



**Universidade de Brasília**

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Classificação de Acromegalia em Imagens Tomográficas  
Computadorizadas de Vértex (L1 - L5) com uso de  
Ensemble de Classificadores Binários**

Clerimar Paulo Bragança

Trabalho apresentado como requisito parcial  
para obtenção do Título de Mestrado em Informática

Orientador

Prof. Dr. Bruno Luigi Machiavello

Coorientador

Prof. Dr. Pedro de Azevedo Berger

Brasília

2018

Universidade de Brasília — UnB  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Programa de pós-graduação em Informática

Coordenador: Prof. Dr. Bruno Luiggi Machiavello

Banca examinadora composta por:

Prof. Dr. Bruno Luiggi Machiavello (Orientador) — CIC/UnB  
Prof. Dr. Flávio de Barros Vidal (Coorientador) — CIC/UnB  
Prof. Dr. Antônio Padilha L. Bó — CIC/UnB

### **CIP — Catalogação Internacional na Publicação**

Paulo Bragança, Clerimar.

Classificação de Acromegalia em Imagens Tomográficas Computadorizadas de Vértex (L1 - L5) com uso de Ensemble de Classificadores Binários / Clerimar Paulo Bragança. Brasília : UnB, 2018.

57 p. : il. ; 29,5 cm.

Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2018.

1. Acromegalia, 2. Tomografia Tomputadorizada, 3. Aprendizado de Máquina.

CDU 004

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil



**Universidade de Brasília**

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Classificação de Acromegalia em Imagens Tomográficas  
Computadorizadas de Vértebras (L1 - L5) com uso de  
Ensemble de Classificadores Binários**

Clerimar Paulo Bragança

Trabalho apresentado como requisito parcial  
para obtenção do Título de Mestrado em Informática

Prof. Dr. Bruno Luigi Machiavello (Orientador)  
CIC/UnB

Prof. Dr. Flávio de Barros Vidal    Prof. Dr. Antônio Padilha L. Bó  
CIC/UnB    CIC/UnB

Prof. Dr. Bruno Luigi Machiavello  
Coordenador do Programa de pós-graduação em Informática

Brasília, 27 de Julho de 2018

# Dedicatória

Dedico esse trabalho aos meus pais Anísio Bragança (*in memoriam*) e Maria Narcisa, com todo meu amor e gratidão, por tudo que fizeram por mim ao longo de minha vida. Desejo poder ter sido merecedor dos esforços dedicados por vocês não somente quanto à minha formação, mas em todos os aspectos.

# Agradecimentos

Primeiro de tudo, gostaria de agradecer a Deus por me guiar, iluminar e me dar tranquilidade para seguir em frente com os meus objetivos e não desanimar com as dificuldades.

Agradeço ao meu orientador, Professor Doutor Bruno Luiggi Machiavello, por toda a paciência, empenho com que sempre me orientou neste trabalho e em todos aqueles que realizaram durante os seminários do mestrado. Muito obrigado por ter me corrigido quando necessário sem nunca me desmotivar.

Desejo igualmente agradecer ao meu coorientador Professor Doutor Pedro de Azevedo Berger e aos Professores que estiveram em minha banca examinadora, Doutor Flávio de Barros Vidal e Doutor Antônio Padilha L. Bó.

Agradeço também à minha esposa Danyele, por todo amor, companheirismo, compreensão e colaboração, muito obrigado por ter estado ao meu lado nessa fase da minha vida.

Por último, agradeço aos meus amigos e familiares, por compreenderem minha ausência em certos momentos e principalmente pelo constante apoio em minha caminhada.

# Resumo

A acromegalia é uma doença crônica e insidiosa. Sua causa, em grande parte, se dá via adenomas hipofisários (tumores benignos) secretores do Hormônio do Crescimento. Pode causar várias manifestações clínicas, como, por exemplo, o crescimento atípico dos ossos e tecidos moles. Exames de imagens obtidos com a Tomografia Computadorizada (TC) são utilizados para observar as características tumorais e acompanhamento radiológico em triagem de condições acromegálicas. Com base no princípio da automação, o presente estudo teve por objetivo diagnosticar a doença acromegalia a partir do processamento de imagens digitais e do aprendizado de máquinas, reconhecendo padrões radiológicos em imagens de TC e identificando possíveis alterações ósseas acromegálicas sofridas especificamente nas vértebras L1-L5 da coluna vertebral. Como viés metodológico, fez-se uso de um *ensemble* – combinação de classificadores. Os classificadores individuais e binários utilizados na construção do *ensemble* foram: SVM, KNN, Naive Bayes e AdaBoost. O método utilizou-se de 54 imagens de tomografia, sendo 27 imagens de indivíduos acromegálicas e 27 imagens de indivíduos livres da doença acromegalia. O aprendizado de máquinas se deu por meio de características calculadas a partir textura das imagens no domínio do espaço e da frequência. O resultado classificatório atingiu taxas de 95% de sensibilidade, 96% de especificidade e acurácia de 95%. Embasado na relevância dos resultados obtidos, fica a possibilidade de novas pesquisas para que se desenvolvam ferramentas que auxiliem no diagnóstico da acromegalia.

**Palavras-chave:** Acromegalia, Tomografia Tomputadorizada, Aprendizado de Máquina.

# Abstract

Acromegaly is a chronic and insidious disease. Their cause are largely associated with pituitary adenomas (benign tumors) secreting Growth Hormone. It can cause various clinical manifestations, such as atypical bone and soft tissue growth. Computed tomography (CT) imaging tests are used to observe tumor characteristics and also to monitor the radiological screening of acromegalic conditions. Based on the principle of automation, the present study aimed to diagnose the acromegaly disease from digital image processing and machine learning, recognizing radiological patterns in CT images and identifying possible acromegalic bone alterations specifically suffered in the vertebrae L1- L5 of the spine. As a methodological bias, a it ensemble - combination of classifiers was used. The individual and binary classifiers used in the construction of it ensemble were: SVM, KNN, Naive Bayes and AdaBoost. The method used 54 tomography images, 27 images of acromegalic individuals and 27 images of individuals free of acromegaly. The machine learning was elaborated by means of characteristics calculated from the texture of the images in the domain of space domain and frequency domain. The classificatory result reached rates of 95 % sensitivity, 96% specificity and 95% accuracy. Based on the relevance of the results obtained, it is possible to carry out new research to develop tools to aid in the diagnosis of acromegaly.

**Keywords:** Acromegaly, Computed Tomography and Machine Learning.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Organização . . . . .	2
<b>2</b>	<b>Revisão da Literatura</b>	<b>3</b>
2.0.1	Efeito do GH / IGF-1 no Metabolismo Ósseo . . . . .	3
2.0.2	Reconhecimento Automático da Acromegalia . . . . .	4
<b>3</b>	<b>Referencial Teórico</b>	<b>6</b>
3.1	Tomografia Computadorizada . . . . .	6
3.1.1	Tomografia Computadorizada de Coluna Vertebral . . . . .	8
3.2	Imagem Digital . . . . .	9
3.2.1	Vizinhança entre <i>Pixels</i> . . . . .	10
3.2.2	Componente Conexo . . . . .	10
3.2.3	Distâncias entre <i>Pixels</i> . . . . .	11
3.3	Segmentação de Imagens . . . . .	11
3.3.1	Limiarização . . . . .	11
3.3.2	Otsu . . . . .	11
3.4	Morfologia Matemática . . . . .	12
3.4.1	Erosão . . . . .	14
3.4.2	Dilatação . . . . .	15
3.4.3	Abertura . . . . .	15
3.4.4	Fechamento . . . . .	15
3.5	Descritores de Características de Imagens . . . . .	16
3.5.1	Descritores no Domínio Espacial . . . . .	16
3.5.2	Descritores no Domínio da Frequência . . . . .	19
3.6	Análise de Componentes Principais . . . . .	23
3.7	Aprendizado de Máquinas . . . . .	23
3.7.1	Máquina de Vetores de Suporte . . . . .	24
3.7.2	K-Vizinhos mais Próximos . . . . .	25
3.7.3	Naive Bayes . . . . .	26
3.8	Ensemble . . . . .	27
3.8.1	<i>Bagging</i> . . . . .	27
3.8.2	<i>Boosting</i> . . . . .	27
3.8.3	AdaBoost . . . . .	27
3.9	Medidas de Acurácia em Métodos de Classificação . . . . .	28
3.10	Sobreajuste . . . . .	29



3.10.1	Validação Cruzada . . . . .	29
3.10.2	Aumento de Dados . . . . .	30
<b>4</b>	<b>Metodologia Proposta</b>	<b>31</b>
4.1	Banco de Imagens . . . . .	31
4.1.1	Corte Automático da Região de Interesse . . . . .	32
4.1.2	Extração de <i>Features</i> . . . . .	33
4.1.3	Base de Dados Utilizada para Treinamento dos Classificadores . . . . .	35
4.1.4	Pré-teste de Classificação . . . . .	35
4.1.5	Etapas do pré-teste de classificação . . . . .	35
4.2	<i>Ensemble</i> . . . . .	36
4.2.1	Classificação Individual . . . . .	36
4.2.2	Combinação de Classificadores . . . . .	37
4.2.3	Saída do <i>Ensemble</i> . . . . .	39
<b>5</b>	<b>Resultados Alcançados</b>	<b>41</b>
5.0.1	Resultados Utilizando Validação Cruzada <i>K-Fold</i> . . . . .	41
5.0.2	Resultados Utilizando Validação Cruzada <i>leave-one-out</i> . . . . .	42
5.0.3	Resultados Utilizando Aumento de Dados e Validação Cruzada <i>K - Fold</i> . . . . .	42
<b>6</b>	<b>Conclusão</b>	<b>44</b>
	<b>Referências</b>	<b>45</b>

# Lista de Figuras

3.1	Exemplo de Unidades de <i>Hounsfield</i> convertidos para tons de cinza, adaptado de [18]. . . . .	7
3.2	Coluna vertebral - Adaptado de [20]. . . . .	8
3.3	a) imagem TC paciente acromegálico, a) imagem TC paciente não acromegálico	9
3.4	Coordenadas espaciais de uma imagem. . . . .	9
3.5	a) Exemplo vizinhos $N_4(p)$ , b) exemplo vizinhos $N_8(p)$ . . . . .	10
3.6	Exemplo de componentes conectados $N_4(p)$ . . . . .	10
3.7	a) Exemplo de conjunto B, b) exemplo de reflexão $\hat{B}$ , adaptado de [21]. . . . .	13
3.8	a) Exemplo de conjunto B, b) exemplo de translação $B_z$ , adaptado de [21]. . . . .	13
3.9	Exemplo de elemento estruturante, adaptado de [21]. . . . .	14
3.10	a) Conjunto A, b) Elemento estruturante B e c) Resultado de $(A \ominus B)$ . . . . .	14
3.11	a) Conjunto A, b) Elemento estruturante B e c) Resultado de $(A \oplus B)$ . . . . .	15
3.12	Exemplo de histograma com imagem de 16 bits. . . . .	16
3.13	Exemplo das orientações $\theta$ da GLCM. . . . .	17
3.14	a) Matriz de pixels, b) GLCM . . . . .	18
3.15	a) Figura retangular, b) espectro de Fourier da Figura retangular. . . . .	20
3.16	Exemplo de matriz de coeficientes DCT com leitura em zig-zag DCT. . . . .	21
3.17	a) exemplo de vários hiperplanos, b) exemplo de hiperplano ótimo. . . . .	24
3.18	Exemplo de K-Vizinhos mais próximos. . . . .	26
3.19	Exemplo de matriz de confusão. . . . .	28
3.20	Exemplo de validação cruzada. . . . .	29
3.21	Exemplo de validação <i>leave-one-out</i> . . . . .	30
4.1	Diagrama do sistema proposto. . . . .	31
4.2	Região de interesse da imagem de TC. . . . .	32
4.3	Fluxograma do corte automático da ROI. . . . .	33
4.4	Exemplo das classes A e B. . . . .	37
4.5	Resultado do classificador h1 . . . . .	38
4.6	Resultados de classificação de h2. . . . .	38
4.7	Exemplo de aumento da especificidade. . . . .	39
4.8	Exemplo de um <i>ensemble</i> . . . . .	40
5.1	Matriz de confusão com validação <i>K-Fold</i> . . . . .	41
5.2	Matriz de confusão com validação <i>leave-one-out</i> . . . . .	42
5.3	Matriz de confusão dados aumentados . . . . .	43

# Lista de Tabelas

3.1	Substâncias e valores de unidades de <i>Hounsfield</i> , adaptado de [18]. . . . .	7
4.1	Elementos do vetor $Q$ . . . . .	34
4.2	Seleção de <i>Features</i> PCA com <i>Software Genes</i> - UFV. . . . .	34
4.3	Resultado dos classificadores com o pré-teste . . . . .	36
4.4	Resultado dos classificadores individuais, <i>leave-one-out</i> . . . . .	37
4.5	Resultado dos classificadores individuais. . . . .	37
5.1	Desvio padrão dos testes de classificação realizados . . . . .	42
5.2	Resultado do <i>ensemble</i> utilizando <i>leave-one-out</i> . . . . .	42
5.3	Desvio padrão dos testes de classificação realizados com aumento de dados . . . . .	43

# Capítulo 1

## Introdução

A acromegalia é um distúrbio do crescimento esquelético e tecidual. É uma doença rara, com incidência anual de três a quatro novos casos por milhão de habitantes [1]. A enfermidade está relacionada à produção excessiva do Hormônio do Crescimento (em inglês, Growth Hormone – GH).

Em mais de 90% dos casos, a hipersecreção de GH tem origem em um tumor benigno (adenoma) da glândula pituitária, também denominada hipófise [2]. A secreção de GH pela hipófise na corrente sanguínea estimula o fígado a produzir outro hormônio denominado Fator de Crescimento Semelhante à Insulina (em inglês, Insulin Growth Factor 1 – IGF-1), sendo este o real causador do crescimento do tecido no corpo.

Altos níveis de IGF-I também causam alterações no metabolismo da glicose (açúcar) e lipídios (gordura), podendo levar a diabetes, à pressão alta e às doenças cardíacas. A aparência de gigantismo pode ocorrer em crianças e adolescentes. Nos adultos que atingiram a sua altura máxima, as mãos, os pés e a mandíbula inferior podem ficar aumentados [2].

O diagnóstico da acromegalia requer demonstração de secreção desregulada e aumentada de GH, bem como níveis elevados de (IGF-1) [3].

Embora estudos de Densidade Mineral Óssea (DMO), tenham apresentado resultados conflitantes, a acromegalia pode ter um efeito deletério na microarquitetura óssea trabecular [4]. Devida a tais alterações ósseas, a acromegalia posiciona-se como uma das causas da osteoporoze secundária, (doença causada pela baixa DMO, tornando os ossos frágeis e quebradiços)[5].

A deformação óssea também pode afetar a coluna vertebral. Pesquisas revelaram que a prevalência estimada de envolvimento da coluna vertebral com a acromegalia é cerca de 40% a 50% [1]. O remodelamento ósseo é estimulado, o osso cortical espessa e a massa óssea trabecular pode ficar diminuída, normal ou aumentada.

Na triagem de complicações acromegálicas, faz-se importante realizar um estudo radiográfico da coluna vertebral com acompanhamento médico no momento do diagnóstico e durante todo o tratamento [6]. O estudo radiográfico pode ser obtido pela Tomografia Computadorizada (TC) da região do abdome.

A variedade de manifestações da doença acromegalia pode levar os pacientes a buscarem aos médicos de diferentes especialidades, inclusive em situações na qual se recomenda o exame de TC do abdome. Por outro lado, se estes médicos tivessem mecanismos para detecção automática da acromegalia, como exemplo, a detecção automática por meio das imagens tomográficas, provavelmente poderiam se beneficiar destes algoritmos e proporcionar um diagnóstico precoce da doença.

Levando em consideração o comprometimento da coluna vertebral com a doença acromegalia e a sua gravidade, este trabalho visa por meio do processamento de imagens de TC e aprendizado de máquinas, diagnosticar a acromegalia. Antes de apresentar a metodologia, o estudo proposto busca responder ao problema de pesquisa: "É possível identificar padrões da acromegalia em imagens tomográficas de vértebras lombares L1- L5 e classificar as imagens quanto a presença ou ausência da doença?"

Tendo a pesquisa analisado em trabalhos correlacionados a influência do comprometimento ósseo pela acromegalia, e observando que a utilização destes conceitos pode trazer recursos para o aprendizado de máquinas, a hipótese levantada por esta pesquisa é que: A agromegalia causa modificações ósseas que podem ser notadas em imagens tomográficas com a ajuda do aprendizado de máquinas e reconhecimento de padrões.

## Objetivo Geral

O objetivo geral é possibilitar um diagnóstico acromegálico por meio da análise de imagens tomográficas.

## Objetivos Específicos

- Reconhecer padrões da doença acromegalia em imagens de TC da região do abdome nas vértebras (L1-L5),
- Fazer a extração de um conjunto de *features* da região de interesse das imagens tomográficas por meio de descritores no domínio do espaço e frequência, os *features* por meio do cálculo de análise de componentes principais,
- Desenvolver um *ensemble* classificador composto com os classificadores binários: *Support Vector Machines* SVM, *K-Nearest Neighbor* KNN, *Naive Bayes* e *Adaptive Boosting* AdaBoost,
- Classificar as imagens de TC considerando a presença ou ausência da doença acromegalia.

