Perímetro para (a) Base Fem e (b) Base Masc.

Na Figura 28 é mostrado o gráfico de desenvolvimento para os centros ósseos selecionados com a métrica Comprimento, para as duas bases. Em cores, estão representados os centros, falange distal do indicador (amarelo), falange distal do mínimo (azul escuro) e falange proximal do dedo médio (rosa).



**Figura 28** - Desenvolvimento médio dos centros ósseos selecionados com a métrica Comprimento para (a) Base Fem e (b) Base Masc.

A partir dos gráficos observa-se que os mesmos apresentam uma curva de comportamento crescente e isso só foi possível ser observado, pois a base de dados utilizada nesta pesquisa contém informações sobre o desenvolvimento ósseo para todas as faixas etárias, porém os profissionais da área médica nem sempre contam com todas essas informações para análise. Esse comportamento crescente e bem definido mostrado nos gráficos anteriores oferece a possibilidade de utilização dessas informações como uma forma

de auxílio na estimação de idade óssea, ou seja, observou-se que as falanges distais dos dedos indicador e mínimo (*fdind* e *fdmin*) podem contribuir em processos de estimação. Isso indica que elas merecem ser analisadas com maior cuidado, o que não foi realizado no presente trabalho, pois faz parte das sugestões para trabalhos futuros.

Embora os mineradores tenham selecionado alguns centros ósseos como importantes apenas para uma das bases (Fem ou Masc), as avaliações a respeito do comportamento dos centros selecionados foram realizadas para as duas bases. Tal procedimento foi realizado com objetivo de verificar o comportamento das informações fornecidas pelo processo de mineração, pois observou-se que a utilização "cega" do conjunto de centros ósseos identificados pelos mineradores não produziu resultados significativos quando comparados com a média dos laudos. Apesar dessa observação, essa avaliação realizada para as duas bases, permitiu verificar que os centros ósseos identificados pelos mineradores contêm informações importantes e indicam a possibilidade de aplicação em processos de estimação.

## 8.4 - A aplicação das redes neurais artificiais.

Como já mencionado, nesta pesquisa utilizou-se a rede neural *Perceptron* multicamadas, com os algoritmos *Levenberg-Marquardt e resilient-Backpropagation*. O algoritmo *LVQ* foi testado; porém, a rede não apresentou convergência. Os parâmetros e topologias utilizados pelos algoritmos são apresentados abaixo, no qual *L–M* refere-se ao algoritmo *Levenberg-Marquardt* e *r–B* ao algoritmo *resilient-Backpropagation*.

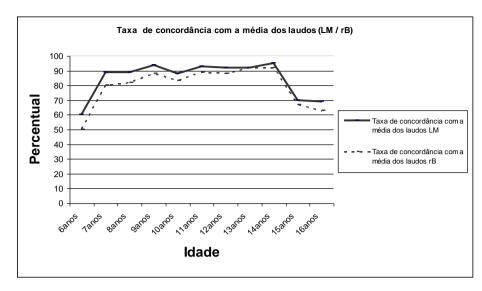
A Tabela 12 apresenta as topologias de rede neurais utilizadas e o seu desempenho. É importante ressaltar que foram utilizadas 11 entradas e 11 saídas para os testes com a Base Geral, que contém todas as faixas etárias utilizadas nesta pesquisa. O mesmo foi alterado de acordo com o teste a ser realizado, por exemplo, quando o classificador neural foi aplicado a cada uma das faixas etárias separadas, utilizou-se 4 neurônios de entrada e 4 de saída.

 Tabela 12 - Parâmetros utilizados nas redes neurais com os algoritmos
 Levenberg-Marquardt e

resilient-Backpropagation.

Algoritmo	Topologia	Matriz de		Fun	ção de	ativação	Taxa de	Erro
							Aprendizagem	
L - M	11 entradas	Matrizes	iniciadas	Saída	dos	neurônios:	0.001	10 <sup>-4</sup>
	1ª camada:	randomicament	te.	Tangen	te Hipe	erbólica		
	30			Saída d	la redo	e neural:		
	neurônios			Linear.				
	2ª camada:							
	20							
	neurônios							
r - B	11 entradas	Matrizes	iniciadas	Saída d	los neu	urônios:	0.001	10 <sup>-4</sup>
	1ª camada:	randomicament	te.	Tangen	te Hip	erbólica		
	30			Saída d	la redo	e neural:		
	neurônios			Linear.				
	2ª camada:							
	20							
	neurônios							

A Figura 30 apresenta um gráfico da análise do desempenho médio de cada uma das topologias, quando aplicadas as informações referentes à Base Geral, conforme a taxa percentual de concordância com a média dos laudos. Foram realizados seis testes, variando épocas e o número de camadas intermediárias. As épocas utilizadas foram 100 e 150; outros testes foram feitos com um número maior de épocas, mas, como o algoritmo não apresentou alteração em seu resultado final, estes números foram mantidos para todos os testes. Quanto ao número de camadas intermediárias, o mesmo teve uma variação de 20 a 30 neurônios, dependendo da quantidade de camadas utilizadas. Por exemplo, quando o teste foi feito com uma camada, utilizaram-se 30 neurônios; com duas camadas, 30 neurônios na primeira e 20 na segunda; com três camadas, utilizaram-se 30 neurônios nas duas primeiras camadas e 20 na terceira camada. O gráfico da Figura 29 ilustra o comportamento levando-se em consideração todos os 6 testes feitos para cada uma das topologias de rede.



**Figura 29** – Gráfico taxa média de concordância dos algoritmos com a média dos laudos nos testes de validação (*LM/rB*) Base Geral.

Como se pode observar na Figura 29, com a aplicação da topologia de rede *LM* obteve-se uma média percentual de 85,0% de concordância, e para o teste *rB* uma média percentual de 80,0%. Esses resultados também foram avaliados com o teste *T-Student*, utilizando uma taxa de significância de 0,05, com objetivo de auxiliar na decisão de escolha da topologia a ser aplicada em testes futuros. A rede *LM* apresentou, de acordo com o teste *T-Student*, uma taxa de significância média de 0,15, como pode ser visto na Tabela 13, enquanto a rede *rB* apresentou uma média de significância de 0,20.

**Tabela 13 -** Resultado da aplicação do teste *T-Student*, para os resultados obtidos com a rede neural *LM* para todas as faixas etárias (Base Geral).

Avaliação	Base Geral
Idade	LM
6 anos	0,12
7 anos	0,08
8 anos	0,18
9 anos	0,17
10 anos	0,09
11 anos	0,19
12 anos	0,18
13 anos	0,14
14 anos	0,18
15 anos	0,13
16 anos	0,20
Média	0,15

As faixas etárias 7 e 10 anos foram as que produziram resultados mais próximos da taxa de significância adotada (0,05), nesta avaliação da rede *LM*. No caso da rede *rB*, apenas a faixa etária de 9 anos produziu resultado próximo ao adotado (0,09).

Não foi realizada nenhuma análise a respeito do tempo gasto para teste e treinamento, pois objetivou-se encontrar caminhos que levassem à definição de novas metodologias e não acerca de custos computacionais. Para todos os treinamentos da rB, o algoritmo converge mais rápido que a LM, porém, a taxa de erro estimado não se aproxima da taxa de erro estabelecida. É interessante destacar que para todos os conjuntos e testes de validação, este algoritmo apresentou o mesmo comportamento.

