

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

Luana Aparecida dos Reis

**Imagens por microscopia óptica não linear para análises de biópsias de
câncer canino**

BELO HORIZONTE
2021

Luana Aparecida dos Reis

**Imagens por microscopia óptica não linear para análises de biópsias de
câncer canino**

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutor em Ciências.

Orientadora: Ana Maria de Paula

Belo Horizonte
2021

Dados Internacionais de Catalogação na Publicação (CIP)

R375i Reis, Luana Aparecida dos.
Imagens por microscopia óptica não linear para análise de biópsias de câncer
canino / Luana Aparecida dos Santos. – 2021.
96f., enc. : il.

Orientadora: Ana Maria de Paula.
Tese (doutorado) – Universidade Federal de Minas Gerais,
Departamento de Física.
Bibliografia: f. 86-96.

1. Óptica não-linear. 2. Biofísica. 3. Análise de imagens. 4. Estatística.
I. Título. II. Paula, Ana Maria de. III. Universidade Federal de Minas Gerais,
Departamento de Física.

CDU – 535 (043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

FOLHA DE APROVAÇÃO

A presente tese, intitulada "**Imagens por microscopia óptica não linear para análises de biópsias de câncer canino**" de autoria de **LUANA APARECIDA DOS REIS** submetida à Comissão Examinadora, abaixo-assinada, foi aprovada para obtenção do grau de **DOUTORA EM CIÊNCIAS**, em vinte e três de setembro de 2021.

Belo Horizonte, 23 de setembro de 2021.

Profa. Ana Maria de Paula

Orientadora da estudante

Departamento de Física/UFMG

Profa. Denise Maria Zezell

Centro de Lasers e Aplicações/IPEN

Prof. Leandro Malard Moreira

Departamento de Física/UFMG

Profa. Jaqueline dos Santos Soares

Departamento de Física/UFOP

Prof. Sebastião José Nascimento de Pádua

Departamento de Física/UFMG



Documento assinado eletronicamente por **Ana Maria de Paula, Professora do Magistério Superior**, em 24/09/2021, às 11:34, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Jaqueline dos Santos Soares, Usuário Externo**, em 24/09/2021, às 13:51, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Denise Maria Zezell, Usuário Externo**, em 24/09/2021, às 15:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leandro Malard Moreira, Professor do Magistério Superior**, em 27/09/2021, às 08:15, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sebastião Jose Nascimento de Padua, Membro de comissão**, em 27/09/2021, às 14:19, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0983516** e o código CRC **6A88BF10**.

Agradecimentos

Chegamos enfim aos agradecimentos da tese! Hora de agradecer a todos os que contribuíram com essa jornada, direta ou indiretamente. Provavelmente tudo o que eu falar aqui não vai ser nem perto do que essas pessoas fizeram por mim, aprendi nesses anos que a linguagem as vezes é pobre para expressar o que sentimos, e que eu provavelmente vou esquecer alguém (a memória já não é mais a mesma, rrsrrs)

Agradeço primeiramente a Deus e a Santíssima Virgem pelo sustento e por guiar meus passos até aqui.

A minha mãe Maria das Graças, pelo amor incondicional, pela paciência e por acreditar em mim quando eu não acreditei. Ao meu irmão Hugo que sempre foi um grande companheiro e incentivador. Ao meu pai e aos meus avós, que mesmo não estando mais presentes fisicamente me ensinaram muito.

Agradeço aos amigos que a física me deu ao longo de todos esses anos. Ao Zé, pela amizade de todas as horas e que é o perseverante que escuta minhas reclamações, choros e desesperos desde o início da graduação, e a Mona por sempre dar corda as minhas ideias giricas desde que viemos para BH (e por incentivar as mudanças no visual). Ao Léo, que me ajudou em muitas coisas, que teve paciência em sentar e estudar comigo, de conversar quando eu surtava e batia na porta da sala com a desculpa do café, agradeço também pelas conversas sobre livros, músicas e tantas coisas aleatórias. A Raphaela que foi um grande presente que a UFMG me deu, obrigada pelas conversas, pelas risadas, pelo carinho e pelos rolês. Agradeço também ao Gustavo, Filó, Cíntia, Wilder, Ludmila, Ronaldo, Beico e Renan que com toda a certeza tornaram os dias mais agradáveis com suas companhias. Agradeço a Sandy e Hithiara pelos anos de amizade e por sempre estarem comigo, mesmo quando as vezes eu estava um pouco ausente, obrigada por me entenderem meninas. A Marília e ao Iury que estão comigo há uma vida inteira. Agradeço ao MUR-BH por me acolher durante esses anos, em especial ao Gou Santa Edith Stein.

Agradeço também à Christie, Shelley, King, Gaiman, Lispector, Aleksiévich e tantos outros que me acompanharam ao longo dessa jornada.

Agradeço aos colegas de laboratório, em especial ao Egleidson, que foi uma “mão na roda” com algumas medidas e a Ana Paula pela paciência em me explicar 1037 vezes as mesma coisas sobre biologia. Agradeço de forma especial a minha orientadora Ana Maria, por toda a paciência quando eu queria surtar com o doutorado e pelo carinho nos momentos complicados da vida. A todos os professores e funcionários do Departamento de Física

UFMG, ao Prof. Geovanni e Ana Paula do Laboratório de Patologia Comparada (ICB) pelas amostras, ao Leonardo Ferreira Calazans pelos scripts utilizados para os boxplots e análises estatísticas e ao Frank e ao Frank Longford da Universidade de Southampton pela colaboração com o desenvolvimento do Software. À CAPES, Fapemig e CNPq pelo apoio financeiro.

“O ideal seria que o homem preservasse sempre uma mente calma e tranquila, e jamais permitisse que uma paixão ou um desejo transitório lhe perturbassem a paz. Não acho que a busca do conhecimento seja uma exceção a essa regra. Se o estudo ao qual nos dedicamos tem a tendência de nos enfraquecer as emoções e destruir nosso gosto pelos prazeres simples que nada pode corromper, então esse estudo é certamente inadequado à mente humana.”

Frankenstein (Mary Shelley, 1818)

Resumo

Segundo dados do Instituto Nacional do Câncer (INCA) o câncer de mama é um tipo de câncer muito comum entre as mulheres, sendo responsável por cerca de 29% dos casos a cada ano. A forma padrão do diagnóstico é feita por meio da análise visual do tecido coletado, que passa por processos de corte e coloração que podem danificar a amostra, tornando-o muitas vezes demorado e dependente da experiência do patologista. Assim, o desenvolvimento de técnicas que possibilitem uma análise quantitativa dos tecidos de forma a obter novas ferramentas de diagnóstico com uma maior agilidade na análise dos dados seria de grande ajuda aos patologistas e benéfico para os pacientes. Neste trabalho, utilizamos as técnicas de microscopia óptica não linear de Geração de Segundo Harmônico (SHG) e de Fluorescência por Excitação de Dois Fótons (TPEF) com laser pulsado para obter imagens em biópsias de neoplasias mamárias em cadela. As imagens de SHG permitem informações da estrutura do colágeno na matriz extracelular que, juntamente com as imagens de TPEF das regiões celulares das biópsias, formam uma base para uma análise de imagem abrangente. Para as análises dessas imagens, um software, que utiliza ferramentas já conhecidas em análises desse tipo de imagem, como o tensor estrutural e a extração de redes de fibras (FIRE), foi desenvolvido a fim de automatizar o processo de análise. O software permitiu obter a separação de regiões fibrosas e celulares das imagens SHG e TPEF. Com essa segmentação foi possível a extração de redes de fibras de colágeno e vários parâmetros específicos para cada região (fibras ou células) formando um conjunto de 29 métricas. Com esse conjunto de métricas conseguimos estudar as alterações causadas nos tecidos devido aos diferentes tipos de neoplasias mamárias, como os tumores mistos benignos, carcinoma em tumor misto, carcinoma ductal *In Situ*, carcinossarcoma, carcinoma micropapilar e carcinoma sólido em comparação com a mama normal, além da análise das alterações sofridas durante a progressão tumoral. Essas alterações foram analisadas por meio de métodos estatísticos análise de variância (MANOVA e ANOVA), para discriminar quais foram as significâncias estatísticas. Além disso, uma análise por discriminante linear (LDA) incluindo todas as métricas extraídas permitiu separar e classificar todos os tipos histológicos estudados. Obtivemos uma sensibilidade de cerca de 90% na classificação entre tecido saudável e canceroso e uma sensibilidade de cerca de 40-60% para comparação entre a mama normal e os subtipos histológicos estudados.

Palavras-chave: Óptica não linear, Geração de segundo harmônico, Câncer de mama, Análise de imagens, Fibras de colágeno, LDA.

Abstract

According to data from the Institute of Cancer (INCA), breast cancer is a very common type of cancer among women, accounting for about 29% of cases each year. The standard diagnosis is given by visual analysis of the collected tissue, which undergoes cutting and coloring processes that can damage the sample, often making it slowly and dependent on the pathologist's experience. Thus, the development of techniques that enable a quantitative tissue analysis in order to obtain new diagnostic tools with greater agility in data analysis would be of great help to pathologists and beneficial to patients. In this work, we used the non-linear optical microscopy techniques of Second Harmonic Generation (SHG) and Two-Photon Excitation Fluorescence (TPEF) with pulsed laser to obtain images in biopsies of mammary neoplasms in female dogs. SHG images provide information of the collagen structure in the extracellular matrix which, that together with TPEF images of cell regions in the biopsies form a basis for comprehensive imaging analysis. For the image analysis, we developed a software, which uses tools already known for this type of image analysis, such as the structural tensor and the extraction of fiber networks (FIRE). The software was developed in order to automate the image analysis process. It allowed to obtain the separation of the fibrous and cellular regions in the SHG and TPEF images. With this segmentation it was possible to extract the collagen fiber networks and several specific parameters for each region (fibers or cells) forming a set of 29 metrics. With this set of metrics we were able to study the changes in the tissues caused by different types of breast cancer, such as benign mixed tumors, carcinoma in mixed tumor, ductal carcinoma *In Situ*, carcinosarcoma, micropapillar carcinoma and solid carcinoma in comparison to normal breast. The changes due to the tumor progression were analyzed using the statistical methods of variance (MANOVA and ANOVA). In addition, a linear discriminant analysis (LDA) including all extracted metrics allowed separating and classifying all histological types studied. We obtained a sensitivity of about 90% in the classification between healthy and cancerous tissue and a sensitivity of about 40-60% for comparison between normal breast and all the studied histological subtypes.

Keywords: Non-linear optics, Second-harmonic generation, Breast cancer, Image analysis, Collagen fibers, LDA.

Sumário

Resumo

Abstract

1	Introdução	12
2	Revisão Teórica	15
2.1	Microscopia óptica não-linear	15
2.1.1	Geração de Segundo Harmônico	15
2.1.2	Fluorescência por Excitação de dois fótons	18
2.2	Mama	19
2.2.1	Patologias da Mama	21
2.2.2	Tipos histológicos	22
2.3	Colágeno	24
2.3.1	SHG no colágeno	25
3	Metodologia Experimental	27
3.1	Amostras	27
3.2	Microscopia por geração de segundo harmônico - SHG	28
3.3	Microscopia de Fluorescência por Excitação de dois fótons - TPEF	30
3.4	Análises das Imagens	32
3.4.1	Extração da Rede de Fibras	33
3.4.2	Segmentação de Fibras e Células	35
3.4.3	Tensor Estrutural	37
3.4.4	Parâmetros Extraídos	38
3.5	Análise de Variância	39
3.6	LDA	39
4	Resultados e Discussão	43
4.1	Análise de Subtipos Histológicos - <i>In Situ</i> , Misto e Sólido	43
4.1.1	Imagem Óptica, SHG e TPEF	44
4.1.2	Imagens geradas pelo PyFibre	45
4.1.3	Resultado das Análises Estatísticas	48

4.1.4	LDA	56
4.2	Análises de Subtipos Histológicos e Progressão Tumoral	58
4.2.1	Subtipos Histológicos	58
4.2.2	Progressão tumoral	64
4.2.3	LDA	68
5	Conclusão e Perspectivas	71
A	PyFibre-estrutura do programa	73
B	Métricas PyFibre	76
B.1	Rede de Fibras	76
B.2	Segmento de Fibras	77
B.3	Segmento de Células	77
C	Gráficos de caixa: Métricas extraídas	79
C.1	Métricas: Tipos histológicos <i>In Situ</i> , Misto e Sólido	79
C.2	Métricas: Tipos histológicos	81
C.3	Métricas: Progressão Tumoral	83
D	Publicações	85
	Referências Bibliográficas	86

Capítulo 1

Introdução

Neoplasias são caracterizadas por um crescimento de células anormais, esse crescimento pode ser lento e organizado ou rápido e desordenado capaz de migrar através do tecido conjuntivo para tecidos distantes de onde a doença se iniciou. No Brasil, segundo dados do INCA (Instituto Nacional de Câncer) [1], cerca de 66 mil casos de câncer de mama (ou neoplasias malignas) são registrados por anos, isto é 29,7% dos casos de tumores em mulheres. Essas porcentagens são apresentadas com exceção dos casos de neoplasias de pele não melanoma que é o tipo mais comum em ambos os sexos, (dados retirados do site do INCA, referentes a coleta de dados para o ano de 2020). Em Medicina Veterinária, cerca de 50% das neoplasias em cadelas são neoplasias mamárias malignas [2], isso ocorre em grande parte devido ao diagnóstico tardio. O tecido mamário canino oferece uma boa comparação com o tecido mamário humano, com semelhanças epidemiológicas, clínicas, biológicas e genéticas [2-9], o que representa uma vantagem no estudo devido a sua alta ocorrência, além de não apresentar grande diferenciação do conteúdo de colágeno no tecido ao longo da vida do animal, como ocorre no caso humano.

O diagnóstico padrão para o câncer ainda é realizado de forma visual na amostra de tecido coletada por meio de biopsia, tornando a análise dependente da experiência dos patologistas, que podem se deparar com situações em que a diferenciação do tecido seja complicada, como o caso de neoplasias benignas ou malignas em estágio inicial. Além disso, a amostra coletada passa por processamentos demorados e dispendiosos, como o corte e a coloração, além de estarem sujeitas a artefatos de processamento. Dessa forma, o desenvolvimento de técnicas que quantifiquem as modificações ocorridas nos tecidos neoplásicos de maneira mais rápida, eficiente e que não necessitem de um pre-processamento da amostra é importante para uma maior agilidade e padronização no diagnóstico, buscando a redução da morbimortalidade relacionados a essa doença.

O diagnóstico padrão é feito considerando principalmente a diferenciação celular e nas alterações no tecido extracelular adjacente [10,11]. Entretanto, já é relatado na literatura que a interação entre as células e a matriz extracelular (MEC) desempenha um papel importante na progressão e metástase do câncer [12-15]. Dessa forma, estudar o tecido extracelular é de grande importância para conseguirmos entender os processos de

invasividade do tumor [13,14] e a progressão tumoral [16]. Assim, o desenvolvimento de métodos para quantificar mudanças na MEC são importantes ferramentas de diagnóstico. Técnicas de microscopia óptica não linear, como a microscopia por geração de segundo harmônico (do inglês, *second harmonic generation*, SHG) vem sendo amplamente utilizada para a caracterização da MEC em tecidos biológicos, principalmente no estudo da estrutura de organização do tecido [15,17-24]. Já que é conhecido o fato das fibras de colágeno, principal constituinte da MEC, gerarem sinal de SHG [25-28]. Estudos destacaram que a organização das fibras de colágeno como possível biomarcador intrínseco para doenças oncológicas com aplicações clínicas [13,14], além de outras características das fibras de colágeno como, a largura [29,30], o comprimento [31], o ângulo [13] e o alinhamento [21,32-33], podem ser usadas para diferenciar tecidos benignos de malignos, com a agressividade do câncer [34-36] e correlacionadas com a progressão tumoral e a sobrevivência do paciente [31]. Além do fato de ser pouco invasiva, com a ausência da toxicidade causada por marcadores biológicos, e não existir a necessidade de um pré-processamento da amostra, bem como a microscopia de fluorescência por excitação de dois fótons (do inglês, *two photon excited fluorescence*, TPEF).

No entanto, estudos que quantifiquem as propriedades das células e do colágeno de forma simultânea ainda são escassos na literatura. Dessa forma, o objetivo deste trabalho é, caracterizar de maneira quantitativa as mudanças ocorridas no tecido através de microscopias não lineares, avaliando não apenas as fibras de colágeno ou a conformação das regiões celulares, mas as duas de maneira conjunta.

Neste trabalho apresentamos os resultados de imagens multifótons por processos não-lineares de SHG e TPEF obtidos de forma simultânea em regiões intratumorais de neoplasias mamárias em cadela obtidas por procedimento de biópsias padrão coradas com Hematoxilina e Eosina (H&E) de forma a permitir a análise e diagnóstico por patologistas. Como dito anteriormente, as imagens SHG permitem informações da estrutura do colágeno na MEC e as imagens TPEF permitem acessar as regiões celulares da biópsia. Um software para processamento e análise de imagens foi escrito em Python (o PyFibre), de forma a otimizar o processo de análise de um grande número de dados obtidos. Este programa inclui algoritmos de extração de redes de fibras de colágeno e de regiões celulares na imagem, fazendo com que a imagem sejam segmentada em regiões específicas de interesse, dessa maneira todas as propriedades obtidas com o programa, como o número de fibras, número de segmentos de células, o grau de isotropia das fibras e a forma geométrica dos segmentos sejam calculadas após a segmentação da imagem, assim, as métricas extraídas podem ser medidas especificamente para regiões de colágeno e células.

Os resultados obtidos pela análise das imagens coletadas foram analisados por métodos estatísticos que permitiram quantificar e avaliar as propriedades das fibras de colágeno e regiões celulares nas biópsias de câncer, apresentando uma análise de imagem abrangente para identificar as características importantes que são visualizadas pelos patologistas para diagnosticar os tipos histológicos de câncer. Uma análise do discriminante linear (LDA)

permitiu-nos separar claramente tecidos saudáveis e tumorais com uma boa discriminação entre os subtipos de neoplasias estudadas.

Este texto contém seis capítulos organizados da seguinte forma: No capítulo [2](#), apresentamos uma breve descrição teórica da técnica de geração de segundo harmônico e da fluorescência por excitação de dois fótons, uma descrição do colágeno, da mama e suas patologias. No capítulo [3](#) descrevemos as amostras, as montagens experimentais utilizadas na aquisição das imagens, uma descrição dos cálculos utilizados pelo software e as metodologias de análises utilizadas nos dados. No capítulo [4](#) são apresentados e discutidos os resultados experimentais obtidos ao longo do desenvolvimento deste trabalho. No capítulo [5](#) apresentamos as conclusões obtidas. Por fim, nos apêndices encontra-se uma descrição do software, com suas métricas e os resultados obtidos por elas (apêndice [A](#) a [C](#), respectivamente) e uma lista com os trabalhos publicados durante o doutorado (apêndice [D](#)).

Capítulo 2

Revisão Teórica

Neste capítulo apresentaremos uma revisão teórica sobre as técnicas ópticas não lineares e as amostras utilizadas durante a realização deste trabalho, bem como uma revisão sobre as patologias estudadas e quantificadas.

2.1 Microscopia óptica não-linear

Fenômenos ópticos não lineares são aqueles em que a resposta de um material depende de forma não linear com relação a intensidade da luz incidente. A microscopia óptica não-linear faz uso dessa propriedade do material e refere-se a todas as técnicas que utilizem de processos não lineares de interação da luz com o meio material.

Nesta seção descreveremos as duas técnicas de microscopia óptica não linear utilizadas neste trabalho a microscopia por geração de segundo harmônico e a microscopia de fluorescência por excitação de dois fótons.

2.1.1 Geração de Segundo Harmônico

O fenômeno de Geração de Segundo Harmônico, do inglês *Second Harmonic Generation*, ou SHG é uma resposta óptica não linear do material quando este é submetido a um campo elétrico intenso. Esse fenômeno foi observado experimentalmente pela primeira vez em 1961 por Franken *et al.* [37].

Para materiais não lineares podemos expressar a resposta da polarização, $\tilde{P}(t)$, do material na forma de uma série de potência em termos do campo elétrico externo, $\tilde{E}(t)$, na forma [38]:

$$\begin{aligned}\tilde{P}(t) &= \epsilon_0[\chi^{(1)}\tilde{E}(t) + \chi^{(2)}\tilde{E}^2(t) + \chi^{(3)}\tilde{E}^3(t) + \dots + \chi^{(N)}\tilde{E}^N(t)] \\ &= \tilde{P}^{(1)}(t) + \tilde{P}^{(2)}(t) + \dots + \tilde{P}^{(N)}(t),\end{aligned}\tag{2.1}$$

onde $\chi^{(N)}$ é um tensor de susceptibilidade de ordem N.

O termo da polarização responsável pelo sinal de SHG é:

$$\tilde{P}^{(2)}(t) = \chi^{(2)} \tilde{E}^2(t), \quad (2.2)$$

e escrevendo o campo elétrico na forma:

$$\tilde{E}(t) = E e^{-i\omega t} + c.c. \quad (2.3)$$

onde ω é a frequência do campo elétrico incidente e *c.c.* indica o complexo conjugado.

Substituindo 2.3 em 2.2 temos:

$$\begin{aligned} \tilde{P}^{(2)}(t) &= \epsilon_0 \chi^{(2)} [E e^{-i\omega t} + E^* e^{i\omega t}] [E e^{-i\omega t} + E^* e^{i\omega t}] \\ &= 2\epsilon_0 \chi^{(2)} E E^* + \epsilon_0 \chi^{(2)} E^2 e^{-i2\omega t} + c.c. \end{aligned} \quad (2.4)$$

Na equação 2.4 observamos um termo de contribuição de frequência zero, primeiro termo, que não leva a geração de radiação eletromagnética, esse processo é conhecido como retificação óptica, onde um campo elétrico estático é criado através do material não linear. O segundo termo da equação possui uma contribuição com duas vezes o valor da frequência do campo elétrico incidente 2ω , esse termo é a contribuição de segunda ordem que gera o sinal de segundo harmônico.

Na figura 2.1 ilustramos como ocorre o processo de geração de segundo harmônico em um material. Um campo elétrico incidente com frequência ω incide em um material com o tensor de susceptibilidade de segunda ordem não nulo, o resultado da polarização resultante é um termo com frequência ω referente ao termo de polarização de primeira ordem, $\tilde{P}^{(1)}(t)$, e um termo referente a polarização de segunda ordem $\tilde{P}^{(2)}(t)$ [38].

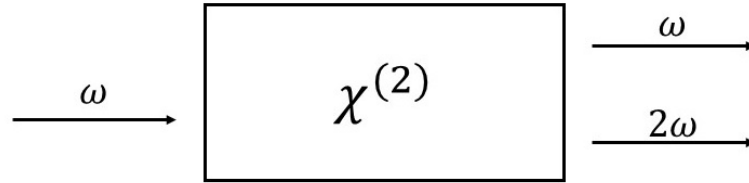


Figura 2.1: Desenho esquemático extraído do livro, [38], da geração do segundo harmônico mostrando a interação da onda de frequência ω com o material descrito pelo tensor $\chi^{(2)}$ gerando o sinal SHG com frequência 2ω .

Uma característica importante do material para que ocorra a geração de um sinal de segundo harmônico é que o material não possua simetria de inversão, ou seja deve ser materiais não centrossimétricos. Isso é necessário porquê para um material centrossimétrico ao trocarmos a direção do campo elétrico incidente, ou seja, $\tilde{E} \rightarrow -\tilde{E}$ precisamos ter alteração no sinal do vetor de polarização, $\tilde{P} \rightarrow -\tilde{P}$. Assim, trocando $\tilde{E}$ por $-\tilde{E}$ na equação 2.2 temos:

$$-\tilde{P}(t) = \epsilon_0 \chi^{(2)} [-\tilde{E}(t)]^2, \quad (2.5)$$

que mostra que:

$$-\tilde{P}(t) = \epsilon_0 \chi^{(2)} \tilde{E}(t)^2, \quad (2.6)$$

portanto, para que $\tilde{P}(t)$ seja igual a $-\tilde{P}(t)$, o termo $\chi^{(2)}$ deve ser nulo. De uma forma geral, todos os materiais centrossimétricos devem possuir seus termos $\chi^{(2N)}$ nulos. Em resumo, para que ocorra o SHG o termo de segunda ordem da polarização deve ser não nulo, o que é característica de materiais que não possuem simetria de inversão, isto é, materiais não centrossimétricos [38]. As fibras de colágeno, estudadas neste trabalho, são um exemplo de material não centrossimétrico que apresenta SHG.

Podemos caracterizar o sinal de SHG em termos quânticos analisando a energia dos fótons incidentes e emitidos.

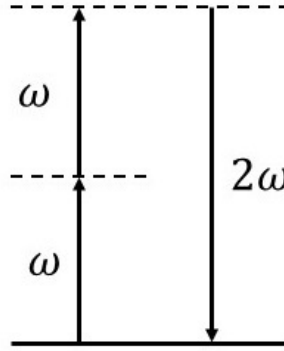


Figura 2.2: Desenho esquemático, extraído do livro [38], da geração do segundo harmônico, onde dois fótons com frequência ω geram um fóton com o dobro da frequência. A linha sólida na figura representa o estado fundamental atômico e as linhas tracejadas representam os níveis virtuais.

Na Figura 2.2 vemos os níveis de energia que descreve o sinal de SHG, onde a linha sólida representa o estado fundamental atômico e as linhas tracejadas representam os níveis virtuais, observamos que dois fótons com frequência ω são destruídos e um fóton com frequência 2ω é criado simultaneamente.