## 1.1 Organização

O texto é organizado inicialmente com um Capítulo sobre os fundamentos da doença acromegalia e o uso da TC como meio de triagens e acompanhamento de complicações acromegálicas. No Capítulo 2, é apresentada uma revisão bibliográfica de pesquisas científicas correlacionadas ao estudo aqui proposto. No Capítulo 3, é apresentado o conceito da tecnologia e das ferramentas utilizadas para o processamento das imagens tomográficas e extração de *features* no domínio do espaço e da frequência. No Capítulo 4, apresenta-se a metodologia utilizada para atingir os objetivos propostos. No Capítulo 5, são apresentados os resultados da classificação das imagens por meio de um *ensemble* classificador. No Capítulo 6, são apresentadas as considerações finais com a conclusão e sugestões para trabalhos futuros.

# Capítulo 2

## Revisão da Literatura

As próximas Seções destacam alguns trabalhos que avaliaram a influência de níveis elevados de GH e IGF-1 na saúde óssea e na DMO em pacientes com acromegalia. Descreve-se a seguir algumas pesquisas correlacionadas ao sistema proposto.

### 2.0.1 Efeito do GH / IGF-1 no Metabolismo Ósseo

Uma pesquisa proposta por Vittorio Locatelli e Bianchi [7] revelou 39 estudos clínicos relatando os efeitos da administração de GH e IGF-1 no metabolismo ósseo em seres humanos osteopênicos e osteoporóticos e na consolidação óssea em pacientes operados com secreção normal de GH. Dezoito estudos clínicos consideraram o efeito do tratamento com GH, quatorze estudos relataram os efeitos clínicos com a administração de IGF-1 e sete relacionados ao efeito de GH / IGF-1 na consolidação óssea. Tanto a administração de GH como de IGF-1 aumentaram significativamente a reabsorção óssea e a formação óssea na maioria dos casos. O estudo concluiu que a terapia com GH e IGF-1 tem um efeito anabólico significativo [7].

Nesta outra pesquisas foram incluídos 33 pacientes com acromegalia, sendo, 14 homens e 19 mulheres. Estes pacientes foram submetidos ao exame de absorciometria por dupla emissão de raios-x (DXA) entre os anos de 2000 a 2014, no Hospital da Universidade Nacional de Seul. Noventa e nove pessoas de diferentes sexos e livres da doença acromegalia foram tidas como grupo controle. Os parâmetros bioquímicos, Escore do osso trabecular, em inglês *Trabecular bone score* TBS da coluna lombar e a DMO foram medidos para todos os envolvidos na pesquisa. O status gonadal foi avaliado no momento do diagnóstico. A pesquisa relatou que a TBS da coluna lombar foi menor nos pacientes com acromegalia comparado às pessoas de ambos os sexos pertencentes ao grupo controle. A microarquitetura esquelética foi deteriorada em pacientes com acromegalia, como avaliado pela TBS, o que parece ser uma consequência do excesso de hormônio de crescimento, bem como do hipogonadismo, especialmente em mulheres[8].

Avaliou-se nesta outra pesquisa que os efeitos do excesso de hormônio de crescimento na renovação óssea e na densidade óssea não são totalmente explicados, mas o eixo GH / IGF-1 desempenha um papel importante na manutenção da massa óssea. Este efeito é comprovado pelos altos níveis de marcadores de renovação óssea na acromegalia ativa, a correlação entre o IGF-I e a remodelação óssea foi comprovada por poucos estudos anteriores [9].

Uma abordagem proposta por ([4]) foi medir a DMO e microarquitetura óssea de um grupo de 45 pacientes acromegálicos com a presença de diabetes mellitus tipo 2 DM2, atividade da doença, status gonadal e comparar esses parâmetros ósseos de pacientes acromegálicos eugo-

nadais com 45 controles saudáveis. Para as medidas, utilizou-se do exame DXA e tomografia computadorizada quantitativa periférica de alta resolução. O resultado da pesquisa mostrou que a acromegalia parece ter um efeito deletério sobre a microarquitetura óssea trabecular e, nessa população específica, o status gonadal pode ser mais importante do que a atividade do DM2 ou da acromegalia na determinação da saúde óssea [4].

Foi analisado por ([10]) o metabolismo ósseo e DMO indicando os principais determinantes desses parâmetros em um grande grupo de pacientes com acromegalia ativa. Um grupo de 121 acromegálicos ativos, com idades entre 23 e 80 anos, de um único centro endocrinológico foi estudado. O estudo mostrou que na acromegalia activa, observa-se um aumento da renovação óssea dependente do IGF-I e um aumento da DMO. Mostrou também que nos acromegálicos hipogonadais, a alta reabsorção óssea diminui a DMO e pode levar à osteoporose e que há um menor aumento na reabsorção óssea e maior aumento na DMO em mulheres com acromegalia do que em homens [10].

Outro estudo concluiu que o hipopituitarismo também pode ser uma causa de perda óssea devido a deficiências de vários hormônios hipofisários, especialmente hormônio de crescimento ou terapia de reposição excessiva [11].

Em estudos da osteoporose, buscou classificar a doença por meio de caracterização da textura de radiografias de regiões ósseas trabeculares com o uso de Máquinas de Vetores de Suporte (SVM). As características utilizadas permitiram uma taxa de precisão de (96,22)% [12].

## 2.0.2 Reconhecimento Automático da Acromegalia

Uma pesquisa de reconhecimento da acromegalia via imagens faciais conduzida por Kong Xiangyi ([13]) analisou as características advindas da doença acromegalia, como o espaçamento alargado dos dentes, prognatismo, aumento frontal ósseo, aumento do nariz, proeminência do arco zigomático, protuberância e proeminência da testa, inchaço dos tecidos moles (lábios, nariz, aumento das orelhas). Com base nessas modificações foi desenvolvido um sistema de reconhecimento automático que permite que médicos ou pacientes rastreiem proativamente as características faciais e detectem acromegalia. Usando 1123 fotografias de rostos, treinou-se e integrou vários métodos de aprendizado de máquina, incluindo Modelos Lineares Generalizados (LM), K-vizinhos mais próximos (KNN), Máquinas de Vetores de Suporte (SVM), Árvores de Regressão (RT) e Rede Neural Convolutiva (CNN), formando um *ensemble* para detecção facial de acromegalia. Além das alterações faciais típicas, os sintomas e sinais da acromegalia também incluem a estimulação do crescimento de muitos tecidos, como pele, tecido conjuntivo, cartilagem, ossos, vísceras e muitos tecidos epiteliais. O melhor resultado dos métodos propostos mostrou uma sensibilidade de 96% e uma especificidade de 96% [13].

Outro estudo buscou o desenvolvimento de um programa de computador que em seus testes separou 24 fotografias faciais de pacientes com acromegalia das de 25 fotografias de indivíduos não acromegálicos. O ponto principal para o desenvolvimento foi usar um banco de dados previamente desenvolvido que consistiu em representações tridimensionais de 200 cabeças de pessoas não acromegálicas. As pessoas foram fotografadas em ângulos frontais e laterais dos rostos. Os pacientes foram agrupados em sujeitos com características faciais leves, moderadas e severas da acromegalia pela impressão geral. Em seguida, analisou-se todas as imagens usando análise de similaridade computadorizada baseada em jatos Gabor e funções geométricas. O software classificou corretamente 71,9% dos pacientes e 91,5% dos controles. A acurácia da classificação dos pacientes pela análise visual foi de 63,2 e 42,1% por especialistas e internistas

gerais, respectivamente. A precisão da classificação para os controles foi de 80,8 e 87,0% por especialistas e internistas, respectivamente. As maiores diferenças em precisão entre software e especialistas e internistas estavam presentes para pacientes com acromegalia leve [14].



# Capítulo 3

## Referencial Teórico

Neste capítulo são apresentados importantes conceitos e embasamento teórico sobre o assunto abordado na pesquisa em questão.

Os tópicos a seguir referem-se, de acordo com a metodologia a ser utilizada para desenvolvimento da pesquisa.

### 3.1 Tomografia Computadorizada

A tomografia computadorizada (TC) é um método de obtenção de imagens por raios-X. Este método foi desenvolvido e introduzido na prática clínica pelo físico Godfrey Hounsfield em 1972 [15]. A tecnologia permitiu aos médicos observar pela primeira vez tecidos do cérebro e posteriormente de outras partes do corpo sem necessidade de cirurgia.

Para obtenção das imagens, uma radiação precisa é emitida pela TC no espaço de interesse do corpo humano, no qual cada estrutura é capaz de fazer absorção da radiação emitida. Maior radiação é absorvida pelas estruturas com maior densidade e, dessa forma, há menor sensibilização dos detectores que estão em sua projeção, formando imagens claras. Pouca radiação é absorvida pelas estruturas com baixa densidade e, dessa forma, há maior sensibilização dos detectores e imagens mais escuras são formadas.

Os detectores são responsáveis pela captação da radiação que conseguir ultrapassar o corpo do paciente, portanto, a radiação que ultrapassa as estruturas atinge os detectores com intensidade característica à densidade das estruturas estudadas [16].

A Unidade de *Hounsfield* (HU) é uma representação numérica em TC das densidades das estruturas do corpo humano. O ponto de referência dos números de TC é a água, tendo esta o valor zero (0), o osso cortical denso tem valor de +700, ou até +3.000, e o ar que produz a menor quantidade de atenuação tem o valor de -1.000 [17]. Entre esses dois extremos existem tecidos e substâncias com números de TC variados de acordo com suas atenuações [18].

A Tabela 3.1 mostra um exemplo de atenuações da escala de Hounsfield.

As imagens são reconstruídas por meio de cálculos matemáticos e auxílio da análise bidimensional de Fourier, que consiste em analisar as funções de tempo e de espaço, somando as frequências e amplitudes. Após a sua reconstrução computacionalmente, cada *pixel* é atribuído a um número de HU [19]. Feito isso, as graduações de tons de cinzas são designadas para os números de TC [17].

Gordura, músculos e órgãos, que se encaixam entre as densidades do osso e do ar, aparecem com uma variação de tonalidades de cinza, como mostra o exemplo dado na Figura 3.1 .

Tabela 3.1: Substâncias e valores de unidades de *Hounsfield*, adaptado de [18].

Substância	HU
Ar	-1000
Pulmão	-500
Gordura	0
Água	15
Fluido cerebrospinal	30
Rim	+30 a +45
Sangue	+10 a +40
Músculo	+37 a +45
Massa cinzenta	+20 a +30
Fígado	+40 a +60
Tecidos moles, Contraste	+100 a +300
Osso	+700 (osso trabecular) a + 3000 (osso denso)

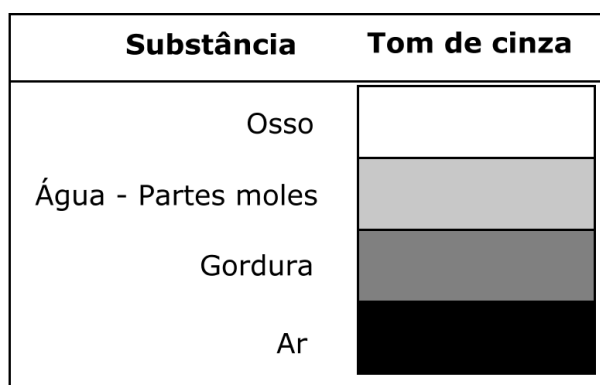


Figura 3.1: Exemplo de Unidades de *Hounsfield* convertidos para tons de cinza, adaptado de [18].

A TC se divide em helicoidal e multislice.

TC helicoidal: o sistema espiral ou helicoidal, utilizado a partir de 1990, é possibilitado em razão da contínua rotação do tubo em volta da área de interesse. Essa tecnologia permite que estudos do tórax e abdome sejam mais precisos e as reconstruções de imagens bidimensionais e tridimensionais sejam feitas com maior eficácia [16].

O tempo de aquisição sempre deve ser considerado na realização do exame, em virtude da relação do deslocamento da mesa pela espessura do corte utilizado (*pitch*), capacidade do tubo, região e patologia de interesse, tempo de contraste e dose de radiação para que assim se otimize o sistema da TC Helicoidal.

TC multislice: neste método, a tecnologia helicoidal é associada a múltiplos conjuntos de anéis detectores, o que permite aquisição simultânea de varias informações.

Meios de contrastes: são substâncias químicas utilizadas nos métodos de diagnóstico por imagem para melhor visualizar estruturas do corpo humano [16].

### 3.1.1 Tomografia Computadorizada de Coluna Vertebral

O exame de TC da coluna vertebral é uma forma de se avaliar os órgãos, articulações, ossos e tecidos moles como músculos, cartilagens e tendões da região vertebral.

A coluna vertebral é composta por 32 vértebras separadas por discos fibrocartilagosos [20]. É composta por 5 setores:

- Cervical; compõe o esqueleto axial do pescoço e suporte da cabeça,
- Torácica; suporta a cavidade torácica,
- Lombar; possibilita a mobilidade entre a pelve e a parte torácica do tronco e suporta a cavidade abdominal,
- Sacral; faz a união da cintura pélvica à coluna vertebral,
- Coccigea; trata-se de uma estrutura rudimentar, mas que tem como função atuar no suporte do assoalho pélvico [20].

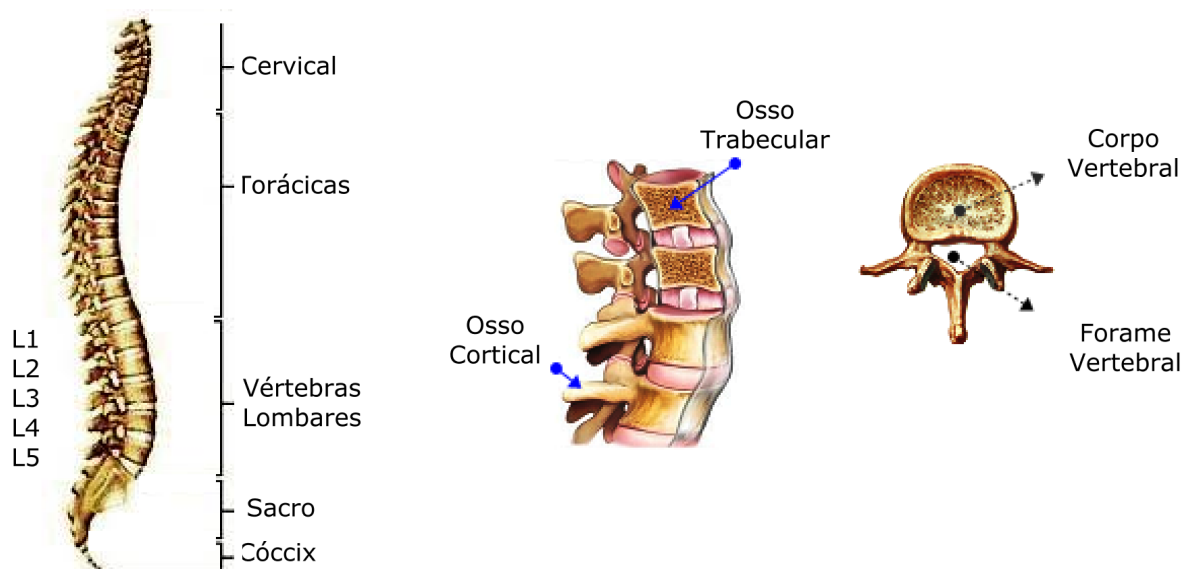


Figura 3.2: Coluna vertebral - Adaptado de [20].

Pequenos corpos formam as vértebras cervicais, sendo maiores as lombares e sacrais, devido ao fato do peso do corpo concentrar-se nas vértebras inferiores [20].

As cinco vértebras lombares (L1-L5) têm o forame vertebral pequeno em comparação com o corpo vertebral, este que consiste basicamente de uma massa cilíndrica com aparência de esponja por ter pequenas pontes chamadas trabéculas. As bordas das superfícies superior e inferior são compostas de osso compacto denominado cortical. Essas vértebras podem variar consideravelmente de tamanho de uma pessoa para outra pessoa [20].

Várias são as doenças que podem acometer regiões ósseas, cita-se aqui a osteoporose, doença classificada como primária ou secundária. A primária pode ser causada pós menopausa ou associada a doenças hereditárias. A osteoporose secundária pode ser decorrente de uma gama

de drogas e doenças, como a acromegalia e outras, podendo se sobrepor a um quadro primário [20].

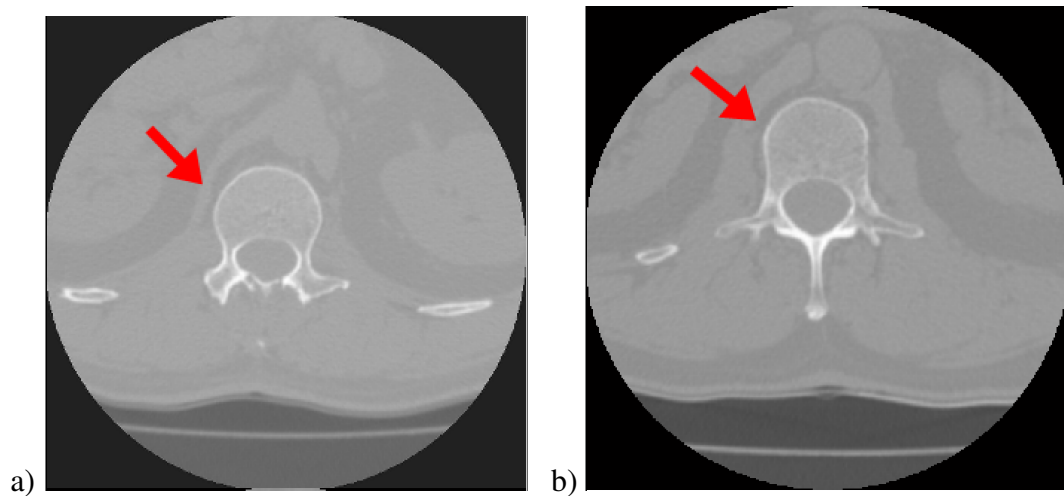


Figura 3.3: a) imagem TC paciente acromegálico, a) imagem TC paciente não acromegálico

As Figuras 3.3- a), e 3.3- b), mostram um exemplo de imagens tomográficas abdominais de uma pessoa saudável e de uma pessoa com acromegalia respectivamente.