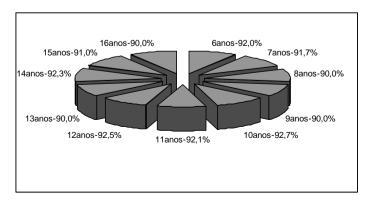
Por outro lado, com relação aos testes feitos com o algoritmo *LM*, obtiveram-se comportamentos superiores ao *rB* e melhores resultados de testes de validação. A Figura 29 mostra esse comportamento.

Ao contrário da topologia *rB*, no caso da topologia *LM*, a taxa de erro tende a se aproximar da taxa de erro estabelecida. Por ter apresentado uma taxa média de significância mais próxima da adotada, a topologia *LM* foi eleita para aplicação em outros testes.

A topologia eleita como a de melhor desempenho é composta de duas camadas neurais intermediárias, com 30 e 20 neurônios respectivamente, e obteve convergência com 100 épocas de treinamento.

A seguir, nas Figuras 30 e 31 são apresentados os gráficos com os percentuais alcançados de acordo com a taxa de concordância com a média dos laudos, para cada faixa etária em cada um dos algoritmos utilizados; para a métrica Área, a mesma avaliação foi feita com Perímetro e Comprimento, porém não são apresentados os resultados.

A Figura 30 apresenta os resultados obtidos com a métrica Área para cada uma das faixas etárias de ambos os sexos; o resultado percentual médio de concordância com a média dos laudos foi de 91,3%.



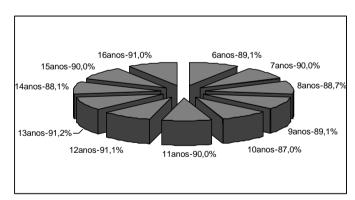
**Figura 30** – Gráfico taxa de concordância com a média dos laudos (Área *Weka*) para cada faixa etária Base Geral.

Nos gráficos referentes ao *Weka* estão os resultados obtidos pelo algoritmo *CSF*, cuja taxa de significância média foi de 0,16, apresentada na Tabela 14. Essa média foi a média de significância mais próxima da adotada obtida entre os algoritmos utilizados.

**Tabela 14** – Resultado da aplicação do teste *T-Student*, para os centros ósseos encontrados com algoritmo *CSF* para todas as faixas etárias Base Geral.

Avaliação	CSF
Idade	
6 anos	0,20
7 anos	0,09
8 anos	0,18
9 anos	0,15
10 anos	0,17
11 anos	0,19
12 anos	0,18
13 anos	0,08
14 anos	0,18
15 anos	0,28
16 anos	0,11
Média	0,16

Na Figura 31 tem-se a mesma análise, feita com o algoritmo *StARminer*; o mesmo apresentou um resultado percentual médio de concordância com a média dos laudos de 89,5%.



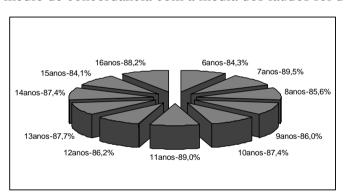
**Figura 31** - Gráfico taxa de concordância com a média dos laudos (Área *StARminer*) para cada faixa etária Base Geral.

O teste com o algoritmo *StARminer* também foi avaliado pelo *T-Student*, resultando em uma taxa média de significância de 0,22, sendo que as faixas etárias de 15 e 13 anos foram as que mais se aproximaram da taxa de significância adotada, como pode ser visto na Tabela 15.

Tabela 15 -	Resultado	da	aplicação	do	teste	T-Student,	para	os	centros	ósseos	encontrados	com
	algoritmo	StA	Rminer pa	ra to	odas a	s faixas etái	rias B	ase	Geral.			

Avaliação	Base Geral
Idade	Área -S
6 anos	0,33
7 anos	0,17
8 anos	0,18
9 anos	0,37
10 anos	0,47
11 anos	0,19
12 anos	0,18
13 anos	0,08
14 anos	0,18
15 anos	0,04
16 anos	0,20
Média	0,22

A seguir, a Figura 32 apresenta um dos resultados obtidos com a métrica Área para as informações da Base Fem. Para a avaliação feita com a aplicação do algoritmo *CSF*, o resultado percentual médio de concordância com a média dos laudos foi de 86,8%.



**Figura 32** - Gráfico taxa de concordância com a média dos laudos (Área *Weka*) para cada faixa etária Base Fem.

Os resultados obtidos com as informações da Base Fem também foram aplicados ao teste *T-Student* e podem ser observados na Tabela 16.

**Tabela 16 -** Resultado da aplicação do teste *T-Student*, para os centros osseos encontrados com algoritmo (*Weka -CSF*) para todas as faixas etárias Base Fem.

Avaliação	Base Fem
Idade	Área - W
6 anos	0,20
7 anos	0,18
8 anos	0,20
9 anos	0,18
10 anos	0,19
11 anos	0,18
12 anos	0,21
13 anos	0,21
14 anos	0,12
15 anos	0,20
16 anos	0,28
Média	0,19

Assim como para os demais testes, os resultados obtidos para a Base Fem (com *Weka*) foram aplicados ao *T-Student*, a taxa de significância média obtida com aplicação teste foi de 0,19 e a faixa etária que produziu resultados mais próximos do adotado foi 14 anos, como pode ser visto na Tabela 16.

O teste também foi aplicado aos resultados da Base Fem com algoritmo *StArminer* a taxa média de significância obtida foi de 0,23 como pode ser visto na Tabela 17. Para a Base Masc, o mesmo procedimento foi repetido.

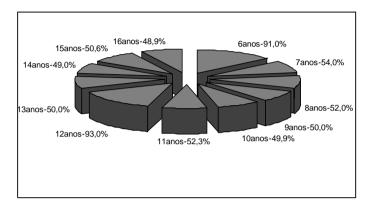
**Tabela 17 -** Resultado da aplicação do teste *T-Student*, para os centros ósseos encontrados com algoritmo (*StARminer*) para todas as faixas etárias Base Fem.

Avaliação	Base Fem
Idade	Área - S
6 anos	0,10
7 anos	0,36
8 anos	0,12
9 anos	0,30
10 anos	0,39
11 anos	0,10
12 anos	0,21
13 anos	0,20
14 anos	0,33
15 anos	0,20
16 anos	0,21
Média	0,23

As faixas etárias que produziram resultados mais próximos da taxa de significância adotada foram 6, 8 e 11 anos.

Ainda foram realizados testes com as Bases Fem e Masc separadas por faixa etária. Os centros ósseos separados para cada uma das faixas etárias foram aplicados também às outras faixas etárias para que o desempenho desses centros fosse analisado.

A Figura 33 apresenta o gráfico dessa análise para a faixa etária de 6 anos da Base Masc. O resultado percentual médio de concordância com a média dos laudos foi de 58,2%. O mesmo teste foi aplicado a cada uma das outras faixas etárias das Bases Fem e Masc.