A resposta não linear de um material a um campo elétrico intenso faz com que a polarização do meio desenvolva novos componentes de frequência que não estavam presentes no campo incidente. Esses novos componentes atuam como novas fontes de frequência do campo eletromagnético. Sabemos que em um material dielétrico [38]

$$\tilde{\mathbf{D}} = \epsilon_0 \tilde{\mathbf{E}} + \tilde{\mathbf{P}}. \quad (2.7)$$

Das Equações de Maxwell para materiais não magnéticos [38]

$$\nabla \cdot \tilde{\mathbf{D}} = \tilde{\rho}. \quad (2.8)$$

$$\nabla \cdot \tilde{\mathbf{B}} = 0. \quad (2.9)$$

$$\nabla \times \tilde{\mathbf{B}} = -\mu_0 \frac{\partial \tilde{\mathbf{D}}}{\partial t} + \mu_0 \tilde{\mathbf{J}}. \quad (2.10)$$

$$\nabla \times \tilde{\mathbf{E}} = -\frac{\partial \tilde{\mathbf{B}}}{\partial t}. \quad (2.11)$$

Se considerarmos as regiões do espaço que não contêm carga e nem corrente livre, assim $\tilde{\rho} = 0$ e $\tilde{\mathbf{J}} = 0$, respectivamente. Se aplicarmos o $\nabla \times$ na equação 2.11 e substituírmos a equação 2.10 na equação resultante, obtemos:

$$\nabla \times \nabla \times \tilde{\mathbf{E}} + \mu_0 \frac{\partial^2 \tilde{\mathbf{D}}}{\partial t^2} = 0. \quad (2.12)$$

Aplicando uma identidade vetorial, $\nabla \times \nabla \times \tilde{\mathbf{E}} = \nabla(\nabla \cdot \tilde{\mathbf{E}}) - \nabla^2 \tilde{\mathbf{E}}$, substituindo a equação 2.7 e descartando os termos de interesse obtemos a equação de onda em termos da polarização 38:

$$\nabla^2 \tilde{\mathbf{E}} - \frac{1}{c^2} \frac{\partial^2 \tilde{\mathbf{E}}}{\partial t^2} = \frac{1}{\epsilon_0 c^2} \frac{\partial^2 \tilde{\mathbf{P}}}{\partial t^2}. \quad (2.13)$$

2.1.2 Fluorescência por Excitação de dois fótons

Os processos de excitação multifótons foram previstos primeiramente por Maria Göppert-Mayer em sua tese de doutorado sobre teoria de transições quânticas de dois fótons em átomos em 1931 39, mas o fenômeno de excitação por dois fótons só foi observado experimentalmente em 1961 por Kaiser and Garrett 40, na publicação sobre a excitação de dois fótons em $CaF_2 : Eu^{2+}$. Mas apenas em 1990 foi desenvolvida a técnica de microscopia de fluorescência por excitação de dois fótons 41.

O processo de excitação de um material por dois fótons consiste em um processo óptico não linear envolvendo a absorção (quase) simultânea de dois fótons por esse dado material quando a energia desses fótons é combinada (os fótons podem ter energias iguais $E_1 + E_1$ ou diferentes $E_1 + E_2$, que pode ser obtida com lasers diferentes) de forma suficiente para induzir uma transição do estado fundamental para um estado excitado. Entretanto, para que esse processo ocorra é necessário que a regra de seleção seja satisfeita 42-44. Na figura 2.3, ilustramos o tipo mais simples de excitação por absorção de dois fótons, onde dois fótons com a mesma energia são combinados.

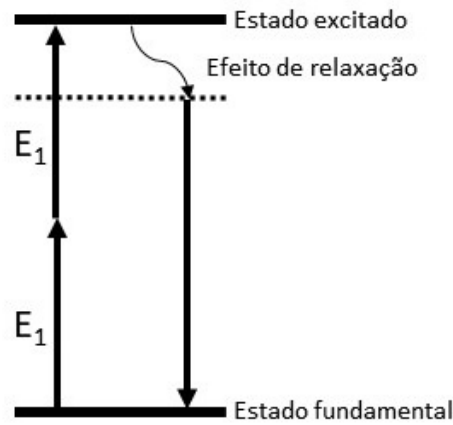


Figura 2.3: Temos então dois fótons com a mesma energia E_1 que são absorvidos pelo material quase simultaneamente e depois decai emitindo a fluorescência. Figura extraída e modificada da referência [42]

O que acontece é que um fóton com energia E_1 interage com a molécula e esta é levada para um estado virtual intermediário e deixada lá por um intervalo de tempo, τ_v , muito curto, cerca de $(10^{-15} - 10^{-16})s$, que pode ser estimado pelo princípio de incerteza. Se durante esse intervalo de tempo um segundo fóton com o mesmo valor de energia do primeiro fóton E_1 , ou com energia E_2 interagir com a molécula ela pode ser levada para o estado excitado [42-44].

A taxa de absorção de dois fótons tem uma dependência quadrática com a intensidade da luz incidente. Assim uma característica importante na fluorescência por excitação de dois fótons é que ela está limitada pelo volume focal, uma vez que a energia de um fóton com comprimento de onda longo não é suficiente para excitar a amostra acima ou abaixo do plano focal e, visto que, no plano focal temos um fluxo maior de fótons. Essa característica da excitação por absorção de dois fótons permite uma melhoria no contraste das imagens reduzindo a região de interação, sem a necessidade de utilizarmos um orifício ótico no caminho de detecção para garantirmos que o sinal coletado provém apenas do plano focal. Um orifício ótico pode ser usado no caminho ótico apenas para melhorar a razão sinal-ruído. Outra característica importante é que a excitação por dois fótons reduz foto-degradação da amostra, uma vez que ao usarmos um laser pulsado para a excitação da amostra a energia média recebida pela amostra é baixa [45].

2.2 Mama

Nesta seção descrevemos de forma breve a anatomia e a histologia de uma glândula mamária normal e em seguida descrevemos algumas patologias mamárias.

Na figura [2.4] vemos um esquema da estrutura anatômica da mama.

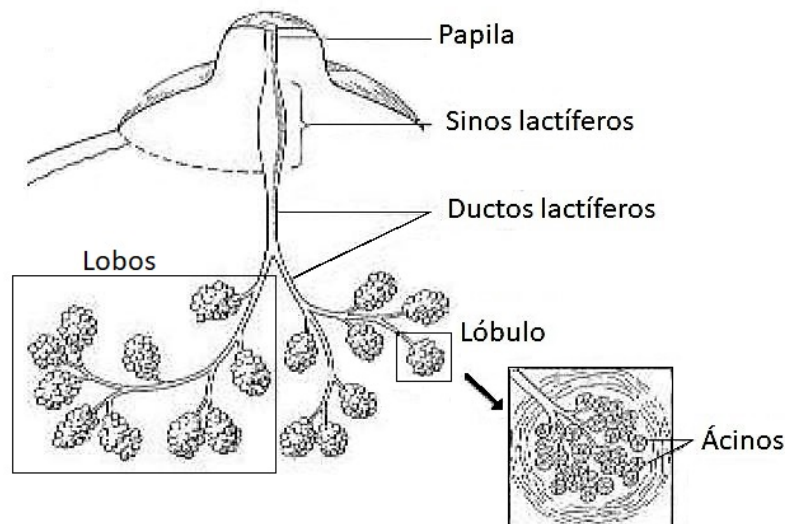


Figura 2.4: Desenho esquemático da estrutura mamária. Essa figura foi retirada da referência [46] e modificada.

A constituição da mama humana é composta de tecido glandular, tecido conjuntivo e tecido adiposo. Sua constituição pode ser separada em uma componente funcional e em uma componente estrutural. A componente funcional (tecido glandular) é a responsável pelas funções da mama como glândula, isto é, produzir e secretar leite. O tecido glandular, composto por lobos e ductos, é conhecido como parênquima mamário e é composto basicamente por tecidos de células epiteliais. Os tecidos epiteliais são formados por células intimamente unidas, com um mínimo de material intercelular entre elas acompanhadas por um tecido conjuntivo, com o qual se ligam por uma fina camada chamada membrana basal. Cada glândula mamária é composta por cerca de 20 a 25 lobos que são regiões de agrupamentos de lóbulos, cada lobo contém entre 20 a 40 lóbulos. Os lóbulos são estruturas de contorno circular formadas por ácinos e ductos. Os ácinos produzem o leite que será secretado e conduzido através dos ductos, drenando assim o lóbulo [46-50].

Já a componente estrutural da mama é composta por células de tecido adiposo e pelo tecido conjuntivo, que é conhecido como estroma mamário e é responsável pela proteção e sustentação da mama, e é dividido em estroma interlobular e intralobular. O estroma é formado por células e moléculas de matriz extracelular (MEC) e é também nesta parte do estroma mamário que o colágeno é sintetizado. Os lóbulos estão embutidos em tecido conjuntivo mais denso, chamado estroma interlobular e distingue-se pela abundância de fibras de colágeno, garantindo à mama, maior capacidade de resistir a graus excepcionais de tensão. Este tipo de tecido pode ser encontrado na forma irregular, cuja forma se assemelha à do tecido conjuntivo frouxo, porém com maior quantidade de fibras colágenas, bem como na forma regular, caracterizada pela predominância de fibras firmemente unidas em feixes paralelos. Já os ácinos estão embutidos em tecido conjuntivo frouxo, chamado estroma intralobular. A constituição dessa componente do tecido pode variar muito na

mama humana ao longo dos anos, aumentando a quantidade de tecido adiposo e, assim, diminuindo a quantidade de tecido conjuntivo, isto é, de fibras do estroma [46-50].

2.2.1 Patologias da Mama

É considerada uma patologia alguma anormalidade que surge no tecido. Os tumores mamários ou neoplasias são as patologias que surgem no tecido na forma de nódulos com tamanho e consistência variados, elas referem-se também a situações associadas ou desenvolvimento do tumor e podem conter ulcerações e reações inflamatórias no local. Os tumores podem ser classificados como benigno ou maligno e uma mesma glândula mamária pode apresentar diferentes tipos de tumores. Apenas tumores malignos são considerados câncer [46].

Tumores benignos são caracterizados por apresentarem um crescimento lento e possuírem o estroma semelhante ao dos tecidos normais, no que se refere a ocorrência de invasão. Já os tumores malignos surgem de células que sofreram alguma mutação em seu DNA, estas células possuem um crescimento rápido e desordenado, além da ocorrência em alguns casos de invasão do tecido adjacente, isto é, metástase. Em alguns casos esse crescimento celular é tão rápido que não ocorre a vascularização necessária para o tecido, ocorrendo necrose. [46].

A classificação histológica do câncer de mama humano é dividida em dois grandes grupos, os carcinomas *in situ* e invasivos. Os carcinomas *in situ* são caracterizados por uma proliferação de células confinadas dentro da membrana basal e podem ser organizadas em padrões, esses tumores podem ser divididos em carcinomas lobulares *in situ* ou carcinomas ductais *in situ*. Lesões *in situ* não são facilmente detectáveis, pois são alterações iniciais [2,4,46]. Já os carcinomas invasivos são caracterizados pelo crescimento das células e migração dessas células para o tecido subjacente.

Devido a alta taxa de incidência do câncer de mama entre as mulheres, e a alta mortalidade devido a um diagnóstico ou tratamento tardio é importante o estudo cada vez mais robusto dessa doença, porém existem algumas limitações associadas aos estudos de tumores que se concentram em avaliar culturas de células ou TMA (tissue microarray) e que muitas vezes não demonstram toda a complexidade e heterogeneidade de um carcinoma mamário. Um método pelo qual os estudos tradicionais do câncer de mama humano pode ser melhorado é integrando modelos animais.

O cão tem sido proposto como modelo animal espontâneo para o câncer de mama humano, com base nas semelhanças entre os carcinomas mamários caninos e humanos, como morfologia, comportamento biológico, genética do tumor e curso clínico dos tumores mamários como a alta incidência, a semelhança na idade relativa de início, os fatores de risco, o padrão metastático e os indicadores de prognóstico [2-9]. Além disso o sequenciamento do genoma canino mostrou que há maiores semelhanças entre as sequências de genes humanos e de cães [51].

Patologia comparada

Dessa forma a oncologia comparativa é de grande importância nesses estudos, pois permite a investigação de cânceres humanos usando cânceres semelhantes que ocorrem espontaneamente em cadelas. Uma das vantagens de estudar cânceres em cadelas em relação aos modelos pré-clínicos atuais está no fato de que esses tumores se desenvolvem naturalmente na presença de um sistema imunológico em funcionamento. Além disso, a vida útil mais curta dos cães comparada com os humanos permite aos pesquisadores um estudo completo da progressão tumoral em um espaço de tempo mais curto suficiente para permitir a comparação das respostas ao tratamento. Outra vantagem em utilizar o cão como um modelo comparativo está no fato de que os tumores mamários caninos são as neoplasias mais comumente diagnosticado em cadelas, quase 50% de todas as neoplasias caninas [2].

As neoplasias mamárias caninas, assim como as humanas, são comumente apresentadas como nódulos circunscritos com tamanho, consistência e mobilidade variáveis e podem estar associadas à ulceração da pele e reações inflamatórias locais. Os tumores mamários caninos podem ser observados em uma única glândula mamária ou envolver múltiplas glândulas mamárias simultaneamente e podem ser de diferentes tipos histológicos [2,11]. As neoplasias mamárias podem ser de origem epitelial ou mesenquimal, podendo haver uma combinação de tecidos epiteliais e mioepiteliais ou tecidos epiteliais e mesenquimais [2,11].

2.2.2 Tipos histológicos

Os tumores mamários benignos caninos são classificados em: adenoma, fibroadenoma, papiloma ductal e tumor misto benigno. Os carcinomas mamários malignos caninos são classificados em: carcinomas *in situ* (ductal e lobular), carcinoma em tumor misto, carcinoma tubular, carcinoma papilar, carcinoma sólido, carcinoma anaplásicos e carcinomas de tipos especiais (carcinoma micropapilar, carcinoma de células escamosas, carcinoma mucinoso, carcinoma secretor, carcinoma rico em lipídios e carcinoma de células fusiformes) [11]. Para o desenvolvimento deste trabalho foram escolhidos alguns tipos histológicos de carcinomas mamários em cadelas. Aqui descreveremos a morfologia dos tipos estudados.

Carcinoma Ductal *In Situ*: O Carcinoma Ductal *In Situ* é caracterizado como uma proliferação de células epiteliais malignas nos ductos intra ou extra-lobulares, onde as células malignas anormais ocupam o lúmen ductal sem descontinuidade ou ausência da membrana basal. É o carcinoma *In Situ* mais comum, frequentemente observado em associação com carcinomas mamários caninos invasivos (carcinomas sólidos, papilares, carcinomas em tumores mistos). A proliferação de células epiteliais exibe uma arquitetura celular atípica, caracterizada por pontes de conexão ao longo do lúmen ductal e uma

camada de células epiteliais associadas a uma camada contínua de células mioepiteliais [11].

Tumor Misto Benigno: Os tumores misto benignos são as neoplasias benignas mais frequentes no cão. Possuem um padrão histológico complexo caracterizados pela proliferação benigna das células morfológicamente semelhantes aos componentes epiteliais (luminal ou mioepitelial), que podem se tornar malignos levando ao desenvolvimento de carcinomas em tumores mistos, e das células que produzem cartilagem e/ou tecido adiposo, possivelmente em combinação com o tecido fibroso. Nessas neoplasias as fibras de colágeno são abundantes, assim como na glândula mamária normal, porém as fibras são mais organizadas [4,11,52].

Carcinoma em Tumor Misto: Os carcinomas em tumores mistos são tumores que contêm focos ou nódulos de células epiteliais com pleomorfismo elevado e mitose atípica, a proliferação desse carcinoma pode invadir ou substituir completamente uma lesão benigna. As células epiteliais malignas exibem um crescimento infiltrativo, que pode ser identificado pela perda de continuidade da membrana basal associada a uma aglomeração de células neoplásicas que penetram no estroma. A ocorrência de proliferações de carcinoma não invasivo (*in situ*) também pode ser observada. A diferenciação entre carcinoma em tumor misto *in situ* e invasivo é possível pela presença de invasão e microinvasão estromal. As áreas de invasão são caracterizadas pela presença de aglomerados de células epiteliais tumorais infiltrativas no estroma periductal próximo aos componentes do carcinoma. [11,52]

Carcinoma Micropapilar: O carcinoma micropapilar é caracterizado por pequenas estruturas papilares que exibem espaços císticos semelhantes a vasos linfáticos difusamente distribuídos por todo o tecido mamário. Dentro desses espaços estão agrupamentos de células epiteliais que exibem um padrão micropapilar. As áreas *in situ* são caracterizadas por uma proliferação epitelial maligna restrita ao ducto (circunscrito por uma membrana basal) contendo ninhos de células epiteliais desprovidos de um núcleo fibrovascular. Por sua vez as áreas invasivas são raras e estão correlacionada a um prognóstico desfavorável [11].

Carcinossarcoma: Os carcinossarcomas são compostos de células morfológicamente semelhantes às células epiteliais malignas (epitelial luminal e/ou mioepitelial) e células semelhantes ao tecido conjuntivo maligno concomitantemente, mas possuem características histológicas extremamente variáveis apresentando vários tipos de diferenciação, incluindo áreas que apresentam abundância de tecido conjuntivo e áreas sólidas. Macroscopicamente os carcinossarcomas são bem delineados e com uma superfície de corte firme e óssea. São tumores com um crescimento rápido e por isso tem um prognóstico pior em comparação com outros tipos de carcinomas e foram caracterizados primeiramente como tumor misto maligno [4,11].

Carcinoma Sólido: O carcinoma sólido é um tipo comum de tumor mamário em cães. São geralmente mal definidos com as células tumorais epiteliais indiferenciadas exibindo núcleos pequenos e hipercromáticos e são organizadas em arranjo sólido, com a formação de cordões, lâminas ou aglomerados. Esse tipo tumoral possui alto índice mitótico e a quantidade de estroma pode variar de moderado a pequeno. Como são tumores muito celulares a vascularização do tecido pode não ocorrer muito bem, causando áreas de necrose [4,11].

2.3 Colágeno

Como mencionado anteriormente, a matriz extracelular (MEC) é o componente não celular do estroma e está presente em todos os tecidos e órgãos. Ela é composta principalmente por proteoglicanos e proteínas fibrosas, dentre as quais o colágeno é a mais abundante, correspondendo a cerca de 30% das proteínas presentes no corpo, o que corresponde a 6% do peso corporal. O colágeno é responsável por fornecer resistência e sustentação ao tecido, além de regular a adesão celular e auxiliar na migração, como mencionado em [53-59]. Desta forma podemos entender que a MEC é um importante objeto de estudo para o entendimento do ambiente tumoral.

Na Figura 2.5 ilustramos a constituição das fibras de colágeno.

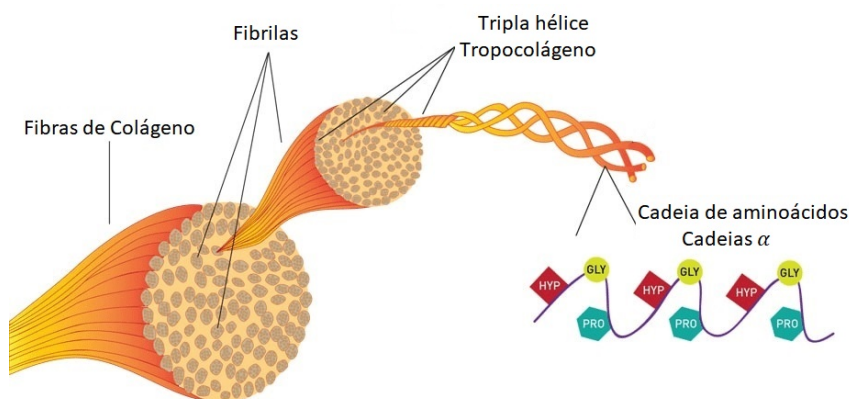


Figura 2.5: Estrutura das fibras de colágeno mostrando desde a estrutura de tripla hélice até o agrupamento na formação das fibras. Figura extraída e modificada do site <https://www.esteticistacomovoce.com.br/o-que-e-colageno/> no dia 16 de agosto de 2021.

Como apresentado nas referências [55,59-62], atualmente cerca de 28 tipos diferentes de colágenos foram identificados cada um com uma função, porém com a característica comum de uma estrutura em tripla-hélice. O colágeno é composto por aminoácidos que se ligam formando uma cadeia, chamada de cadeia α . A principal característica dessa cadeia está na repetição do tripleto de aminoácidos Gly-X-Y, e são diferenciadas apenas pelos aminoácidos nas posições X e Y. As cadeias α se ligam na forma de uma estrutura em tripla-hélice com giro à direita e as diferentes combinações das três cadeias α determina o tipo de colágeno. Cada cadeia α individual é identificada pela seguinte nomenclatura:

$\alpha n(N)$, onde N é o numeral romano que indica o tipo de colágeno e n é o número da cadeia α . Assim, a composição da cadeia do colágeno tipo II, por exemplo, é $[\alpha 1(II)_3]$, enquanto que a do colágeno tipo I, um heterotrímero com duas cadeias $\alpha 1$ idênticas e uma terceira cadeia $\alpha 2$ distinta é $[\alpha 1(I)]_2\alpha 2(I)$ [59]. A tripla hélice é estabilizada pela formação de ligações de hidrogênio entre cadeias, bem como a quantidade total dos aminoácidos prolina (Pro) e hidroxiprolina (Hyp), nas posições X e Y, que juntos constituem cerca de 10% para os colágenos em geral e 20% do total de aminoácidos nos colágenos fibrilares humanos [59,62].

Colágenos formadores de fibras, como por exemplo os colágenos tipo I, II e III, são organizados de forma hierárquica, onde o nível hierárquico mais baixo consiste em moléculas de colágeno em hélice-tripla, conhecidas como tropocolágenos. Os tropocolágenos são encontrados na natureza com cerca de 300 nm de comprimento e aproximadamente 1,5 nm de diâmetro [56,62,63]. As moléculas de tropocolágeno são agrupadas de forma escalonada, com uma periodicidade D, com $D = 67$ nm formando fibrilas com diâmetro de aproximadamente 100-500 nm e com um comprimento na faixa de microns [55,56,60]. Por fim as fibrilas são agrupadas formando as fibras de colágeno.

Na mama os tipos de colágeno mais comuns são os tipo I, III, que são fibrilares, e o tipo IV, que é não fibrilar e envolve a membrana basal. Mas o colágeno tipo I é o principal responsável pelo sinal, uma vez que é o mais abundante, correspondendo a 90% do colágeno encontrado no corpo [64]. O colágeno tipo I é o mais importante, uma vez que ele é o responsável pela sustentação e funcionalidade dos tecidos devido a sua interação com as células presentes na matriz extracelular. As fibras de colágeno tipo I possuem $\sim 1 \mu\text{m}$ de largura, $\sim 10 \mu\text{m}$ de comprimento com uma concentração de Gly de 33%, Pro de 12% e Hyp de 11% [53,55,60,63,64].

2.3.1 SHG no colágeno

Desde a década de 1980 já é conhecido que as fibras de colágeno são uma estrutura não centrossimétrica [65] e, assim, produzem um sinal de segundo harmônico, e desde a década de 2000 Campagnola *et.al.* relataram uma implementação prática desta técnica em tecidos biológicos [66]. Desta forma, microscopia por geração de segundo harmônico vem sendo, desde então, muito utilizada no estudo de tecidos biológicos [13,20,21,24,26,31,67,68], principalmente pelo fato de que na microscopia de SHG não há necessidade de um pré-processamento para a aquisição das imagens, dessa forma tecidos histológicos fixo, tecidos *in vitro* e modelos *in vivo*, podem ser estudados [55].

Como visto na seção anterior [2.1.1], o sinal de SHG surge da polarização induzida no material. Assim, para que o material tenha um sinal de segundo harmônico o termo de segunda ordem da polarização, $\chi^{(2)}$, deve ser não nulo. A susceptibilidade não linear de 2ª ordem, impõe que a não simetria do material esteja na escala de tamanho de λ_{SHG} , as proteínas primárias que atendem a esses requisitos são o colágeno tipo I e II. Embora $\chi^{(2)}$ seja a quantidade medida em um experimento, a propriedade em nível molecular

relacionada a não linearidade é a primeira hiperpolarizabilidade, β . A susceptibilidade não linear de 2ª ordem está relacionada com a hiperpolarizabilidade molecular do material na forma:

$$\chi^{(2)} = N_S \langle \beta \rangle, \quad (2.14)$$

onde N_S é a densidade das moléculas e os colchetes indicam uma média orientacional, mostrando uma necessidade de um ambiente não centrossimétrico [26, 27].

A origem molecular de β no colágeno surge das ligações peptídicas ao longo da molécula e não de um aminoácido específico, e a geração do sinal em cada uma dessas ligações resultando em uma amplificação coerente do sinal não linear [27, 28, 69, 70]. Foi demonstrado experimentalmente que os elementos não nulos da matriz $\chi^{(2)}$, surgem do ângulo de inclinação ($\sim 50^\circ$) das hélices α na molécula de colágeno [27, 28].

O SHG é um processo coerente em cristais, isso significa que para que o SHG tenha um casamento de fase perfeito $\Delta K = K_{2\omega} - 2K_\omega = 0$, onde $K_{2\omega}$ e K_ω são os vetores de onda para o SHG e o fóton incidente, respectivamente. Porém, em tecidos biológicos existe uma aleatoriedade e uma dispersão inerentes que resultam em uma distribuição de valores ΔK diferentes de zero, assim, o SHG em tecidos é melhor descrito como um processo quase coerente [27, 28].

Capítulo 3

Metodologia Experimental

Neste capítulo descreveremos as amostras de tecido mamário canino utilizadas para este estudo, bem como as técnicas de microscopia óptica não linear utilizadas na obtenção das imagens das amostras, que são a microscopia por SHG e a microscopia de TPEF. Aqui também descreveremos o software desenvolvido durante esse trabalho para a análise das imagens coletadas.

3.1 Amostras

Todas as amostras de tecido mamário utilizadas nesse estudo foram obtidas em colaboração com o grupo do professor Geovanni Dantas Cassali e a aluna de doutorado Ana Paula Vargas Garcia, Laboratório de Patologia Comparada do Departamento de Patologia Geral, Instituto de Ciências Biológicas (ICB), UFMG. A utilização das amostras foi aprovada pelo comitê de ética no uso de animais (CEUA) da UFMG nº 251/2018.