## 3.2 Imagem Digital

Pode-se dar a definição de uma imagem por meio de uma função bidimensional  $f(x,y)$ , na qual  $x$  e  $y$  constituem coordenadas espaciais de um plano. A Figura 3.4 exemplifica essa situação. Em qualquer par de coordenadas  $(x,y)$ , a amplitude de  $f$  é denominada nível ou intensidade de cinza da imagem nesta posição.

Uma imagem digital é composta por um número finito de elementos, cada um com localização e valores específicos. Esses elementos são clamados em inglês de (*picture e element pixel*) [21].

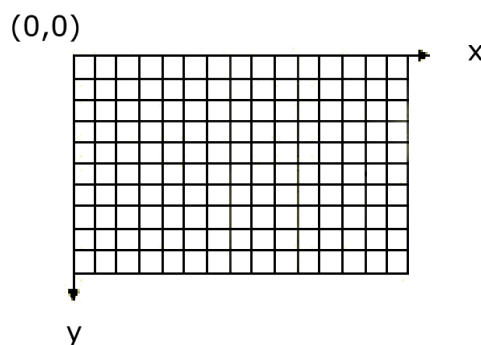


Figura 3.4: Coordenadas espaciais de uma imagem.

Uma imagem pode ter  $2^k$  níveis de cinza, sendo comum referir-se a ela como uma imagem  $k$  bits. Por exemplo, uma imagem com 256 valores discretos possíveis de níveis de cinza é chamada de uma imagem de 8 bits. Portanto, cada *pixel* pode estar associado a um valor da

escala de cinza entre 0 e  $(2^k) - 1$ . Quanto maior o número de *pixels* numa imagem, melhor será sua resolução, permitindo uma melhor diferenciação entre as estruturas [21].

### 3.2.1 Vizinhança entre *Pixels*

Um *pixel*  $p$  na coordenada  $(x, y)$  pode ter quatro vizinhos distribuídos por localizações horizontais e verticais cujas coordenadas são:  $(x + 1, y)$ ,  $(x - 1, y)$ ,  $(x, y + 1)$ ,  $(x, y - 1)$ .

Esses vizinhos são chamados de vizinhos 4 de  $p$ , expresso por  $N_4(p)$ , já outros quatro vizinhos diagonais de  $p$  têm coordenadas:  $(x + 1, y + 1)$ ,  $(x + 1, y - 1)$ ,  $(x - 1, y + 1)$ ,  $(x - 1, y - 1)$ , são os chamados vizinhos 8 de  $p$ , expresso por  $N_8(p)$ .

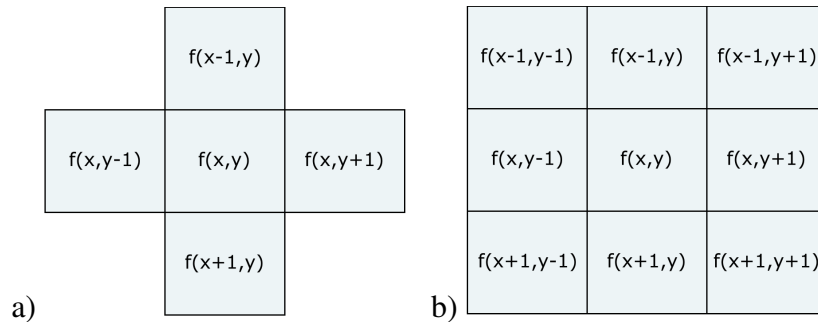


Figura 3.5: a) Exemplo vizinhos  $N_4(p)$ , b) exemplo vizinhos  $N_8(p)$

Os *pixels*  $p$  e  $q$  são 4 adjacentes se, e somente se,  $q \in N_4(p)$ .

Os *pixels*  $p$  e  $q$  são 8 adjacentes se, e somente se,  $q \in N_8(p)$ .

### 3.2.2 Componente Conexo

Dado um subconjunto de *pixels*  $C$ , dois *pixels*  $p$  e  $q$  são ditos conectados em  $C$  se existe um caminho entre  $p$  e  $q$  formado inteiramente por *pixels* em  $C$  [21].

Para cada *pixel*  $p$  em  $C$ , o conjunto de *pixel* conectados a  $p$  define uma componente conectada de  $C$ . Se  $C$  tem apenas uma componente conectada, então ele é chamado de um conjunto conectado.

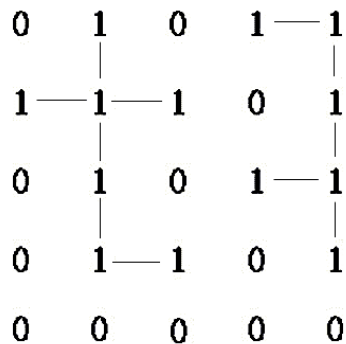


Figura 3.6: Exemplo de componentes conectados  $N_4(p)$

### 3.2.3 Distâncias entre *Pixels*

Seja  $i$  e  $j$  *pixels* com coordenadas  $(x, y)$  e  $(m, n)$ , respectivamente [21]. Os cálculos de distância  $D$  entre  $i$  e  $j$  podem ser calculadas como mostram as Equações a seguir:

distância euclidiana;

$$D(i, j) = \sqrt{(x - m)^2 + (y - n)^2}, \quad (3.1)$$

distância *city block* ou *Manhattan*;

$$D(i, j) = |x - m| + |y - n|, \quad (3.2)$$

distância *chessboard*;

$$D(i, j) = \max |x - m|, |y - n|. \quad (3.3)$$

## 3.3 Segmentação de Imagens

A segmentação subdivide uma imagem em regiões ou objetos não sobrepostos que a compõem. O nível de detalhe em que a subdivisão é realizada depende do problema a ser resolvido [21]. Existem vários métodos para desenvolvimento da segmentação, como por exemplo a segmentação por crescimento de região, segmentação por *watersheds* e o método de limiarização, do qual expõe mais detalhes por ser a base da segmentação do algoritmo Otsu utilizado neste trabalho.

### 3.3.1 Limiarização

Uma simples propriedade que os *pixels* em uma região podem compartilhar é a intensidade. Uma forma natural de segmentar essas regiões é através do limiar. O limiar pode criar imagens binárias a partir de níveis de cinza, transformando todos os *pixels* abaixo de um limiar a zero "0", (preto) e todos os *pixels* acima desse limiar a um "1", (branco).

Uma vez que  $g(x, y)$  constitui uma versão com limiar global  $T$  de  $f(x, y)$ , tem-se a imagem segmentada  $g(x, y)$  por:  $g(x, y) = 1$  se  $f(x, y) > T$  ou  $g(x, y) = 0$  se  $f(x, y) \leq T$ .

O principal problema com o limiar é que neste considera apenas a intensidade, desconsiderando qualquer relação entre os *pixels*, portanto, não há garantia de que os *pixels* identificados pelo processo de limiarização sejam contíguos [21].

Outro problema é que o limiar pode ser comprometido por mudanças na iluminação, uma vez que em toda a cena pode fazer com que algumas partes sejam mais claras (na luz) e algumas partes mais escuras (na sombra), dificultando o processo de segmentação.

Problemas assim podem ser em partes resolvidos, determinando limiares locais. Ou seja, em vez de ter um único limite global, permite-se que o limite em si varie suavemente conforme a imagem [21].

### 3.3.2 Otsu

Levando-se em conta a forma peculiar de embasamento por cálculos feitos no histograma de uma imagem, a variância entre classes encontra no método Otsu um ótimo aliado em sua maximização.

A ideia básica é que as classes com limiares bem estabelecidos sejam distintas em relação aos valores de intensidade de seus pixels e, inversamente, que um limiar que oferece a melhor separação entre as classes em termos de valores de intensidade seja o limiar ótimo [21].

O método Otsu pode ser resumido pelas Etapas seguintes.

1) Calcula-se o histograma normalizado da imagem a ser processada. Denominam-se os componentes do histograma como  $p_i$ , sendo  $i = \{0,2,3,\dots, L-1\}$ .

2) Efetuam-se, utilizando a Equação 3.4, as somas acumuladas,  $P_1(k)$ , para  $k = \{0,1,\dots, L-1\}$ .

$$P_1(k) = \sum_{i=0}^k p_i. \quad (3.4)$$

3) Calculam-se com a Equação 3.5, as médias acumuladas  $m(k)$ , para  $k=0,1,\dots, L-1$ ,

$$m(k) = \sum_{i=0}^k i p_i. \quad (3.5)$$

4) A intensidade média global,  $m_G$ , é dada pela Equação 3.6,

$$m_G = \sum_{i=0}^{L-1} i p_i. \quad (3.6)$$

5) O cálculo da variância,  $\sigma_B^2(k)$ , deve ser efetuado para  $k = \{0,1,\dots, L-1\}$ , utilizando a Equação 3.7,

$$\sigma_B^2(k) = \frac{[m_G P_1(k) - m(k)]^2}{P_1(k)[1 - P_1(k)]}. \quad (3.7)$$

6) Obtenha o limiar de Otsu,  $k^*$ , utilizando e avaliando a Equação 3.8, para todos os inteiros de  $k$ , cuide-se para que a condição  $0 < P_1(k) < 1$  seja mantida e selecione o valor  $k$  que produz o máximo  $\sigma_B^2(k)$ . Se a máxima não for única, obtenha-se  $k^*$  pela média dos valores de  $k$  que correspondem aos diversos valores máximos detectados.

$$\sigma_B^2(k^*) = \max \sigma_B^2(k), 0 \leq k \leq L-1. \quad (3.8)$$

7) Uma vez obtido  $k^*$ , a imagem a ser processada  $f(x,y)$  é então segmentada conforme a Equação 3.9, sendo  $h(x,y)$  uma versão do limiar *Otsu*.

$$\begin{aligned} h(x,y) &= 1 \text{ se } f(x,y) > k^*, \\ h(x,y) &= 0 \text{ se } f(x,y) \leq k^*. \end{aligned} \quad (3.9)$$

### 3.4 Morfologia Matemática

A morfologia matemática é a parte do processamento de imagem que tem por objetivo extrair características associadas à geometria dos objetos.

Tendo-se A e B como conjunto  $Z^2$ , cujos elementos do conjunto consistem de um vetor bidimensional com coordenadas  $(x,y)$  de um *pixel* preto ou branco, segundo convenção de uma imagem.

Caso a imagem seja representada em níveis de cinza, pode-se representá-la como conjuntos cujos componentes estão em  $Z^3$ . Assim, dois componentes de cada elemento do conjunto referem-se às coordenadas de um *pixel*, e o terceiro componente corresponde ao seu valor discretizado de intensidade [21].

A **reflexão** de um conjunto B indicado por  $\hat{B}$  é definido como:

$$\hat{B} = \{w | w = -b, \text{ para } (b \in B)\}. \quad (3.10)$$

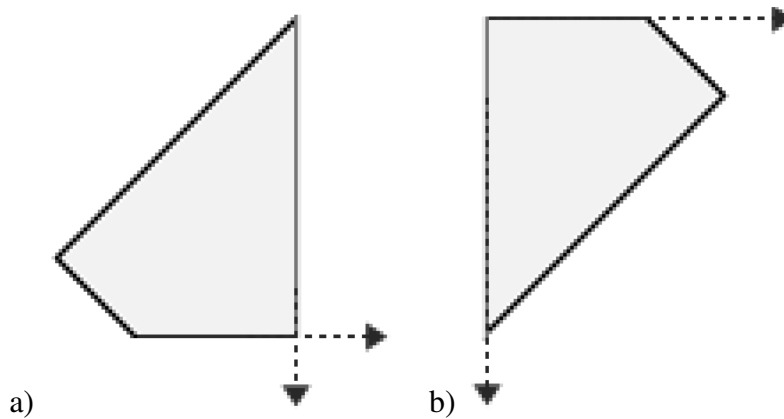


Figura 3.7: a) Exemplo de conjunto B, b) exemplo de reflexão  $\hat{B}$ , adaptado de [21].

Se B é o conjunto de *pixels*, que representam um objeto em uma imagem, então  $\hat{B}$  é simplesmente o conjunto dos pontos em B cujas coordenadas  $(x,y)$  foram substituídas por  $(-x,-y)$ . A reflexão reflete todos os elementos de B em torno da origem desse conjunto [21].

A **translação** de um conjunto B no ponto  $Z = (z_1, z_2)$ , indicada por  $(B)_z$  é definida como:

$$(B)_z = \{c | c = b + Z, \text{ para } (b \in B)\}. \quad (3.11)$$

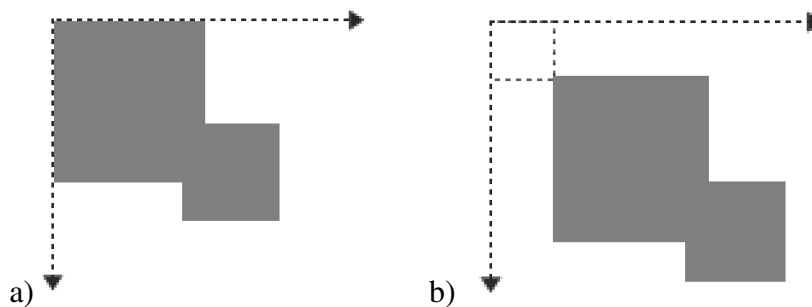


Figura 3.8: a) Exemplo de conjunto B, b) exemplo de translação  $(B)_z$ , adaptado de [21].

Se B é conjunto de *pixels* que representa um objeto em uma imagem, então  $(B)_z$  é o conjunto de pontos em B, cujas coordenadas  $(x,y)$  foram substituídas por  $(x + z_1, y + z_2)$ . A translação transfere a origem de B para o ponto Z.



A reflexão e a translação de conjuntos são amplamente empregadas na morfologia para formular operações baseadas nos chamados elementos estruturantes. Esses elementos estruturantes são pequenos conjuntos ou subimagens utilizadas para examinar outra imagem buscando propriedades de interesse. A Figura 3.9 mostra um exemplo de elemento estruturante [21].

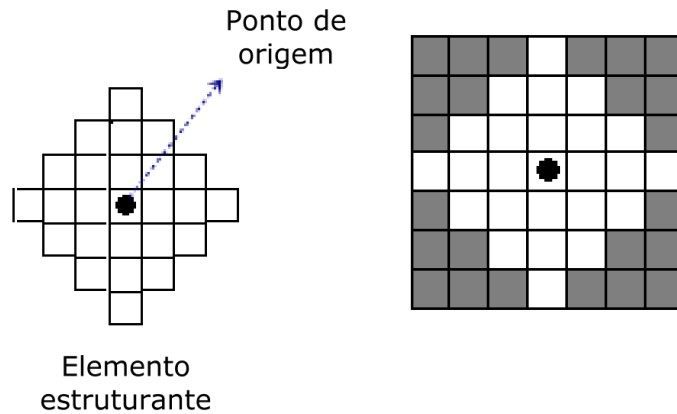


Figura 3.9: Exemplo de elemento estruturante, adaptado de [21].

A aplicabilidade utilizando-se elementos estruturantes sobre conjuntos pode se dar pelos operadores morfológicos, erosão e dilatação.

### 3.4.1 Erosão

Tendo  $Z^2$  os conjuntos  $A$  e  $B$ , a erosão de  $A$  por  $B$  é indicada por:

$$A \ominus B = \{Z | (B)_Z \subseteq A\}. \quad (3.12)$$

Indica a Equação 3.12 que a erosão de  $A$  por  $B$ , constitui conjunto da totalidade dos pontos  $Z$  de maneira que  $A$  contém  $B$  transladado por  $Z$ . Exemplo apresentado na Figura 3.10.

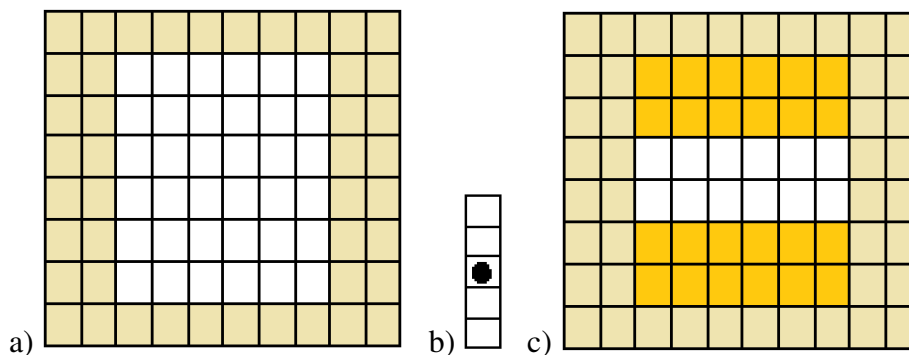


Figura 3.10: a) Conjunto  $A$ , b) Elemento estruturante  $B$  e c) Resultado de  $(A \ominus B)$ .

A erosão é uma operação de diminuição ou afinamento dos objetos da seguinte forma:

1) Considerando a Figura 3.10, primeiramente, crie um novo conjunto fazendo varredura de  $B$  ao longo de  $A$ , para que a origem de  $B$  passe por todos os elementos de  $A$ .

1) Em cada localização da origem de  $B$ , marque o lugar como um membro do novo conjunto (sombreado) se  $B$  estiver completamente contido em  $A$  [21].