**Figura 33** - Taxa de concordância com a média dos laudos para cada faixa etária, utilizando centros ósseos selecionados para 6 anos (Base Masc - Área)

A Tabela 18 apresenta os resultados referentes à aplicação do *T-Student*, para os resultados obtidos com a faixa etária de 6 anos da Base Masc onde pode-se observar que nenhuma das faixas etárias obteve taxa de significância próxima do adotado.

**Tabela 18 -** Resultado da aplicação do teste *T-Student*, para os centros ósseos encontrados para a faixa etária de 6 anos com (*Weka -CSF*) para todas as faixas etárias Base Masc.

Avaliação	Base Masc
Idade	
6 anos	0,20
7 anos	0,18
8 anos	0,20
9 anos	0,18
10 anos	0,19
11 anos	0,18
12 anos	0,21
13 anos	0,21
14 anos	0,18
15 anos	0,20
16 anos	0,21
Média	0,19

Os centros ósseos mais vezes selecionados no processo de mineração, em cada uma das faixas etárias, para cada umas das métricas e bases também foram avaliados. Os

resultados dessa avaliação podem ser observados nas Tabelas 19 e 20 e seus comportamentos apresentados nas Figuras 34 e 35.

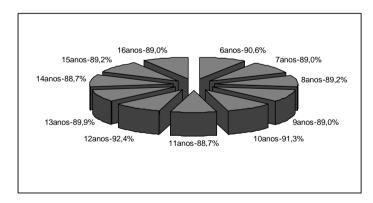
Tabela 19 – Centros ósseos mais vezes selecionados em todas as faixas etárias para Base Fem.

Área	Perímetro	Tamanho	
a_atr1_fdmin	p_atr7_fpanu	t_atr5_fdanu	
a_atr5_fdanu	p_atr9_fdmed	t_atr7_fpanu	
a_atr8_manu	p_atr11_fmmed	t_atr12_mmed	
a_atr10_fmmed	p_atr13_fdind	t_atr13_fdind	
	p_atr19_fppol	t_atr17_fdpol	

**Tabela 20** – Centros ósseos mais vezes selecionados em todas as faixas etárias para Base Masc.

Área	Perímetro	Tamanho	
a_atr1_fdmin	p_atr1_fdmin	t_atr1_fdmin	
a_atr4_mmin	p_atr10_fmmed	t_atr4_mmin	
a_atr6_fmanu	p_atr17_fdpol	t_atr6_fmanu	
a_atr7_fpanu		t_atr10_fmmed	

A Figura 34 mostra o comportamento de cada uma das faixas etárias da Base Fem, quando aplicadas ao classificador neural utilizando os centros ósseos mais vezes selecionados. O resultado percentual médio de concordância com a média dos laudos obtidos foi de 89,7%.



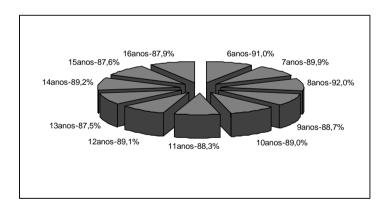
**Figura 34** - Taxa de concordância com a média dos laudos para cada faixa etária, quando aplicadas aos centros ósseos mais vezes selecionados (Base Fem – Área).

Na Tabela 21, pode-se observar os resultados obtidos pelos centros ósseos mais vezes selecionados da Base Fem quando aplicadas ao teste *T-Student*. A taxa média de significância produzida foi de 0,17, sendo que as faixas etárias de 12 e 13 anos foram as que produziram os resultados mais próximos da taxa adotada.

**Tabela 21** – Resultado da aplicação do teste *T-Student*, para os centros ósseos mais vezes selecionados com algoritmo (*Weka* ) para todas as faixas etárias Base Fem.

Avaliação	Base Fem
Idade	Relevantes
6 anos	0,20
7 anos	0,18
8 anos	0,20
9 anos	0,18
10 anos	0,19
11 anos	0,18
12 anos	0,12
13 anos	0,09
14 anos	0,18
15 anos	0,16
16 anos	0,21
Média	0,17

A mesma avaliação foi realizada com a Base Masc, e pode ser vista na Figura 35, o resultado percentual médio de concordância com a média dos laudos foi de 89,1%.



**Figura 35** – Taxa de concordância com a média dos laudos para cada faixa etária, quando aplicadas aos centros ósseos mais vezes selecionados (Base Masc – Área).

Esses resultados da Base Masc, assim como os anteriores, foram aplicados ao teste *T-Student* e a taxa média de significância obtida foi de 0,18, como pode ser visto na Tabela 22, sendo que as faixas etárias 6, 8, 11 e 15 anos atingiram valores próximos ao da taxa de significância adotada.

**Tabela 22 -** Resultado da aplicação do teste *T-Student*, para os centros ósseos mais vezes selecionados com algoritmo (*Weka*) para todas as faixas etárias Base Masc.

Avaliação	Base Masc
Idade	Relevantes
6 anos	0,10
7 anos	0,17
8 anos	0,11
9 anos	0,26
10 anos	0,17
11 anos	0,11
12 anos	0,28
13 anos	0,28
14 anos	0,23
15 anos	0,12
16 anos	0,17
Média	0,18

Observando todas as análises gráficas anteriores e os resultados referentes ao teste T-Student, é necessária uma avaliação sobre os resultados percentuais obtidos com o trabalho dos mineradores. Percebe-se que os mesmos, quando aplicados a todas as métricas da base CarpEven selecionaram vários centros ósseos, inclusive centros aplicados em outras metodologias, porém não alcançaram os resultados esperados (comparação com a média dos laudos). Foram então necessárias novas formas de combinação entre os centros de ossificação. A base CarpEven foi separada por sexo e aplicada aos mineradores, o que gerou a identificação de centros ósseos distintos daqueles encontrados anteriormente, porém os resultados ainda não puderam ser considerados satisfatórios - sempre comparando com a média dos laudos. Analisou-se também as Bases separadas por métricas (Área, Perímetro e Comprimento), com os resultados também não alcançando a taxa de significância no intervalo adotado (0,05), apesar da seleção de diferentes centros ósseos e da indicação que a métrica Comprimento poderia ser melhor avaliada devido aos resultados apresentados. A fim de refinar o processo de mineração as Bases Fem e Masc foram analisadas por faixa etária. Os centros ósseos selecionados como potencialmente discriminantes poderiam ser melhor observados e combinados entre si ou a outros. Analisando as combinações e a contribuição de cada centro ósseo, com a aplicação do teste *T-Student*, iniciou-se um processo de interferência nas combinações. Isso foi realizado com base nas metodologias existentes e nos centros ósseos reconhecidos como importantes, a fim de se selecionar um conjunto, ou conjuntos, de centros ósseos que fossem capazes de produzir resultados mais precisos, ou seja, resultados mais próximos da média dos laudos.

## 8.5 - Combinações centros ósseos.

Foram efetuadas combinações entre os centros ósseos mais relevantes encontrados com a aplicação dos mineradores, levando-se em consideração os resultados obtidos pela rede neural utilizada e o teste *T-Student*. Buscou-se identificar centros ósseos que fossem capazes de estimar a idade com taxa de concordância elevada, com relação à média dos laudos e taxa de significância dentro do adotado (0,05).

Para a formação dos conjuntos, a avaliação foi baseada na análise dos centros que mais vezes foram selecionados e nas combinações utilizadas pelas metodologias anteriores.