As amostras foram colhidas a partir de biópsias de tumores de ocorrência espontânea em cadelas. Os tecidos foram obtidos no hospital veterinário da UFMG em procedimentos de mastectomia simples, mastectomia regional, mastectomia radical unilateral ou bilateral dependendo do tamanho do tumor, estadiamento clínico, drenagem linfática e localização do tumor, pelo hospital veterinário da UFMG e foram enviadas ao laboratório de patologia do ICB embebidas em formol onde foram processadas. As amostras selecionadas neste estudo foram coletadas entre os anos de 2009 a 2019 e classificadas de acordo com os critérios descritos no Consenso para o Diagnóstico, Prognóstico e Tratamento de Tumores Mamários Caninos [11]. Todas passaram pelo procedimento padrão de montagem em lâmina para diagnóstico. Primeiro é feito uma macroscopia, onde o profissional direciona qual parte da amostra passará pelo processo de inclusão em parafina. Logo após a parafinização a amostra passa por cortes em blocos de 4 mm de lado. As amostras foram cortadas por micrótomo em fatias bem finas e uniformes, com espessura comumente de 4 a 6 μm e coradas com Eosina e Hematoxilina (H&E) e foram montados em lâminas de microscópio padrão e cobertos por uma lamínula de vidro para leitura e diagnóstico. Essas amostras foram avaliadas por um patologista experiente e especializado em câncer

de mama canino, o professor Geovanni Dantas Cassali e utilizadas no diagnóstico histopatológico do tumor. Para as fotografias das regiões selecionadas como áreas representativas do tipo tumoral mostradas nesse estudo foi utilizado um microscópio Olympus BX41 com luz branca e câmera acoplada, onde as imagens de microscopia de campo claro foram gravadas. Podemos ver um exemplo de imagem padrão de diagnóstico na figura 3.1, essa imagem é de um tecido de hiperplasia de células colunares. A região com coloração em roxo é referente ao corante hematoxilina, que se liga ao núcleo celular, nesse caso um ducto lactífero com hiperplasia. A hiperplasia é caracterizada por um crescimento ordenado e contínuo das células que revestem o ducto, com os núcleos permanecendo normais não existe nenhuma alteração neoplásica. Já região com coloração rósea é referente ao corante eosina que se liga ao estroma, região contendo o colágeno, e a membrana citoplasmática.

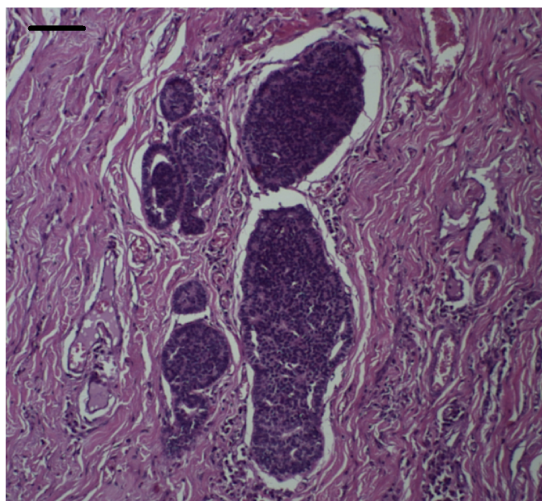


Figura 3.1: Imagem de microscopia de campo claro de uma lâmina de diagnóstico padrão corada com H&E em um tecido de hiperplasia mamária de células colunares. Imagem coletada com uma objetiva de 4x, a escala corresponde a 200 μm .

3.2 Microscopia por geração de segundo harmônico - SHG

As imagens de microscopia por SHG são obtidas usando um microscópio FV-300 da Olympus modificado para permitir o acoplamento de um laser pulsado para a excitação dos processos multifótons. Usamos um laser pulsado de Ti-safira, sintonizável (680 - 1080 nm) com pulsos de 140 fs (modelo Chameleon, do fabricante Coherent) e com taxa de repetição de 80 MHz. O laser é alinhado no microscópio BX61-WI vertical por espelhos externos, colinearmente com o laser contínuo do microscópio confocal para permitir medidas em todos os detetores (fotomultiplicadoras) do microscópio confocal. A potência do laser que incide na amostra é ajustada através de uma placa de meia-onda e um polarizador, antes de ser direcionado ao microscópio.

A figura 3.2 mostra um desenho esquemático da montagem experimental utilizada para obtenção das imagens SHG com o caminho do laser a partir da unidade de escaneamento.

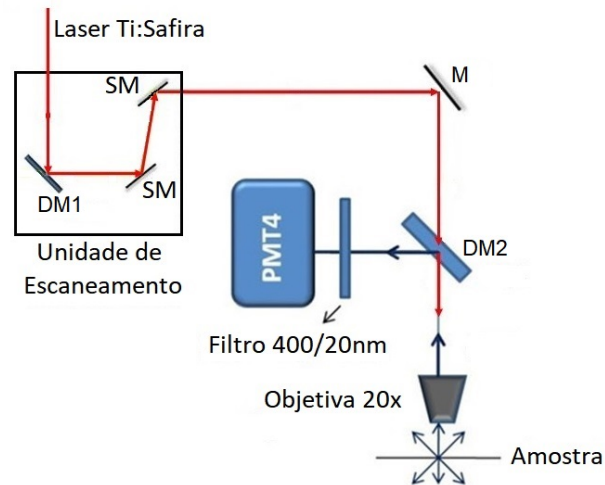


Figura 3.2: Esquema da montagem experimental para o experimento de SHG, onde, DM significa espelho dicróico, PMT significa fotomultiplicadora e SM significa espelho de varredura.

Após o caminho inicial para um ajuste de potência, o feixe de excitação entra na unidade de escaneamento. Na unidade de escaneamento o feixe de laser é refletido por um espelho dicróico (DM1) para o par de espelhos (SM) que podem ser movimentados de forma a variar os ângulos de inclinação do mesmo, isso permite um movimento de varredura em xy na posição da amostra. Dessa forma podemos excitar a amostra ao longo de uma área e obter uma imagem ponto a ponto.

Ao sair da unidade de escaneamento a luz passa por um espelho dicróico (DM2) que está posicionado logo acima da objetiva. Esse espelho dicróico transmite o infravermelho (680 – 1080 nm), faixa do comprimento de onda de excitação da amostra, e reflete o visível (abaixo de 680 nm), faixa do comprimento de onda de emissão do SHG da amostra. Utilizamos uma objetiva de 20X de aumento, com imersão em água e abertura numérica (N.A) de 0,90, que focaliza a luz sobre a amostra. Após a interação da luz com a amostra, coletamos o sinal do segundo harmônico, por retroespalhamento, pela mesma objetiva de focalização. A luz retro espalhada que sai da objetiva passa novamente pelo espelho dicróico que reflete o sinal para uma fotomultiplicadora sensível para a coleta do segundo harmônico. Para excitar a amostra usamos o laser sintonizado em 800 nm, que gera um sinal de segundo harmônico em 400 nm. Um filtro passa-banda centrado em 400 nm com largura de banda de 20 nm é colocado na frente da fotomultiplicadora PMT4 para filtrar o laser espalhado. As imagens foram feitas com luz circularmente polarizada com uma potência de laser na amostra de 7 mW com uma fluência de aproximadamente de $7 \times 10^{-4} \text{ J/cm}^2$ e com a energia do laser por pulso em cerca de 0,1 nJ. Obtemos imagens fazendo uma varredura de 512×512 pixels, pixel de $0,921 \mu\text{m}$. Assim a área da imagem é de $0,471 \text{ mm} \times 0,471 \text{ mm}$.

Através de um controlador piezoelétrico podemos movimentar a objetiva de focalização na direção da propagação do feixe de excitação, movimentando dessa forma o plano focal

e obtendo imagens em diferentes profundidades. Em nossos experimentos, o sinal de SHG foi coletado para três planos diferentes, separados por $1\ \mu\text{m}$. Dessa forma, garantimos um melhor foco e as 3 imagens coletadas são utilizadas nos processos de filtragem de ruído no software de análise.

3.3 Microscopia de Fluorescência por Excitação de dois fótons - TPEF

Para a obtenção das imagens por microscopia de TPEF, foi utilizado o mesmo aparato experimental da microscopia por SHG, com algumas modificações, retirando o DM2, que serão descritas a seguir. O feixe de laser com comprimento de onda de 800 nm percorre o mesmo caminho até chegar a unidade de escaneamento. Após sair da unidade de escaneamento ele passa por um espelho e é focado na amostra pela objetiva. A fluorescência é coletada pela mesma objetiva de focalização por retroespalhamento. Após passar pela objetiva ela passa novamente pelo espelho dicróico DM1, responsável, agora, por refletir o laser de excitação e transmitir a fluorescência. O sinal de fluorescência passa por uma lente e por uma íris que barram a luz proveniente de espelhamento e de outros planos, dessa forma garantimos que apenas a luz proveniente do plano focal chegue até as fotomultiplicadoras PMT1 e PMT2, que são usadas para obter imagens de diferentes regiões da banda de fluorescência.

A figura 3.3 mostra um desenho esquemático da montagem experimental utilizada para a coleta da fluorescência por excitação de dois fótons.

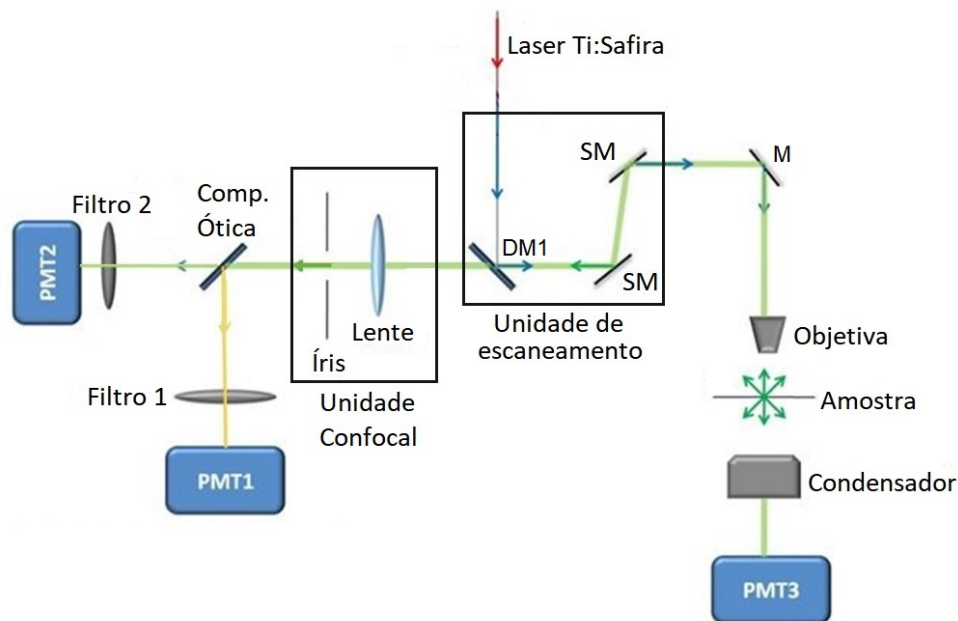


Figura 3.3: Esquema da montagem experimental para o experimento microscopia de fluorescência por excitação de dois fótons. Na imagem também está indicado o condensador e a fotomultiplicadora (PMT3) onde o sinal de transmissão é coletado.

Após passar pela região confocal, temos um componente ótico (identificado na [3.3](#) como Comp. ótico) que pode ser um espelho dicróico que reflete os comprimentos de onda menores que 505 nm (DM), um espelho dicróico que reflete comprimentos de onda menores que 570 nm (DM), ambos separam o sinal da fluorescência em bandas espectrais que serão medidos nas fotomultiplicadoras PMT1 e PMT2, ou ainda um espelho metálico que reflete toda a fluorescência para PMT1. Em nossos experimentos utilizamos o espelho dicróico DM1 em 505 nm. Um filtro passa banda (560-600 nm) é utilizado antes da fotomultiplicadora para eliminar o restante de laser espalhado e selecionar a banda de fotoluminescência a ser coletada pela PMT2. Esta faixa de comprimento de onda foi escolhida para medirmos a fluorescência da eosina, que é o corante utilizado na marcação das biópsias.

Obtivemos também uma imagem de transmissão através da luz transmitida pela amostra que é coletada pelo condensador e focalizada em uma fotomultiplicadora (PMT3).

As imagens de SHG, TPEF e Transmissão são obtidas para as mesmas posições da amostra permitindo uma comparação direta das posições nas imagens. O resultado é a formação de imagens ópticas não-lineares por varredura no plano focal da amostra como vemos na figura [3.4](#). Nessa figura temos a imagem de SHG, [3.4\(A\)](#), a imagem de TPEF, [3.4\(B\)](#), e a imagem de Transmissão, [3.4\(C\)](#), da mesma amostra de tecido de hiperplasia de células colunares. Nas imagens de SHG o que vemos são as fibras de colágeno, como descrito anteriormente, e nas imagens de TPEF vemos a fluorescência do corante eosina, que mostra tanto a região de parênquima quanto as fibras de colágeno. Pela imagem de luz transmitida também temos uma visão tanto das região estromal do tecido quanto do parênquima.

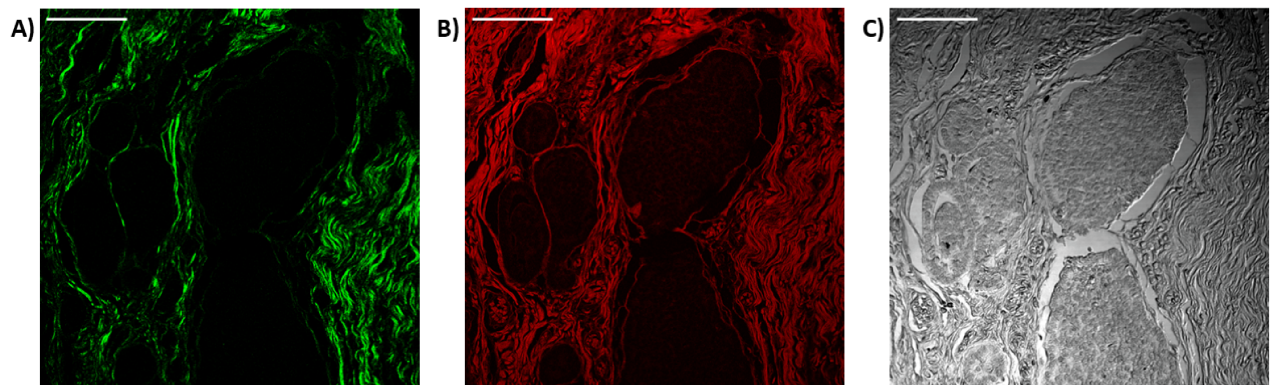


Figura 3.4: Imagens em escala de cinza, coloridas artificialmente, dos sinais de intensidade do SHG, (A), TPEF, (B), e Transmissão, (C), para o tecido de hiperplasia mamária de células colunares. As imagens têm 512×512 pixels e a barra de escala é de $100 \mu\text{m}$.

3.4 Análises das Imagens

Depois da aquisição das imagens de SHG, TPEF e Transmissão, queremos agora, estudar as modificações sofridas pelo tecido mamário canino nos diferentes subtipos histológicos estudados. Para a análise dessas imagens foi desenvolvido um software em colaboração com a Universidade de Southampton, o PyFibre (Python Fibrous Image Analysis Toolkit), que foi escrito pelo pós-doc Frank Longford. O PyFibre é escrito em linguagem de Python e foi desenvolvido de forma a otimizar a análise para um número muito grande de dados e tornar automatizado a quantificação de imagens de tecido fibroso, além de gerar um banco de dados com vários parâmetros extraídos que podem ser utilizados em análises estatísticas futuras. O software está disponível em [71].

A análise de imagem realizada pelo PyFibre visa segmentar uma imagem de tecido em regiões de fibras e células, para isso, o software reúne vários algoritmos de processamento de imagem que já são conhecidos na literatura, como FIRE, o CurveAlign e métodos de agrupamento como o K-means [30, 36, 72–74], mas foi feito um aprimoramento desses algoritmos de forma a melhorar a extração de parâmetros. Assim, os cálculos podem ser feitos separadamente para cada região segmentada (tanto para regiões de fibras quanto de células) como pode ser feito também uma média para a imagem global.

Em sua versão atual, o programa carrega imagens em formato TIFF (*Tagged Image File Format*) e analisa de forma automatizada imagens multicanal que contenham sinais SHG, TPEF e Transmissão. Dessa forma, o programa começa lendo as imagens de SHG, TPEF e Transmissão e em algumas etapas de pré-processamento redimensiona suas intensidades, a fim de remover ruídos e realçar características importante. No pré-processamento das imagens utilizamos um “filtro tubular” [75], o algoritmo é baseado em uma combinação dos autovalores da matriz Hessiana e feito de forma a destacar as estruturas que são semelhantes a fibras ao mesmo passo em que atenua regiões ruidosas. A matriz Hessiana $\mathbf{H}(x, y)$ é dada pela equação 3.1

$$\mathbf{H}(x, y) = \begin{pmatrix} f''_{xx}(x, y) & f''_{xy}(x, y) \\ f''_{yx}(x, y) & f''_{yy}(x, y) \end{pmatrix}, \quad (3.1)$$

onde $f(x, y)$ é a imagem bidimensional e $f''(x, y)$ é a derivada segunda direcional.

Com a intensidade das imagens normalizadas e com as características fibrosas destacadas, passamos para as etapas de processamento onde as regiões de fibras e células serão identificadas e separadas para cálculos posteriores. O software utiliza o fato de imagens de SHG conterem apenas sinal de colágeno, o que ajuda a identificar de forma eficiente as regiões fibrosas do tecido. Dessa forma, os sinais de TPEF e Transmissão são utilizados para refinar as fronteiras entre as regiões de fibras e de células. Abaixo será descrito todas as ferramentas utilizadas pelo PyFibre para a segmentação de imagens bem como algumas métricas calculadas que serão abordadas no estudo. No apêndice A será feita uma apresentação sobre a estrutura e metodologia da utilização do software.

3.4.1 Extração da Rede de Fibras

A extração das fibras é feita com base no conceito de redes. Uma rede, $N = V, X, F$, consiste em um conjunto de n nós (ou vértices), $V = 1, 2, \dots, n$, com coordenadas $X = \vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ e um conjunto de m arestas $F = \vec{f}_1, \vec{f}_2, \dots, \vec{f}_m$ que conectam esses nós, o grau de cada nó determina o número de arestas que ele possui [76]. Cada vértice e cada aresta possuem um vetor tridimensional, o primeiro indicando a localização do vértice no espaço, referindo-se a coordenada local de origem, e o segundo indicando os vértices pelos quais a aresta passa [76, 77], como indicado na figura 3.5. Para a criação da rede em nosso trabalho utilizamos uma versão modificada do algoritmo do software FIRE, esse algoritmo é projetado para traçar uma rede em uma imagem com textura fibrosa e extrair a estrutura geométrica de imagens de colágeno [72]. A primeira etapa de funcionamento do FIRE consiste em transformar uma imagem em escala de cinza em binária, fazendo com que os pixels correspondentes as fibras recebam o valor 1 e os pixels correspondentes ao fundo, isto é, pixels onde não há fibras, recebam o valor 0 [72, 77].

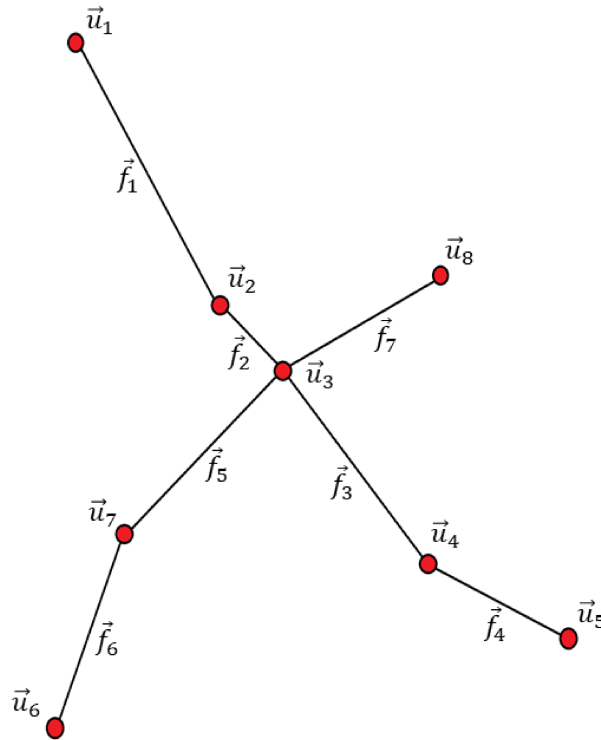


Figura 3.5: Ilustração de uma rede simples, construída como descrita no início da seção. Nela vemos um conjunto de 8 nós (vértices) com coordenadas $\vec{u}_n$ e um conjunto de arestas $\vec{f}^m$. Figura modificada do artigo [77].

Em seguida, uma matriz de distância é calculada representando o número de pixels entre o primeiro plano e o segundo, depois um filtro gaussiano é aplicado de forma a suavizar essa matriz de distância discreta para dados contínuos reduzindo assim o número de valores de máximos redundantes, uma vez que esses pontos de máximos da imagem suavizada serão os pontos de nucleação, esse conjunto primário de pontos de nucleação

são chamados de nós pais. Depois de identificar os pontos de nucleação, a próxima etapa é traçar as arestas (fibras) que se estendem a partir dos pontos de nucleação. Para isso, definimos o conjunto de Pontos Máximos Locais (LMPs), os LMPs ocorrem onde há um máximo local dentro de uma determinada região de pesquisa de cada nó pai, esses novos pontos de nucleação são conhecidos como nós filhos, as fibras são então ligadas a esses nós. Após a atribuição inicial de todos os nós pais e filhos primários, a propagação da rede continua iterativamente em cada LPM até que nenhum outro candidato a nó filho possa ser encontrado [24, 71, 72, 77].

As principais diferenças entre o algoritmo FIRE original e a nossa implementação está no uso de um limite no comprimento da aresta $\vec{r}_0$ e na capacidade de criar uma aresta entre qualquer nó existente, enquanto que no FIRE apenas nós filhos com mesmo nó pai poderiam ser conectados. O uso desse limite do comprimento da aresta nos permite mover os LMPs após sua contribuição original e também controlar a resolução na qual cada fibra pode ser rastreada sem alterar o tamanho da janela de pesquisa especificada para o LMP local. Depois que todos os nós filhos foram propagados uma limpeza é aplicada para remover quaisquer artefatos, em geral, essa limpeza visa remover nós redundantes ou não conectados, bem como redes pequenas (≈ 4 nós) que podem corresponder a ruídos [24, 71].

Como resultado do melhoramento do algoritmo do FIRE, o programa PyFibre nos permite extrair detalhes de fibras com uma maior resolução, e assim traçar uma rede em cima da imagem que contém uma estrutura fibrilar de forma a extrair a rede de fibras como ilustra a figura 3.6.

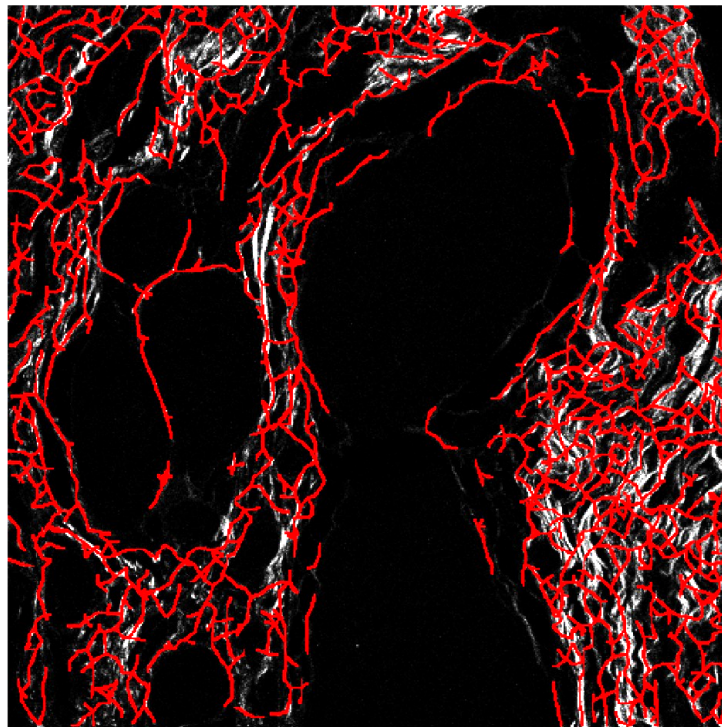


Figura 3.6: Imagem das redes extraídas de uma amostra de hiperplasia mamária de células colunares com o uso do programa PyFibre. Rede de fibras em vermelho sobreposta com a imagem SHG em escala de cinza.

Essa rede bruta gerada pela implementação e melhoramento do FIRE pode ser manipulada de forma a termos uma visão geral de sua estrutura e conectividade de fibra ou para obter informações sobre fibras individuais. No processo de identificação das fibras individuais aplicamos uma abordagem semelhante à já utilizada no algoritmo FIRE e atribuímos uma regra à definição de fibra. Primeiro identificamos todos os nós que contêm apenas uma aresta; isto indica que eles residem no início de uma fibra. Esses nós tornam-se o nó pai da fibra e seu primeiro nó filho será o único nó onde eles compartilham uma aresta. O algoritmo então rastreia de volta todos os nós conectados ao nó filho ativo e reconstrói uma fibra [24].

O resultado visual da extração de fibras individuais na rede bruta pode ser visto na figura 3.7.

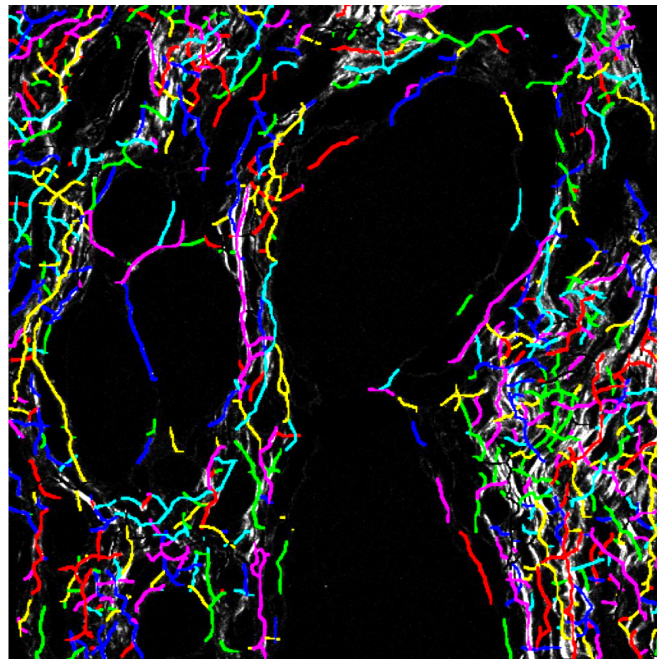


Figura 3.7: Imagem das fibras únicas extraídas com o uso do programa PyFibre em uma amostra de hiperplasia mamária de células colunares. Fibras individuais em cores aleatórias, sobre a imagem SHG em escala de cinza.