### 3.4.2 Dilatação

Seja novamente com  $A$  e  $B$  conjuntos de  $Z^2$ , a dilatação de  $A$  por  $B$  é indicada por:

$$A \oplus B = \bigcup_Z (B_Z \cap A) \neq \emptyset. \quad (3.13)$$

A Equação 3.13 tem por base a reflexão de  $B$  ao redor de sua origem, a qual é seguida da translação dessa reflexão por  $Z$ . É então a dilatação de  $A$  por  $B$  o conjunto de todos os deslocamentos  $Z$  de maneira que se sobreponham  $\hat{B}$  e  $A$  por ao menos um elemento. Um exemplo de dilatação é apresentada na Figura 3.13 .

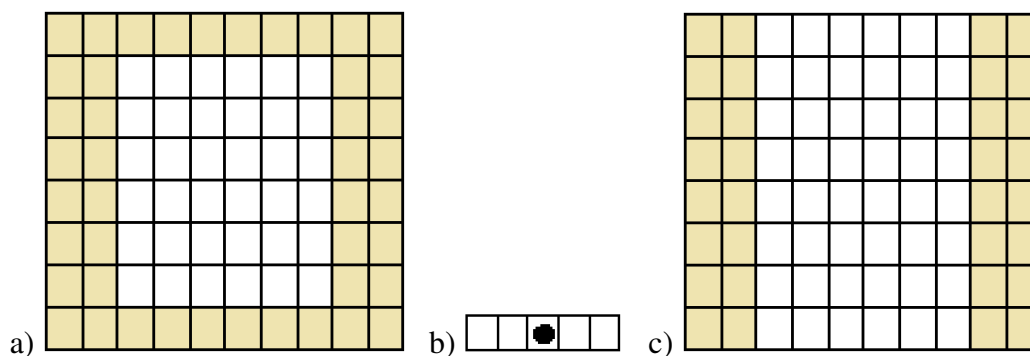


Figura 3.11: a) Conjunto  $A$ , b) Elemento estruturante  $B$  e c) Resultado de  $(A \oplus B)$ .

A dilatação é uma operação que aumenta ou engrossa os objetos da seguinte forma:

- 1) Considerando a Figura 3.11, primeiramente crie um novo conjunto fazendo varredura de  $B$  ao longo de  $A$ , para que a origem de  $B$  passe por todos os elementos de  $A$ ;
- 2) Caso  $B$  não esteja completamente contido em  $A$ , marque o lugar como não sendo um membro do novo conjunto ( não sombreado) [21].

### 3.4.3 Abertura

Indica-se a abertura do conjunto  $A$  pelo estruturante elemento  $B$  desta forma:

$$A \circ B = (A \ominus B) \oplus B. \quad (3.14)$$

A abertura de  $A$  por  $B$  é a erosão de  $A$  por  $B$ , seguida de uma dilatação do resultado por  $B$ .

### 3.4.4 Fechamento

Indica-se o fechamento do conjunto  $A$  pelo estruturante elemento  $B$  assim:

$$A \bullet B = (A \oplus B) \ominus B. \quad (3.15)$$

Esta operação é representada pela dilatação de  $A$  por  $B$ , seguida pela erosão do resultado por  $B$  [21].

## 3.5 Descritores de Características de Imagens

Vários são os métodos utilizados como descritores de características de imagens, mas as linhas que se seguem tiveram como foco considerar as análises de textura e descrição de características com base no domínio da frequência com as transformadas: Transformada Discreta de Fourier (em inglês *Discrete Fourier Transform – DFT*), Transformada Wavelet Discreta (em inglês *Discrete Wavelet Transform – DWT*) e Transformada Discreta de Cosseno (em inglês *Discrete Cosine Transform – DCT*), considerando-se ainda a análise no domínio espacial de momentos de Histograma e na Matriz de Coocorrência de Níveis de Cinza (em inglês (*Gray Level Co-occurrence Matrix – GLCM*)).

### 3.5.1 Descritores no Domínio Espacial

Dizem respeito ao processamento no próprio plano da imagem. Nesta categoria as abordagens têm fundamento na direta manipulação dos *pixels* que compõem as imagens [21]. Neste cenário, empregaram-se duas abordagens no domínio espacial a fim de descrever a textura das imagens, sendo: 1) O estudo dos momentos do histograma e, 2) O emprego de descritores tendo como ponto de partida a GLCM.

#### Histograma

Histograma representa a distribuição estatística de cinza quanto a seus níveis. Assim, trata-se de uma representação gráfica da quantidade de *pixels* em associação a cada graduação de cinza existente na imagem, que pode ser expressa também em percentual da quantidade total de *pixels* existentes.

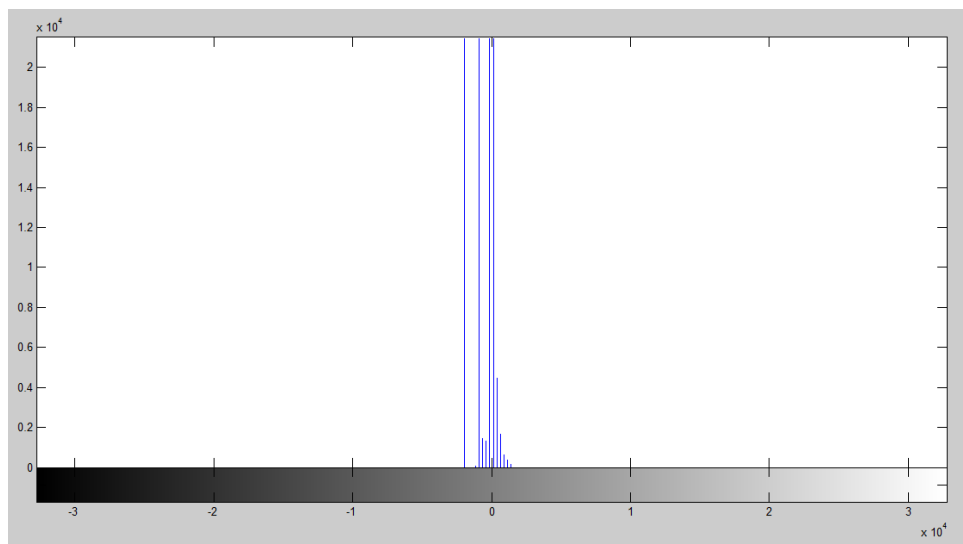


Figura 3.12: Exemplo de histograma com imagem de 16 bits.

A extração de características a partir do histograma de imagens utilizada neste trabalho pode ser descrita conforme se segue: seja  $z$  uma variável aleatória denotando a intensidade discreta de uma imagem e seja  $p(z_i)$ ,  $i = \{1, 2, \dots, L - 1\}$  o histograma correspondente, sendo  $L$  o número

de níveis distintos de intensidade: o n-ésimo momento de  $z$  em torno da média é dado pela Equação 3.16.

$$\mu_n(z) = \sum_{i=0}^L ((z_i - m)^n) p_{z_i}. \quad (3.16)$$

Onde:  $(m)$  é o valor médio de  $z$ , ou seja, a intensidade média dos *pixels* da imagem,

$$m = \sum_{i=0}^L z_i p(z_i). \quad (3.17)$$

A uniformidade da imagem pode ser obtida pela Equação 3.18, a seguir,

$$uniformidade = \sum_{i=0}^{L-1} p^2(z_i). \quad (3.18)$$

A entropia – medida de variabilidade nos níveis de intensidade das imagens – é dada pela Equação 3.19,

$$entropia(z) = \sum_{i=0}^{L-1} p(z_i) (\log_2) p(z_i). \quad (3.19)$$

Variância é representada por,  $\alpha^2(z) = \mu(z)$ , uma medida de contraste de intensidade que pode ser atribuída aos descritores de suavidade relativa.

Mediana se refere à medida de tendência central que mostra o valor do meio em uma amostra de informações [21].

### Matriz de coocorrência de Níveis de Cinza

A GLCM é um método estatístico para extração de características com capacidade para analisar as coocorrências existentes entre pares de *pixels*. O algoritmo para geração de uma GLCM a partir de uma imagem em nível de cinza faz uma varredura na matriz que representa a imagem e determina, para cada combinação de níveis  $i$  e  $j$ , a quantidade de coocorrências que existem entre eles considerando uma distância  $d$  e uma orientação  $\theta$ .

A variável  $\theta$  pode assumir as orientações de  $(0^\circ, 45^\circ, 90^\circ$  e  $135^\circ)$ , a Figura 3.13 exemplifica os valores de orientações  $\theta$  [?].

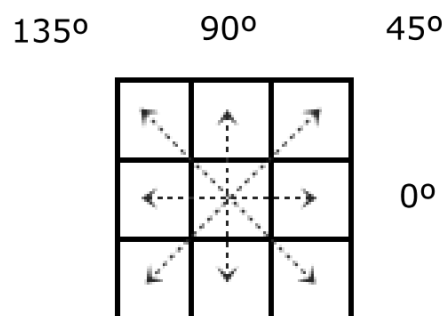


Figura 3.13: Exemplo das orientações  $\theta$  da GLCM.

O exemplo apresentado na Figura 3.14- a) mostra uma matriz de *pixels* de tamanho (5 x 8). Considere que esta matriz seja normalizada contendo apenas 4 níveis de cinza representados pelos valores das células. A sua GLCM é representada pela Figura 3.14- b), uma matriz de tamanho referente à quantidade de níveis de cinza da matriz de *pixels*.

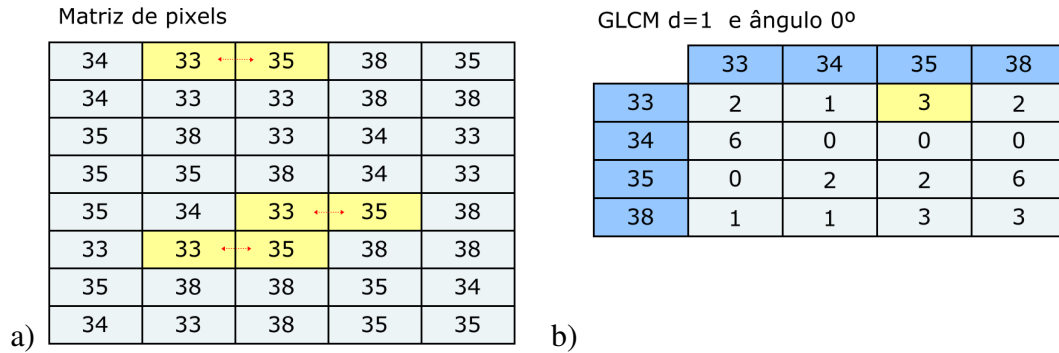


Figura 3.14: a) Matriz de pixels, b) GLCM

A matriz apresentada na Figura 3.14- b) é uma contagem de quantas vezes os pares de *pixels* coocorrem na matriz representada pela Figura 3.13- a), por exemplo: os valores hachurados em amarelo na Figura 3.13- a) representam as coocorrência de *pixels* de valor 33 seguido de *pixels* de valor 35, considerando uma orientação  $\theta = 0^\circ$  e distância  $d=1$ . A contagem das coocorrências são transferidas para a GLCM, por exemplo: os *pixels* de valor 33 seguidos de *pixels* de valor 35 coocorrem por 3 vezes, como mostra a GLCM representada pela Figura 3.14- b).

Uma vez que a GLCM é preenchida, aplica-se a ela cálculos estatísticos para levantar as características mais relevantes da textura. A partir da GLCM, Haralick e colaboradores, estabeleceram 14 descritores de estatísticas para extração de textura. Das medidas estatísticas propostas pelo autor, foram utilizadas as que se seguem: (correlação, contraste, uniformidade, entropia, homogeneidade e diferença inversa) [22].

**Descritores de haralick:** seja  $P(i, j)$  a GLCM, e  $N_g$  o número de níveis discretos de intensidade de uma imagem.  $\mu$  é a média de  $P(i, j)$ ,  $\mu_x(i)$  representa a média da linha  $[i]$ ,  $\mu_y(j)$  representa a média da linha  $[j]$ ,  $\alpha_x(j)$  é igual ao desvio padrão da linha  $i$ , e  $\alpha_y(j)$  é igual ao desvio padrão da coluna  $[j]$ .

Correlação – medida de quão relacionado está um *pixels* com seu vizinho em toda imagem:

$$correlao = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijP(i, j) - \mu_x(i)\mu_y(j)}{\alpha_x(i)\alpha_y(j)}. \quad (3.20)$$

Contraste – medida de intensidade entre um *pixel* e seu vizinho em toda imagem:

$$contraste = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i - j]^2 P(i, j). \quad (3.21)$$

Energia ou uniformidade – medida no intervalo [0,1]. A uniformidade é (1) para uma imagem constante:

$$uniformidade = \sum_{i=1}^{n-1} \sum_{j=0}^{n-1} P(i, j)^2. \quad (3.22)$$

Entropia – mede aleatoriedade dos elementos da matriz de coocorrência:

$$entropia = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \log_2[P(i, j)]. \quad (3.23)$$

Homogeneidade – medida de proximidade espacial da distribuição de elementos da matriz de coocorrência, na diagonal. O intervalo de medida é (0,1):

$$homogeneidade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j)}{1 + |i - j|}. \quad (3.24)$$

Diferença inversa da amostra – (ivglcm):

$$D - inversa = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j)}{1 + \left(\frac{|i-j|^2}{N^2}\right)}. \quad (3.25)$$

### 3.5.2 Descritores no Domínio da Frequência

Consiste, primeiro, em transformar uma imagem no domínio da transformada, realizar o processamento nesse domínio e obter a transformada inversa para retornar os resultados ao domínio espacial [21].

**Transformada de Fourier:** é uma ferramenta matemática fundamental na solução de problemas de processamento de imagens digitais [23]. Uma classe particularmente importante de transformadas lineares bidimensionais  $T(u, v)$  pode se expressar na forma geral da Equação 3.26

$$T(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) r(x, y, u, v), \quad (3.26)$$

onde  $f(x, y)$  é uma imagem de entrada,  $r(x, y, u, v)$  é chamado de *kernel* de transformação direta e a Equação 3.26 é calculada para  $u = 0, 1, 2, \dots, M - 1$  e  $v = 0, 1, 2, \dots, N - 1$ .

As variáveis  $x$  e  $y$  são ainda espaciais, enquanto  $M$  e  $N$  são dimensões de linhas e colunas de  $f$ . As variáveis  $u$  e  $v$  são respectivamente as variáveis da transformada.

$T(u, v)$  é chamado de transformada direta de  $f(x, y)$ . Dado então  $T(u, v)$ , é possível recuperar  $f(x, y)$  utilizando-se da transformada inversa de  $T(u, v)$ , como mostra a Equação 3.27

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u, v) s(x, y, u, v), \quad (3.27)$$

para  $x = 0, 1, 2, \dots, M - 1$  e  $y = 0, 1, 2, \dots, N - 1$ , onde  $s(x, y, u, v)$  é chamada de *kernel* de transformação inversa. As Equações 3.26 e 3.27 juntas são chamadas de um par de transformadas.

Diz-se que o *kernel* de transformação direta é separável se:

$$r(x, y, u, v) = r_1(x, u) r_2(y, v). \quad (3.28)$$

Além disso, diz-se que o *kernel* é simétrico se  $r_1(x, y)$  for funcional igual a  $r_2(x, y)$ , de forma que:

$$r(x, y, u, v) = r_1(x, u)r_1(y, v). \quad (3.29)$$

A transformada bidimensional de Fourier apresentam os seguintes kernels direto e inverso:

$$r(x, y, u, v) = e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})}, \quad (3.30)$$

e respectivamente,

$$s(x, y, u, v) = \frac{1}{MN} e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})}. \quad (3.31)$$

Em sua forma polar de representação, a DFT bidimensional se dá por:

$$F(u, v) = |F(u, v)| e^{j\phi(u, v)}. \quad (3.32)$$

O espectro de Fourier ou magnitude é dado por:

$$|F(u, v)| = [R^2(u, y) + I^2(u, v)]^{\frac{1}{2}}. \quad (3.33)$$

A Figura 3.15 mostra exemplos de espectro de Fourier calculados para duas pequenas imagens.

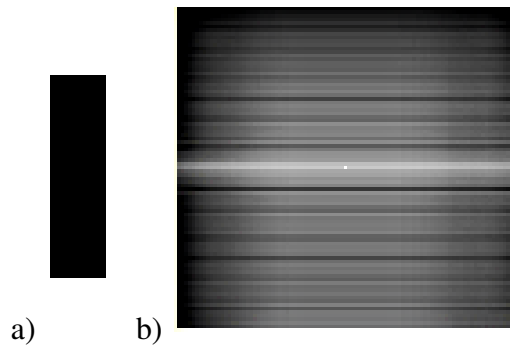


Figura 3.15: a) Figura retangular, b) espectro de Fourier da Figura retangular.

**Estimador por periodograma:** O conceito de frequência está relacionado com a variação da intensidade da função em relação ao espaço. Se a variação da intensidade dos pixels de uma região for grande, a frequência da região será alta.[23]

Espectro é a descrição do sinal, consistindo de componentes senoidais em várias frequências. O estimador por periodograma da densidade espectral de potência (PSD) de um dado processo  $Y$  é definido como

$$\gamma_{Y,P}(\epsilon^{jw}) = \frac{1}{L} \left| \sum_{n=0}^{L-1} y(n) \epsilon^{-jwn} \right|^2, \quad (3.34)$$

onde o subscrito  $P$  se origina do nome do estimador. É equivalente a aplicar uma janela retangular (vetor de características) sobre o intervalo  $0 \leq n \leq (L - 1)$  do sinal  $y(n)$ , elevar ao quadrado

o módulo da transformada de Fourier da sequência truncada e normalizar o resultado por um fator  $L$ , para obter então a medida da densidade espectral de potência .[23]

**Transformada DCT:** Uma das transformações utilizadas com mais frequência para a compressão de imagens é a DCT. Esta transformada é obtida substituindo os *kernels* iguais nas Equações 3.28, 3.29.

$$r(x, y, u, v) = \alpha(u)\alpha(v) \cos\left[\frac{(2x+1)u\pi}{2n}\right] \cos\left[\frac{(2y+1)v\pi}{2n}\right], \quad (3.35)$$

sendo  $r(x, y, u, v) = s(x, y, u, v)$ ,

$$\alpha(u) = \sqrt{\frac{1}{n}} \text{ para } u = 0,$$

$$\alpha(u) = \sqrt{\frac{2}{n}} \text{ para } u = 1, 2, 3, \dots, n-1.$$

Para aplicar a DCT, geralmente a imagem é dividida em blocos de  $8 \times 8$  *pixels* e a transformada de cosseno é aplicada para cada bloco, como resultado, tem-se uma matriz  $8 \times 8$  de componentes de frequência  $F(u, v)$ .