Os centros de ossificação utilizados nas análises são formados pelos metacarpos dos dedos indicador e anular (*mind*, *manu*), falange proximal do dedo médio (*fpmed*) e falanges distais dos dedos polegar e médio (*fdpol,fdmed*).

Para o primeiro conjunto analisado foram utilizados os metacarpos dos dedos indicador e anular (*mind* e *manu*) e a falange proximal do dedo médio (*fpmed*).

Os testes foram feitos com os valores dos centros de ossificação considerados em *pixel*. O objetivo disso é facilitar o trabalho de análise, pois em algumas situações, pode ocorrer ausência de informação de como a imagem foi digitalizada e qual a sua resolução.

A seguir são apresentadas as equações das combinações, os resultados referentes a cada uma das análises feitas e os gráficos que foram gerados.

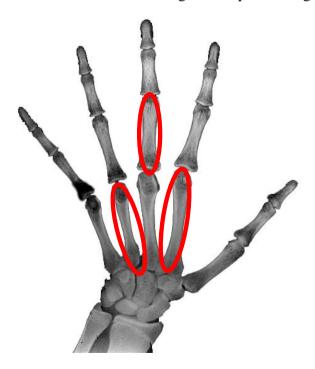


Figura 36 – Centros ósseos utilizados para compor o conjunto da Análise I.

Para essa primeira análise os valores dos comprimentos dos metacarpos foram elevados ao cubo com objetivo de amplificar a discrepância entre os valores dos centros de ossificação em cada fase do desenvolvimento. Posteriormente foram divididos por 1000, para que essa discrepância se mantivesse dentro de uma faixa de valor, que facilitasse as análises, conforme pode-se observar na equação 11.

Dimensão 
$$(pixel)^2 = \left(\frac{(\min d^3 + (manu^3))}{fpmed}\right)$$
 (11)

na qual:  $mind - \acute{e}$  o metacarpo do dedo indicador, manu - metacarpo do dedo anular e fpmed - falange proximal do dedo médio.

Com a aplicação da equação 11, foram construídas novas tabelas de relação *pixels* versus faixa etária, em que a faixa etária é tomada pela média da idade de cada um dos centros de ossificação.

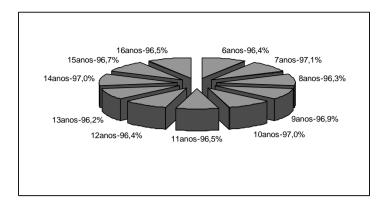
Na Tabela 23 são mostradas as relações para a Base Fem.

**Tabela 23** – Valores dimensionais (*pixels*) <sup>2</sup> para cada idade óssea (Base Fem) – Análise I, conforme equação 11.

Análise I		Análise I	
	Dimensão		Dimensão
Idade (anos)	(pixels) <sup>2</sup>	Idade (anos)	(pixels) <sup>2</sup>
6	40.13	11.25	75.15
6.25	41.45	11.5	76.97
6.5	42.89	11.75	79.13
6.75	44.35	12	80.91
7	45.75	12.25	82.79
7.25	47.43	12.5	85.05
7.5	48.87	12.75	86.89
7.75	50.53	13	89.11
8	52.1	13.25	91.19
8.25	53.71	13.5	93.35
8.5	55.35	13.75	95.48
8.75	56.9	14	97.66
9	58.69	14.25	99.87
9.25	60.38	14.5	102.07
9.5	62.23	14.75	104.35
9.75	63.88	15	106.61
10	65.78	15.25	109.25
10.25	67.47	15.5	111.89
10.5	69.29	15.75	113.88
10.75	71.26	16	114.69
11	73.03		

A seguir são apresentados os gráficos para cada uma das bases, em cada uma das faixas etárias que essa pesquisa abrange, com os centros ósseos que compõem a Análise I.

Para a Base Fem, como pode ser visto na Figura 37 a taxa percentual média de concordância com a média dos laudos é de 96.6%.



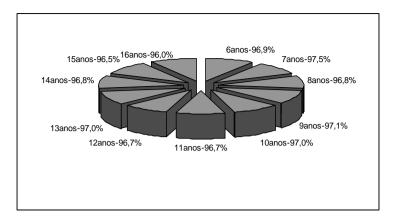
**Figura 37 -** Taxa de concordância com a média dos laudos para cada faixa etária, quando aplicadas aos centros ósseos utilizados na Análise I (Base Fem).

Na Tabela 24, são apresentados os resultados obtidos com a aplicação do teste *T-Student* para a Análise I-Base Fem. Nota-se que a taxa média de significância 0,06 está próxima do adotado e apenas as faixas etárias de 7, 8, 13 e 16 anos apresentaram valores fora da taxa de significância (0,05).

**Tabela 24** – Resultado da aplicação do teste *T-Student*, para os centros ósseos utilizados na Análise I para todas as faixas etárias Base Fem.

Avaliação	Base Fem	
Idade	Análise I	
6 anos	0,04	
7 anos	0,06	
8 anos	0,09	
9 anos	0,04	
10 anos	0,04	
11 anos	0,04	
12 anos	0,02	
13 anos	0,09	
14 anos	0,06	
15 anos	0,03	
16 anos	0,10	
Média	0,06	

Na Figura 38 apresenta-se o gráfico com as taxas percentuais para cada uma das faixas etárias da Base Masc-Análise I. Como pode ser observado para a Base Masc, a taxa percentual média de concordância com a média dos laudos é de 96,8%.



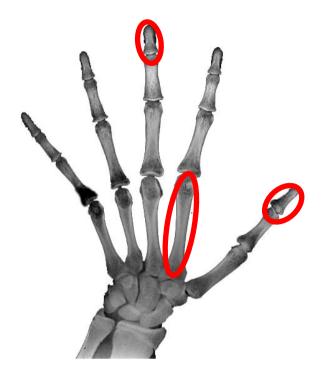
**Figura 38** - Taxa de concordância com a média dos laudos para cada faixa etária, quando aplicadas aos centros ósseos utilizados na Análise I (Base Masc).

Assim como foi realizado com a Base Fem, os resultados obtidos com a aplicação dos centros ósseos da Análise I - Base Masc, também foram submetidos ao teste *T-Student* e a taxa média de significância ficou em torno de 0,05, ou seja, dentro do adotado. Apenas as faixas etárias 6, 8 e 10 anos obtiveram valores acima.

**Tabela 25 -** Resultado da aplicação do teste *T-Student*, para os centros ósseos utilizados na Análise I para todas as faixas etárias Base Masc.

Avaliação	Base Masc	
Idade	Análise I	
6 anos	0,09	
7 anos	0,03	
8 anos	0,06	
9 anos	0,01	
10 anos	0,08	
11 anos	0,05	
12 anos	0,06	
13 anos	0,06	
14 anos	0,04	
15 anos	0,04	
16 anos	0,04	
Média	0,05	

Para o segundo conjunto analisado foram utilizados os seguintes centros ósseos: metacarpo do dedo indicador (*mind*) e as falanges distais dos dedos polegar e médio (*fdpol* e *fdmed*).