3.4.2 Segmentação de Fibras e Células

Com a extração de redes de fibras na imagem conseguimos uma boa estimativa da região de fronteira entre as regiões de fibras e de células da amostra. Mas podemos melhorar essa separação realizando um processo de segmentação com base no algoritmo *BDcreationHE* do software CurveAlign [30,32]. A rotina envolve o agrupamento de vetores de pixel em imagens RGB que é gerada a partir de 3 canais “RGB” referentes aos sinais de SHG, TPEF e de transmissão, que correspondem aos canais vermelho, azul e verde, respectivamente. Assim, as características do colágeno aparecem como vetores unitários com grandes componentes vermelhos, enquanto as características celulares resultam em vetores unitários contendo componentes azuis e verdes. Cada vetor de pixel RGB é então

normalizado em um vetor unitário de forma a reduzirmos a dependência da intensidade em cada imagem. A segmentação da imagem é realizada agrupando os componentes RGB de cada pixel, independentemente de sua localização. Logo após a identificação desses agrupamento de cores principais na imagem composta RGB, identificaremos se eles devem ser classificados como segmentos fibrosos ou segmentos celulares, isso é feito com base em seu vetor de unidade centróide $\vec{c}$ e valores de intensidade média $\bar{I}$ identificado pelo algoritmo K-means. Definimos um vetor simples $v(r, g, b, i)$, descrito na equação 3.2, que indica a fronteira entre os segmentos fibrosos e celulares, um segmento é considerado celular se cada componente desse vetor tiver um valor inferior a esse limite [24, 71];

$$\varphi(\vec{c}, \bar{I}) = \Pi \begin{cases} 1 \text{ se } \arcsin(\vec{c}_R) < v_r, \text{ senão } 0 \\ 1 \text{ se } \arcsin(\vec{c}_G) < v_g, \text{ senão } 0 \\ 1 \text{ se } \arcsin(\vec{c}_B) < v_b, \text{ senão } 0 \\ 1 \text{ se } \bar{I} < v_i, \text{ senão } 0 \end{cases} \quad (3.2)$$

Os pixels presente nesse agrupamento são então combinados para formar um “filtro de células” binário. É importante ressaltar que os valores padrão de $v(r, g, b, i)$ usados pelo software são arbitrários e parametrizados com base nos dados disponíveis e, portanto, são unidades relativas, em vez de absolutas, não esperamos, portanto, que o algoritmo seja eficaz na segmentação de outros conjuntos de dados sem recalculá-los. No entanto, essas imagens combinadas permitiram uma boa separação das regiões contendo fibras e células [24, 71].

Como resultados para a segmentação obtemos imagens como mostrada na figura 3.8, onde (A) mostra os segmentos celulares e (B) mostra os segmentos de fibras extraídos pelo PyFibre para uma biopsia de Hiperplasia de células colunares.

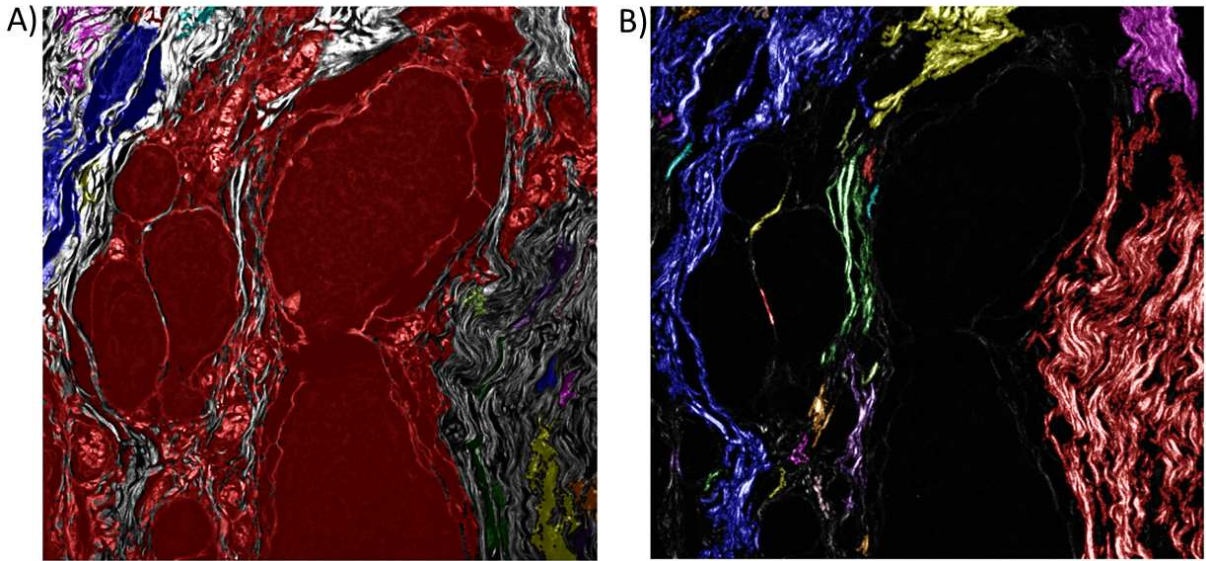


Figura 3.8: Imagem dos segmentos celulares (A) e segmentos de fibras (B) de uma amostra de hiperplasia mamária de células colunares com o uso do programa PyFibre. As cores são escolhidas de forma aleatória apenas mostrando os segmentos conectados.

3.4.3 Tensor Estrutural

Para quantificar o padrão de organização das fibras de colágeno nas imagens de segundo harmônico das biópsias de câncer de mama em cadelas foi usado o conceito de tensor estrutural [78]. Essa é uma metodologia já bem consolidada na literatura no uso de análise de imagens SHG [36,73,79-81]. As definições para o tensor estrutural serão discutidas com base em imagens de SHG contendo fibras de colágeno, mas essa metodologia de análise de imagem pode ser aplicada também em imagens de TPEF.

Consideremos agora uma imagem de SHG como uma função bidimensional de intensidade, $f(x, y)$, em uma escala de cinza. Se as fibras de colágeno estão orientadas ao longo de uma direção preferencial, a variação de intensidade do sinal SHG mais forte, será ao longo da direção ortogonal a orientação das fibras. O tensor estrutural é uma matriz 2×2 , para uma imagem 2D, com derivadas parciais ao longo de x e y , e que contém informações sobre a orientação de todos os pixels da imagem [36,73,79-81]. O tensor estrutural é definido como:

$$\bar{J} = \begin{bmatrix} \langle f_x, f_x \rangle_w & \langle f_x, f_y \rangle_w \\ \langle f_x, f_y \rangle_w & \langle f_y, f_y \rangle_w \end{bmatrix},$$

onde f_x e f_y , são as derivadas parciais espaciais da imagem $f(x, y)$, ao longo de x e y , respectivamente; elas contêm informações sobre a orientação de todos os pixels na imagem [36,73,79] e o produto interno é definido como:

$$\langle g, h \rangle_w = \iint_{R^2} w(x, y) g(x, y) h(x, y) dx dy, \quad (3.3)$$

onde $w(x, y)$ é uma função que especifica resolução espacial da área de interesse.

A derivada é maximizada ao longo da direção que corresponde ao autovalor máximo da matriz do tensor estrutural, λ_{max} , e a direção ortogonal corresponde ao autovalor mínimo dessa matriz, λ_{min} , isto nos permite obter a direção de alinhamento das fibras [79]. Os autovalores máximos e mínimos do tensor estrutural, $\bar{J}$, e seus correspondentes podem ser usados para medir o grau de isotropia da imagem, isto é, a organização estrutural das fibras na imagem e os ângulos de orientação das fibras.

O ângulo de orientação das fibras, θ , na região de interesse da imagem é dado por:

$$\theta = \frac{1}{2} \arctan \left(2 \frac{\langle f_x, f_y \rangle_w}{\langle f_x, f_x \rangle_w - \langle f_y, f_y \rangle_w} \right). \quad (3.4)$$

Como resultado da extração dos ângulos de orientação das fibras, o PyFibre gera um mapa de cores da distribuição do ângulo da fibra no tecido como mostrado na figura 3.9

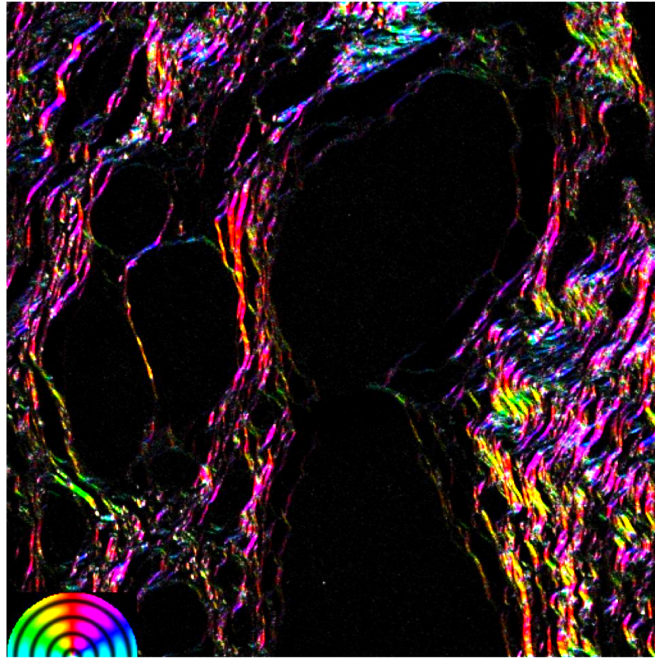


Figura 3.9: Mapa de cores com a distribuição do ângulo da fibra no tecido de hiperplasia mamária de células colunares gerados pelo cálculo do tensor estrutural (mapa de cores do ângulo no lado esquerdo inferior) com o uso do programa PyFibre. As cores indicam a orientação preferencial das fibras.

O grau de isotropia, também conhecida como coerência, é definido como a razão entre a diferença e a soma dos autovalores do tensor [36, 73],

$$\begin{aligned}
 C &= \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} = \\
 &= \frac{\sqrt{(\langle f_y, f_y \rangle_w - \langle f_x, f_x \rangle_w)^2 + 4 \langle f_x, f_y \rangle_w}}{\langle f_y, f_y \rangle_w + \langle f_x, f_x \rangle_w}.
 \end{aligned} \tag{3.5}$$

O grau de isotropia pode ser calculado em uma escala global ou local e os valores variam de 0 a 1. Quando as fibras estão distribuídas aleatoriamente o valor do grau de isotropia é igual a 0 e quando as fibras estão orientadas na mesma direção o valor do grau de isotropia é igual a 1.

3.4.4 Parâmetros Extraídos

Ao todo, utilizando o PyFibre calculamos um conjunto de 29 métricas, entre métricas de regiões de fibra e de células. Essas métricas podem ser agrupadas baseado na forma como elas foram extraídas que são a rede de fibras, o segmento de fibras e o segmento celular e um parâmetro global. Assim, as análises obtidas para as propriedades das fibras e das regiões celulares podem ser trabalhadas para cada segmento individualmente, além de o software gerar um banco de dados com as mesmas 29 métricas de forma global, isto é, relacionados a imagem inteira, são eles: No. de Fibras e de Segmentos Celulares,

Ondulação da Fibra, Comprimento da Fibra, Ângulo SDI (Índice de Desvio Padrão) da Fibra, Grau da Rede, Autovalor da Rede de Fibras, Conectividade da Rede, Densidade de Cross-link, Área do Segmento de Fibras e Celular, Área de Cobertura de Fibras e Celular, Forma do Segmento de Fibras e Celular, Excentricidade do Segmento de Fibras e Celular, Ângulo SDI do Segmento de Fibras e Celular, Grau de Isotropia do Segmento de Fibras e Celular, Grau de Isotropia Local do Segmento de Fibras e Celular, Média da Intensidade do Sinal de SHG e TPEF e STD do Segmento de Fibras e Celular. No apêndice [B](#) descreveremos cada uma das métricas e suas funções.

3.5 Análise de Variância

Para avaliar o conjunto de dados obtido pelas análises do PyFibre utilizamos um teste de hipótese para comparar se as distribuições dos dados obtidos possuem diferenças reais ou apenas ocasionadas pelo acaso. Um dos principais valores fornecidos pelo teste é o valor p. Esse valor indica a probabilidade de considerar que as amostras são diferentes quando na realidade elas são iguais. De uma forma geral, considera-se que um valor p menor que 0,05 é o aceitável para rejeitarmos que as distribuições são iguais. Aplicamos primeiro a análise de variância multivariada (MANOVA), comparando as distribuições em função de todas as métricas. Caso encontremos que o valor p menor do que 0,05 prosseguimos para a aplicação da análise de cada métrica comparando os subtipos histológicos par a par. Para essa análise utilizamos a análise de variância unilateral (ANOVA one-way). Dessa forma conseguimos comparar as distribuições obtidas para os grupos de amostras e determinar se as diferenças observadas entre as distribuições são devidas as modificações sofridas pelo tecido ou apenas referentes ao acaso. Essas análises e a apresentação dos resultados em gráficos de caixa (boxplot) foram realizadas utilizando as rotinas em Python. A escrita dos scripts de análise foram feitas com a colaboração de Leonardo Ferreira Calazans.

3.6 LDA

A fim de estudarmos todo o conjunto de dados gerado pelo PyFibre, avaliando não apenas as 29 métricas para um dado tipo histológico, mas também fazer uma comparação entre esses tipos, utilizamos um método de redução de dimensionalidade e classificação de dados onde podemos avaliar se todos os parâmetros obtidos na análise das imagens nos permitem separar os diagnósticos. O método de análise que escolhemos para essa parte do trabalho foi a Análise de Discriminante Linear (Linear Discriminant Analysis - LDA), já utilizado em alguns trabalhos com tecidos de mama [\[82,83\]](#). O LDA é um método de agrupamento de dados supervisionado, isto é, ele leva em consideração os rótulos das classes (neste trabalho os diagnósticos dados pelo patologista).

Como técnica de redução da dimensão dos dados, o LDA, procura remover parâmetros muito correlacionados e redundantes, projetando os dados em um espaço de dimensão

inferior de forma a maximizar a separação entre as classes [84–86]. Para essa separação o primeiro ponto é montar uma matriz com todos os dados fornecidos pelo programa, da forma [87]:

$$\begin{bmatrix} P_1 A_1 & P_2 A_1 & \cdots & P_n A_1 \\ P_1 A_2 & P_2 A_2 & \cdots & P_n A_2 \\ \vdots & \vdots & \ddots & \vdots \\ P_1 A_m & P_2 A_m & \cdots & P_n A_m \end{bmatrix}, \quad (3.6)$$

onde P_n é o parâmetro calculado para a amostra A_m . É importante notar que os parâmetros medidos devem ser reescalados de forma a termos uma distribuição gaussiana entre -1 e 1, além de montarmos uma matriz com os rótulos, diagnósticos, fornecidos pelo patologista:

$$\begin{bmatrix} \text{Diag } 1 \\ \text{Diag } 2 \\ \vdots \\ \text{Diag } N \end{bmatrix}. \quad (3.7)$$

Depois desse pré-processamento dos dados calculamos a matriz de dispersão de separabilidade dos dados dentro das próprias classes, isto é a distância entre a média e a amostra em cada classe, que também é chamada de variância dentro da classe [87,88]:

$$S_d = \sum_{i=1}^c \sum_{j=1}^n (\bar{P} - \bar{m}_i)(\bar{P} - \bar{m}_i)^T, \quad (3.8)$$

onde $\bar{m}_i$ é o vetor médio das amostras nas diferentes classes (diagnósticos) e c é o número de classes no problema.

Queremos também calcular a matriz de dispersão dos dados entre classes, isto é, a distância entre a média das diferentes classes, que também é chamada de variância entre as classes, e é dada por [87,88]:

$$S_e = \sum_{i=1}^c N_i (\bar{m}_i - \bar{m})(\bar{m}_i - \bar{m})^T, \quad (3.9)$$

onde $\bar{m}$ é a média geral e N_i é o tamanho das diferentes classes (diagnósticos).

O próximo passo é calcular os autovetores e autovalores da matriz $S_e S_d^{-1}$ [87]. Os autovetores associados aos maiores autovalores, que possuem as informações mais relevantes, irão indicar os novos eixos com a dimensionalidade dos dados reduzida e a melhor separabilidade das classes.

O último passo é então a projeção dos dados nesse novo espaço, no qual os novos eixos são os maiores autovetores. Essa projeção é chamada de critério de Fisher [88].

$$A_{lda} = \arg \max_A \left| \frac{A^T S_e A}{A^T S_d A} \right|. \quad (3.10)$$

Dessa forma esperamos um resultado como esquematizado na figura [3.10](#):

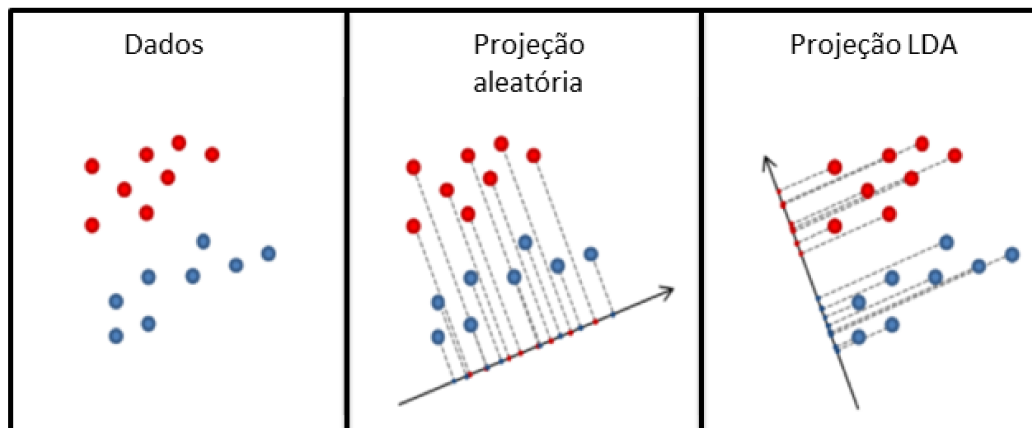


Figura 3.10: Ilustração da projeção de um conjunto de dados feita em um espaço com dimensão menor (no caso uma linha), tanto para uma projeção direta dos dados e para uma projeção com o LDA. Figura extraída e modificada do site <https://nirpyresearch.com/classification-nir-spectra-linear-discriminant-analysis-python/> no dia 17 de agosto de 2021.

Nesse trabalho, o objetivo de utilizar o LDA é transformar o espaço de 29 métricas extraídas pelo software, que formam um espaço de entrada de 29 dimensões, em um espaço de dimensão reduzida de forma a maximizar a variância entre os diagnósticos e minimizar a variância dentro de cada diagnóstico sem perder informações. Os novos eixos que maximizam a separação das classes será dado pelos maiores autovetores associados aos maiores autovalores da matriz do espaço de entrada com as 29 métricas fornecidas pelo PyFibre.

O LDA pode ser utilizado como método de aprendizado de máquina (do inglês machine learning) para fazer previsões em amostras. Como dito anteriormente o LDA leva em consideração os rótulos das classes para a separação, dessa forma podemos pensar que essa técnica não é interessante para classificação de dados, porém o LDA é treinado apenas por uma parte do conjunto de dados, o chamado conjunto de treino, e testado por um outro conjunto de dados, o conjunto de teste. O conjunto de teste é então usado para verificar a precisão da classificação, isso geralmente é feito comparando a variância e a covariância dos dados em cada um dos grupos conhecidos, assim, naquele dado de treino que possui a maior covariância e a menor variância existe a maior possibilidade de classificá-lo como pertencente aquele grupo [\[89\]](#). Geralmente a separação feita para o conjunto de treino e teste é tal que cerca de 80% dos dados formem o conjunto de treino e 20% o conjunto de teste, com a escolha do conjunto de dados feita aleatoriamente [\[90\]](#). Após a separação do conjunto total de dados utilizamos o método estratificado de validação cruzada K-fold(CV), onde o conjunto de treino é dividido em K conjuntos menores de forma que

$K - 1$ “*fold*” é usado como treino e o restante dos dados é usado como teste, como podemos ver na figura 3.11. A medida da precisão é dada pela estratificado de CV, e é a média dos valores calculados nesse loop [91]. Vale ressaltar que para nosso trabalho a divisão feita nos dados foi feita separando todo o conjunto de dados em 6 “*folds*” com 10 repetições.

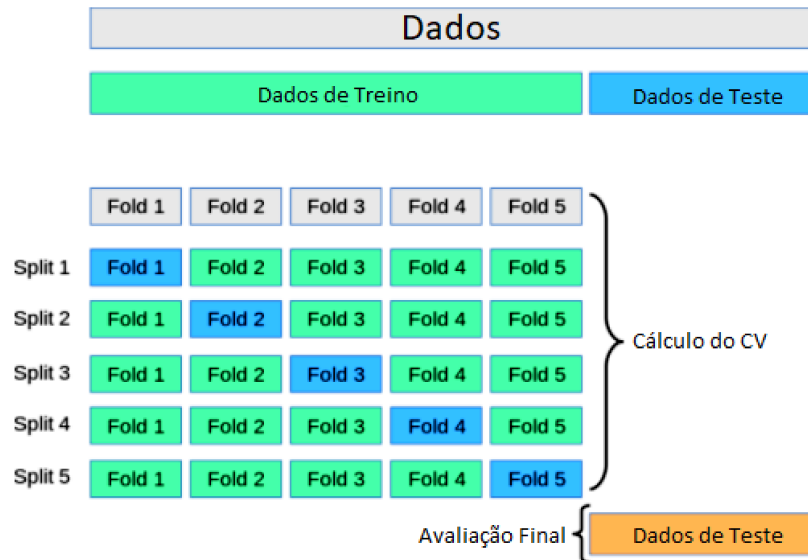


Figura 3.11: Diagrama de aplicação do K-fold. Figura extraída e modificada do site https://scikit-learn.org/stable/modules/cross_validation.html no dia 17 de agosto de 2021.

Também para obter as métricas mais relevantes nos modelos LDA, aplicamos uma eliminação recursiva de parâmetros com validação cruzada (RFECV). Esse método seleciona o melhor conjunto de parâmetros, removendo os parâmetros mais fracos, que nos dá uma maior precisão para os dados de teste [92].

Para os dados de câncer de mama utilizamos os métodos descritos acima para investigar as 29 métricas gerados pelo PyFibre. Na estratificação CV [91], todo o conjunto é dividido 6 vezes em conjuntos de dados de teste e treinamento, considerando uma representação proporcional de cada tipo de câncer e RFECV [92] para manter apenas os recursos relevantes em nossos modelos de LDA. Essas análises foram obtidas utilizando rotinas em Python. Os scripts de análises estão disponíveis no github A.

Capítulo 4

Resultados e Discussão

Neste capítulo apresentamos os resultados obtidos através das técnicas de microscopia óptica não linear (SHG e TPEF), bem como os resultados obtidos pelo processamento dessas imagens usando o software PyFibre, para os diferentes tipos de neoplasias em mama de cadela. Apresentamos também os gráficos de caixa (BoxPlot) e as análises estatísticas do teste de hipótese (MANOVA e ANOVA) de algumas métricas selecionadas para o estudos de características do tecido mamário. No apêndice desta tese apresentamos o restante dos resultados obtidos em gráfico de caixa.

Os resultados estão separados em duas seções. No primeiro conjunto de dados, referente a seção 4.1 utilizamos apenas 3 tipos histológicos e a mama normal para avaliarmos as modificações sofridas pelo tecido, utilizando o software de análise de imagens que esta descrito na seção 3.4 de Metodologia Experimental. Foi nessa primeira etapa de desenvolvimento do trabalho que o PyFibre foi desenvolvido, assim, além dos resultados obtidos após a confecção do software, também utilizamos algumas dessas imagens para a própria construção do software, além de imagens já estudadas de tecido de próstata, parte de outros trabalhos realizados pelo nosso grupo que podem ser encontradas em 21.

Como sequência do trabalho fizemos comparações entre os tipos tumorais já apresentados e mais três novos tipos referente a seção 4.2. Estes resultados foram também utilizados com o objetivo final de comparar os dados coletados e analisados utilizando o PyFibre e a sobrevida dos animais, que pode ser encontrado no artigo 31. Nesse estudo apenas as amostras de mama normal foram as mesmas utilizadas no estudo anterior, os demais tipos tumorais que coincidem (carcinoma em tumor misto e carcinoma sólido) foram coletados de lâminas e animais diferentes.

4.1 Análise de Subtipos Histológicos - *In Situ*, Misto e Sólido

Ao todo foram analisadas, nessa primeira etapa do trabalho, 55 casos (pacientes), divididos em 13 casos de tecido de glândula mamária normal medidos em 20 lâminas histológicas com 100 imagens coletadas, 12 casos de carcinoma ductal *In Situ* medidos em

17 lâminas histológicas com 78 imagens coletadas, 11 casos de carcinoma em tumor misto medidos em 17 lâminas histológicas com 128 imagens coletadas e 19 casos de carcinoma sólido medidos em 24 lâminas histológicas com 110 imagens coletadas.

4.1.1 Imagem Óptica, SHG e TPEF

Na figura 4.1 temos uma composição das imagens obtidas pelas técnicas experimentais descritas na seção 2.1. Cada uma das imagens foram escolhidas pelo patologista de forma a melhor representar os diagnósticos estudados (Normal, *In Situ*, Misto e Sólido) e organizadas de acordo com a agressividade do tipo tumoral.