O primeiro coeficiente de frequência, na posição (0,0) da matriz de saída, é chamado coeficiente *DC*. O coeficiente *DC* é uma medida do valor médio dos 64 *pixels* de entrada. Os outros 63 elementos da matriz de saída são chamados de coeficientes *AC*. Exemplo representado na Figura 3.16.

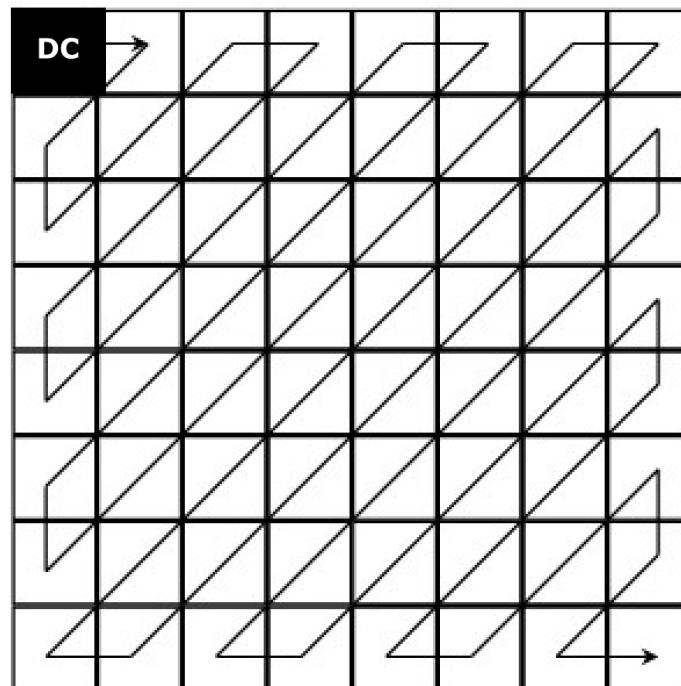


Figura 3.16: Exemplo de matriz de coeficientes DCT com leitura em zig-zag DCT.



Quando se percorre uma matriz de coeficientes DCT do primeiro coeficiente para o 64º coeficiente de frequência, está se movendo da informação de frequência baixa para a informação de frequência alta. Os coeficientes de frequência mais alta são gradativamente menos importantes para a qualidade percebida da imagem.

Exceto para o coeficiente DC, que neste trabalho é descartado, os 63 coeficientes AC são lidos usando a varredura zig-zag, transformando a matriz em um fluxo de dados ajustado para cálculos estatísticos.

**Transformada Wavelet:** as wavelets diferem da transformada de Fourier uma vez que permitem, para além de obter informação do domínio da frequência, também obter informação do domínio do tempo.

Para decompor uma imagem bidimensional pode-se usar a transformada wavelet em duas dimensões, para isso são necessárias uma função de escala bidimensional,  $\varphi(x,y)$  e três wavelets bidimensionais  $\psi^H(x,y)$ ,  $\psi^V(x,y)$  e  $\psi^D(x,y)$ .

Essas wavelets medem variações de intensidade das imagens ao longo das direções:  $\psi^H$ , mede variações ao longo de colunas,  $\psi^V$  mede variações ao longo de linhas e  $\psi^D$  corresponde a variações ao longo de diagonais.

Dadas as funções wavelet de escala, estende diretamente a DWT unidimensional para a DWT bidimensional [21].

Primeiramente definem-se as funções de base escalonadas e transladadas:

$$\varphi_{j,m,n}(x,y) = 2^{\frac{1}{2}} \varphi(2^j x - m, 2^j y - n) \quad (3.36)$$

e

$$\psi_{j,m,n}^i(x,y) = 2^{\frac{1}{2}} \psi(2^j x - m, 2^j y - n). \quad (3.37)$$

Sendo  $i = H, V, D$ , índice que identifica as wavelets separáveis e direcionalmente sensíveis, tem-se que:

- $\psi^H(x,y) = \psi(x), \varphi(y)$ ,
- $\psi^V(x,y) = \varphi(y), \psi(x)$ ,
- $\psi^D(x,y) = \psi(x), \varphi(y)$ .

A transformada wavelet discreta da imagem  $f(x,y)$  de tamanho  $M \times N$  é dada como:

$$W\varphi(j_0 m, n) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{x=0} \sum_{n=0}^{y=0} f(x,y) \varphi_{j_0, m, n}(x,y), \quad (3.38)$$

e

$$W\psi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{x=0} \sum_{n=0}^{y=0} f(x,y) \psi_{j, m, n}^i(x,y), \quad (3.39)$$

onde  $j_0$  é uma escala inicial arbitrária e os coeficientes  $W\varphi(j_0 m, n)$  definem uma aproximação de  $f(x,y)$  na escala  $j_0$ . Definições horizontais, verticais e diagonais de  $f(x,y)$  para escalas  $j \geq j_0$  é dado por  $W\psi^i(j, m, n)$ . Geralmente define-se  $j_0$  e seleciona-se  $N=M=2^J$ , de forma que  $J = 0, 1, 2, \dots, J-1$  e  $m=n = 0, 1, 2, \dots, 2^j-1$ . Dado portanto,  $W\varphi$  e  $W\psi^i(j, m, n)$  nas Equações 3.38, 3.39 é possível obter  $f(x,y)$  utilizando a transformada wavelet inversa dada pela Equação:

$$f(x,y) = \frac{1}{\sqrt{MN}} \sum_m \sum_n W\varphi(j_0 m, n) \varphi_{j_0, m, n}(x,y), \quad (3.40)$$

somada a ,

$$+\frac{1}{\sqrt{MN}} \sum_{i=H,V,D} \sum_{j=j_0}^{\infty} \sum_m \sum_n W_{\psi}^i(j,m,n) \psi_{j,m,n}^i(x,y). \quad (3.41)$$

### 3.6 Análise de Componentes Principais

A Análise de Componentes Principais, ( em inglês *Principal Component Analysis* – PCA), é um método que projeta os dados multivariados em um espaço de dimensão menor reduzindo a dimensionalidade do espaço do conjunto dos dados destacando a relação das informações mais importantes.

O método PCA está fundamentado no conceito de covariância entre as variáveis. Consiste em fatorar a matriz de dados  $X$ , de modo que  $X = TL^T + E$ , onde  $L$  é a matriz dos pesos (mostram as relações entre as variáveis),  $T$  corresponde às coordenadas das amostras no novo sistema de eixos (escores) e  $E$  a matriz dos resíduos [24]. O símbolo  $T$  é o operador de transposição de matriz.

Inicialmente, deve-se calcular a matriz de covariância,  $C$ .

$$(C) = E_0^T E_0 / (n - 1), \quad (3.42)$$

sendo  $E_0$  a matriz pré-processada e  $n$  é o número de linhas da matriz. Em seguida calculam-se os autovalores e autovetores normalizados de  $C$ .

$$CL = \Delta L, \quad (3.43)$$

onde  $\Delta$  é a matriz diagonal dos autovalores. Cada autovetor  $L$  é um vetor de pesos de uma componente principal. Cada autovalor  $\lambda$  fornece a quantidade de variância explicada pela respectiva componente, de modo que  $PC1$  tem o maior autovalor,  $PC2$  o segundo maior, e assim sucessivamente [25].

### 3.7 Aprendizado de Máquinas

Uma das tarefas do aprendizado de máquinas se aplica ao reconhecimento de padrões, sendo este definido formalmente como o processo pelo qual um padrão/sinal recebido é incluído a uma classe dentre um número predeterminado de classes [26].

Um classificador realiza o reconhecimento de padrões passando inicialmente por uma etapa de treinamento, durante a qual se apresenta ao classificador um conjunto de padrões de entrada com a categoria à qual cada padrão particular pertence. Em seguida, apresenta-se ao classificador um novo padrão ainda não visto, mas que pertence à mesma população de padrões utilizada para treinar o classificador. Portanto, o classificador deve ser capaz de identificar a classe daquele padrão particular baseando-se na informação que ele extraiu dos dados de treinamento [26].

Estudos apresentam uma grande variedade de elementos de aprendizagem de máquinas. Para compreendê-los, é útil ver como sua estrutura é afetada pelo contexto em que eles irão operar.

O campo de aprendizagem de máquinas normalmente distingue três casos: aprendizado por reforço, aprendizado não supervisionado, e supervisionado [26].

**Aprendizagem por reforço:** este aprendizado, em vez de ser informado sobre o que fazer por um instrutor, este inclui o problema de aprender como o ambiente funciona, é realizado pela interação contínua com o ambiente, visando sempre minimizar um índice escalar de desempenho.

**Aprendizagem não supervisionada:** envolve a aprendizagem de padrões na entrada, quando não são fornecidos valores de saídas específicos. Portanto, não são fornecidos ao classificador exemplos rotulados da função a ser aprendida.

**Aprendizagem supervisionada:** envolve a aprendizagem de uma função a partir de exemplos de entradas e saídas, envolve o fornecimento de uma resposta desejada ao classificador por cada vetor de treinamento, ou seja, há exemplos rotulados da função a ser aprendida pelo classificador.

Um modelo de aprendizagem supervisionada consiste de três componentes inter-relacionados:

1 ) Ambiente, o qual fornece um vetor  $x$  com uma função de distribuição cumulativa fixa, mas desconhecida  $F_x(x)$ .

2 ) Professor, fornece uma resposta desejada  $d$  para cada vetor de entrada  $x$  recebido do ambiente.

3 ) Classificador, este consiste no algoritmo de aprendizado o qual é capaz de implementar um conjunto de funções de mapeamento de entrada e saída descritas por:  $y = F(x, w)$ , onde  $y$  é a resposta real produzida pelo classificador em resposta à entrada  $x$ , e  $w$  é um conjunto de parâmetros livres selecionados do espaço de parâmetros (pesos).

O problema da aprendizagem supervisionada é selecionar a função  $F(x, w)$  que aproxima a resposta desejada  $d$  de uma forma ótima.

Por escopo do decorrente estudo, a seguir, são descritos os algoritmos supervisionados utilizados por esta pesquisa.

### 3.7.1 Máquina de Vetores de Suporte

Máquina de vetores de suporte (em inglês *Support Vector Machine* – SVM) é uma técnica de aprendizado supervisionado estudada e desenvolvida por Vladimir Vapnik[27].

A ideia da SVM é encontrar hiperplanos que separam os dados o máximo possível dentre os geralmente inúmeros hiperplanos que possam separá-los.

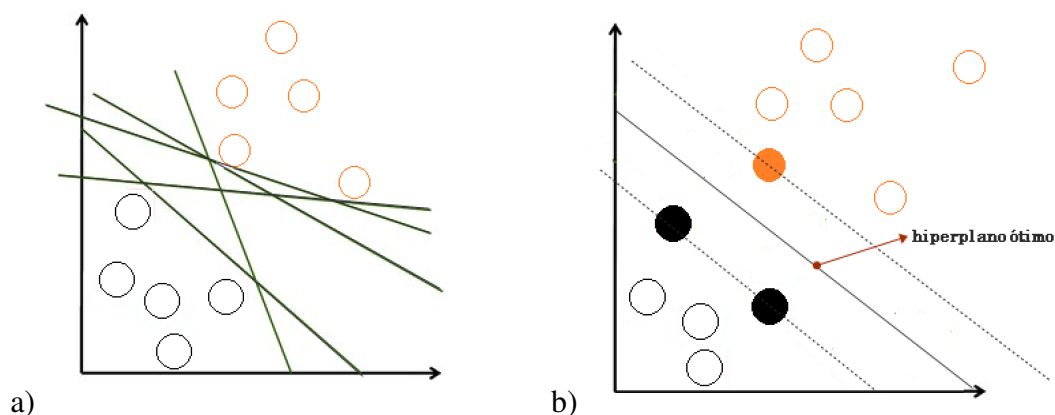


Figura 3.17: a) exemplo de vários hiperplanos, b) exemplo de hiperplano ótimo.

O ideal é que os hiperplanos de separação tenham desempenho tão bom nos dados de treinamento quando na generalização de desempenho em novos dados de entrada [28]. A largura do hiperplano interfere diretamente na probabilidade de ocorrer um erro na classificação. Quanto maior for a distância entre as margens de separação, mais preciso tende a ser o algoritmo de classificação. Isso mostra que o principal objetivo do SVM é encontrar a maior separação entre as margens das classes analisadas [29].

Durante o treinamento das SVMs, são escolhidos entre todos os vetores de treinamento, aqueles mais importantes no processo de definição dos limites de cada classe. Esses vetores definem o hiperplano ótimo de separação. A obtenção deste hiperplano ótimo está definido por meio da minimização da função quadrática descrita conforme a Equação 3.44.

$$f(x) = \sum_{i=1}^N \alpha_i y_i (h(x), h(x_i)) + \beta 0. \quad (3.44)$$

onde  $\alpha$  e  $\beta$  representam os parâmetros encontrados durante o treinamento, e  $(h(x), h(x_i))$  constituem os vetores de características e  $y$  é a definição de classes [30].

Para os dados não linearmente separáveis, a SVM trabalha para que estes sejam linearmente separáveis usando funções *kernel*. Essas funções dependem de duas operações matemáticas:

- 1) O mapeamento não linear de um vetor de entrada para um espaço de características de alta dimensionalidade.
- 2) A construção de um hiperplano ótimo para separar as características adquiridas no passo 1.

Várias são as funções que podem ser disponibilizadas para o *kernel*, mas, por escopo deste trabalho, cita-se a função Linear, recomendada quando a separação linear dos dados for direta e a função de Base Radial (RBF), utilizada em casos de não linearidade de separação dos dados.

### 3.7.2 K-Vizinhos mais Próximos

Os algoritmos K-Vizinhos mais Próximos (em inglês *K-Nearest Neighbor* – KNN) estão entre os mais simples dos algoritmos de aprendizado de máquina. A ideia é memorizar o conjunto de treinamento e depois prever o rótulo de qualquer nova instância com base nos rótulos de seus vizinhos mais próximos no conjunto de treinamento. A lógica por trás de tal método é baseada na suposição de que os recursos que são usados para descrever os pontos de domínio são relevantes para suas rotulações de uma maneira que faz com que pontos próximos tenham o mesmo rótulo[31]. Portanto, o KNN é baseado na medição das distâncias entre os dados de teste e cada um dos dados de treinamento para decidir a saída final da classificação.

Funções de distância: a métrica geralmente utilizada pelo KNN para o cálculo de distância  $D$  entre dois pontos é a distância euclidiana.

A escolha de  $K$  define a localidade do KNN. A Figura 3.18 apresenta um exemplo de classificação com dois rótulos de classe e  $K = 7$ . No exemplo, são aferidas as distâncias de uma nova amostra, representada por um quadrado, as demais amostras de treinamento são representadas pelas bolinhas azuis e amarelas. A variável  $K$  representa a quantidade de vizinhos mais próximos que serão utilizados para analisar de qual classe a nova amostra pertence. Com isso, das sete amostras de treinamento mais próximas da nova amostra, cinco são do rótulo A e duas do rótulo B. Portanto, como existem mais vizinhos do rótulo A, a nova amostra receberá o rótulo A [31].

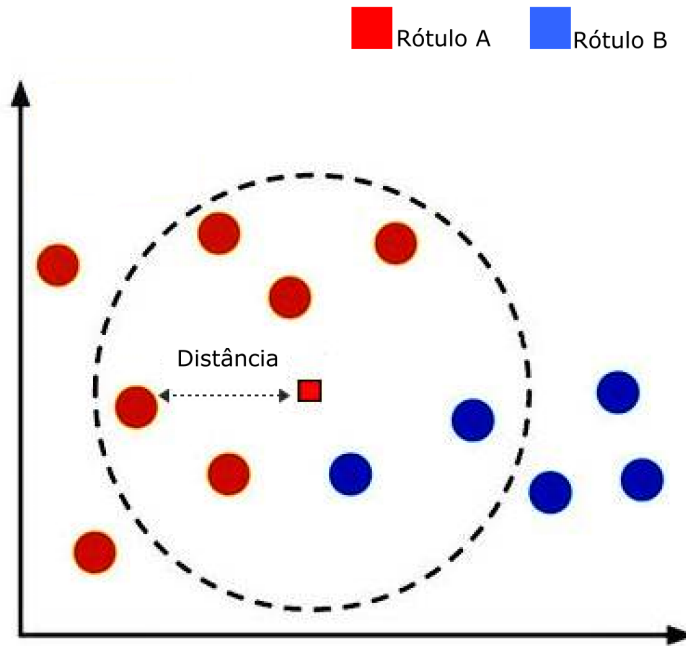


Figura 3.18: Exemplo de K-Vizinhos mais próximos.

### 3.7.3 Naive Bayes

O objetivo do classificador Naive Bayes é por meio de análises estatísticas das informações coletadas, verificar se uma amostra analisada pertence ou não a uma determinada classe [32].

O Naive Bayes é fundamentado na Equação 3.45.

Exemplo da aplicação do algoritmo Naive Bayes: sejam  $x = (x_1, x_2, \dots, x_d)$  amostras  $\in \mathfrak{X}$  e sejam as classes  $c_1, c_2, \dots, c_n$ .