**Figura 39 -** Centros ósseos utilizados para compor o conjunto da Análise II. A Análise II obedece à equação 12:

Dimensão (pixel)= 
$$\left(\frac{\min d}{(fdpol + fdmed)}\right)$$
 (12)

na qual:  $mind - \acute{e}$  o metacarpo do dedo indicador, fdpol - falange distal do dedo polegar e fdmed - falange distal do dedo médio.

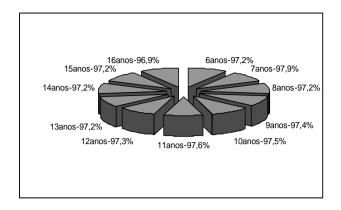
Como foi realizado nos casos anteriores, também para a Análise II foi gerada uma tabela de relação dimensional (*pixels* versus faixa etária).

A seguir na Tabela 26 podem ser observados os valores de relação para as informações referentes à Base Masc.

Tabela 26 - Valores dimensionais (pixels) para cada centro óssseo (Base Masc) – Análise II.

Anális	e II	Anális	se II
	Dimensão		Dimensão
Idade (anos)	(pixels)	Idade (anos)	(pixels)
6	192,79	11,25	256,43
6.25	192.79	11.5	260.09
6.5	192.79	11.75	265.78
6.75	199.65	12	271.46
7	206.52	12.25	277.15
7.25	213.4	12.5	282.84
7.5	220.29	12.75	284.35
7.75	221.05	13	285.87
8	221.81	13.25	287.38
8.25	222.57	13.5	288.89
8.5	223.33	13.75	293.52
8.75	225.41	14	298.16
9	227.49	14.25	302.8
9.25	229.57	14.5	307.46
9.5	231.66	14.75	311.28
9.75	235.11	15	315.11
10	238.56	15.25	318.95
10.25	242.02	15.5	322.8
10.5	245.49	15.75	323.75
10.75	249.13	16	324.7
11	252.78		

Como pode ser visto na Figura 40, a taxa percentual média de concordância em relação à média dos laudos, para a Base Fem com os valores da Análise II, é de 97,3%.



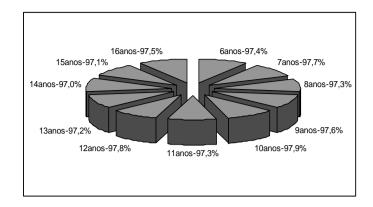
**Figura 40 -** Taxa de concordância com a média dos laudos para cada faixa etária, metacarpo do dedo indicador (*mind*) e as falanges distais dos dedos polegar e médio (*fdpol e fdmed*) - Análise II (Base Fem).

Assim como foi realizado na Análise I, os resultados obtidos com a Análise II para a Base Fem foram submetidos ao teste *T-Student* e a taxa média de significância ficou em torno de 0,04, dentro da taxa adotada. Como pode ser observado também, apenas as faixas etárias para 14 e 16 anos ficaram acima.

**Tabela 27 -** Resultado da aplicação do teste *T-Student*, para os centros ósseos utilizados na Análise II para todas as faixas etárias Base Fem.

Avaliação	Base Fem	
Idade	Análise II	
6 anos	0,03	
7 anos	0,02	
8 anos	0,03	
9 anos	0,01	
10 anos	0,04	
11 anos	0,04	
12 anos	0,02	
13 anos	0,04	
14 anos	0,06	
15 anos	0,03	
16 anos	0,09	
Média	0,04	

Para a Base Masc a taxa percentual média de concordância com a média dos laudos foi de 97,4%, quando aplicadas aos centros ósseos utilizados na Análise II, como pode ser visto na Figura 41.



**Figura 41** - Taxa de concordância com a média dos laudos para cada faixa etária, metacarpo do dedo indicador (*mind*) e as falanges distais dos dedos polegar e médio (*fdpol e fdmed*) - Análise II (Base Masc).

Os resultados produzidos por essa avaliação também foram aplicados ao teste *T-Student* e a taxa média de significância obtida foi 0,04. Apenas a faixa etária de 12 anos ficou acima do adotado, e neste caso específico, ocorreu devido ao número de casas decimais utilizado na apresentação dos resultados.

**Tabela 28 -** Resultado da aplicação do teste *T-Student*, para os centros ósseos utilizados na Análise II para todas as faixas etárias Base Masc.

Avaliação	Base Masc
Idade	Análise II
6 anos	0,03
7 anos	0,03
8 anos	0,03
9 anos	0,01
10 anos	0,05
11 anos	0,05
12 anos	0,06
13 anos	0,04
14 anos	0,04
15 anos	0,04
16 anos	0,04
Média	0,04

Com base nos resultados obtidos, de concordância percentual e *T-Student*, a Análise II pode ser eleita para compor uma nova metodologia. Nota-se que a taxa de concordância com a média dos laudos para essa Análise II, oscila entre 97,3% e 97,4% e a taxa média de significância obtida foi de 0,04, ou seja dentro da faixa adotada (0,05).

## 8.6 - Considerações a respeito dos resultados.

Neste capítulo, foram apresentados os resultados obtidos pelos testes, com os centros ósseos que se destacaram no processo de mineração e aplicação da rede neural. Observando os gráficos da seção anterior, conclui-se, que:

- A utilização de ferramentas de mineração na base CarpEven foi necessária devido ao volume de informações gerado com a extração das métricas. Isso permitiu que indicadores importantes fossem encontrados.
- A análise da base CarpEven sob diferentes abordagens possibilitou uma melhor análise das informações encontradas. Primeiramente foi aplicada aos mineradores de forma completa, depois separada por sexo e por métrica (Área, Perímetro e Comprimento). A métrica Comprimento apresentou resultados mais relevantes, mas as taxas de significância encontradas pelas avaliações ficaram fora da faixa adotada. Devido a esse fato os centros ósseos foram aplicados aos mineradores separados por faixa etária. Com esse procedimento observou-se avanço nos resultados, porém ainda não alcançando a taxa de significância adotada.
- O processo de mineração apontou novos centros de ossificação que ainda não haviam sido analisados em metodologias anteriores, como potencialmente discriminantes, o que resultou em uma observação mais profunda desses centros.
- As informações que resultaram de suas aplicações foram analisadas, baseando-se nos resultados do teste *T-Student*, nas metodologias existentes e no comportamento conhecido de alguns centros ósseos. Combinados com os centros que se destacaram como potencialmente discriminantes, a fim de se encontrar uma combinação de centros com capacidade para estimar a idade óssea, observou-se significativo progresso nos resultados.
- A combinação de centros associou centros já utilizados em metodologias anteriores (metacarpos), identificados como importantes, e combinados com falanges distais e proximais, selecionadas pelo processo de mineração, como potencialmente discriminantes.
- A Análise I (combinação de metacarpos e falange proximal) apresentou resultados significativos, e taxa média de significância dentro do estabelecido, porém a Análise II (combinação metacarpos e falanges distais), apresentou resultados superiores, tanto

- percentuais quanto na taxa de significância, e foi eleita para compor a nova metodologia.
- Finalmente, o fato da Análise que compõe a nova metodologia utilizar-se de apenas três centros ósseos e medidas em *pixel*, faz com que a mesma possa ser empregada para a estimação de idade óssea no auxilio ao diagnóstico clínico médico, e também oferece a possibilidade para aplicações automatizadas dirigidas à grandes bases de dados.