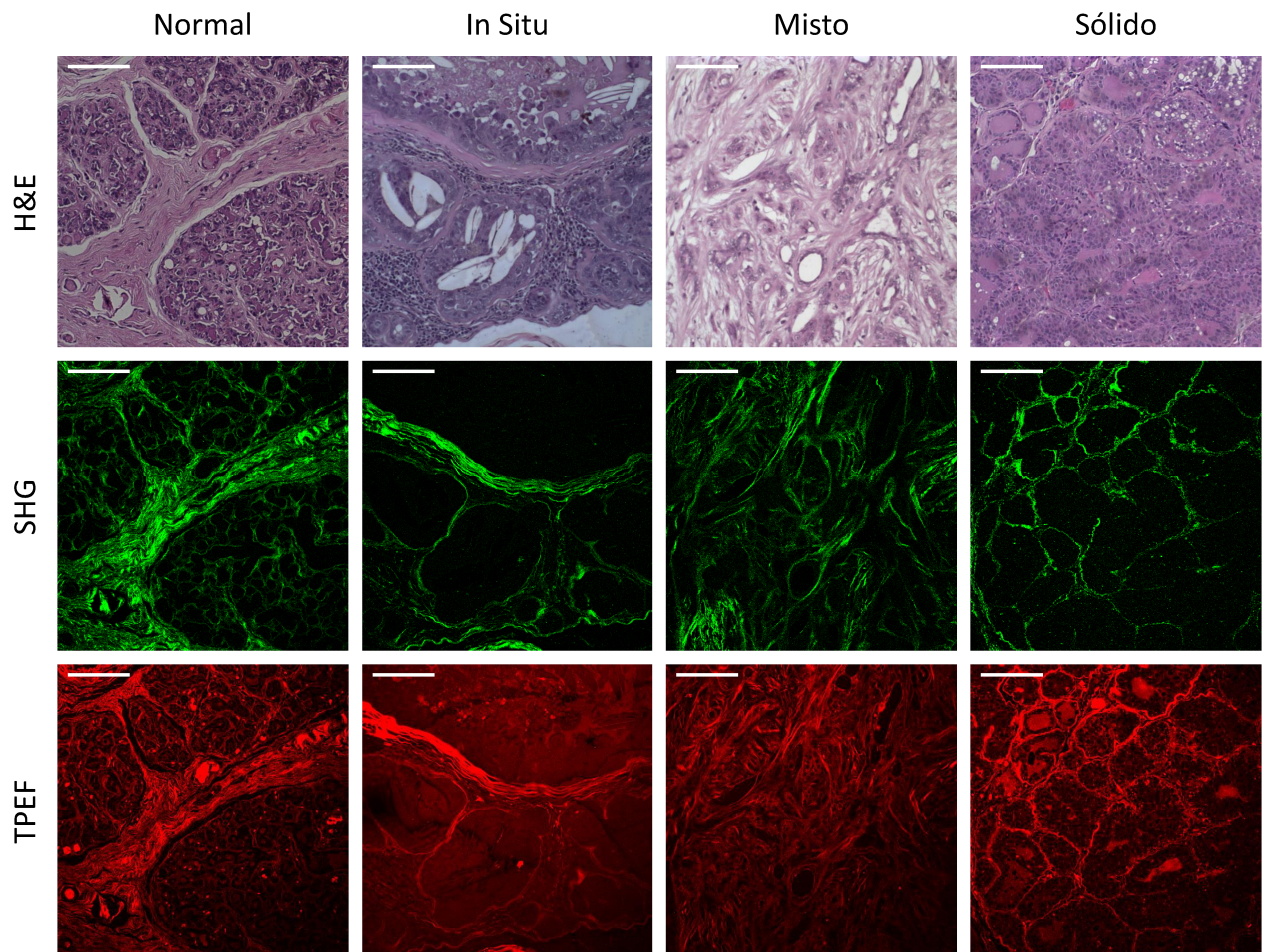


Figura 4.1: Figuras com as imagens obtidas por microscopia de campo claro (primeira linha) coradas com H&E, imagens em escala de cinza de microscopia de SHG coradas digitalmente (segunda linha) e imagens em escala de cinza de microscopia por TPEF coradas digitalmente (terceira linha) com os diagnósticos estudados organizados pelas colunas (Normal, *In Situ*, Misto e Sólido, respectivamente). A barra da escala é de 100 μm .

Na primeira linha temos as imagens de campo claro obtidas das lâminas histológicas coradas com H&E que foram utilizadas no procedimento padrão de diagnóstico. Os pontos com coloração mais escura nas imagens são do corante hematoxilina, que marca o núcleo

celular e não possui fluorescência, já o restante do tecido é marcado pela eosina, que é um corante fluorescente. A segunda e terceira linha são as imagens de SHG e TPEF, respectivamente, obtidas nas mesmas regiões que as imagens de campo claro. É importante ressaltar que nas imagens de SHG temos apenas o sinal gerado pelas fibras de colágeno, que evidenciam algumas diferenças entre os tecido da glândula mamária para os diferentes tipos histológicos. Para o tecido de mama normal as fibras estão mais dispersas e de uma maneira mais desorganizada, porém com lóbulos e ácinos bem delimitados por feixes de fibras. No carcinoma *In Situ*, aparentemente, temos uma menor quantidade de fibras em relação a mama normal e, portanto, menos dispersas pelo tecido. Para o carcinoma em tumor misto, observamos que as fibras do tecido não possuem mais separação distinta entre ácinos e os lóbulos, não temos mais fibras contornando essas regiões de forma a bem separá-las. Já no carcinoma sólido, vemos uma quantidade de fibras muito menor em comparação com os outros tipos, observamos também o rompimento das fibras em algumas regiões, o que indica invasão celular. As imagens de TPEF, são geradas pela fluorescência da eosina. Assim, além de vermos o colágeno existente no tecido, conseguimos também ver o citoplasma das células.

4.1.2 Imagens geradas pelo PyFibre

As imagens de SHG e TPEF obtidas experimentalmente são processadas pelo software PyFibre. Em uma primeira análise mostramos um mapa de cores com a distribuição de ângulos (orientação) das fibras, que é construído utilizando o tensor de estrutura, e uma imagem com redes (networks) crescidas sobre as fibras de colágeno, mostradas na primeira e segunda linhas da figura 4.2, respectivamente, para os casos da figura 4.1. Pelo mapa de cores da distribuição angular das fibras, podemos avaliar de maneira visual a organização das fibras de colágeno no tecido. Para a mama normal temos uma imagem bem colorida, com poucas e pequenas regiões com cores dominantes, isso indica que as fibras do tecido de mama normal são mais desorganizadas, o que corrobora com o encontrado na literatura [13, 20, 25, 27, 34, 36, 67, 72, 73, 79, 93–99]. Para os demais tecidos observamos que as fibras possuem uma direção preferencial, isso pode ser visto pelo fato de que para os tecidos neoplásicos temos sempre uma cor que se sobressai na imagem, o que indica um maior alinhamento das fibras de colágeno nos tecidos tumorais.

Outras imagens geradas pelo software de tratamento de dados apresentadas na figura 4.2 são as imagens de segmentos de fibras e de segmentos celulares, terceira e quarta linhas, respectivamente. As imagens de segmentos de fibras são apresentadas sobrepostas às imagens em tons de cinza do SHG e as imagens de segmentos celular são apresentadas sobrepostas às imagens em escala de cinza do TPEF. Cada um desses segmentos é apresentado com cores aleatórias que identificam cada um dos diferentes segmentos. Das imagens de segmentos de fibras, observamos que para o tecido de mama normal o número de segmentos de fibras é maior do que para os tecidos neoplásicos. Um detalhe interessante que conseguimos observar é que tanto o carcinoma em tumor misto quanto o carcinoma

sólido apresentam uma ruptura em alguns dos segmentos de fibras o que indica áreas de invasão celular no tecido, assim temos um segmento de célula maior nessas regiões.

Na última linha da figura 4.2 temos a extração de fibras individuais. Ao observarmos essas imagens podemos perceber que o tecido de mama normal possui mais fibras individuais e fibras menores em comparação aos demais tecidos neoplásicos. Observamos essa diferença, principalmente, entre a mama normal e o carcinoma sólido que é tipo tumoral mais agressivo.

Com a segmentação da imagem, separando de forma clara e automática a região de fibras da região de células (que inclui o lúmen) e com a extração de fibras individuais, são gerados em conjunto um banco de dados que será usado afim de avaliar especificamente parâmetros de interesse para cada uma das regiões e para as fibras individuais. É importante ressaltar que as cores geradas pelo processamento das imagens com o software são aleatórias e são apenas para identificar cada segmento.

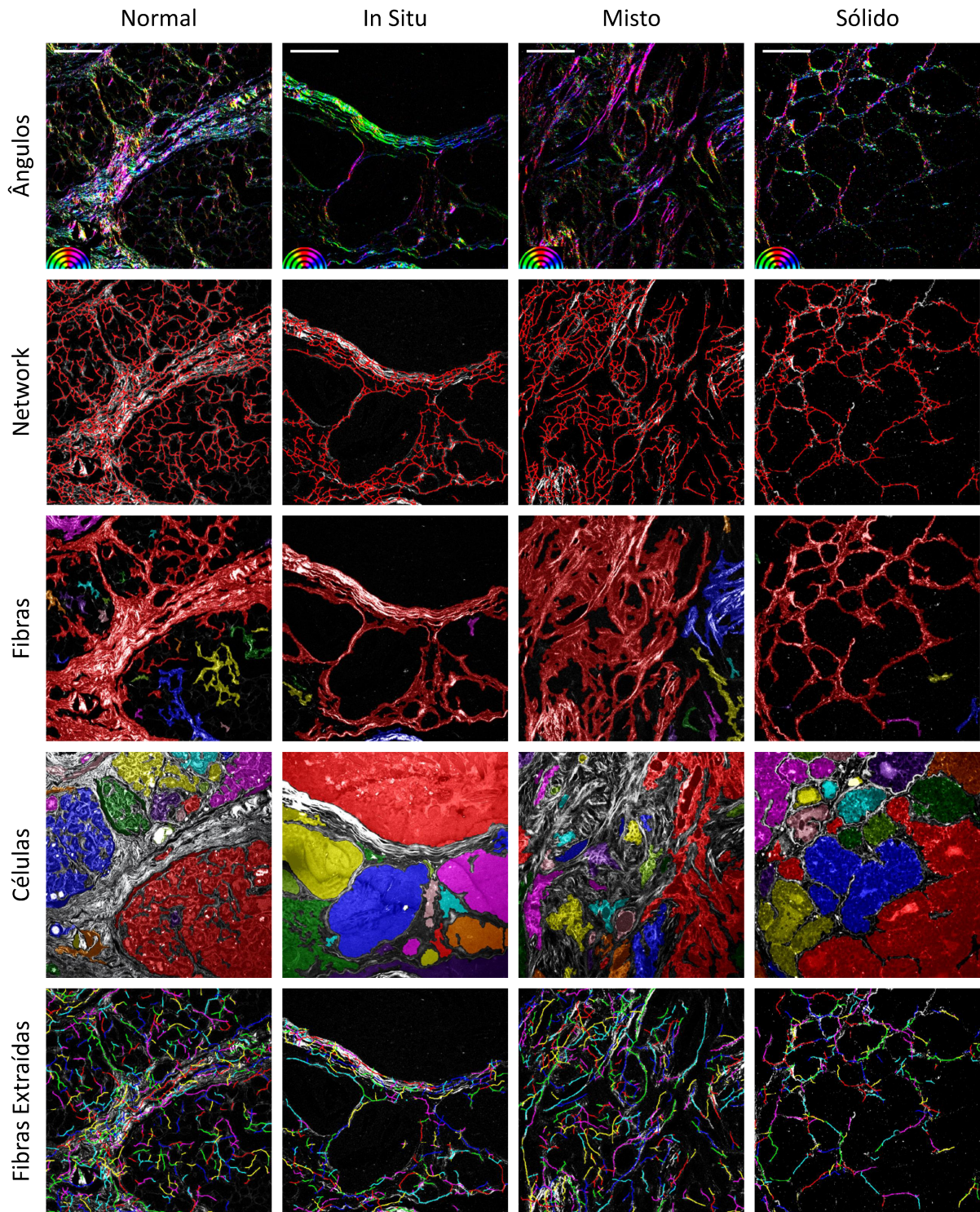


Figura 4.2: Imagens geradas pelo PyFibre para os diagnósticos estudados organizados pelas colunas (Normal, *In Situ*, Misto e Sólido, respectivamente). Mapa de cores com a distribuição do ângulo da fibra no tecido (primeira linha), rede de fibras extraídas, crescidas sobre a imagem de SHG (segunda linha), segmento de fibras e segmento celular (terceira e quarta linha, respectivamente) com as cores geradas de forma aleatória são construídos sobrepostos com a imagem de intensidade de SHG em escala de cinza e extração de fibras únicas (última linha) também apresentadas em cores aleatórias. A barra da escala é de 100 μm .

4.1.3 Resultado das Análises Estatísticas

Nesta seção apresentamos os resultados obtidos através de algumas das métricas que podem ser extraídas das imagens de SHG, TPEF e Transmissão que foram tratadas pelo software PyFibre para a análise do tecido da glândula mamária. Com base nas métricas separadas podemos identificar algumas que nos ajudam na diferenciação entre os tipos de tecido de mama. Aqui apresentaremos apenas algumas métricas geradas pelo PyFibre, estas métricas foram selecionadas com base nos artigos publicados na literatura [13-21, 25, 27, 34-36, 67, 72, 73, 79, 93-99], em parâmetros analisados pelo patologista ao fornecer um diagnóstico (como por exemplo a organização do tecido no entorno do tumor e a proliferação celular) e algumas métricas que não haviam sido apresentadas ainda na literatura. As demais métricas extraídas pelo software estão apresentadas no apêndice [C](#).

Para quantificar as modificações ocorridas no tecido de mama apresentamos os resultados utilizando gráficos em caixa (Boxplot). O Boxplot nos fornece uma análise visual de características importante dos dados, tais como a dispersão, a mediana e assimetria dos dados. O tamanho da caixa é chamado de interquartil e os dados nos gráficos são divididos em quatro partes, o quartil. O primeiro quartil é marcado pela linha localizada no canto inferior da caixa, onde aproximadamente 25% dos dados estão abaixo dela, o segundo quartil é evidenciado pela mediana, que é a linha vermelha traçada dentro da caixa e marca o valor no qual 50% dos dados estão abaixo dele. Já o terceiro quartil é a linha localizada no canto superior da caixa, onde aproximadamente 75% dos dados estão abaixo dela. O local onde a linha vertical começa (de baixo para cima) indica o valor mínimo do conjunto de dados dentro de uma faixa de 1,5 interquartil e onde a linha termina indica o valor máximo do conjunto de dados dentro da faixa de 1,5 interquartil. Os pontos fora da caixa são chamados de *outliers*, os *outliers* indicam possíveis valores discrepantes [100].

Para comparar os resultados obtidos entre os diferentes tipos de tecidos neoplásicos, aplicamos, primeiramente, o teste MANOVA. Com ele comparamos as distribuições das amostras de mama normal e das amostras de tumores mamários para todas as métricas. Como resultados tivemos um valor $p < 0,001$. Como o resultado foi significativo, passamos para o teste ANOVA one-way, onde comparamos cada distribuição das amostras por cada métrica. Dessa forma, conseguimos estudar as alterações no tecido para cada métrica específica. Os resultados desse teste e gráficos em caixa com essas distribuições serão apresentados a seguir. Um valor de $p < 0,05$ é considerado pela literatura estatisticamente significativo, isto é, um p menor que esse valor nos indica que os dados não vieram de uma mesma distribuição. Neste trabalho, valores de p menores que 0,001 são denotados por (***), p menor que 0,01 são denotados por (**) e p menor que (0,05) são denotados por (*).

Comparação Normal x Câncer

Como uma primeira análise das amostras separamos elas em dois grupos, como vemos na figura 4.3, amostras de mama normal e amostras de tecidos neoplásicos, que são compostas pelos tecidos de carcinoma *In situ*, carcinoma em tumor misto e carcinoma sólido, e verificamos quais das métricas as separam de forma mais simples, dessa forma temos uma primeira verificação dos dados.

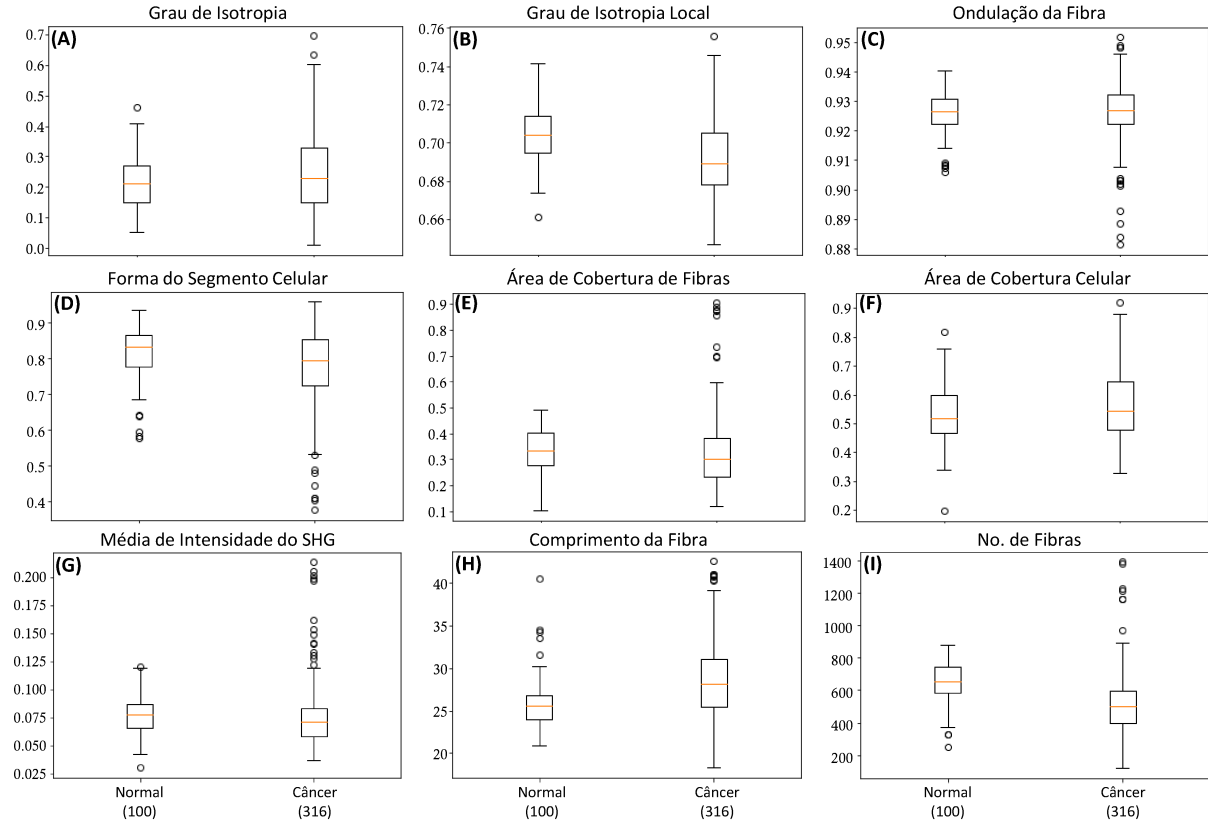


Figura 4.3: Gráficos de boxplot para as métricas indicadas nos títulos de cada gráfico para os diagnósticos de tecidos mamários normal e tumoral. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

Grau de Isotropia: O grau de isotropia dos feixes de fibras é uma medida do grau de alinhamento das fibras de colágeno no tecido mamário na imagem inteira, ela é calculada a partir do tensor de estrutura, definido em 3.4.3. O resultado final para cada imagem é dado pela média dos valores de alinhamento de cada segmento de fibra extraído pelo software. Esse valor é dado entre 0 para um arranjo de fibras completamente desorganizado, isto é áreas isotrópicas, e 1 para um arranjo de fibras completamente organizado, isto é áreas anisotrópicas.

Pelo gráfico (A) da figura 4.3 vemos que, de uma forma geral, as fibras de colágeno nas neoplasias são mais organizadas, mais alinhadas, do que as amostras de mama normal. Isso pode ser visto pelo aumento no valor da mediana para as amostras de neoplasias da glândula mamária (0,228) em relação as amostras de mama normal (0,211), o que

corroborar com os dados descritos na literatura [24, 31, 36]. Para a comparação entre as amostras o valor p foi significativo, como podemos ver pela tabela 4.1.

Grau de Isotropia Local: O grau de isotropia local, assim como a métrica anterior, também é uma medida do grau de alinhamento das fibras, mas para essa métrica calculamos a média do tensor de estrutura para uma janela de 3x3 pixels nas regiões de segmentos de fibras, nos dando assim um grau de isotropia local.

Pelo gráfico (B) da figura 4.3 observamos uma queda no valor da mediana para as amostras neoplásicas (0,689) em relação as amostras de mama normal (0,704). O valor p para essa comparação foi significativo como podemos ver na tabela 4.1.

A diferença no comportamento dos gráficos de Grau de Isotropia e Grau de Isotropia Local, onde ambos são calculados da mesma maneira, está no fato de usarmos uma janela de 3x3 pixels, e portanto, avaliarmos a métrica anisotropia de forma local. Dessa forma, como o tensor avalia os pixels pela diferença de intensidade, quando as fibras de colágeno estão, globalmente, dispersas pelo tecido, como é o caso da mama normal, ao avaliarmos uma janela de 3x3 pixels temos pouca variação na intensidade entre um pixel e seu vizinho, fazendo com que tenhamos valores maiores, e portanto, o que veremos é uma maior organização local. Já para o caso de tecidos neoplásicos, onde existe uma maior organização das fibras em torno do tumor, a variação de intensidade entre um pixel e seus vizinhos pode ser muito maior, uma vez que ao escolhermos um pixel da fibra que esteja na borda do tumor, o pixel vizinho terá um valor de intensidade muito diferente do pixel da fibra, fazendo com que tenhamos valores mais baixos, e assim, o que veremos é uma maior desordem local.

Ondulação da Fibra: Esta métrica corresponde as métricas extraídas da rede de fibras, mais especificamente a média dos valores de ondulação das fibras únicas. Para calcular a ondulação das fibras extraídas, calculamos a razão entre a distância linear de um ponto a outro em uma fibra (dentro do tamanho da imagem) pelo comprimento total da fibras. Na figura 4.4 definimos L_0 , linha tracejada, como a distância ponto-a-ponto da fibra e L_1 , linha contínua, como o comprimento total do feixe que percorre toda a fibras.

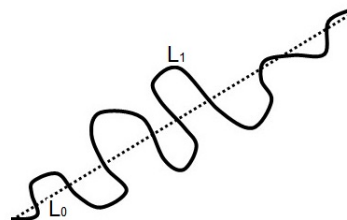


Figura 4.4: Definição dos parâmetros de comprimento pont-a-ponto da fibra, L_0 e comprimento total do feixe, L_1 , linhas tracejada e contínua, respectivamente.

Assim a ondulação das fibras é dada por:

$$P_s = \frac{L_0}{L_1}. \quad (4.1)$$

O P_s está limitado entre 0 e 1, onde quanto mais ondulada é a fibra menor o valor de P_s [73].

Pelo gráfico (C) da figura 4.3 observamos que não existe uma variação na distribuição das amostras tumorais em relação a distribuição para a mama normal, com os valores da mediana praticamente iguais para ambos os tipos, (0,927) para as amostras neoplásicas e (0,926) para as amostras de tecido normal.

Forma do Segmento Celular: Esta métrica indica o quão circular ou alongada é o segmento celular, ou seja, a “forma” do segmento celular. Ela é obtida considerando a razão entre a circunferência de um círculo com a mesma área do segmento celular e o perímetro deste segmento celular. Desta forma, avaliamos que para valores próximos de 0 o segmento celular é completamente circular e em 1 o segmento é completamente linear. Assim, para a imagem toda, a forma do segmento é a média de todos os segmentos extraídos.

Pelo gráfico (D) da figura 4.3 observamos uma queda no valor da mediana das amostras de tecido neoplásicos (0,794) em relação a mama normal (0,832) e a comparação entre as amostras nos deu um valor p significativo, como podemos ver pela tabela 4.1. O que esperaríamos para essa métrica se observamos a do segmento celular pela figura 4.1 era que as amostras de tecido normal tivessem um valor que indicassem uma maior circularidade do que as amostras de tecido neoplásicos, porém o valor dessa métrica para esse grupo sugere que eles são mais alongados. Uma explicação para isso pode ser vista na figura 4.2, embora visualmente as regiões celulares da imagem do tipo normal sejam relativamente circulares, os segmentos celulares identificados pelo software de análise possuem muitas irregularidades que levam a vários contornos, devido a presença das estruturas de colágeno dentro das regiões detectadas como celulares. Esses contornos levam a um valor muito maior para o perímetro do segmento do que seria esperado visualmente. Já para os tecidos neoplásicos as regiões celulares exibem menos estruturas de colágeno em seu interior e, portanto, os segmentos celulares são mais amplos e não possuem muitos contornos.

Área de Cobertura de Fibras: Cálculo da área de cobertura de fibras na imagem. É determinado pela proporção de pixels do segmento de fibras contido na imagem dividido pela área total da caixa delimitadora da imagem.

Vemos pelo gráfico (E) da figura 4.3 uma queda no valor da mediana para as neoplasias mamárias (0,303) em relação ao valor da mediana para as amostras de mama normal (0,332), o que já era esperado, visto que uma característica dos tumores é ser mais celular e com um tecido menos fibroso [11,101], porém aqui não tivemos um valor p significativo, como pode ser visto na tabela 4.1.

Área de Cobertura Celular: Cálculo da área de cobertura do segmento celular na imagem. É determinado pela proporção de pixels do segmento celular contido na imagem dividido pela área total da caixa delimitadora da imagem.

Como esperado, essa métrica mostra uma tendência oposta a da área de cobertura de fibras, embora isso não seja garantido pelo processo de segmentação. Pelo gráfico (F) da figura 4.3 podemos observar um aumento no valor da mediana na área coberta pelo segmento celular para os tecidos neoplásicos (0,545) em relação ao valor da mediana das amostras de mama normal (0,519), o que era esperado [11, 101]. Apesar de a diferença não ser aparentemente significativa, temos um valor p significativo para essa comparação como visto na tabela 4.1.

Média da Intensidade do SHG: Para essa métrica determinamos uma estimativa da densidade de fibras pelo cálculo da razão entre a intensidade do sinal SHG na imagem pela intensidade na área total da imagem, porém esses sinais são normalizados no pré-processamento da imagem.

Pelo gráfico (G) da figura 4.3 observamos uma queda no valor da mediana para as amostras tumorais (0,071) em relação a mama normal (0,077) o que já era esperado, visto que tecidos tumorais possuem uma área coberta por fibras menor do que a mama normal, porém essa comparação não teve um valor p significativo, como pode ser observado na tabela 4.1.

Comprimento da Fibra: Para essa métrica calculamos o comprimento médio de cada fibra extraída (consideradas como fibras individuais) pelo software. Essa métrica é calculada conforme a distância euclidiana percorrida ao longo da fibra extraída.

Pelo gráfico (H) da figura 4.3 observamos um aumento médio no tamanho das fibras para as amostras neoplásicas (28,12) em relação ao tamanho da fibra para a mama normal (25,52). Para essa métrica tivemos um valor p significativo como visto na tabela 4.1.

No. de Fibras: Aqui computamos o número médio de fibras individuais extraídas para cada imagem.

Pelo gráfico (I) da figura 4.3 observamos uma queda na quantidade de fibras individuais extraídas nas amostras neoplásicas (497,5) em relação as amostras de mama normal (653,0), isso está de acordo com o que vimos na figura (E), uma vez que tendo uma área coberta por fibras para as neoplasias mamárias é menor do que para a mama normal é de se esperar que o número de fibras extraídas pelo programa também seja menor. Tivemos um valor p significativo para essa métrica como visto na tabela 4.1.

De maneira geral, 6 das 9 métricas selecionadas separam bem amostras neoplásicas de amostras de mama normal, mesmo considerando o fato de que para as amostras tumorais temos 3 tipos distintos de neoplasias mamárias, temos dessa forma uma primeira separação muito boa. Podemos agora fazer uma avaliação entre os diagnósticos.