A abordagem Bayesiana supõe que as probabilidades de cada classe  $P(c_i)$  e as densidades de probabilidade condicionais  $p(x|c_i)$  de  $x$  com respeito a cada uma das classes  $c_i$ , sendo  $i = \{1, 2, \dots, n\}$ , sejam conhecidas.

Na ausência de qualquer outra informação, pode-se classificar uma amostra  $x$  como sendo da classe  $c_i$  de maior probabilidade. Porém, dado que  $x$  foi observado, o classificador acerta a classificação com probabilidade  $P(\omega_i)$ , mas erra com probabilidade  $\sum_{j \neq i} P(\omega_j)$ .

Como tem-se as condicionais, pode-se utilizar o teorema de Bayes e calcular a probabilidade  $P(\omega_i|x)$ , ou seja,

$$P(\omega_i|x) = \frac{P(\omega_i)p(x|\omega_i)}{P(x)}. \quad (3.45)$$

na qual  $P(c_i)$  é a *priori*,  $p(x|c_i)$  é a densidade condicional ou verossimilhança,  $p(x) = \sum_j P(c_j) p(x|c_j)$  é a evidência e  $P(c_i|x)$  é a *posteriori*. Por fim o classificador toma a sua decisão baseando-se nos valores a *posteriori*.

## 3.8 Ensemble

Um *ensemble* de preditores é uma junção de vários classificadores cujas decisões são combinadas por alguma regra para classificar os exemplos de teste [33]. Vários são os meios de programar a combinação de classificadores, como exemplo, citam-se os métodos seguintes.

- Métodos de combinação linear: as saídas fornecidas pelos classificadores são combinadas linearmente. Como exemplos de métodos de combinação linear têm-se a soma, a média, a mediana [34].

- Métodos de combinação não linear: as saídas dos classificadores são combinadas de forma não linear, como o que ocorre no uso do voto majoritário [35].

- Métodos baseados em estatísticas: utilizam-se probabilidades ou combinações estatísticas nas saídas geradas pelos classificadores.

Um conjunto de classificadores geralmente mostra um desempenho melhor do que os classificadores individuais que o compõem. A explicação para essa melhoria é dada a seguir.

Suponha que os classificadores binários  $\{c_1, c_2, \dots, c_n\}$  sejam testados tendo como entrada para classificação o conjunto de dados  $x$ , o qual possui elementos a serem classificados.

Quando a saída de um dos classificadores  $c$  estiver errada para um dos elementos de  $x$ , a maioria dos outros classificadores pode estar correta. Dessa forma, em caso de votação majoritária prevalecem a maioria dos acertos para o elemento classificado, o que pode acarretar em melhorias para o classificador [35].

A função do uso de *ensemble* se faz quando os classificadores individuais não são suficientes para classificar completamente todos os exemplos de teste desconhecidos.

### 3.8.1 Bagging

Para formação do *ensemble* com o método *Bagging*, os classificadores são treinados de forma independente por diferentes conjuntos de treinamento conforme o método de inicialização. Para construí-los é necessário criar  $k$  conjuntos de treinamento idênticos e replicar esses dados de treinamento de forma aleatória para construir  $k$  classificadores independentes por amostragem com reposição. Em seguida, devem-se agregar os classificadores por um método de combinação apropriado, tal como a maioria de votos, ou outros modelos de combinações, por exemplo, média e soma [35].

### 3.8.2 Boosting

A abordagem por *Boosting* foi proposta por Schapire em 1990. A principal diferença em relação ao *Bagging* é que os conjuntos de dados reamostrados são construídos especificamente para gerar aprendizados complementares e a importância do voto é ponderada com base no desempenho de cada modelo, em vez da atribuição de mesmo peso para todos os votos [36].

### 3.8.3 AdaBoost

O modelo *Adaptive Boosting* – *AdaBoost* é uma implementação do método *Boosting*, portanto, é um algoritmo de aprendizagem que tem por finalidade combinar por algum modo as saídas de vários classificadores fracos buscando alcançar a performance de um preditor forte [36].

O AdaBoost chama um classificador fraco repetidamente em iterações  $t = 1, \dots, T$ . Para cada chamada a distribuição de pesos  $D_t$  é atualizada para indicar a importância do exemplo no conjunto de dados usado para classificação. Inicialmente todas as instâncias têm os mesmos pesos e, estes são alterados durante o treinamento. Os padrões que são incorretamente classificados têm os seus pesos incrementados, no entanto, pesos associados a padrões corretamente classificados sofrem um decréscimo nos seus pesos correspondentes [36]. Dessa forma, o algoritmo de aprendizagem consegue concentrar esforços nos padrões mais difíceis de serem aprendidos.

O método para se obter a hipótese final  $H(x)$  fornecida pelo *AdaBoost* se faz por uma combinação ponderada das diversas saídas das iterações, assim, uma votação ponderada das  $T$  hipóteses geradas é realizada para gerar a saída final do classificador [36].

### 3.9 Medidas de Acurácia em Métodos de Classificação

Um classificador pode expressar os resultados apresentados na matriz de confusão exemplificada na Figura 3.19. A matriz de confusão é uma ferramenta para avaliação de modelos estatísticos. As linhas na matriz de confusão representam os valores previstos para o modelo, sendo que as colunas representam os valores atuais.

		Classe esperada	
		VP	VN
Classe prevista	VP	VERDADEIRO POSITIVO VP	FALSO POSITIVO FP
	VN	FALSO NEGATIVO FN	VERDADEIRO NEGATIVO VN

Figura 3.19: Exemplo de matriz de confusão.

Com base nos resultados da matriz de confusão é possível calcular as métricas de sensibilidade, especificidade e acurácia [37].

Estas medidas podem ser usadas na classificação de qualquer entidade, são medidas importantes para seleção e interpretação de testes de diagnóstico usados na prática clínica.

O resultado de um classificador pode ser dado como Positivo (P) ou Negativo (N), as amostras, cujos resultados são contrários aos grupos que pertencem, são definidas como Falso Negativo (FN) ou Falso Positivo (FP).

A sensibilidade de um método reflete o quanto este é eficaz em identificar corretamente, dentre todos os indivíduos avaliados, ou seja, mede quão eficaz é o resultado positivo para identificar aqueles com a doença:

$$sensibilidade = \frac{VP}{VP + FN} \quad (3.46)$$

A especificidade do teste resume quão eficaz é o resultado de um teste negativo na identificação daqueles sem doença:

$$especificidade = \frac{VN}{VN + FP}. \quad (3.47)$$

Acurácia é a proporção de acertos, ou seja, o total de verdadeiros positivos e verdadeiros negativos em relação à amostra estudada:

$$acurcia = \frac{VP + VN}{VP + VN + FP + FN}. \quad (3.48)$$

### 3.10 Sobreajuste

Sobreajuste, (em inglês *overfitting*) é uma terminologia que se refere a um modelo classificador que se ajusta aos seus dados de treinamento, mas, esses conceitos não se aplicam a novos dados de entrada, afetando negativamente a capacidade de generalização dos modelos [38].

Existem técnicas adicionais que são utilizadas para ajudar a encontrar um modelo com maior generalização e diminuir o sobreajuste. Citam-se aqui os métodos de reamostragem por validação cruzada e aumento de dados.

#### 3.10.1 Validação Cruzada

A técnica de reamostragem mais popular é a validação cruzada *k-fold*, exemplificado na Figura 3.21. Este método permite que se treine e teste um modelo *k* vezes em diferentes subconjuntos de dados de treinamento e crie uma estimativa do desempenho de um modelo de aprendizado de máquina em dados não vistos. Primeiramente, divide-se o conjunto de dados *N* em *k* subconjuntos,  $k > 1$ , presumindo que *k* é divisível por *N*. O modelo é treinado com todos os subconjuntos, exceto um, e o procedimento é repetido para um total de *k* tentativas, fazendo uso, por vez, de um subconjunto diferente para a validação [30].

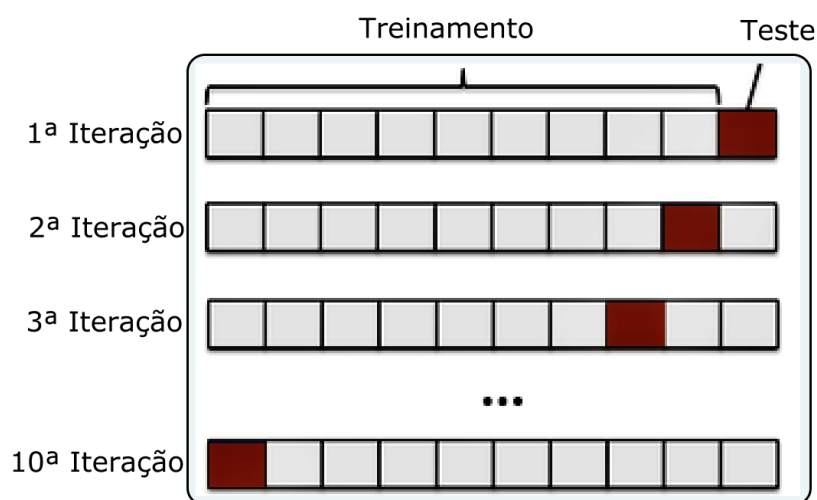


Figura 3.20: Exemplo de validação cruzada.



## Leave-One-Out

Outro método de validação cruzada é o método *leave one out*, em que um conjunto de dados com  $[N]$  exemplos, realizam-se  $[N]$  testes; para cada teste, usa-se  $N - 1$  exemplos para treinamento e um exemplo como entrada para o classificador.

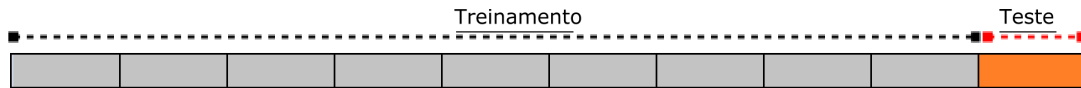


Figura 3.21: Exemplo de validação *leave-one-out*.

### 3.10.2 Aumento de Dados

O aumento de dados, em inglês (*Data Augmentation*), é uma técnica amplamente usada em muitas tarefas de aprendizado de máquina para aumentar virtualmente o tamanho do conjunto de dados de treinamento e evitar o sobreajuste[39]. As técnicas tradicionais de aumento de dados para tarefas de classificação de imagens criam novas amostras a partir dos dados de treinamento originais, por exemplo, invertendo, distorcendo, adicionando uma pequena quantidade de ruído ou buscando por um recorte de uma imagem original.

# Capítulo 4

## Metodologia Proposta

A metodologia proposta foi dividida em quatro etapas principais para a classificação da acromegalia, sendo: 1) Extração de características com o uso da DFT, DWT [40] e DCT [41]; 2) Extração de características de texturas fazendo uso dos momentos de Histogramas e GLCM; 3) Seleção de características fazendo uso da análise de PCA,[24]; e, 4) Classificação das imagens com o uso de um *ensemble* formado por classificadores binários SVM [27], KNN [42], *Naive Bayes* [29] e AdaBoost [36].

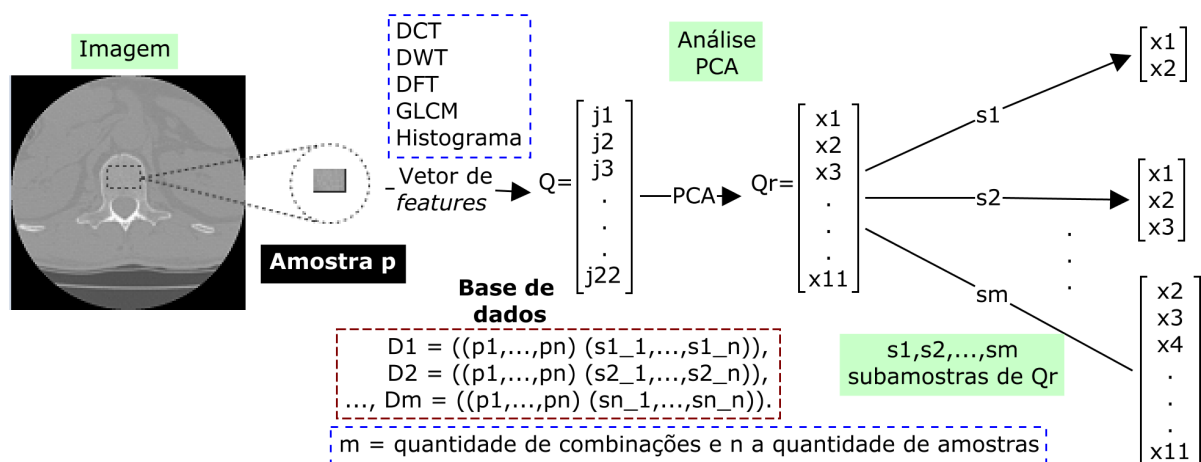


Figura 4.1: Diagrama do sistema proposto.

### 4.1 Banco de Imagens

O banco de imagens utilizado no presente estudo foi composto por duas classes, uma contendo 27 imagens de pacientes com acromegalia, e outra contendo 27 imagens de pacientes não acromegálicos. Todas as imagens do banco em questão são de 16 bits em níveis de cinza, obtidas por TC em cortes axiais da região do abdome. Ambos grupos de imagens são compostos por 13 indivíduos do sexo masculino e 14 indivíduos do sexo feminino.

As imagens foram fornecidas pelo Hospital Universitário de Brasília – HUB, localizado em Brasília, Distrito Federal, Brasil. A licença de uso das imagens para pesquisa se deu pelo conselho de ética do HUB.

A base de dados utilizada para treinamento e aprendizado de máquinas se deu por vetores de *features*. É um conjunto de descritores estatísticos calculado a partir do espectro e dos coeficientes de transformadas, de histograma e da matriz de coocorrência da Região de Interesse (em inglês, *Region Of Interest* – ROI) de cada imagem tomográfica. Aqui, a ROI foi representada pela região óssea do corpo vertebral de vértebras lombares (L1- L5), exemplificado na Figura 4.2.

Como as estruturas das vértebras obviamente mudam de tamanho de uma pessoa para outra, naturalmente também variam o tamanho da ROI de uma imagem para outra imagem. Portanto, para amostrar essa região, foram feitos testes com as imagens, buscando estabelecer um tamanho de corte da ROI que fosse o máximo possível de se amostrar respeitando as variações de tamanhos das regiões de interesses correspondentes à cada imagem da base de dados. Ao final dos testes, fixou-se o tamanho da amostra em  $(48 \times 48)$  pixels.

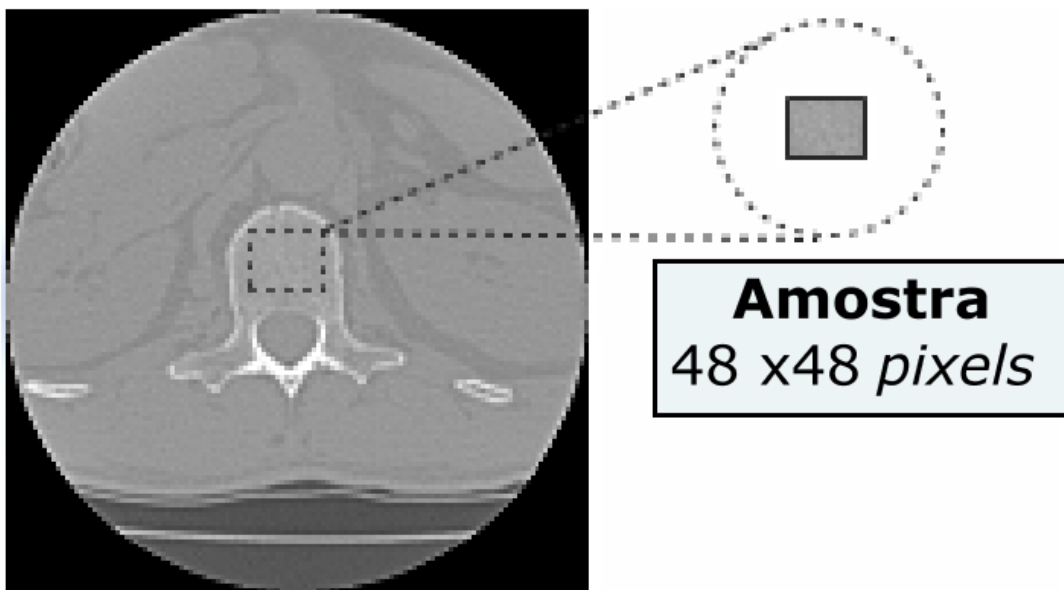


Figura 4.2: Região de interesse da imagem de TC.

#### 4.1.1 Corte Automático da Região de Interesse

O conhecimento da escala de tons de cinza existente nas imagens tomográficas possibilitou chegar ao isolamento da região óssea e ao corte automático da região de interesse via cinco etapas:

1) A partir da imagem original, criou-se uma máscara, a qual foi utilizada para extrair a localização exata do corte.

2) Para ressaltar a região óssea, converteu-se por limiar os valores das imagens para valores binários. O limiar foi proporcionado pelo algoritmo Otsu[43], descrito na Seção 3.3.2. Como o número de HU para regiões ósseas é maior que o limiar estabelecido pelo algoritmo Otsu, ressaltou-se em branco a região óssea fotografada, sendo esta uma parte interessante para o corte da ROI conforme exemplificado na Figura 4.3b;

3) Para fundir pequenas descontinuidades que restaram da etapa anterior, aplicou-se um fechamento morfológico sobre a imagem máscara. Para a realização do fechamento, fez-se uso

de um elemento estruturante quadrático de tamanho  $[3 \times 3]$  *pixels*. O resultado dessa operação é apresentado na Figura 4.3c;

4) Recursivamente buscou-se pela região central da maior área de componentes conexos utilizando vizinhança  $N_8(p)$  apresentado na Seção 3.2.2. Neste momento, aquela região representa a parte central do corpo vertebral, o exemplo é apresentado na Figura 4.3d;

5) Após encontrar o centroide do corpo vertebral apresentado na Figura 4.3e, suas coordenadas espaciais são mapeadas, a fim de lograr a orientação para o corte da amostra da ROI na imagem original.