Desenvolver técnicas para estimação de idade óssea que encontrem informações que auxiliem e reduzam o trabalho dos profissionais da área médica, com uso de aplicações computacionais, para geração de diagnósticos precisos, rápidos e sem subjetividades presentes nos métodos clássicos, é um fator de grande relevância.

Com intuito de contribuir com as metodologias existentes, esta pesquisa investigou as associações entre as métricas extraídas das imagens carpais da Base *CarpEven*. Para tanto, foram segmentados todos os centros ósseos da Base *CarpEven*, com o objetivo de garantir precisão na análise dos dados.

A criação de uma base de dados constitui uma forma de auxiliar novas pesquisas, já que a sua constituição é baseada em métricas (Área e Perímetro), que representam um complemento às métricas existentes (Comprimento). Todas as imagens utilizadas para a geração da base de dados métrica estão disponíveis para consulta, assim como uma imagem ilustrativa do procedimento para obtenção das métricas. Devido ao grande volume de dados gerado, foi necessário aplicá-los a algoritmos que fossem capazes de encontrar informações discriminantes, "os mineradores".

O trabalho dos mineradores apontou centros ósseos potencialmente discriminantes, o que permitiu que esses centros fossem combinados com outros, baseando-se no comportamento já observado desses centros ósseos nas metodologias existentes e nos resultados obtidos com o teste *T-Student*. Formas de combinação tais como, proporções, foram utilizadas, porém não apresentaram resultados discriminantes para todas as faixas etárias.

Os resultados obtidos, com a combinação dos centros ósseos, vistos na Análise II e eleitos para compor a nova metodologia, tornam viável a aplicação destes centros em procedimentos de estimação da idade óssea, tendo em vista que apresentaram significativa taxa de concordância e significância em relação à média dos laudos (obtidos por E&R, G&P e T&W), sendo possível estimar a idade óssea baseando-se apenas nas análises das falanges distais e metacarpo.

A nova metodologia permite que imagens sejam analisadas, com significativo nível de confiança e de forma simples, apenas com a demarcação dos três centros ósseos, sem a necessidade de muitas informações a respeito da imagem, já que trabalha com medidas em

*pixel*, onde não há necessidade do prévio conhecimento a respeito da resolução em que a imagem foi digitalizada.

O fato de não utilizar uma maior quantidade de centros de ossificação, permite que as avaliações sejam feitas de forma prática, viabilizando ainda mais sua aplicação e criando possibilidades para que haja correções de desordens de crescimento. Auxilia na busca por metodologias simples que forneçam informações precisas. Não exige que os profissionais se utilizem de outras informações da imagem, como por exemplo, a resolução da mesma.

Esta pesquisa obteve êxito em dar continuidade ao trabalho que vêm sendo realizado e abrange os vários aspectos utilizados pelos métodos clássicos e outros métodos que podem ser associados a estes.

## 9.1 - Trabalhos futuros.

A seguir são listadas propostas para a continuidade desta pesquisa e sugestões para trabalhos futuros.

- Pesquisas para investigar a presença de número áureo em outros ossos da mão, utilizando a base *CarpEven* como fonte de informação, analisando as proporções existentes entre as falanges, não só em relação ao Comprimento, mas também utilizando Área e Perímetro.
- Análise baseada em regras estatística capazes de estimar idade óssea.
- Análise do comportamento das epífises.
- Análise das proporções das falanges e estudo desse desenvolvimento em relação à maturação óssea.
- Aplicação das informações da base *CarpEven* a outros mineradores.
- Aplicação dos centros ósseos utilizados para compor a nova metodologias às métricas Área e Perímetro.

A fim de situar no tempo a literatura utilizada para elaboração desta pesquisa, apresenta-se a seguir um gráfico referente à literatura utilizada de acordo com o ano de publicação e quantidade, como pode ser observado na Figura 42.

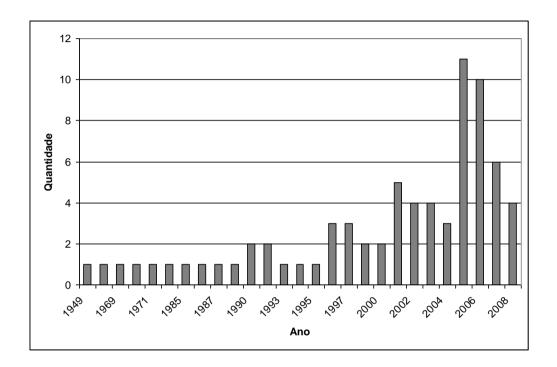


Figura 42 - Artigos e periódicos utilizados na pesquisa de acordo com ano de publicação.

ACHESON, R. M.; VINICIUS, J. H.; FOWER, G. B. (1966). Studies in the reliability of assessing skeletal maturity from X rays. Part III. Greulich-Pyle and Tanner- Whitehouse method contrasted. Hum Biol, v. 38, n. 3, páginas 204-218.

ALMEIDA, L.M.; PADILHA, T. P. P; OLIVEIRA, F. L.; PREVIERO, C. A.(2003) *Uma Ferramenta para Extração de Padrões*, Revista Eletrônica de iniciação científica, publicação SBS, ano III, volume III, número IV, dez. 2003, ISSN 1519-8219

AVANCINI, E. B.; FAVARETTO, J. A. (1997). *Uma abordagem evolutiva e ecológica*. Vol. 1 Ed. Moderna. São Paulo

BALIGA A. B. (2004) *Face illumination normalization with shadow consideration*. Dissertação de Mestrado, CARNEGIE MELLON UNIVERSITY, 2004

BASU, S.; MELVILLE;(2005) Prem. "Weka Tutorial". University of Texas at Austin. Disponível em: < http://www.cs.utexas.edu/users/ml/tutorials/Weka-tut/sld002.htm >. Acesso em 8 de outubro 2008.

BORUCKI, S.; BOCZAR,T.; CICHOŃ, A., (2006) "The application of the resilient backpropagation algorithm and power spectrum density for recognizing the acoustic emission signals generated by basic partial discharge forms using artificial neuron networks", Technical University of Opole ul. Sosnkowskiego 31, 45-272 Opole, Poland

BOSQUIEIRO,M. R.. et al.(2001). *Determination of skeletal and estimate of the age through carpal radiographics*, Rvev. Cons. Reg. Odontol. Pernambuco, v.4, n.1, p. 59-66, Janeiro/Junho.

CALLES, A.; SANNOMIYA, E. K; (2005) Comparação da idade óssea com a cronológica em indivíduos portadores da síndrome de Down pelo índice de Eklöf & Ringertz, por meio de radiografias de mão e punho. Cienc. Odontol. Bras. p. 39-44, abr/jun;8(2), 2005.

CASANOVA, M. S.; ORTEGA, A. I.; HAITER NETO,F. (2006) ALMEIDA, S. M. Análise comparativa da maturação óssea determinada pelo método de Grave-Brown entre imagens convencionais e digitalizadas, R Dental Press Ortodon Ortop Facial 104 Maringá, v. 11, n. 5, p. 104-109, set./out. 2006

COX JR, F., (2000) Programando para web com PHP /MySql,-UPE - POLI - Engenharia Eletrônica, agosto 2000.