Comparação Tipos Histológicos

Após a análise das diferenças entre o tecido da glândula mamária normal e neoplásica, separamos os grupos das neoplasias mamárias que foram utilizados nesta parte do trabalho e analisamos o comportamento de cada um dos grupos de amostras comparando-os com a mama normal e cada um dos tipos entre sim. Novamente utilizamos o teste MANOVA. Com ele comparamos as distribuições das amostras de mama normal e os três diferentes tipos tumorais (*In Situ*, Misto e Sólido) simultaneamente para todas as métricas. Além da comparação entre cada dois diferentes tipos tumorais para todas as métricas. Como resultados para todas as avaliações do MANOVA, tivemos um valor $p < 0,001$. Como os resultados foram significativo, passamos para a o teste ANOVA one-way, onde comparamos as distribuições das amostras par a par para cada métrica. Dessa forma analisamos as diferenças significativas para cada tipo tumoral.

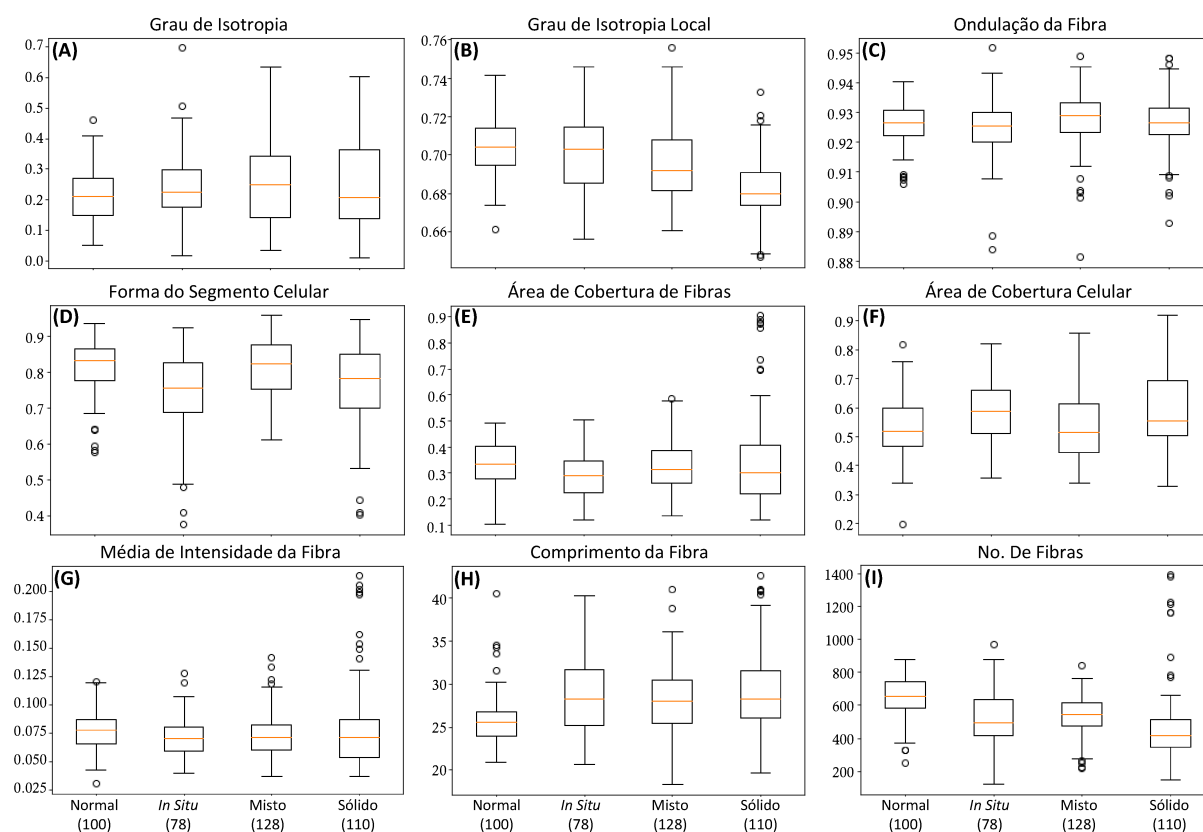


Figura 4.5: Gráficos de boxplot para as métricas indicadas nos títulos dos gráficos para os diagnósticos: Normal, carcinoma *In Situ*, carcinoma em tumor misto e carcinoma sólido. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

Grau de Isotropia: Pelo gráfico (A) da figura 4.5 observamos um leve aumento no valor da mediana para as neoplasias mamárias (exceto para o carcinoma sólido (0,207)) em relação a mama normal (0,210), porém como vemos na tabela 4.1 apenas a comparação entre Normal x Misto possui um valor p significativo. Provavelmente a razão para conse-

guirmos uma separação entre o carcinoma em tumor misto e a mama normal analisando a organização das fibras da imagem inteira é devido ao fato de conseguirmos, para uma dada região, observar tanto regiões de invasão e rompimento das fibras quanto regiões *in situ* (regiões de crescimento onde não houve rompimento de fibras e invasão no tecido).

Já para os demais diagnósticos, embora visualmente conseguimos observar uma maior organização das fibras, as fibras ainda tendem a se organizar em torno do lóbulo, como ocorre no carcinoma *In Situ* ou a neoplasia é de um tipo tumoral muito celular como é o caso do carcinoma sólido, onde neste trabalho só conseguimos fazer as medidas em regiões mais do entorno do tumor, ou quando conseguimos medir uma região com todo o tumor as fibras estão circundando o tumor. Ao calcularmos o tensor de estrutura apenas para as fibras e na imagem como um todo muitas das direções se cancelarão mutuamente.

Grau de Isotropia Local: Pelo gráfico (B) da figura 4.5, observamos um comportamento inverso ao comportamento do gráfico anterior, e principalmente, com uma queda acentuada no valor da mediana da anisotropia do carcinoma sólido (0,680) em relação ao da mama normal (0,704). Para essa métrica obtivemos valores p significativos para todas as comparações, como pode ser visto na tabela 4.1, exceto para a comparação entre o carcinoma *In Situ* e a mama normal.

Ondulação da Fibra: Pelo gráfico (C) da figura 4.5 observamos um aumento no valor da mediana do carcinoma em tumor misto (0,929) em relação a mama normal e aos demais tipos histopatológicos, principalmente o carcinoma *In Situ*. Essa variação pode ser observada pelo valor p para essa comparação, como visto na tabela 4.1.

Forma do Segmento Celular: Pelo gráfico (D) da figura 4.5 vemos uma queda para o valor da mediana dos diagnósticos histopatológicos em relação a mama normal (0,832), principalmente para o carcinoma *In Situ* (0,755) e o carcinoma sólido (0,783).

Ao observarmos a estrutura do tecido para o carcinoma *In Situ*, vemos que, apesar de não existir regiões de invasão o lóbulo se deforma ficando mais alongado e maior do que um lóbulo saudável. Já para o carcinoma sólido, como ele é um tipo tumoral onde o tecido é muito celular e as fibras que circundam o tumor foram em grande parte rompidas e assim os aglomerados de células não tem uma forma muito definida o valor dessa métrica é pequeno. Os valores p para as comparações entre os diferentes diagnósticos podem ser vistos na tabela 4.1.

Área de Cobertura de Fibras: Podemos perceber pelo gráfico (E) da figura 4.5 que existe uma queda na área de cobertura de fibras para as neoplasias mamárias em relação a mama normal, o que era esperado visto que os tumores possuem uma maior proliferação celular comprimindo assim o tecido extracelular em torno do tumor [11,101]. Pela tabela 4.1 podemos ver para quais comparações obtivemos um valor p significativo.

Área de Cobertura Celular: Aqui, observamos pelo gráfico (F) da figura 4.5 que a área de cobertura de células é maior para as neoplasias mamárias do que para a mama normal. De forma especial o carcinoma sólido apresenta um valor maior, o que era esperado, visto que ele é o tipo tumoral mais celular [11, 101]. Pela tabela 4.1 podemos ver para quais comparações obtivemos um valor p significativo.

Média de Intensidade do SHG: Pelo gráfico (G) na figura 4.5 percebemos uma queda nos valores da mediana para os diagnósticos tumorais em relação a mama normal, mesmo com os sinais normalizados no pré-processamento da imagem. Para o carcinoma *In Situ* temos o menor valor (0,070), o que corrobora com a descrição do gráfico (E) da figura 4.5, indicando assim que este tipo tumoral, apesar de não apresentar invasão celular também é um tipo tumoral menos fibroso e mais celular. Infelizmente para as amostras aqui trabalhadas apenas a comparação entre o carcinoma *In Situ* e a mama normal teve um valor p significativo, como visto na tabela 4.1.

Comprimento da Fibra: Pelo gráfico (H) da figura 4.5, observamos uma diferença muito grande o valor da mediana dos diagnósticos de neoplasias mamárias em relação a mama normal, e que esta possui um valor médio de comprimento de fibras bem menor em relação aos tecidos tumorais. Isso pode ocorrer devido ao fato de que nas neoplasias o crescimento celular faz com que as fibras do tecido se tornem mais estiradas ajudando assim em seu rompimento para a invasão celular [31, 96]. Os valores p significativos são mostrados na tabela 4.1.

No. de Fibras: Pelo gráfico (I) da figura 4.5 observamos uma queda significativa para o número médio de fibras extraídas para os diagnósticos de neoplasias mamárias em relação ao número médio de fibras para a mama normal. O carcinoma sólido possui o menor valor de mediana para o número médio de fibras extraídas (416), corroborando com o fato de que essa neoplasia mamária é a mais celular e, assim, a menos fibrosa. Os valores p significativos são mostrados na tabela 4.1.

Essa métrica também está de acordo com as demais que nos indicam a forma do tecido tumoral. Ela, juntamente com o comprimento da fibra (H) indica que mesmo para o caso onde as fibras estão mais alinhadas, o número de fibras é menor, isso pode estar diretamente ligado ao fato de o carcinoma sólido ser invasivo (o que alinha as fibras e as torna maiores) e muito mais celular (que faz com que o número médio de fibras seja menor).

Em resumo: O Grau de Isotropia Local foi a métrica com mais separações entre as comparações, visto que 5 das 6 comparações tiveram valor p significativo, seguido da Forma do Segmento Celular, da Área de Cobertura Celular e do No. de Fibras, cada um com 4 comparações com valor p significativo, como pode ser visto na tabela 4.1. Entre as comparações apenas para a *In Situ* x Sólido tivemos poucas métrica que tiveram um valor p significativo (Grau de Isotropia Local e Área de Cobertura de Fibras). Isso pode ocorrer

devido ao fato de que o carcinoma sólido é um tipo muito celular e por isso existe uma dificuldade em fazer as medidas com ele. Por isso métricas como a Forma do Segmento Celular e a Área de Cobertura Celular não possuem grande diferenças em seus valores de mediana.

Tabela 4.1: Tabela de valores p com a comparação das métricas entre pares de diagnósticos. A significância denotada por $p < 0,05$ são denotados como (*), $p < 0,01$ são denotados como (**), e $p < 0,001$ são denotados como (***)

Métricas	Normal X Câncer	Normal X <i>In Situ</i>	Normal X Misto	Normal X Sólido	<i>In Situ</i> X Misto	<i>In Situ</i> X Sólido	Misto X Sólido
Grau de Isotropia	*		**				
Grau d Isotropia Local	***		***	***	*	***	***
Ondulação da Fibra					*		
Forma do Segmento Celular	**	***		***	***		***
Área de Cobertura de Fibras		**			**	*	
Área de Cobertura Celular	**	***		***	*		**
Média de Intensidade do SHG		*					
Comprimento da Fibra	***	***	***	***			
No. de Fibras	***	***	***	***			***

4.1.4 LDA

Reunimos agora todos os dados gerados pelo PyFibre e fazemos a avaliação de classificação dos dados, apresentando a análise do LDA para as 29 métricas. Na figura [4.6](#), vemos o desempenho do modelo conforme variamos o número de métricas.

Embora as pontuações ótimas de CV possam ser relatadas para conjuntos de dados contendo menos de 29 métricas, a diferença geral na precisão entre os dados com o melhor conjunto de métricas e o conjunto das 29 métricas é insignificante, ($\pm 0,2\%$), comparada à variação entre os modelos treinados em diferentes “*fold*”. Portanto, mantemos o conjunto completo de métricas.

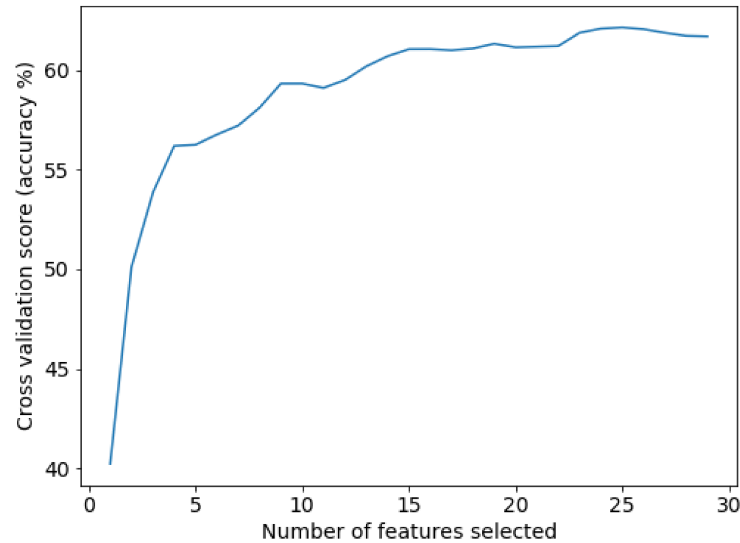


Figura 4.6: Validação cruzada para um modelo de LDA treinado usando dados do conjunto de treinamento K-fold, em relação ao número de métricas incluídas no modelo.

A figura 4.7 ilustra os resultados do LDA para os dois componentes mais discriminativos do conjunto de dados, isto é os dois maiores autovetores. Os círculos cheios são para o conjunto de treinamento e as cruzes são para o conjunto de teste. A media de cada grupo de caso é indicada pela estrela amarela e as elipses mostram um desvio padrão dos dados de treinamento. Podemos observar que a região do tecido mamário normal se separa bem das regiões neoplásicas, obtivemos uma precisão de $91,0 \pm 2,8\%$ de acerto com nossos dados, o que já era esperado pelos resultados apresentados nos gráficos do boxplot da figura 4.3 e pelos valores significativos de p apresentados na tabela 4.1. Para o carcinoma em tumor misto, a elipse mostra uma sobreposição parcial com as elipses do carcinoma *In Situ* e do carcinoma sólido. Isso é algo esperado, já que esses tumores podem exibir crescimento *In Situ* ou invasivo. No entanto, há também uma boa separação entre o carcinoma *In Situ* e o carcinoma sólido, o que não acontece se observamos para cada métrica de forma individual. Esse é um achado importante, pois a progressão do carcinoma *In Situ* para o carcinoma invasivo é uma passo importante na evolução do câncer de mama. Como esperado da sobreposição entre grupos de carcinoma, a precisão do nosso conjunto de dados de teste ao distinguir entre todos os diagnósticos, $61,0 \pm 5,0\%$, não é o ideal, no entanto é um bom valor 24.

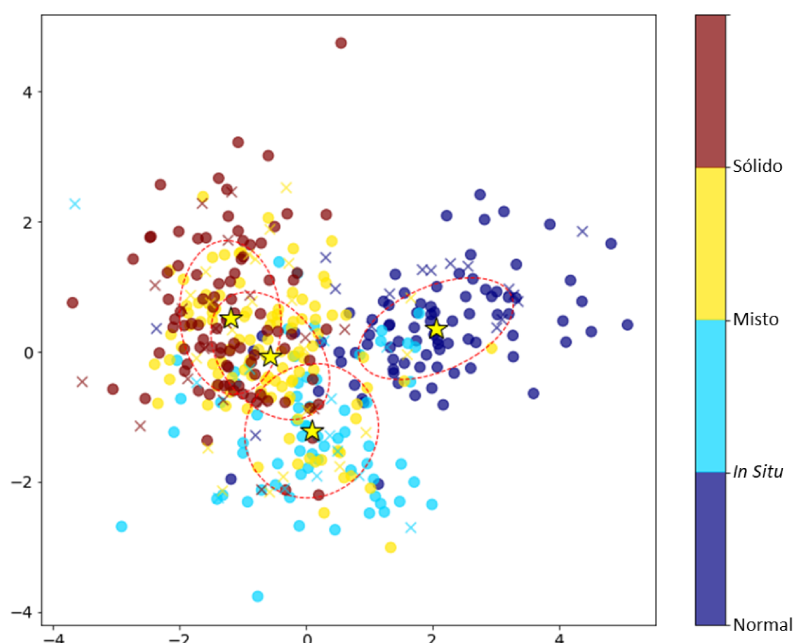


Figura 4.7: Gráfico LDA considerando todas as métricas analisadas com os tipos histológicos indicados pela barra de cores. Os círculos cheios são para o conjunto de treinamento e as cruzes são para o conjunto de teste. A média de cada grupo de caso é indicada pela estrela amarela e as elipses mostram um desvio padrão dos dados de treinamento.

4.2 Análises de Subtipos Histológicos e Progressão Tumoral

Nesta seção apresentamos resultados incluindo diferentes tipos tumorais em cadelas. Foram selecionados casos de pacientes com informações sobre o acompanhamento clínico dos animais. Os casos foram selecionados apenas com um único tipo por paciente. Parte desses resultados foram utilizados por Ana Paula Vieira Garcia, do Instituto de Ciências Biológicas da UFMG, para estudos de correlação com a sobrevivência do animal [31]. Na primeira parte da seção, 4.2.1, apresentamos os resultados para todos os tipos tumorais e na segunda parte, 4.2.2, avaliamos as mudanças das fibras de colágeno durante a progressão tumoral.

4.2.1 Subtipos Histológicos

Aqui foram analisados 45 casos (pacientes) ao todo, divididos em 6 casos de tumor misto benigno (TMB) medidos em 11 lâminas histológicas com 108 imagens coletadas, 16 casos de carcinoma em tumor misto (CTM) medidos em 23 lâminas histológicas com 291 imagens coletadas, 7 casos de carcinoma micropapilar (CMp) medidos em 11 lâminas histológicas com 112 imagens coletadas, 4 casos de carcinossarcoma (CSS) medidos em 6 lâminas histológicas com 83 imagens coletadas e 12 casos de carcinoma sólido (CS) medidos em 13 lâminas histológicas com 142 imagens coletadas. Todos os tipos estão descritos na seção de amostras 3.1. Nos estudos de comparação entre os subtipos histo-

lógicos foram utilizadas amostras diferentes das utilizadas na seção 4.1, mesmo para os casos com diagnóstico igual ao da análise feita anteriormente. Isso foi feito dessa forma, pois a continuação desse estudo que visava correlacionar parâmetros de sobrevida com o diagnóstico necessitava que os animais investigados tivessem de forma prioritária, como diagnóstico, aquele tipo específico de tumor. Na primeira etapa deste trabalho alguns animais foram diagnosticados com diferentes tipos tumorais o que dificultaria uma correlação específica do tipo tumoral com a sobrevida do animal, além de escolhermos lâminas com regiões maiores de massa tumoral de maneira a conseguirmos medir varias imagens representativas do tumor em uma mesma lâmina.

As amostras de tecido de mama normal foram as mesmas utilizadas na primeira parte do trabalho. Todas as imagem de carcinoma em tumor misto, carcinoma micropapilar, carcinossarcoma e carcinoma sólido são de áreas intra tumorais, as amostras de tumor misto benigno possuem áreas peri e intra tumorais.

Na figura 4.8 temos imagens representativas dos cortes histológicos de amostras de tumor misto benigno, carcinoma micropapilar e carcinossarcoma, que são os subtipos tumorais que ainda não foram apresentados. A primeira linha mostra as imagens de campo claro das amostras coradas com H&E, na segunda linha estão as imagens de SHG e na terceira linha estão as imagens de TPEF.

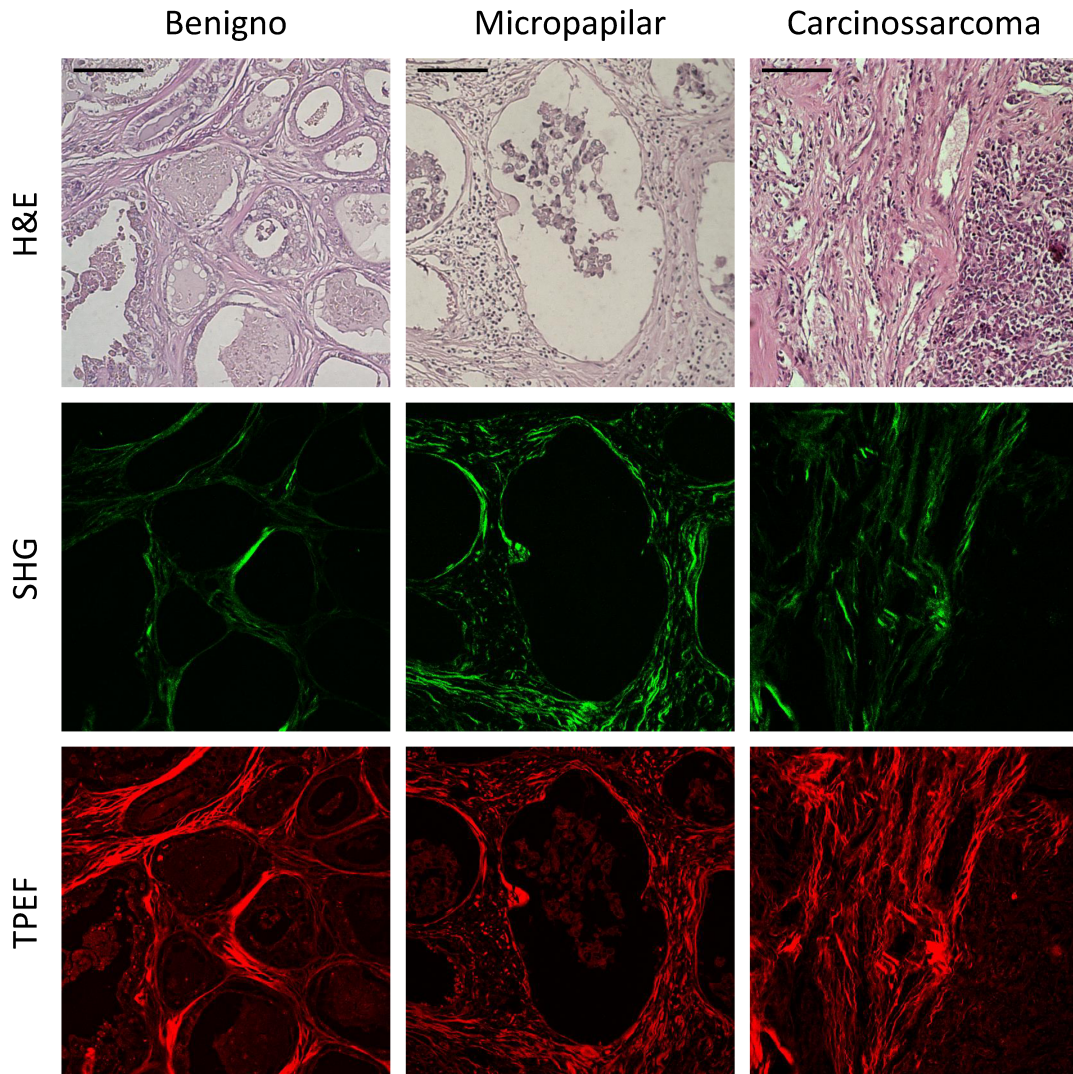


Figura 4.8: Figuras com as imagens obtidas por microscopia de campo claro (primeira linha) coradas com H&E, imagens em escala de cinza de microscopia de SHG coradas digitalmente (segunda linha) e imagens em escala de cinza de microscopia por TPEF coradas digitalmente (terceira linha) com os diagnósticos estudados organizados pelas colunas (Benigno, Micropapilar e carcinossarcoma, respectivamente). A barra da escala é de 100 μm .

Na figura [4.9](#) temos as imagens geradas com o PyFibre com parâmetros extraídos, já descritos anteriormente. Na primeira linha temos a imagem da distribuição dos ângulos de orientação das fibras de colágeno em um mapa de cores. Na segunda e terceira linhas temos o seguimentos de fibras conectadas e o seguimento celular, respectivamente. Na quarta linha temos as imagens das fibras individuais extraídas.