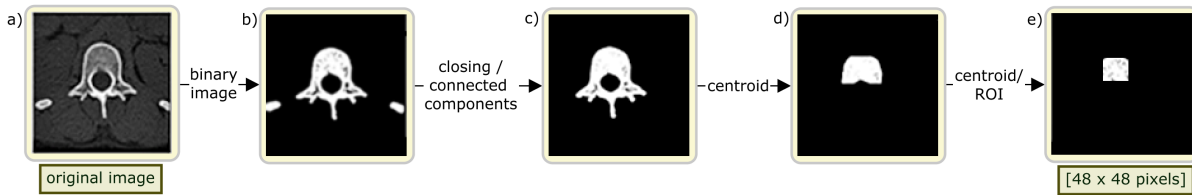


Figura 4.3: Fluxograma do corte automático da ROI.

#### 4.1.2 Extração de *Features*

A extração de *features* das amostras da ROI se deu no domínio espacial sobre o histograma, e ainda sobre o domínio das transformadas DFT, DWT e DCT.

As medidas de texturas calculadas a partir do histograma sofrem limitações de não carregarem informações sobre a posição relativa dos *pixels* em relação uns aos outros [44]. Como um modo de trazer tais informações ao processo de análise de texturas e sanar a deficiência em questão, fez-se uso de uma segunda abordagem espacial para a extração de *features*, acrescentando cálculos com o uso da GLCM.

As seguintes Etapas descrevem como se fez a extração de características das amostras retiradas da ROI das imagens tomográficas.

1) No domínio espacial, calculou-se sobre o histograma das amostras da ROI, um vetor (*vrh*), contendo os seguintes *features*: (entropia, média, mediana, uniformidade e variância).

2) Aplicou-se a DWT sobre as amostras das imagens, e a partir dos coeficientes resultantes calculou-se um vetor (*vrw*) contendo os mesmos *features* do vetor (*vrh*), porém no domínio da transformada DWT.

3) Aplicou-se a transformadas DFT sobre as amostras, e sobre o espectro resultante calculou-se um vetor (*vrf*) que também contém os mesmos elementos do vetor (*vrh*), acrescido do *feature* (periodograma).

4) Calculou-se a transformada DCT sobre as amostras da ROI, em seguida, calculou-se sobre os coeficientes AC um vetor unitário (*vtd*) contendo o *feature* (entropia).

5) Sobre a GLCM, extraiu-se um vetor de *features* (*vrg*), contendo os seguintes *features*: (correlação, contraste, diferença inversa, homogeneidade, entropia e uniformidade). Os cálculos foram realizados com o operador “um pixel à direita e um pixel abaixo”, ou seja, orientação  $\theta = 135^\circ$  com um pixel de distância, ( $d = 1$ ).

Ao final de todos os cálculos, formou-se um vetor *Q* contendo a união de todos os *features* calculados.  $Q = \{vrh \vee vrw \vee vrf \vee vrd \vee vrg\}$ . O resultado desta união é um vetor *Q* contendo 23 *features* descritos na Tabela 4.1.

Tabela 4.1: Elementos do vetor  $Q$ .

<b>Domínio</b>	<b>Features calculados</b>
Domínio espacial – histograma	entropia, variância, uniformidade, média, mediana
Domínio espacial - GLCM	correlação, contraste, d-inversa, homogeneidade, entropia e uniformidade
Domínio da frequência - DFT	entropia, variância, uniformidade, média, mediana, periodograma
Domínio da frequência - DWT	entropia, variância, uniformidade, média, mediana
Domínio da frequência - DCT	entropia

Mais importante que a quantidade de *features* obtida está a importância que cada uma dessas características contribui para o objetivo de classificar as imagens tomográficas. No entanto, o vetor  $Q$  pode conter redundância em que se refere à quantidade de informações carregadas por seus elementos. Tendo vista a escolha dos *features* mais representativos, fez-se uso da análise de PCA, a qual apresenta em ordem de importância os componentes com maior carga em seus autovetores.

A partir do cálculo de PCA, o primeiro componente principal é aquele de maior variância, que explique o máximo de variabilidade dos dados, o segundo componente de maior importância, o que apresenta a segunda maior variância e assim sucessivamente, até o componente principal de menor importância .

Neste trabalho, utilizou-se os componentes principais que explicam pelos menos 80% da variação total dos dados, como mostra a Tabela 4.2.

Tabela 4.2: Seleção de *Features* PCA com *Software Genes* - UFV.

<b>Autovalor</b>	<b>Importância</b>	<b>Destaque</b>	<b>Feature</b>
1	6,18	22,9 %	pc1
2	3,29	11,77 %	pc2
3	2,01	7,19%	pc3
4	1,89	6,76 %	pc4
5	1,77	6,34 %	pc5
6	1,59	5,71 %	pc6
7	1,33	4,78 %	pc7
8	1,26	4,53 %	pc8
9	1,17	4,20 %	pc9
10	1,01	3,64 %	pc10
11	0,95	3,41 %	pc11

O resultado da aplicação da análise de PCA ao conjunto  $Q$ , resultou em um conjunto reduzindo  $Q_r$ , o qual contém os elementos mostrados na Tabela 4.2.

Foi identificado pelo cálculo de PCA, que todos os elementos do conjunto  $Q_r$  podem ser importantes para a classificação das imagens, no entanto, resta saber qual ou quais são as melhores combinações destes elementos para serem utilizadas para o aprendizado dos classificadores. Sendo assim, combinaram-se os elementos de  $Q_r$  por meio de uma análise combinatória

simples da seguinte forma: considere que a variável  $n$ , representa a quantidade de elementos de cada novo subconjunto que se deseja adquirir, e  $q$  representa a quantidade total de elementos de  $Qr$ . A quantidade de combinações possíveis se faz pela Equação 4.1. Os valores de  $n$  variam de  $\{2, \dots, 10\}$ .  $n!$  e  $p!$  são números fatoriais ( números naturais inteiros positivos, o fatorial de um número é calculado pela multiplicação desse número por todos os seus antecessores até chegar ao número 1).

$$\chi_{(n,q)} = \frac{n!}{(n-q)!q}. \quad (4.1)$$

### 4.1.3 Base de Dados Utilizada para Treinamento dos Classificadores

Conforme cada combinação que se formou com os elementos de  $Qr$  formou-se uma base de dados  $D$ .

#### Divisão da Base de Dados

Conforme cada uma das combinações de  $Qr$ , separou-se um percentual de 25% da base de dados  $D$  para a formação de um conjunto  $P$ . O conjunto  $P$  foi utilizado para um pré-teste de classificação aplicado a todos os modelos de classificadores ( SVM, KNN, NBayes, AdaBoost).

Após a criação do conjunto  $P$ , fez-se a divisão do restante da base  $D$  em dois conjuntos, um conjunto de Treinamento ( $TR$ ) e outro conjunto de Testes ( $TS$ ).

Geralmente, quando se separa um conjunto de dados em um conjunto de treinamentos e um conjunto de testes, a maior parte dos dados é usada para treinamento e uma parte menor dos dados é usada para teste. Aqui os dois conjuntos foram divididos com os seguintes tamanhos: o conjunto  $TR$  com 64% da informação da base  $D$  e o conjunto  $TS$  com 36% restantes da base  $D$ .

### 4.1.4 Pré-teste de Classificação

Com o propósito de identificar qual o melhor subconjunto de  $Qr$  a ser utilizado para o aprendizado de máquinas e classificação, os classificadores SVM, KNN, Naive Bayes e AdaBoost foram submetidos a um pré-teste de classificação.

O pré-teste foi aplicado utilizando-se de conjuntos  $P$  formados conforme cada subconjunto de  $Qr$ .

### 4.1.5 Etapas do pré-teste de classificação

As Etapas do pré-teste forma efetuadas como se segue:

1 ) Aplicou-se o teste de classificação por meio da validação cruzada *leave-one-out*. Fez-se isso com todos os classificadores individuais - SVM, KNN, Naive Bayes e Adaboost. Mensurou-se para cada um dos classificadores a taxa de sensibilidade e especificidade aos testes de pré-classificação realizados;

2 ) Aplicou-se a Etapa anterior para todas as subamostras formadas pelas combinações dos elementos de  $Qr$ ;

3 ) Para cada modelo classificador SVM, KNN, Naive Bayes e Adaboost, selecionou-se o subconjunto de  $Qr$  considerando a maior sensibilidade e especificidade aos testes realizados.

Os resultados do pré-teste são apresentados na Tabela 4.3 .

Tabela 4.3: Resultado dos classificadores com o pré-teste

Classificador	sensibilidade	especificidade	acurácia
SVM - (Linear)	100	0,14	0,57
SVM - (RBF)	0,14	0,85	0,49
KNN	0,71	0,71	0,71
N-Bayes	0,14	0,85	0,49
AdaBoost	0,85	0,71	0,78

A construção de um classificador que obtenha um bom desempenho exige uma base de dados significativa, de preferência que apresente todas as instâncias,mas, isso nem sempre é realidade. Ainda que trabalhando com poucos dados, esta Etapa buscou analisar o comportamento dos modelos classificadores diante de uma pequena parcela dos dados, as métricas de sensibilidade, especificidade e acurácia foram medidas, e conforme expressos na Tabela 4.3.

A ideia é que os resultados do pré-teste sejam generalizados para testes maiores com um maior quantitativo de amostras de imagens. É uma forma de verificar o desempenho dos classificadores individuais, para que posteriormente estes sejam utilizados na formação de um *ensemble*.

## 4.2 Ensemble

Buscando melhorias na classificação das imagens, as saídas dos classificadores individuais foram combinadas formando um *ensemble*.

Antes da elaboração do *ensemble*, os classificadores (SVM - KNN , Naive Bayes e AdaBoost) receberam os parâmetros obtidos na Seção 4.1.5 e foram submetidos a testes individuais para análise de desempenho. A ideia é que os classificadores se comportem e apresentem resultados de forma semelhante aos resultados do pré-teste.

### 4.2.1 Classificação Individual

A classificação individual se fez aplicando o conjunto TR para o treinamento dos modelos de classificadores, em seguida, aplicou-se o conjunto de dados TS como entrada para testes de classificação. Fez-se isso para todos os modelos de classificadores - (SVM, KNN, Naive Bayes e Adaboost).

#### Classificação Individual Com uso da Validação Cruzada *leave-one-out*

A Tabela 4.4 mostra os resultados individuais de classificação utilizando a validação cruzada *leave-one-out* tendo como entrada a união dos conjuntos TR e TS.

#### Classificação Individual Com uso da Validação Cruzada *K-Fold*

Os resultados desta classificação são apresentados na Tabela 4.5, consistindo da média e do desvio padrão de 10 testes realizados por meio da validação cruzada *k-fold*.

Tabela 4.4: Resultado dos classificadores individuais, *leave-one-out*.

Classificador	sensibilidade	especificidade	acurácia
SVM Linear	0,95	0,15	0,55
SVM RBF	0,15	0,90	0,52
KNN	0,70	0,55	0,62
N-Bayes	0,20	0,90	0,55
AdaBoost	0,60	0,65	0,62

Tabela 4.5: Resultado dos classificadores individuais.

Classificador	sensibilidade	especificidade	acurácia
SVM Linear	0,90	0,20	0,55
Desvio padrão	0,05	0,05	0,03
SVM RBF	0,20	0,91	0,55
Desvio padrão	0,06	0,06	0,05
KNN	0,67	0,67	0,67
Desvio padrão	0,05	0,05	0,03
N-Bayes	0,25	0,84	0,54
Desvio padrão	0,18	0,13	0,07
AdaBoost	0,74	0,67	0,68
Desvio padrão	0,15	0,18	0,5

Os resultados dos classificadores apresentaram desempenhos semelhantes aos resultados do pré-teste, ou seja, apresentaram boa generalização para os novos dados.

## 4.2.2 Combinação de Classificadores

O teste classificatório ideal, com 100% de sensibilidade e especificidade raramente existe na prática, pois, geralmente o aumento da sensibilidade de um teste classificador, pode acarretar em diminuir a sua especificidade ou vice versa. Esse desbalanceamento pode acontecer considerando apenas quando se trabalha com um único classificador. Portanto, contando que é aceitável a combinação de classificadores, é possível criar vertentes de aumento da sensibilidade e vertentes de aumento da especificidade, como exemplificado com as classes fictícias A e B (não doentes e doentes) a seguir:

$$\begin{array}{l}
 \text{Classe A} \\
 \text{Elementos} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\} \quad \overbrace{\{0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\}}^{\text{Rótulos}} \\
 \\
 \text{Classe B} \\
 \text{Elementos} = \{z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10}, z_{11}\} \quad \{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\}
 \end{array}$$

Figura 4.4: Exemplo das classes A e B.



Seja (A) a classe de não doentes contendo 11 elementos, e todos esses elementos rotulados por 0.

Seja (B) a classe de doentes contendo 11 elementos, e todos esses elementos rotulados por 1. As classes A e B são exemplificadas pela Figura 4.4.

Suponha que um classificador  $h1$  tenha classificado as classes (A e B), e seus resultados sejam os valores apresentados na Figura 4.5.

	Classe - A	Classe - B	Sensibilidade	Especificidade	Acurácia
Resultado $h1 =$	0 1 1 0 1 1 1 0 1 0 1	1 1 1 1 1 0 1 1 1 1 1	0,90	0,36	0,63

Figura 4.5: Resultado do classificador  $h1$

A Figura 4.5 mostra que o classificador modelo  $h1$  classificou corretamente 4 elementos da classe A e 10 elementos da classe B, apresentando alta sensibilidade.

Suponha que um classificador  $h2$  também tenha classificado as classes (A e B), e seus resultados sejam os valores apresentados na Figura 4.6.

	Classe - A	Classe - B	Sensibilidade	Especificidade	Acurácia
Resultado $h2 =$	1 0 0 1 0 0 0 1 0 1 0	0 1 1 1 1 1 1 1 1 1 1	0,90	0,63	0,77

Figura 4.6: Resultados de classificação de  $h2$ .

A Figura 4.6 mostra que o classificador  $h2$  classificou corretamente 7 elementos do conjunto A e 10 elementos do conjunto B, apresentando alta sensibilidade e valor mediano de acurácia.

Analisando os dois classificadores  $h1$  e  $h2$ , viu-se que o classificador  $h1$  apresentou sensibilidade de 90%, em contrapartida, apresentou uma especificidade de 36% e acurácia mediana, já o classificador  $h2$  obteve 90% de sensibilidade e valor mediano quanto às métricas especificidade e acurácia.

A preferência entre os dois classificadores apresentados se dá ao  $h2$  que apresentou melhor acurácia. No entanto, com a combinação destes dois classificadores, é possível usar o bom desempenho de  $h1$  quanto a sensibilidade, e criar uma vertente para melhorar a especificidade do classificador mediano  $h2$  e consequentemente melhorar a sua acurácia. Este procedimento é descrito a seguir.

### Vertente para Aumento da Especificidade

De acordo com o contexto clínico, quanto mais sensível um teste diagnóstico, maior é seu valor preditivo negativo (maior a segurança de que a pessoa com teste negativo não tem a doença), isso permite aceitar que os resultados negativos (N) obtidos com o classificador  $h1$  sejam confiáveis, logo, é possível criar uma vertente de aumento da especificidade de um teste da seguinte forma:

**Se  $h_2$  classificar como Positivo - P, aceite,**  
**Se  $h_1 \vee h_2$  classificar como Negativo - N, aceite.**

Esta regra é aplicada coluna por coluna das tabelas de resultados como mostra a Figura 4.7

	Classe - A										Classe - B										Sensibilidade	Especificidade	Acurácia	
h1	0	1	1	0	1	1	1	0	1	0	1	1	1	1	1	1	0	1	1	1	1	0,90	0,63	0,77
h2	1	0	0	1	0	0	0	1	0	1	0	0	1	1	1	1	1	1	1	1	1	0,81	0,90	0,85
<b>Resultado =</b>	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	0	1	1	1	1				

Figura 4.7: Exemplo de aumento da especificidade.

A Figura 4.7 mostra o classificador  $h_1$  corrigindo erros de classificação de  $h_2$  e aumentando de 0,63 para 0,90 a especificidade no resultado final da combinação de  $h_1$  e  $h_2$ . Com o aumento da especificidade, aumenta-se também a acurácia final.

### Vertente para Aumento da Sensibilidade

Suponha que outros classificadores ( $h_3$  e  $h_4$ ) tenham classificado as classes (A e B) e  $h_3$  tenha apresentado alta sensibilidade e valor mediano de especificidade e  $h_4$  tenha apresentado alta especificidade e baixa sensibilidade. Portanto, da mesma forma que se fez combinando os dois classificadores ( $h_1$  e  $h_2$ ), para aumento da especificidade, pode se fazer também com os classificadores ( $h_3$  e  $h_4$ ) para aumento da sensibilidade. Para isso, primeiramente, deve-se considerar que no contexto clínico, quanto mais específico for um teste diagnóstico, maior é seu valor preditivo positivo (maior a segurança de que a pessoa com teste positivo tem a doença), isso permite aceitar que os resultados Positivos (P) obtidos com o classificador  $h_4$  de alta especificidade sejam confiáveis, logo, é possível criar uma vertente de aumento da sensibilidade da seguinte forma:

**Se  $h_3$  classificar como Negativo - N, aceite,**  
**Se  $h_3 \vee h_4$  classificar como Positivo - P, aceite.**

Considerando que as saídas dos classificadores sejam não correlacionadas, ao final da combinação dos classificadores - ( $h_1, h_2$ ) e ( $h_3, h_4$ ), tem-se duas vertentes, uma com a sensibilidade aumentada e a outra com especificidade aumentada.