DUARTE, E.; NEVES, T. Determinação da idade óssea, no serviço de imagiologia do HUC, Disponível em :< http://www.huc.min-saude.pt/imagiologia/area\_tecnica/determinacao\_idade\_ossea/determinacao\_idade\_ossea.swf > Acesso em 15.set.2008.

EKLOF, O.; RINGERTZ, H. (1967). "A method for assessment of skeletal maturity". Annals Radiology. vol. 10, pp 330-336.

ENLOW, D.H. (1993). Crescimento facial. 3.ed. São Paulo: Artes Médicas, 1993. p. 553.

FAYYAD, U.; PIATETSKY-SHAPIRO G.; SMYTH P.(1996) "Data mining and knowledge discovery in databases: an overview". Comm. of the ACM, vol. 39, n.11.

FAYYAD, U. M. & UTHURUSAMY, R. (2002)"Evolving Data Mining into Solutions for Insights", Comminications of the ACM: vol.45, N° 8, p. 28-31, 2002.

FERNANDES, L.G.; PORTUGAL, M.S., (1995) "Redes Neurais Artificiais e Previsão de Séries Econômicas: Uma Introdução". Universidade Federal do Rio Grande do Sul.

FERRO, M.;LEE, H. D.(2001) O Processo de KDD – Knowledge Discovery in Database para Aplicações na Medicina., SEMINC,2001.

FRANÇA, B.H.S. et al. (1998). *Estimativa de Idade em Indivíduo Adulto Jovem*, Congresso Brasileiro de Odontologia, Recife, Brasil.

FRITZ, S.L.; LIVINGSTON, W.H. (1985). The effect of anode curvature on radiographic heel effect, Med. Phys 12 (4): p.443-446

GARCIA, R. M.; *Introdução à Mineração de Dados*. [Material de pesquisa usado em aula sobre mineração de dados], UFSCar, 25-09-2008.

GASPAR, T.L.; (2006). *Reconhecimento de faces humanas usando redes neurais MLP*. Dissertação (mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo.

GREULICH, W. W.; PYLE, S. I. (1992). *Radiographic Atlas of Skeletetal Development of the Hand and Wrist*. 2.ed., Ed. University Press.

GOEDTEL, A.(2007), "Estimador neural de velocidade para motores de indução trifásicos". Tese de Doutorado, USP- Escola de Engenharia de São Carlos.

HAGAN M. T.; MENHAJ ,M.B.; (1994) *Training Feedforward Networks with the Marquardt Algorithm* IEEE Transactions On Neural Networks, Vol. 5, No. 6, November 1994 p.989-993

HAITER N., F.; ALMEIDA, S. M.; LEITE, C. C. (2000). *Comparative Study of the Greulich & Pyle and Tanner & Whitehouse Methods for Estimating Skeletal Age*, Revista Odontológica Brasileira, v. 14, n. 4, p. 378-384,

HAITER, F.; KURITA, L. M.; CASANOVA, M. S. (2006); "Skeletal age assessment: A comparison of 3 methods", American Journal of Orthodontics and Dentofacial Orthopedics, October, 2006.

HAYKIN, S. (2001), Redes Neurais: princípios e prática, Editora Bookmam.

HEBB, D. O. (1949), *The Organization of Behavior: A Neuropsychological Theory*, Science Editions Inc.

HINES, G.D., RAHMAN Z., D. J., JOBSON, WOODELL, G. A., (2004) Single-Scale Retinex Using Digital Signal Processors, Global Signal Processing Expo (GSPX)

LAND, E. H., and MCCANN, J. J. (1971). *Lightness and retinex theory*. Journal of the Optical Society of America, 61, 1-11.

LIMA, R.S.A.; *Data Mining e KDD*. Disponível em <a href="http://www.nugen.uece.br/arquivos/apresentacoes/jornada\_bf3/data\_mining\_kdd.pdf">http://www.nugen.uece.br/arquivos/apresentacoes/jornada\_bf3/data\_mining\_kdd.pdf</a> Acesso em 15 set 2008

LIPPMAN, R. P. (1987), "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, pgs. 4-22.

MARQUES, A. M. S et al. (2001). *On Determining a Signature for Skeletal Maturity*, Proceedings of the 14th Brazilian Symposium on Computer Graphics and Image Processing, p. 246-251.

MERCADANTE, M.M.N. (1996). *Radiografia de mão e punho*. In: FERREIRA, F.V. Ortodontia: diagnóstico e planejamento clínico. São Paulo: Artes Médicas, 1996. p. 188-216.

MITTCHEL, T. M. (1997) "Machine Learning". McGraw-Hill, New York, 1997

MICHAEL, D. J.; NELSON, A. C. (1989). "HANDX: a model based system for automatic segmentation of bones from digital hand radiographs". IEEE Transactions on Medical Imaging, vol. 8, n. 1.

MCAULIFFE, M. J. (2007) "MIPAV - Medical Image Processing, Analysis and Visualization"- Center for Information Technology (CIT) National Institutes of Health (NIH), version:3.1.1 (2007-02-02), Biomedical Imaging Research Services Section, ISL, CIT, NIH

MORAES, M.E.L. et al. (2003). *Reliability of Greulich & Pyle and Eklof & Ringertz methods for skeletal age evaluation in brazilian children.* Rev. Odontol. UNESP, São Carlos, v. 32, n. 1, p. 9-17, Jan/Jun 2003.

NETLER, F.H. (1990). *The ciba collection of medical illustrations. Part I- Anatomy*. Physiology and metabolic disorders, vol. 8, Novatis medical education program, Califórnia,

NIEMEIJER, M. (2002). *Automating Skeletal Age Assessment*, Master's Thesis, University Utrecht,

NI, Z.; LI, F.; YANG,S.; LIU,X.; ZHANG,W.; LUO,O.;(2007) "Attributes Reduction Based on GA-CFS Method" School of Management, Hefei University of Technology, Hefei 230009, China, Computer Science June.

OLIANI, R.S.;(2005) "*Redes neurais artificiais*" USP- EESC, Departamento de Engenharia Elétrica, São Carlos, 2005. 43p. (Apostila).

OLIVETE JR, C.; (2005a). Estimativa de idade óssea através da análise carpal baseada na simplificação do método de Eklof e Ringertz. Dissertação de Mestrado, USP- Escola de Engenharia de São Carlos.

OLIVETE JR, C.; RODRIGUES, E. L. L.(2005b), O Efeito da Correção do Efeito Heel em Imagens Radiográficas da Mão. Revista Brasileira de Física Médica, v. 1, n. 1, p. 38-51, 2005.

OLIVETE JR, C.; RODRIGUES, E. L. L.; QUEIROZ, A. C. (2005c). *Metodologia para estimativa da idade óssea através da análise carpal simplificada*. In: X Congresso Brasileiro de Física Médica, 2005, Salvador, 2005

OLIVETE JR, C.; RODRIGUES, E. L. L. (2005d). *Contornos ativos snakes para extração de características de imagens carpais*. In: XX Congresso Brasileiro de Engenharia Biomédica, 2006, São Pedro-SP. XX Congresso Brasileiro de Engenharia Biomédica, 2006. p. 129-132.