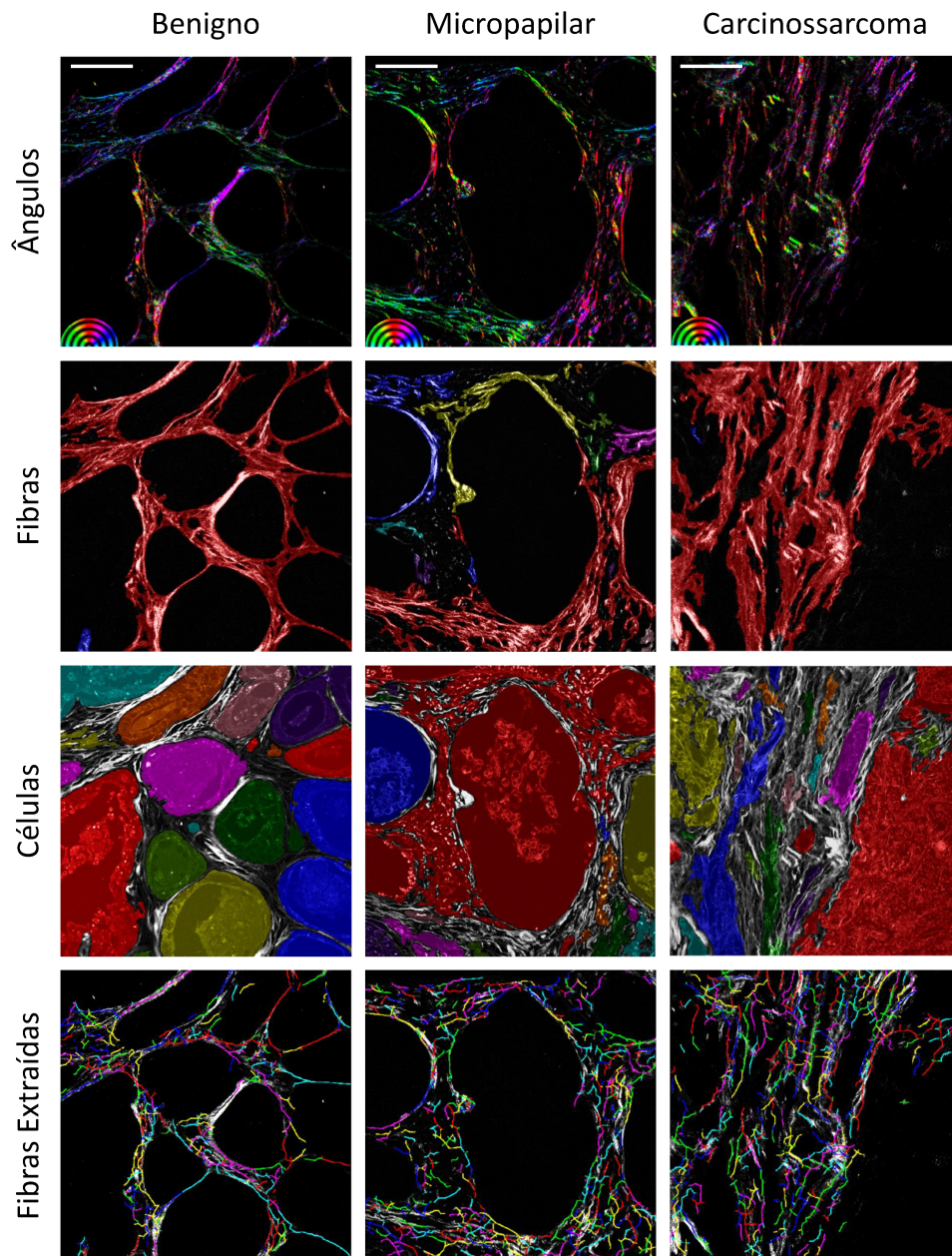


Figura 4.9: Imagens geradas pelo PyFibre para os diagnósticos estudados organizados pelas colunas (Benigno, Micropapilar e carcinossarcoma, respectivamente). Mapa de cores com a distribuição do ângulo da fibra no tecido (primeira linha), rede de fibras extraídas, crescidas sobre a imagem de SHG (segunda linha), segmento de fibras e segmento celular (terceira e quarta linha, respectivamente) com as cores geradas de forma aleatória são construídos sobrepostos com a imagem de intensidade de SHG em escala de cinza e extração de fibras únicas (última linha) também apresentadas em cores aleatórias. A barra da escala é de 100 μm .

Análise estatística

Apresentamos os resultados estatísticos para o tecido da glândula mamária normal e neoplásica, separados pelo tipo de neoplasia. Novamente utilizamos o teste MANOVA. Com ele comparamos as distribuições das amostras de mama normal e as 5 diferentes

neoplasias mamárias (tumor misto benigno, carcinoma em tumor misto, carcinoma micropapilar, carcinosarcoma e carcinoma sólido) simultaneamente para todas as métricas. Além da comparação entre cada duas diferentes neoplasias para todas as métricas. Como resultados para todas as avaliações do MANOVA, tivemos um valor $p < 0,001$. Como os resultados foram significativo, passamos para o teste ANOVA one-way, onde comparamos as distribuições das amostras par a par para cada métrica discutida na seção 4.1.3. o grau de isotropia das fibra de colágeno em escala total e local, a ondulação das fibras, a forma do segmento celular, a área de cobertura de fibras e células, a intensidade média das fibras de colágeno, o comprimento e o número de fibras, para todos os tipos histológicos estudados. Esses dados são apresentados na figura 4.10 e as correlações estatísticas são apresentadas na tabela 4.2.

Para os parâmetros de grau de isotropia para a imagem toda, gráfico (A) da figura 4.10, vemos que em todos os tipos histológicos as fibras de colágeno estão mais alinhadas em relação a mama normal, como vimos nos dados analisados na primeira etapa do trabalho, e temos uma separação entre o tecido normal e o neoplásico, como visto na tabela 4.2. Isso está de acordo com os resultados apresentados na literatura [13-21,25,27,34-36,67,72,73,79,93-99]. Já para o grau de isotropia local, gráfico (B) da figura 4.10, percebemos, novamente em concordância com os resultados obtidos anteriormente, uma queda dos valores dos tecidos neoplásicos em relação a mama normal, porém a tumor misto benigno possui um aumento no valores da mediana em relação às demais amostras, isso está ligado ao fato de que o tumor misto benigno possui um estroma mais acentuado, isto é, é mais fibroso e com as fibras delimitando bem o crescimento celular, o que o caracteriza como benigno.

Assim como nos resultados anteriores a ondulação das fibras de colágeno, gráfico (C) da figura 4.10, tem pouca variação entre os diagnósticos. Pela tabela 4.2 vemos que apenas as comparações com o carcinoma sólido tem alguma separação.

A forma do segmento celular é apresentada no gráfico (D) da figura 4.10, vemos que com exceção do carcinosarcoma, todos os valores diminuem em relação a mama normal, principalmente o carcinoma micropapilar e sólido, que está de acordo com os resultados obtidos anteriormente.

Para os parâmetros de cobertura do segmento de fibra e célula, gráficos (E-F) da figura 4.10, observamos que o tumor misto benigno e o carcinosarcoma apresentam uma maior área coberta por colágeno nos dados coletados, corroborando com [4,102,103], essa aparente discrepância com relação aos demais tipos histológicos e a mama normal se deve ao fato de que tanto o tumor misto benigno quanto o carcinosarcoma possuem um crescimento acentuado do estroma, a diferença está no fato de que no tumor misto benigno esse crescimento é controlado e não existe áreas infiltrativas. Além disso, os resultados mostram uma área mais coberta por células no carcinoma sólido em comparação com os outros tipos histológicos, em concordância com a característica dos carcinomas sólidos que apresentam crescimento expansivo altamente celularizados [9]. Ambos os parâmetros

estão de acordo com os resultados obtidos na primeira etapa deste trabalho.

A intensidades média das fibras de colágeno, gráfico (G) da figura 4.10, possuem valores menores que as amostras de tecido de mama normal, com exceção do tumor misto benigno e do carcinossarcoma.

O comprimento das fibras e o número total de fibras extraídas, gráficos (H) e (I) da figura 4.10, respectivamente, mostram que as fibras de colágeno apresentam menor comprimento no caso das neoplasias mamárias mais agressivas, como é o caso do CMp, CSS e CS, e que a quantidade de colágeno diminui em todos os tecidos neoplásicos, exceto neste caso, para o tumor misto benigno. Esses resultados levam à hipótese de que entre os subtipos tumorais, os carcinomas mamários menos agressivos apresentam fibras de colágeno menores. Esse resultado foi o que motivou a continuação do trabalho com relação a sobrevida dos animais, visto que o comprimento das fibras teve uma correlação com a sobrevida, essa parte do trabalho pode ser vista no artigo 31.

Com todas as métricas analisadas podemos chegar a conclusão de que, em geral, as neoplasias mamárias apresentam fibras mais organizadas, maiores e em menor quantidade em relação a mama normal, como pode ser visto nos gráficos (A), (E), (G), (H) e (I) da figura 4.10. Apenas o tumor misto benigno e o carcinossarcoma são exceção em algumas dessas métricas, porém devido a grande quantidade de matriz extracelular característica desses tipos histológicos observamos essa aparente discrepância dos dados, mas que na verdade está de acordo com a literatura 4,102,103.

Em um apanhado geral sobre as comparações que apresentam valores p significativos mostradas na tabela 4.2, vemos que o Grau de Isotropia Local foi a métrica com mais separação entre as comparações dos diagnósticos, com 13 comparações com valor p significativo em um total de 15 comparações. Métricas como a Forma do Segmento Celular, a Área de Cobertura Celular, o Comprimento da Fibra e do No. de Fibras também tiveram bons resultados nas separações entre os diagnósticos, cada um com 12 comparações com valor p significativo. Porém os parâmetros de Ondulação da Fibra e Grau de Isotropia não tiveram tantas separações entre os diagnósticos, com 6 e 5 separações, respectivamente. Uma observação importante de se fazer é que para o Grau de Isotropia tivemos apenas valores p significativos entre as comparações de mama normal e as neoplasias mamárias. Entre as comparações, Normal x CMp e TMB x CS foram as que tiveram maiores números de parâmetros diferenciáveis, com um total de 8 parâmetros cada (Ondulação da Fibra não teve valor p significativo para a primeira e o Grau de Isotropia para a segunda), em contraponto, Normal x CTM tiveram poucas métrica com valor de p significativo (apenas o Grau de Isotropia, o Comprimento da Fibra e o No. de Fibras).

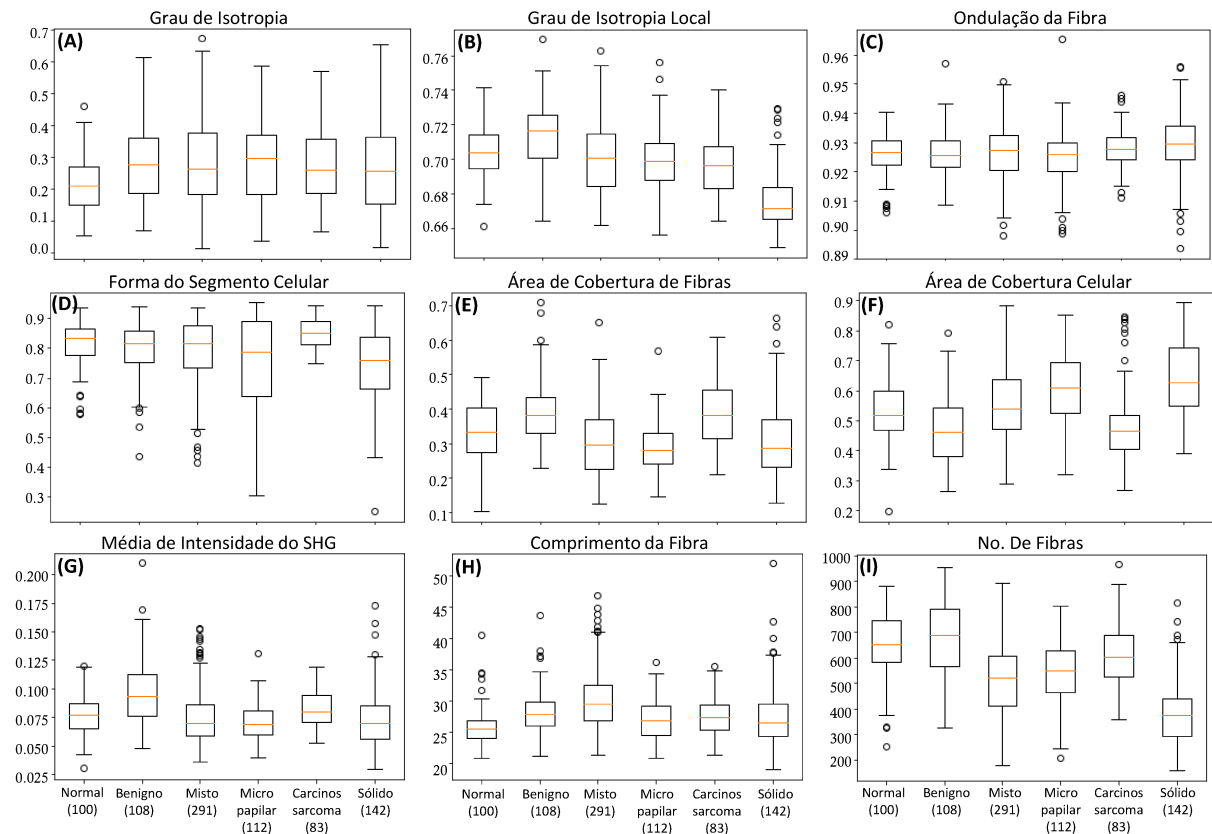


Figura 4.10: Gráficos de boxplot para as métricas indicadas nos títulos dos gráficos para os diagnósticos: Normal, tumor misto benigno, carcinoma em tumor misto, carcinoma micropapilar, carcinossarcoma e carcinoma sólido. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

Tabela 4.2: Tabela de valores p com a comparação das métricas entre pares de diagnósticos. A significância denotada por $p < 0,05$ são denotados como (*), $p < 0,01$ são denotados como (**) e $p < 0,001$ são denotados como (***)

Métricas	Normal X TMB	Normal X CTM	Normal X CMP	Normal X CSS	Normal X CS	TMB X CTM	TMB X CMP	TMB X CSS	TMB X CS	CTM X CMP	CTM X CSS	CTM X CS	CMP X CSS	CMP X CS	CSS X CS
Grau de Isotropia	***	***	***	***	**										
Grau de Isotropia Local	**		*	**	***	***	***	***	***	*	**	***		***	***
Ondulação da Fibra					*			*	**			**	*	**	
Forma do Segmento Celular	*		***	***	***		*	***	***	***	***	***	***		***
Área de Cobertura de Fibras	***		***	***		***	***		***	***	***		***	*	***
Área de Cobertura Celular	*		***	*	***	***	***		***	***	**	***	***		***
Média de Intensidade do SHG	***		**	*		***	***	***	***	***		**	***		**
Comprimento da Fibra	***	***	**	***	**	**	**	*	*	***	***	***			
No. de Fibras		***	***		***	***	***	*	***		**	***	***	***	***

4.2.2 Progressão tumoral

Ao avançarmos com o trabalho, estudamos as mudanças sofridas pelo tecido à medida em que um tumor progride. Para essa análise utilizamos as amostras de carcinoma em

tumor misto, avaliadas de maneira total na sessão acima, separando-as em três regiões distintas, 24 imagens de regiões com áreas benignas (CTMb), 110 imagens de regiões com áreas de crescimento tumoral *in situ* (CTMis) e 157 imagens de regiões com áreas de crescimento tumoral invasoras (CTMi), foram avaliadas no total. Imagens representativas dessas três regiões distintas do carcinoma em tumor misto podem ser vistas na figura [4.11](#).

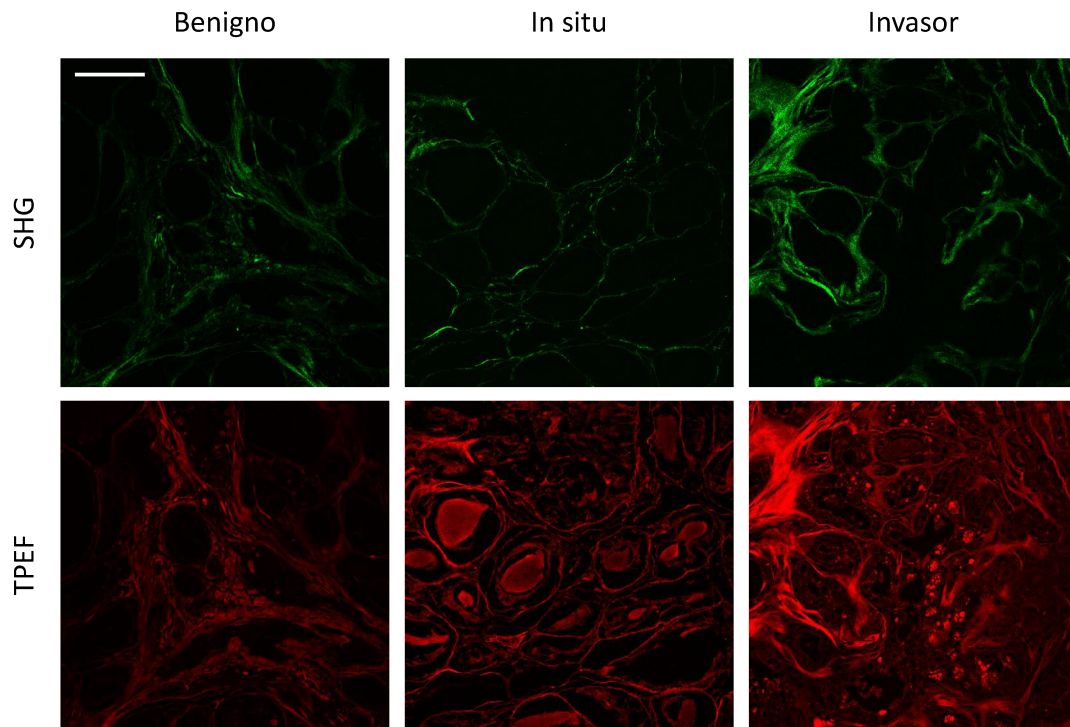


Figura 4.11: Figuras com as imagens em escala de cinza de microscopia de SHG coradas digitalmente (primeira linha) e imagens em escala de cinza de microscopia por TPEF coradas digitalmente (segunda linha) com regiões das amostras de Carcinoma em Tumor Misto estudadas organizadas pelas colunas (Benigna, *In Situ* e Invasor, respectivamente). A barra da escala é de 100 μm .

Análise Estatística

Apresentamos os resultados estatísticos para o tecido da glândula mamária normal e neoplásica, separadas pela região do tecido de carcinoma em tumor misto. Novamente utilizamos o teste MANOVA. Com ele comparamos as distribuições das amostras de mama normal e as 3 diferentes regiões do carcinoma em tumor misto (benigna, *in situ* e invasora) simultaneamente para todas as métricas. Além da comparação entre cada duas diferentes regiões para todas as métricas. Como resultados para todas as avaliações do MANOVA, tivemos um valor $p < 0,001$. Como os resultados foram significativo, passamos para o teste ANOVA one-way, onde comparamos as distribuições das amostras par a par para cada métrica. Os gráficos com as distribuições dos dados dos subgrupos do carcinoma em tumor misto para a progressão tumoral são apresentados na figura [4.12](#) e suas correlações estatísticas são apresentadas na tabela [4.3](#).

Os parâmetros de grau de isotropia para a imagem total, gráfico (A) da figura 4.12, e grau de isotropia local, gráfico (B) da figura 4.12, seguem o mesmo padrão qualitativo dos resultados obtidos anteriormente, isto é, para a imagem toda temos um aumento no valor da mediana de todos os subgrupos em relação a mama normal e uma queda para os valores na mediana dos subgrupos na análise da isotropia local, porém para essas amostras observamos um pequeno aumento no CTMis em relação a mama normal. Por este resultado é possível dizer que durante a progressão tumoral as fibras de colágeno estão mais alinhadas com o desenvolvimento da neoplasia, o que facilita a migração das células tumorais para o tecido extracelular [7,104-108].

Para o parâmetro de ondulação das fibras de colágeno, gráfico (C) da figura 4.12, podemos perceber uma queda para os valores de CTMb e CTMis em relação a mama normal, principalmente para o CTMis, por outro lado temos um aumento no valor da mediana para o CTMi.

Para a forma do segmento celular é apresentada, gráfico (D) da figura 4.12, observamos que todos os valores das medianas dos subgrupos diminuem em relação a mama normal.

Nos parâmetros de cobertura do segmento de fibra e célula, gráficos (E-F) da figura 4.12, temos uma queda para todos os subgrupos do CTM em relação a mama normal, principalmente para o CTMis. Observamos também um aumento nas áreas de cobertura de células. Esse comportamento pode estar ligado ao fato de que no crescimento *in situ*, já existe uma componente maligno, isto é, um aumento na proliferação celular, muito maior que no caso do benigno, dessa forma as fibras se expandem de maneira a tentar comportar esse crescimento das células.

A intensidades média das fibras de colágeno, gráficos (G) da figura 4.12, possuem valores p significativos para alguns subgrupos do CTM.

O comprimento das fibras e o número total de fibras extraídas, (H) e (I), respectivamente, mostram que as fibras de colágeno apresentam maior comprimento para os subgrupos, mas é importante notar que para o CTMi o comprimento da fibra tem uma ligeira queda em relação ao CTMb e ao CTMis. O número total de fibras de colágeno diminui para todos os subtipos em relação a mama normal, porém temos um pequeno aumento para o CTMi. Podemos dizer que, no geral, durante a progressão tumoral o comprimento das fibras de colágeno aumenta à medida que a neoplasia se desenvolve. Observamos que no crescimento tumoral *in situ* as fibras de colágeno mais alongadas e menos fragmentadas, limitando o crescimento da neoplasia, porém, quando as células rompem a membrana basal e passam a ter um crescimento tumoral invasivo, as fibras tornam-se mais curtas e mais fragmentadas em relação ao estado anterior. O que mais uma vez nos leva a considerar que para um subtipo tumoral mais agressivo, as fibras de colágeno são mais curta, mesmo que para o caso geral de um CTM, que não é um tipo muito agressivo, as fibras tenham comprimento maior do que o normal.

Em um resumo geral sobre os valores p significativos apresentados na tabela 4.3, vemos que o parâmetro Comprimento da Fibra teve um valor p significativo em todas as com-

parações entre os subgrupos realizadas nesse estudo. O Grau de Isotropia Local, a Área de Cobertura de Fibras e a Área de Cobertura Celular também tiveram bons resultados nas separações dos subgrupos, cada um com 5 comparações com valor p significativo das 6 comparações realizadas. O parâmetros com o menor número de separações foi a Ondulação da Fibra com apenas 2 comparações com valor p significativo. Novamente, o Grau de Isotropia teve uma boa separação apenas nas comparações entre a mama normal e os demais subgrupos do CTM. Entre as comparações, Normal x CTMb e CTMb x CTMi foram as que tiveram o maior número de parâmetros diferenciáveis, com um total de 7 parâmetros cada, em contraponto, Normal x CTMi foi a comparação com o menor número de parâmetros com valor de p significativo, 5 ao todo.

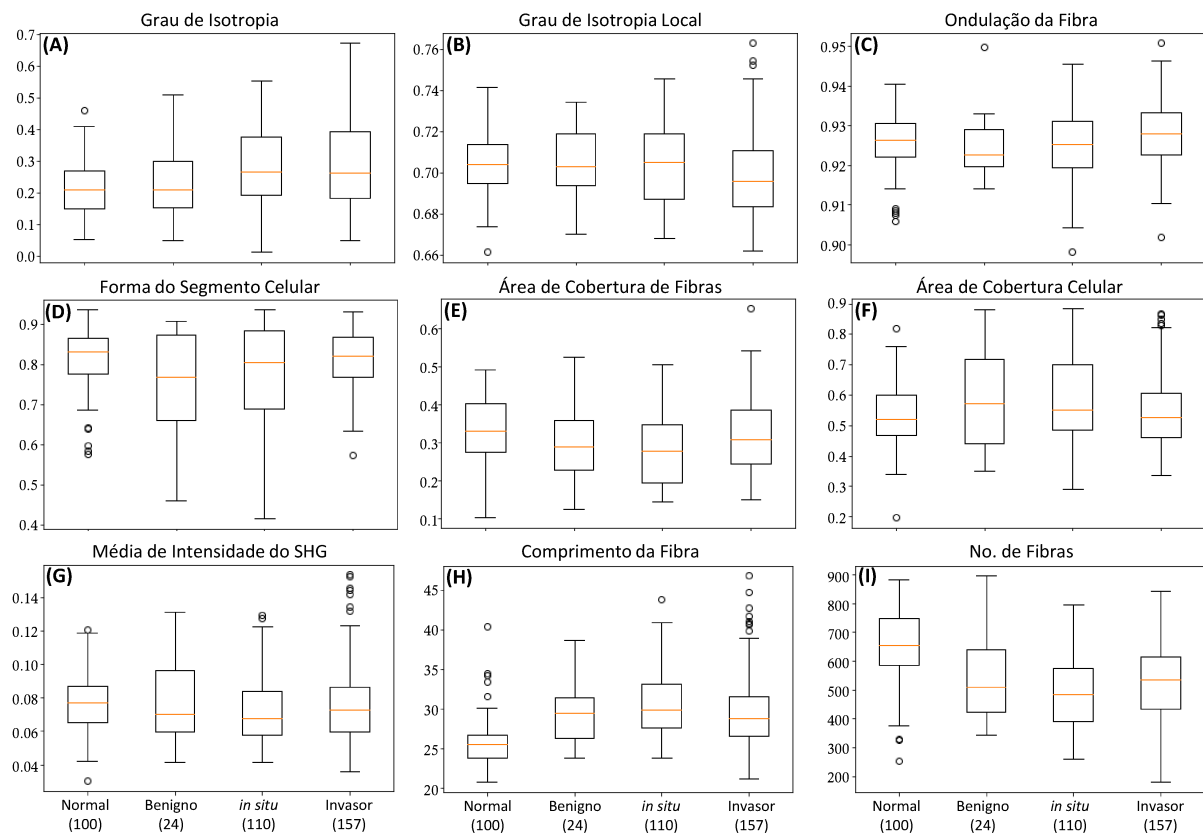


Figura 4.12: Gráficos de boxplot para as métricas indicadas nos títulos dos gráficos para os diagnósticos: Normal, regiões benignas de carcinoma em tumor misto, regiões *in situ* em carcinoma em tumor misto e regiões invasoras em carcinoma em tumor misto. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

Tabela 4.3: Tabela de valores p com a comparação das métricas entre pares de diagnósticos. A significância denotada por $p < 0,05$ são denotados como (*), $p < 0,01$ são denotados como (**) e $p < 0,001$ são denotados como (***).

Métricas	Normal X CTMb	Normal X CTMis	Normal X CTMi	CTMb X CTMis	CTMb X CTMi	CTMis X CTMi
Grau de Isotropia	***	***	***			
Grau de Isotropia Local	**		**	**	***	*
Ondulação da Fibra					*	*
Forma do segmento Celular	***	***	***			
Área de Cobertura de Fibras	***	***		***	***	***
Área de Cobertura Celular	**	***		***	***	**
Média de Intensidade do SHG	***			***	***	
Comprimento da Fibra	***	***	***	***	*	*
No. de Fibras		***	***	***	***	

4.2.3 LDA

Apresentamos a seguir os gráficos de LDA para todos os tipos histológicos e para as amostras da progressão tumoral. As análises foram feitas da mesma forma que na seção anterior, seção 4.1.4. Além de uma análise com apenas as neoplasias malignas e uma análise com apenas as regiões do CTM.

LDA Tipos Histológicos

Na figura 4.13(a), apresentamos o gráfico do LDA, considerando todos os parâmetros obtidos, para os tipos histológicos estudados nessa parte do trabalho. Podemos observar que existe uma separação nítida entre os dados da mama normal e os dados das neoplasias mamárias, além de uma boa separação entre os dados do carcinoma sólido e as demais neoplasias, que apresentam uma sobreposição entre os dados. Para a separação entre as amostras de tecido normal e as neoplasias mamárias obtivemos uma precisão de $93 \pm 2\%$, o que é um resultado um pouco melhor do que o que obtivemos na primeira etapa do trabalho e considerando ainda um número maior de tipos histológicos. No entanto, considerando agora a discriminação entre todos os tipos, a precisão foi de apenas $43 \pm 3\%$, o que é um resultado pior do que o obtido anteriormente, porém com uma margem de erro bem menor, isso pode ser devido ao fato de que o tumor misto benigno, o carcinoma em tumor misto e o carcinosarcoma possuem algumas características parecidas em se tratando de matriz extracelular, o que faz com que o LDA tenha dificuldade na previsão de algumas

dessas amostras.