### 4.2.3 Saída do Ensemble

Para a saída do ensemble, fez-se uso de uma votação ponderada, aplicando pesos aos classificadores de maior sensibilidade e especificidade constatados durante a fase de pré-teste. Utilizou também vertentes de aumento da sensibilidade e especificidade como exemplificado a Seção 4.2.2.

O Ensemble se fez como mostra a Figura 4.8.

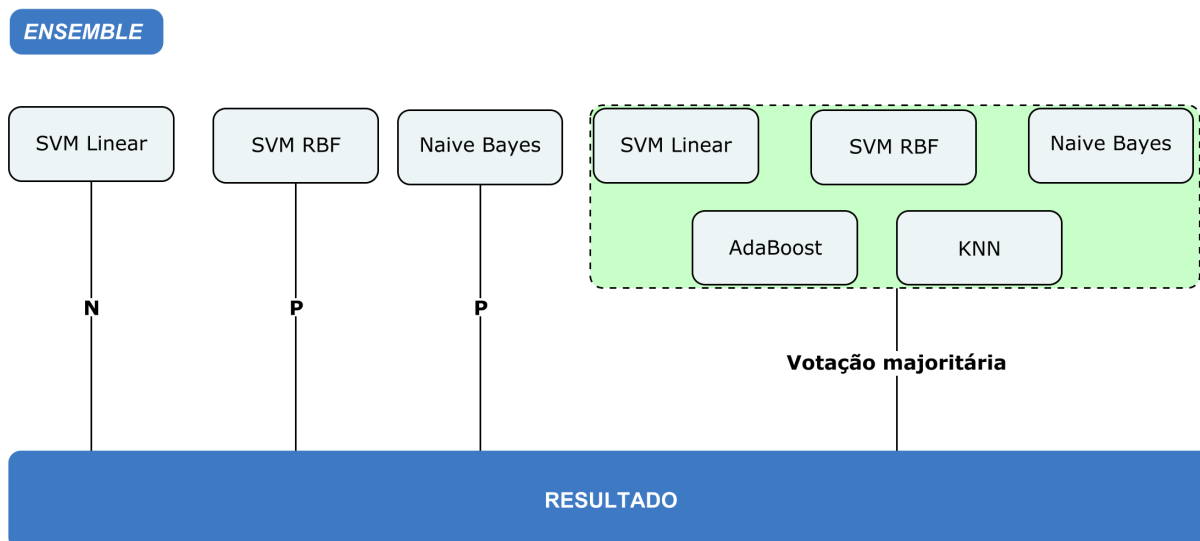


Figura 4.8: Exemplo de um *ensemble*.

### Combinação dos Classificadores que Compõe o *Ensemble*

1) - Devido a alta especificidade em que os classificadores SVM- RBF e Naive Bayes apresentaram durante a fase de pré-teste e, sabendo-se que quanto mais específico for um teste diagnóstico maior é seu valor preditivo positivo, classifica-se em Positivo P, se o classificador SVM- RBF ou o classificador Naive Bayes classificar como P.

2) - Devido a alta sensibilidade em que o classificador SVM -Linear apresentou durante a fase de pré-teste e sabendo que quanto mais sensível for um teste diagnóstico maior é seu valor preditivo negativo, classifica-se em Negativo N, se o classificador SVM - Linear classificar como N.

3) - Em casos contrários às condições 1 e 2, o resultado do *ensemble* se faz pela votação majoritária dos classificadores (SVM Linear, SVM RBF, KNN, Naive Bayes e AdaBoost).

# Capítulo 5

## Resultados Alcançados

Apesar de geralmente os sistemas de combinação de classificadores apresentarem melhores resultados em comparação aos classificadores que o compõe, não há garantias que isto ocorra sempre. Aqui é notório que o *ensemble* possibilitou um aumento considerável da acurácia final da classificação quando se compara com os resultados das saídas individuais dos classificadores. Isso se deu devido à forma como foram combinados os classificadores e também por estes não terem os resultados de suas saídas correlacionadas. As Seções seguintes apresentam os resultados alcançados com a aplicação do *ensemble*.

### 5.0.1 Resultados Utilizando Validação Cruzada *K-Fold*

A Figura 5.1 apresenta uma matriz de confusão com a média dos resultados de 10 testes do *ensemble* obtidos por meio da validação cruzada *k-fold*.

		Classe esperada	
		VP	VN
Classe prevista	VP	95%	3,7%
	VN	5%	96,3%

Figura 5.1: Matriz de confusão com validação *K-Fold*

A Tabela 5.1 mostra pelo valor de desvio padrão que os resultados foram razoavelmente uniformes aos testes de validação cruzada.

As métricas de sensibilidade, especificidade e acurácia foram calculadas para os testes.

Sensibilidade 0,96%; significa que se o resultado do classificador for positivo, há 96% de chances de o indivíduo ser acromegálico e 4% (100-96%) de chances de ser não acromegálico.

Especificidade 0,95%; significa que se o resultado do classificador for negativo, há 95% de chances de o indivíduo ser não acromegálico e 5% (100-95%) de chances de ser acromegálico.

Tabela 5.1: Desvio padrão dos testes de classificação realizados

	<b>sensibilidade</b>	<b>especificidade</b>	<b>acurácia</b>
Resultado	95%	96%	95%
Desvio padrão	0.05	0.07	0,04

### 5.0.2 Resultados Utilizando Validação Cruzada *leave-one-out*

A Figura 5.2 apresenta uma matriz de confusão dos resultados de testes de classificação utilizando a validação *leave-one-out*.

		Classe esperada	
		VP	VN
Classe prevista	VP	95%	10%
	VN	5%	90%

Figura 5.2: Matriz de confusão com validação *leave-one-out*

A Tabela 5.2 mostram as métricas calculadas para o teste com validação *leave-one-out*..

Tabela 5.2: Resultado do *ensemble* utilizando *leave-one-out*

	<b>sensibilidade</b>	<b>especificidade</b>	<b>acurácia</b>
Resultado	95%	90%	92%

Sensibilidade 0,90%; significa que se o resultado do classificador for positivo, há 90% de chances de o indivíduo ser acromegálico e 10% (100-90%) de chances de ser não acromegálico.

Especificidade 0,94%; significa que se o resultado do classificador for negativo, há 94% de chances de o indivíduo ser não acromegálico e 6% (100-94%) de chances de ser acromegálico.

### 5.0.3 Resultados Utilizando Aumento de Dados e Validação Cruzada *K - Fold*

Os resultados até o momento apresentados foram efetuados sobre um grupo pequeno de imagens.

Os resultados seguintes são testes do comportamento do *ensemble* classificador após a implementação do aumento de dados - *data augmentation*, o qual foi implementado por meio do incremento de pequena quantidade de ruídos gaussiano às amostras e também por variedades de corte manuais da região interesse.

Os dados foram aumentados como ferramenta para tratamento de sobreajuste. Com o aumento foi possível chegar vertiginosamente a 390 amostras de pacientes acromegálicos e 390 imagens de não acromegálicos.

Para o aprendizado de máquinas, utilizou-se 70% da informação para treinamento dos classificadores e 30% para testes.

A Figura 5.3 apresenta uma matriz de confusão da classificação utilizando 30% da informação como testes. A classificação se fez por meio de 10 testes de validação cruzada *k-fold* utilizando o *ensemble* classificador.

		Classe esperada	
		VP	VN
Classe prevista	VP	94,2%	1,7%
	VN	5,8%	98,3%

Figura 5.3: Matriz de confusão dados aumentados

Tabela 5.3: Desvio padrão dos testes de classificação realizados com aumento de dados

	<b>sensibilidade</b>	<b>especificidade</b>	<b>acurácia</b>
Resultado	94%	98%	96%
Desvio padrão	0.052	0.005	0.024

Métricas calculadas para testes com aumento de dados.

Sensibilidade 0,98%; significa que se o resultado do exame for positivo, há 98% de chances de o indivíduo ser acromegálico e 2% (100-98%) de chances de ser não acromegálico.

Especificidade 0,94%; significa que se o resultado do exame for negativo, há 94% de chances de o indivíduo ser não acromegálico e 6% (100-94%) de chances de ser acromegálico.

### Considerações Finais

A classificação final com o aumento de dados apresentou desempenho próximo aos resultados da classificação com uso dos dados reais.

O uso do *ensemble* reduziu o risco de escolha por um classificador com desempenho fraco. Outra abordagem que possibilitou a classificação das imagens acromegálicas foi o uso de diferentes parâmetros de treinamento para diferentes classificadores. Em resumo, todo o procedimento da classificação se deu pela escolha dos classificadores, a escolha dos descritores de características utilizados e consideravelmente a escolha do método de combinação dos classificadores na formação do *ensemble*.

# Capítulo 6

## Conclusão

A presente dissertação teve o objetivo de classificar imagens tomográficas por meio do reconhecimento de padrões e aprendizado máquinas. A doença classificada foi acromegalia, uma doença rara, e por este motivo dificultou os estudos, uma vez que uma grande quantidade de imagens radiológicas é difícil de ser encontrada.

Ainda que a quantidade de imagens para o estudo em questão seja pequena para se tratar de aprendizado de máquinas, este trabalho apresentou dados relevantes quanto à possibilidade de classificação da acromegalia por meio de reconhecimento de padrões em imagens tomográficas.

Foi realizada uma revisão atualizada, abordando diversos conceitos ao estudo em questão, porém, existe uma escassez de trabalhos científicos correlatos ao estudo proposto.

Quanto aos resultados obtidos, estes alcançaram os objetivos desejados, ou seja, o estudo mostrou que é possível com o auxílio do aprendizado de máquinas, classificar imagens de TC de vértebras (L1- L5) conforme a presença ou ausência da doença acromegalia. Ainda mais importante que os resultados conseguidos e apresentados, fica a expectativa de novas pesquisas relacionadas ao tema, aprimorando técnicas de reconhecimento de padrões acromegálicos e ao mesmo tempo possibilitando o crescimento de um banco de imagens tomográficas em questão.

Esta pesquisa abre também um leque para que outros estudos sejam feitos, considerando-se outras doenças ou até mesmo o uso de imagens de outros exames radiológicos para detectar padrões em doenças ósseas como, por exemplo, a osteoporose.

Em função da indisponibilidade de imagens e do tempo de conclusão desta dissertação, recomenda-se para trabalhos futuros, que se façam testes da metodologia desenvolvida em um conjunto maior de imagens tomográficas.

Por fim, sugere-se também a possibilidade de testes de outras formas de combinação dos classificadores utilizados na construção do *ensemble* para possíveis melhorias no sistema de classificação aqui desenvolvido.

# Referências

- [1] Philippe Chanson. Acromegaly. *Orphanet Journal of Rare Diseases*, 2008. 1
- [2] American Brain Tumour Association. Pituitary tumours. *American Brain Tumour Association*, 2013. 1
- [3] Shlomo Melmed. Acromegaly pathogenesis and treatment. *The Journal of Clinical Investigation*, 2009. 1
- [4] Miguel Madeira, Leonardo Vieira Neto, Francisco de Paula Paranhos Neto, Inayá Corrêa Barbosa Lima, Laura Maria Carvalho de Mendonça, Mônica Roberto Gadelha, and Maria Lucia Fleiuss de Farias. Acromegaly has a negative influence on trabecular bone, but not on cortical bone, as assessed by high-resolution peripheral quantitative computed tomography. *The Journal of Clinical Endocrinology Metabolism*, 2013. 1, 3, 4
- [5] B. Frisch R. Bartl. Osteoporosis: Diagnosis prevention therapy - second ed. *Springer*, 2009. 1
- [6] Giuseppina Padova, Graziella Borzì, Laura Incorvaia, Guido Siciliano, Valentina Migliorino, Mario Vetri, and Patrizia Tita. Prevalence of osteoporosis and vertebral fractures in acromegalic patients. *Clinical cases in mineral and bone metabolism : the official journal of the Italian Society of Osteoporosis, Mineral Metabolism, and Skeletal Diseases*, 2011. 1
- [7] Vittorio Locatelli and E. Bianchi. Effect of gh/igf-1 on bone metabolism and osteoporosis. *Springer Berlin Heidelberg, International Journal of Endocrinology*, 2014. 3
- [8] Kim SW Kim SY Shin CS Hong AR, Kim JH. Trabecular bone score as a skeletal fragility index in acromegaly patients. *Springer Berlin Heidelberg, International Journal of Endocrinology*, 2015. 3
- [9] Lenka Sterančáková Zdenko Killinger, Martin Kužma and Juraj Payer. Osteoarticular changes in acromegaly. *Springer Berlin Heidelberg, International Journal of Endocrinology*, 2012. 3
- [10] Wojciech Zgliczyński, Magdalena Kochman, Waldemar Misiorowski, and Piotr Zdunowski. In acromegaly, increased bone mineral density (bmd) is determined by gh-excess, gonadal function and gender. *Springer Berlin Heidelberg, Neuro endocrinology letters*, 2007. 4



- [11] Jowita Halupczok Marek Bolanowski and Aleksandra Jawiarczyk-Przybyłowska. Pituitary disorders and osteoporosis. *Springer Berlin Heidelberg, International Journal of Endocrinology*, 2015. 4
- [12] I. Slim, H. Akkari, A. Ben Abdallah, I. Bhourri, and M. Hedi Bedoui. Trabecular bone radiograph characterization using fractal, multifractal analysis and svm classifier. *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical year = 2018, journal = World Academy of Science, Engineering and Technology*,. 4
- [13] Xiangyi et al Kong. Automatic detection of acromegaly from facial photographs using machine learning methods. *Springer Berlin Heidelberg*, 2017. 4
- [14] Günther M Roemmler J Stalla GK Sievers C Reincke M Schopohl J Würtz RP Schneider HJ1, Kosilek RP. A novel approach to the detection of acromegaly: accuracy of diagnosis by automatic face classification. *J Clin Endocrinol Metab*, 2011. 5
- [15] Sutton David. Radiology and imaging for medical students. 7th ed. *Churchill Livingstone, Inc*, 1999. 6
- [16] Alexandre Araújo dos Santos. *Tomografia Computadorizada Princípios Físicos e Aplicações*. 2007. 6, 7
- [17] K.L. Bontrager and J.P. Lampignano. *Tratado de Posicionamento Radiográfico e Anatomia Associada*. 2015. 6
- [18] Perry J Horwich. Wavelet methods for time series analysis. *Cambridge University*, 2000. x, xi, 6, 7
- [19] S. White and M.J Pharoah. *Radiologia Oral: Princípios e Interpretação*. 2015. 6
- [20] Jamil et al Natour. *Coluna Vertebral, conhecimentos básicos*. 2004. x, 8, 9
- [21] R.C. Gonzalez and R.E Woods. Digital image processing. *Pearson*, 2010. x, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 22
- [22] R. Haralick, K. Shanmugam, and I Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973. 18
- [23] P.S.R. Diniz, E.A.B. da Silva, and S.L Netto. *Processamento Digital de Sinais - 2.ed.: Projeto e Análise de Sistemas*. 2013. 19, 20, 21
- [24] Wellington da Silva Lyra, Edvan Cirino da Silva, Mario Cesar Ugulino de, Wallace Duarte Fragoso, and Germano Veras. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. *scielo*, 2010. 23, 31
- [25] H Hotteling. Analysis of a complex statistical variables into principal components. 1933. 23
- [26] S.J. Russell, S.J. Russell, P. Norvig, and E. Davis. *Artificial Intelligence: A Modern Approach*. 2010. 23

- [27] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 1995. 24, 31
- [28] Sanjeev R. Kulkarni and Gilbert Harman. Statistical learning theory: A tutorial. *Computational Statistics*, 2011. 25
- [29] R.O. Duda and P.E Hart. Pattern classification and scene analysis. *Wiley*, 1973. 25, 31
- [30] S Haykin. Redes neurais: Princípios e prática. *Armed*, 2007. 25, 29
- [31] Oliver Kramer. Dimensionality reduction with unsupervised nearest neighbors. *Intelligent Systems Reference Library*, 2013. 25
- [32] T.M Mitchell. Machine learning. *McGraw-Hill Science/Engineering/Mathy*, 1997. 26
- [33] T Dietterich. Machine learning research: Four current directions. *The AI Magazine*, 1998. 27
- [34] L.I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. 2014. 27
- [35] S.W. Lee and A. Verri. *Support Vector Machine Ensemble with Bagging*. 2003. 27
- [36] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Springer Berlin Heidelberg*, 1995. 27, 28, 31
- [37] Akobeng Anthony K. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Springer Berlin Heidelberg, Acta Paediatrica*, 2006. 28
- [38] S. Theodoridis and K Koutroumbas. Pattern recognition. *Springer Berlin Heidelberg, Elsevier Science*, 2008. 29
- [39] Hiroshi Inoue. Data augmentation by pairing samples for images classification. *Springer Berlin Heidelberg, Images Classification*, 2018. 30
- [40] P.S Addison. The illustrated wavelet transform handbook: Introductory theory and applications in science, engineering, medicine and finance. *CRC Press*, 2017. 31
- [41] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, 1974. 31
- [42] K. Fukunage and P. M Narendra. A branch and bound algorithm for computing k-nearest neighbors. *IEEE Trans. Comput*, 1975. 31
- [43] N Otsu. A threshold selection method from graylevel histograms. *automatica*. 1975. 32
- [44] R.C. Gonzalez and R.E Woods. Digital image processing. *Pearson*, 2000. 33