OLIVETE, C. J.; QUEIROZ, A. C.; RODRIGUES, E. L. L. (2005e). Simplificação do método de EKLOF & RINGERTZ para a estimativa da idade óssea através da análise de imagens radiográficas de mão. In: WORKSHOP DE INFORMÁTICA MÉDICA, 2005, Porto Alegre. CD-ROM

OLIVETE, J. C.; RODRIGUES, E. L. L. (2006). Estimador da Idade Óssea via Análise Carpal / Anacarp — Software para Estimação da Idade Óssea Baseado no Método de Eklof & Ringertz, Wim — Workshop de Informática Médica, 2006

OLIVETE JR, C., RODRIGUES, E. L. L. (2006a). *Anacarp: uma Ferramenta para estimativa da idade óssea voltada ao diagnóstico médico*. In: XX Congresso Brasileiro de Engenharia Biomédica, 2006, São Pedro-SP. XX Congresso Brasileiro de Engenharia Biomédica, 2006. p. 644-647

OLIVETE JR, C.; RODRIGUES, E. L. L. (2006b). Software automático para determinação da idade ossea baseado na simplificação do Método de Eklof & Ringertz. Revista Brasileira de Física Médica, Aprovado para Publicação, 2006.

OLIVETE JUNIOR, C; RODRIGUES, E.L.L,(2006c) Modelos de contornos deformáveis para extração de características em imagens carpais, WVC - Workshop de Visão Computacional, 2006, São Carlos.

OLIVETE JR, C.; (2007). Novas metodologias automáticas para estimativa da idade óssea baseadas no processamento de imagens radiográficas da mão .Qualificação de Doutorado, USP- Escola de Engenharia de São Carlos.

ORTEGA,A.V.(2002), "Implementação de sistemas inteligentes para dimensionamento de leds hi-bright em equipamentos automotivos. Dissertação(mestrado),, Faculdade de Engenharia da Unesp, Bauru-SP.

OTSU, N. A. (1979) Threshold Selection Method from Gray-Level Histograms. IEEE Transaction on Systems, Mas, and Cybernetics, p. 62-66, 1979

QUEIROZ, A. C.; (2006). Metodologia de extração automática de característica da mão para estimação de idade óssea utilizando redes neurais artificiais no processo de decisão. Dissertação de Mestrado, USP- Escola de Engenharia de São Carlos.

QUINLAN, J. R. (1992). *Learning with continuous classes*. *In* <i>Proceedings of the 5th Australian Joint Conference on Artificial Intelligence</i>, Hobart, Australia (pp. 343-348). Singapore: World Scientific.

RAYMUNDO, E.M.; RODRIGUES,E.L.L.; LAUREANO,G.T.; MENDONÇA,M.M.; GUIMARÃES,L.T.G.; *Metodologia adaptativa de processamento de imagem radiográfica carpal com realce de bordas*. IV SIIM 2007, São Carlos - SP, 2007, CD-ROM.

RAYMUNDO, E.M.; RODRIGUES,E.L.L.(2008); *Metodologia de processamento utilizando* o algoritmo retinex em imagens radiográficas carpais, XXI CBEB2008, Congresso brasileiro de engenharia biomédica, Salvador, Brasil.

RIBEIRO, M. X. (2008), Suporte a sistemas de auxílio ao diagnóstico e de recuperação de imagens por conteúdo usando mineração de regras de associação Tese de Doutorado, USP-Escola de Engenharia de São Carlos.

RODRIGO, L. G.; FERNÁNDES, M. M.; ARRIBAS, J.I. (2003); A Fully Automatic Algorithm for Contourn Detection of Bones in Hand Radiographs using Active Contourns, ICIP 2003 – IEEE, 2003.

RODRIGUES, E. L. L.; OLIVETE JR, C. (2006h). Estimador da Idade Óssea via Análise Carpal/Anacarp - Software para Estimação da Idade Óssea Baseado no Método de Eklof & Ringertz. In: VI Workshop on Medical Informatics (WIM´2005), 2006, Vila Velha - ES, 2006.

RUMELHART, D. E., Hinton, G. E., Williams, R. J. (1986), "Learning Internal Representations by Error Propagation", Parallel Distributed Processing, vol. 1, MIT Press, pgs. 318-362.

SANTOS, L.; *Introdução aos SGBD's ISI* – "*Introdução aos Sistemas de Informação Semestre de Verão 2005/2006*", ISEL-DEETC-SSIC. Disponível em : <a href="http://www.deetc.isel.ipl.pt/sinfoconhecimento/isi/">http://www.deetc.isel.ipl.pt/sinfoconhecimento/isi/</a> Acesso em 15set. 2008.

SIMPSON, P. (1990), Artificial Neural Systems: Foundations, Paradigms, Applications and Implementations, Pergamon Press Inc.

SOUZA, A.; ROCHA, N.;(2005) *O processo KDD em sistemas de mineração de dados*, 2005 Disponível em <a href="http://www.marcelo.perotto.nom.br/arquivos/artigos/kdd/Artigo\_KDD.pdf">http://www.marcelo.perotto.nom.br/arquivos/artigos/kdd/Artigo\_KDD.pdf</a> Acesso em 15 set 2008

ŠTRAVS, L. & BRILLY, M. (2007) "Development of a low-flow forecasting model using the M5 machine learning method" Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova 2, SI-1000 Ljubljana, Slovenia, Copyright © 2007 IAHS Press

TAFFAREL, B.; WANGENHEIN, A. V.; SILVA, P. C. A.; (2003). Processamento de imagens para análise e estimativa de problemas de crescimento ósseo em crianças, III Workshop de informática aplicada à Saúde-CBComp 2003, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

TANNER, J. M.; WHITEHOUSE, R. W.; HEALVY (1969). A New System for Estimating Skeletal Maturity from Hand and Wrist, with Standarts Derived From a Study Of 2600 Healthy British Children, Departament of Growth and Development Institute of Child Health. University of London; and Departament of Statistics, Rothamsted Experimental Station, Harpenden.

TAVANO, O. (2001). Radiografias carpal e cefalométrica como estimadores da idade óssea e do crescimento e desenvolvimento, Bauru, Brasil, 2001.

TAVANO, O. (2004). *A radiografia carpal como estimador da idade óssea*. In:FREITAS, A. et al. Radiologia Odontológica. 6.ed. São Paulo: Artes Médicas, 2004. cap.29, p. 603.

TORRES, R. (1970); "Crescimiento armonioso y la divina proporcion". Divulgação Cultural Odontológica, volume 162, páginas 3-13, Março./Abril, 1970.

TRUCCO, E; VERRI, A.; (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice Hall: New Jersey, 1998.

VALENTIM, P. R. B.; PIMENTA, A.S.; SANTOS, D. M. B.; MARQUES NETO, M.I. C. (2005) "Uma solução software livre para mineração de dados em grades computacionais" Faculdade Ruy Barbosa, – Salvador – BA

XU, C.; PRINCE, J. L. (1997), *Gradient Vector Flow: A New External Force for Snakes*, IEEE Proc. Conf. On Comp. Vis. Patt. Recog., p. 66-71, 1997.

WILDEROM, B. P.; & FRANK, M.. (2002) Delphi6: Cliente/Servidor com Firebird/Interbase: Aplicação PDV e Controle de Estoques . ed. Érica, 2002

WITTEN, I. H.; FRANK E.; (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.