Para melhor visualizar as separações entre os tipos tumorais, na figura 4.13(b), retiramos os dados das amostras de mama normal e tumor misto benigno e avaliamos apenas a separabilidade para os casos de tumores malignos. Para essa separação observamos que os dados do carcinoma sólido continuam apresentando uma boa separabilidade entre as demais neoplasias, mas agora também podemos observar uma separação do carcinoma micropapilar com as demais neoplasias, a precisão do nosso conjunto de dados de teste em discriminar entre os diagnósticos é de $57 \pm 2\%$, isto é, obtivemos uma melhor separação entre esses tipos, mas novamente vemos uma sobreposição nos dados do carcinoma em tumor misto e no carcinosarcoma.

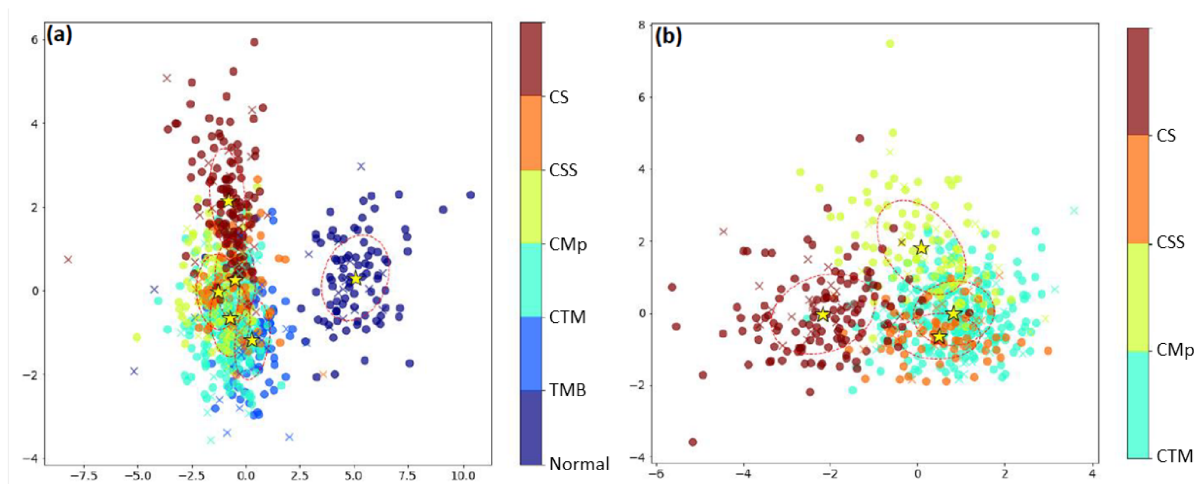


Figura 4.13: Gráficos de LDA considerando todas as métricas analisadas com os tipos histológicos indicados pela barra de cores. No gráfico (a) consideramos todos os tipos histológicos, já no gráfico (b) consideramos apenas os tipos histológicos malignos. Os círculos cheios são para o conjunto de treinamento e as cruzes são para o conjunto de teste. A média de cada grupo de caso é indicada pela estrela amarela e as elipses mostram um desvio padrão dos dados de treinamento.

LDA Progressão Tumoral

Ao explorarmos mais as variações na progressão do câncer considerando os dados de carcinoma em tumor misto considerando seus subgrupos (regiões benignas, *in situ* e invasoras), apresentamos na figura 4.14(a) o modelo LDA com as imagens de tecido normal e progressão tumoral no carcinoma em tumor misto. A precisão obtida para a discriminação entre o tecido normal e o carcinoma em tumor misto é de $97 \pm 2\%$, que é um excelente valor. Avaliando os dados de todas as amostras também obtivemos uma boa separação $65 \pm 2\%$, indicando que nosso método de análise é sensível aos detalhes da ruptura do colágeno característica das regiões invasivas. Porém ao fazermos o LDA dos dados sem as mostras de mama normal, figura 4.14(b), a precisão dos dados de teste foi de $57 \pm 3\%$ o que é consideravelmente menor, isso provavelmente se deve ao fato de o número de amostras das regiões benignas para o carcinoma em tumor misto é muito baixo.

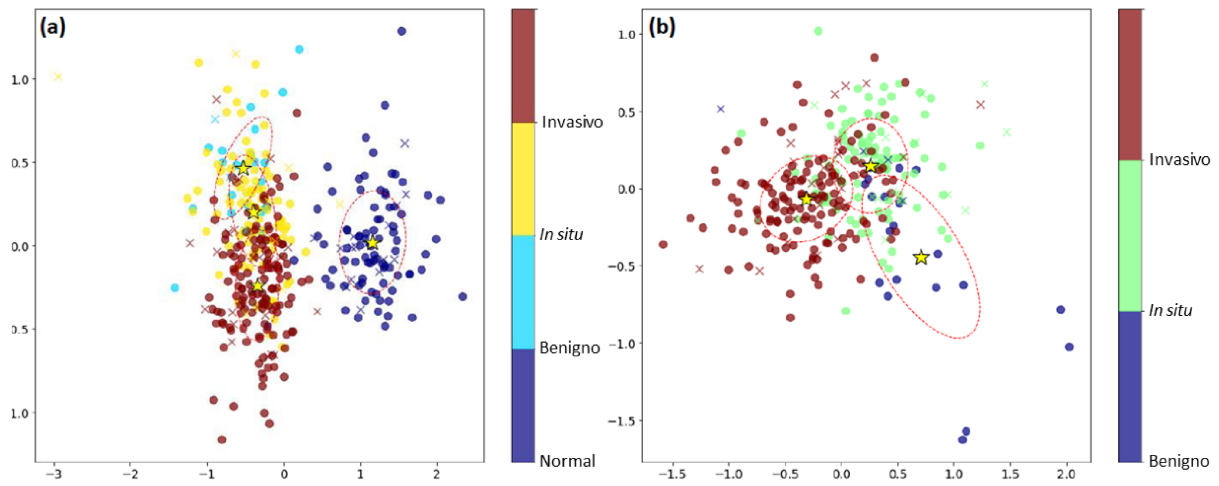


Figura 4.14: Gráficos de LDA considerando todas as métricas analisadas com as regiões do carcinoma em tumor misto indicadas pela barra de cores. No gráfico (a) consideramos todas as regiões e a mama normal, já no gráfico (b) consideramos apenas as regiões selecionadas nas amostras de carcinoma em tumor misto. Os círculos cheios são para o conjunto de treinamento e as cruzes são para o conjunto de teste. A média de cada grupo de caso é indicada pela estrela amarela e as elipses mostram um desvio padrão dos dados de treinamento.

Capítulo 5

Conclusão e Perspectivas

Neste trabalho utilizamos técnicas de microscopia óptica não linear, SHG e TPEF, para analisar tecidos mamários normais e neoplasias com desenvolvimento natural em cadela. Com essas técnicas conseguimos avaliar a organização do tecido estromal, com as imagens SHG obtivemos informações do colágeno e as imagens TPEF permitiram acessar as regiões celulares da biópsia, em áreas intratumorais para diferentes tipos histológicos. O trabalho foi desenvolvido observando primeiro um grupo pequeno de subtipos tumorais, carcinoma ductal *In Situ*, carcinoma em tumor misto e carcinoma sólido, sem a preocupação em termos apenas um subtipo por paciente. E na sequência foi utilizado um número maior de subtipos histológicos, como o tumor misto benigno, o carcinoma em tumor misto, o carcinoma micropapilar, o carcinosarcoma e o carcinoma sólido, apenas um tipo por paciente. Realizamos também um estudo relacionando as diferenças no tecido para um mesmo tipo histológico mas olhando diferentes regiões desse tecido, de forma a analisar a progressão tumoral para o carcinoma em tumor misto, que nos permite analisar regiões benignas, *in situ* e invasoras em uma mesma amostra.

Para a análise das imagens um software que possibilita uma análise simultânea das imagens de SHG e TPEF foi desenvolvido e com ele conseguimos um estudo abrangente sobre várias modificações causadas nas neoplasias mamárias. O software utiliza-se de vários algoritmos de processamento de imagem já conhecidos na literatura, porém um aprimoramento desses algoritmos é feito de forma a melhorar a extração dos parâmetros de interesse. O software foi desenvolvido para obter parâmetros baseados nas observações que os patologistas geralmente utilizam para dar um diagnóstico. O PyFibre nos permitiu identificar e separar de maneira rápida e eficientemente as regiões de fibras e células nas imagens e com essas análises, conseguimos extrair das imagens parâmetros de fibras individuais de colágeno, quantificar as orientações das fibras e extrair parâmetros específicos para as regiões de fibras e células nas imagens.

Com os resultados obtidos pelo PyFibre conseguimos quantificar separadamente cada métrica para cada subtipo tumoral e através de análises de variância avaliar suas distribuições. Esses resultados, juntamente com o LDA, mostraram-se eficazes na separação das amostras da glândula mamária normal e dos tecidos neoplásicos, com uma clara dis-

criminação entre os tipos histológicos. Obtivemos uma sensibilidade de cerca de 90% na classificação entre tecido saudável e canceroso e uma sensibilidade de cerca de 40-60% para comparação entre a mama normal e os subtipos histológicos estudados.

Esses resultados indicam que essas novas ferramentas de análise podem se tornar úteis na complementação do trabalho do patologista. Pois temos um procedimento mais rápido e quantitativo no estudo das amostras.

Como continuidade para o trabalho, podemos estudar diferentes subtipos tumorais de neoplasias mamárias, tanto para as regiões intra quanto peritumoral. Para ambas as regiões essa análise pode ser feita em tecido fresco avaliando não apenas o colágeno, mas também a auto-fluorescência das amostras. Além da implementação de imagens de THG (Geração de Terceiro harmônico), que permitirão estudar as modificações sofridas pelos núcleos celulares em diferentes tipos de neoplasias.

Em um viés mais estatístico, podemos utilizar outros métodos de aprendizado por máquina nas análises dos dados, buscando melhores métodos de classificação das amostras.

Apêndice A

PyFibre-estrutura do programa

O PyFibre, como descrito anteriormente, é um software desenvolvido para análise de imagens de tecido fibroso. O software pode ser executado em um terminal (CLI) ou em uma interface gráfica (GUI). Ele foi projetado para funcionar com imagens com o formato Targged Image File Format (TIFF) em um formato multicanal contendo sinais de SHG e TPEF.

O PyFibre é hospedado no Enthought Deployment Manager (EDM), que deve ser instalado antes da instalação do programa. Os passos para a instalação do PyFibre, após a instalação do ambiente EDM, pode ser encontrada na página do GitHub [71].

Após a instalação do PyFibre o programa pode ser executado tanto no CLI quanto em um GUI. Para executar o PyFibre pela linha de comando basta seguir os comando indicados no GitHub [71]. E veremos a execução do programa apenas por comandos. Se a opção para executar o PyFibre for pela interface gráfica chamamos o executável *PyFibre – GUI* na linha de comando e a interface gráfica será iniciada. Um painel de exibição será iniciado como mostrado na figura A.1. O painel de exibição de arquivos contém opções para carregar e organizar arquivos de imagem para análises.

No lado esquerdo da imagem podemos ver um diretório com as pastas disponíveis no computador, dessa forma conseguimos encontrar facilmente os dados a serem analisados. Após abrir a pasta com os dados basta adicionar os arquivos a serem analisados (Add File), que também podem ser procurados utilizando uma palavra-chave (Key) e o filtro (Filter), esses arquivos ficarão visíveis em uma lista rolável no lado esquerdo. Caso algum arquivo indesejado tenha sido carregado existe a opção de remove-lo, basta selecionar e clicar em ‘Remove File’. Após carregar todos os arquivos a serem trabalhados é só executar o programa clicando em ‘Run’ no canto superior esquerdo.

O que aparecerá em seguida é a imagem selecionada, com todas as imagens geradas pelo programa como o mapa de intensidade (Tensor Image), a Rede de Fibras (Network Image), a extração de fibras individuais (Fibres), os segmentos de fibras e celular (Fibre e Cell Segments), respectivamente, podendo ser vistas apenas alterando as abas na parte superior da imagem. Após a conclusão da análise, o ícone ‘Save Database’ no canto superior esquerdo exibe uma janela de arquivo na qual é possível salvar uma cópia agrupada de

todas as métricas de imagem carregadas.

Atualmente, apenas análises de arquivos TIFF para técnicas de imagem de SHG e TPEF criados pelos programas Olympus FluoView e ImageJ são analisadas pelo PyFibre. O PyFibre é capaz de reconhecer e carregar automaticamente as imagens de SHG e PL e agrupá-las, se elas contiverem um esquema de nomes de arquivo apropriado, contendo palavras-chaves (sem distinção entre maiúsculas e minúsculas) como no exemplo: $\{\text{diretório}\}/\{\text{prefixo}\}-\{\text{palavra-chave}\}\{\text{sufixo}\}.tif$. Os arquivos aparecerão como: $\{\text{diretório}\}/\{\text{prefixo}\}$. Espera-se que os arquivos contendo a palavra-chave SHG representem uma imagem em tons de cinza (o sinal do segundo harmônico), os arquivos contendo a palavra-chave PL devem representar duas imagens em tons de cinza (sinal de TPEF e o de transmissão). Os arquivos contendo a palavra-chave SHG-PL combinados representarão três imagens em tons de cinza (SHG, PL, Trans).

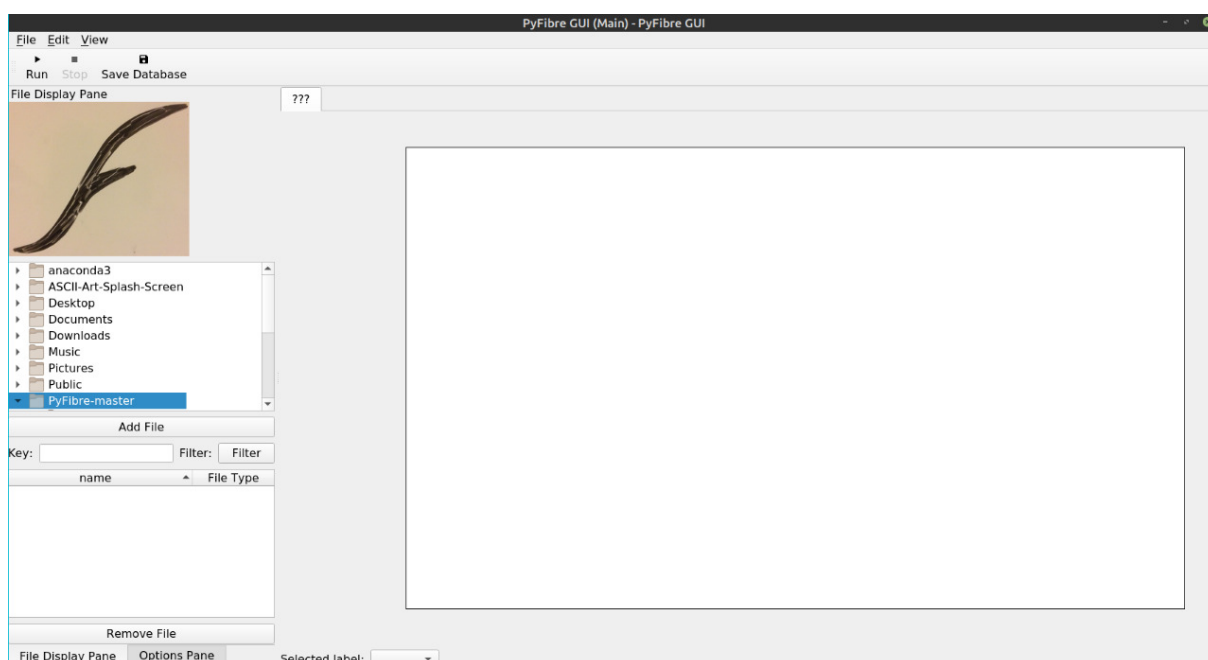


Figura A.1: Tela inicial do PyFibre.

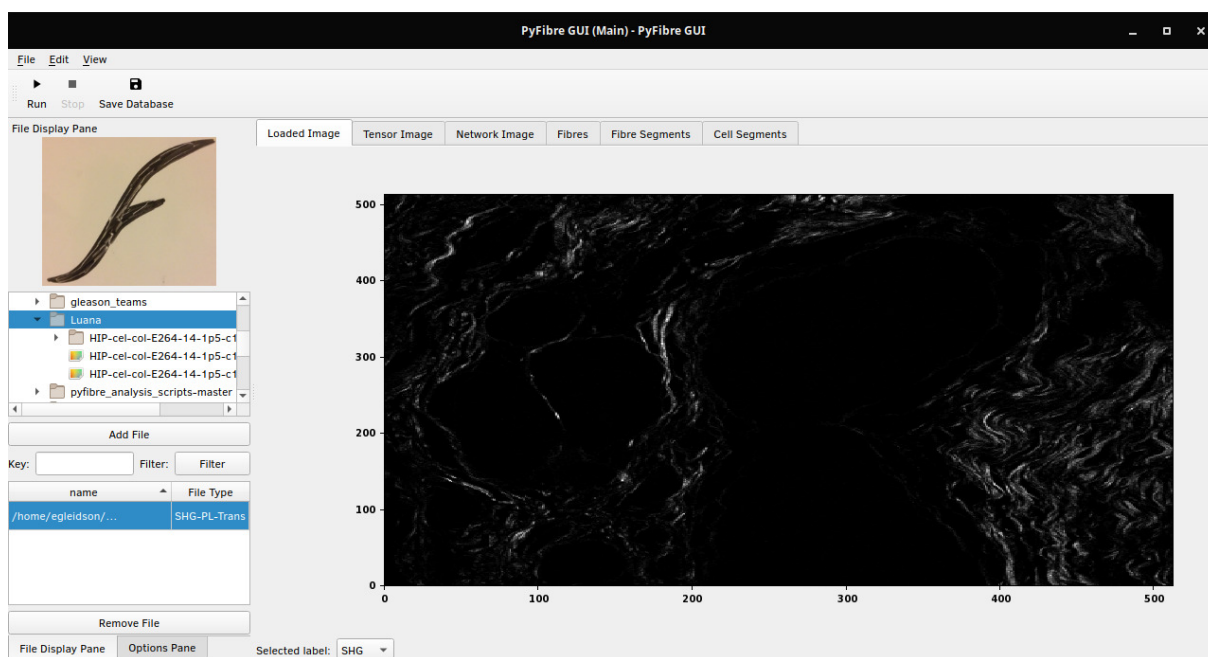


Figura A.2: Tela Com imagem carregada e executada pelo PyFibre.

Apêndice B

Métricas PyFibre

O PyFibre fornece ao todo 29 parâmetros agrupadas em 4 grupos de acordo com a forma pela qual esses parâmetros foram calculados. São eles: rede de fibras, segmento de fibras, segmento de células e imagem global. Aqui, apresentaremos todos os 29 parâmetros calculados, com os nomes em inglês, como pode ser encontrado no GitHub [\[71\]](#) e com a tradução livre, entre parêntesis, utilizada para as métricas apresentadas nesse trabalho.

B.1 Rede de Fibras

Nesse grupo temos 8 métricas calculadas para cada rede de fibras extraídas pelo algoritmo.

- **No.Fibre** (Número de Fibras): Número de fibras extraídas;
- **Mean Fibre Waviness** (Ondulação da Fibra): Ondulação média da fibra (comprimento/deslocamento);
- **Mean Fibre Length** (Comprimento da Fibra): Comprimento médio da fibra em pixel;
- **Fibre Angle SDI (Standard Deviation Index)**: SDI da distribuição de ângulos (méd/máx) da fibra;
- **Fibre Network Degree**: Número médio de arestas da rede de fibra por nó;
- **Fibre Network Eigenvalue**: Autovalor máximo da matriz de adjacência da rede;
- **Fibre network Connectivity**: Média da conectividade da rede de fibras;
- **Fibre Network Cross-Link Density**: Média de ligações cruzadas por fibras.

B.2 Segmento de Fibras

Nesse grupo temos 10 métricas calculadas para cada segmento de fibra identificado pelo algoritmo no processo de segmentação.

- **Fibre Segment Area:** Área do segmento de fibras em pixels;
- **Fibre Segment Coverage** (Área de Cobertura de Fibras): Proporção da área do segmento de fibra/área da caixa delimitadora;
- **Fibre Segment Circularity:** Medida do quão circular é o segmento de fibra;
- **Fibre Segment Eccentricity:** Medida da forma da excentricidade do segmento de fibra;
- **Fibre Segment SHG Angle SDI:** SDI da distribuição de ângulos (méd/máx) para todos os pixels de imagem SHG no segmento de fibra;
- **Fibre Segment SHG Anisotropy** (Grau de Isotropia): Anisotropia calculada pelo tensor de estrutura, média para todos os pixels da imagens SHG no segmento de fibra;
- **Fibre Segment SHG Local Anisotropy** (Grau de Isotropia Local): Anisotropia calculada pelo tensor de estrutura para todos em regiões de 2×2 pixel, média para todos os pixels da imagens SHG no segmento de fibra;
- **Fibre Segment SHG Mean (Média da Intensidade do SHG):** Intensidade média de SHG para os pixels do segmento de fibra na imagem;
- **Fibre Segment SHG STD:** Desvio padrão da intensidade SHG para os pixels do segmento de fibra na imagem;
- **Fibre Segment SHG Entropy:** Entropia de Shannon média dos pixels do segmento de fibra na imagem SHG;

B.3 Segmento de Células

Nesse grupo temos 11 métricas calculadas para cada segmento de célula identificado pelo algoritmo no processo de segmentação.

- **No.Cell:** Número de segmentos de células extraídas;
- **Cell Segment Area:** Área do segmento de célula em pixels;
- **Cell Segment Coverage** (Área de Cobertura Celular): Proporção da área do segmento de célula/área da caixa delimitadora;

- **Cell Segment Circularity** (Forma do Segmento Celular): Medida do quão circular é o segmento de célula;
- **Cell Segment Eccentricity**: Medida da forma da excentricidade do segmento de célula;
- **Cell Segment PL Angle SDI**: SDI da distribuição de ângulos (méd/máx) para todos os pixels de imagem PL no segmento de célula;
- **Cell Segment PL Anisotropy**: Anisotropia calculada pelo tensor de estrutura, média para todos os pixels da imagens PL no segmento de célula;
- **Cell Segment PL Local Anisotropy**: Anisotropia calculada pelo tensor de estrutura para todos em regiões de 3×3 pixel, média para todos os pixels das imagens PL no segmento de célula;
- **Cell Segment PL Mean**: Intensidade média de PL para os pixels do segmento de fibra na imagem;
- **Cell Segment PL STD**: Desvio padrão da intensidade PL para os pixels do segmento de fibra na imagem;
- **Cell Segment PL Entropy**: Entropia de Shannon média dos pixels do segmento de célula na imagem PL;

Com todos esses parâmetros o PyFibre gera uma banco de dados para cada região selecionada e segmentada para a imagem e para a imagem global. Três arquivos principais contendo todas as métricas são gerados com a extensão .pkl e .xls, além das imagens geradas salvas com a extensão .png.

Apêndice C

Gráficos de caixa: Métricas extraídas

Neste apêndice apresentaremos os demais gráficos de caixa obtidos através das métricas calculados pelo PyFibre.

C.1 Métricas: Tipos histológicos *In Situ*, Misto e Sólido

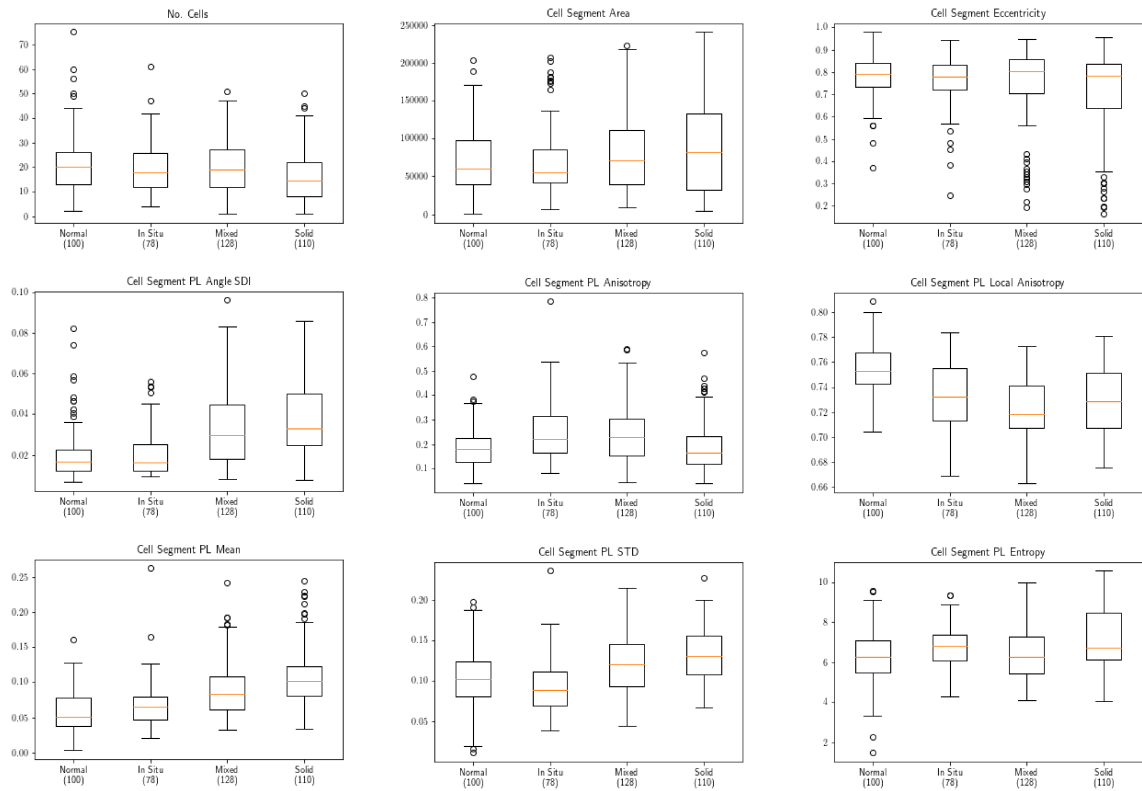


Figura C.1: Gráficos de caixa para as métricas relacionadas à região celular, com as métricas indicadas nos títulos de cada gráfico, para os diagnósticos de tecidos mamários normal e tumoral. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

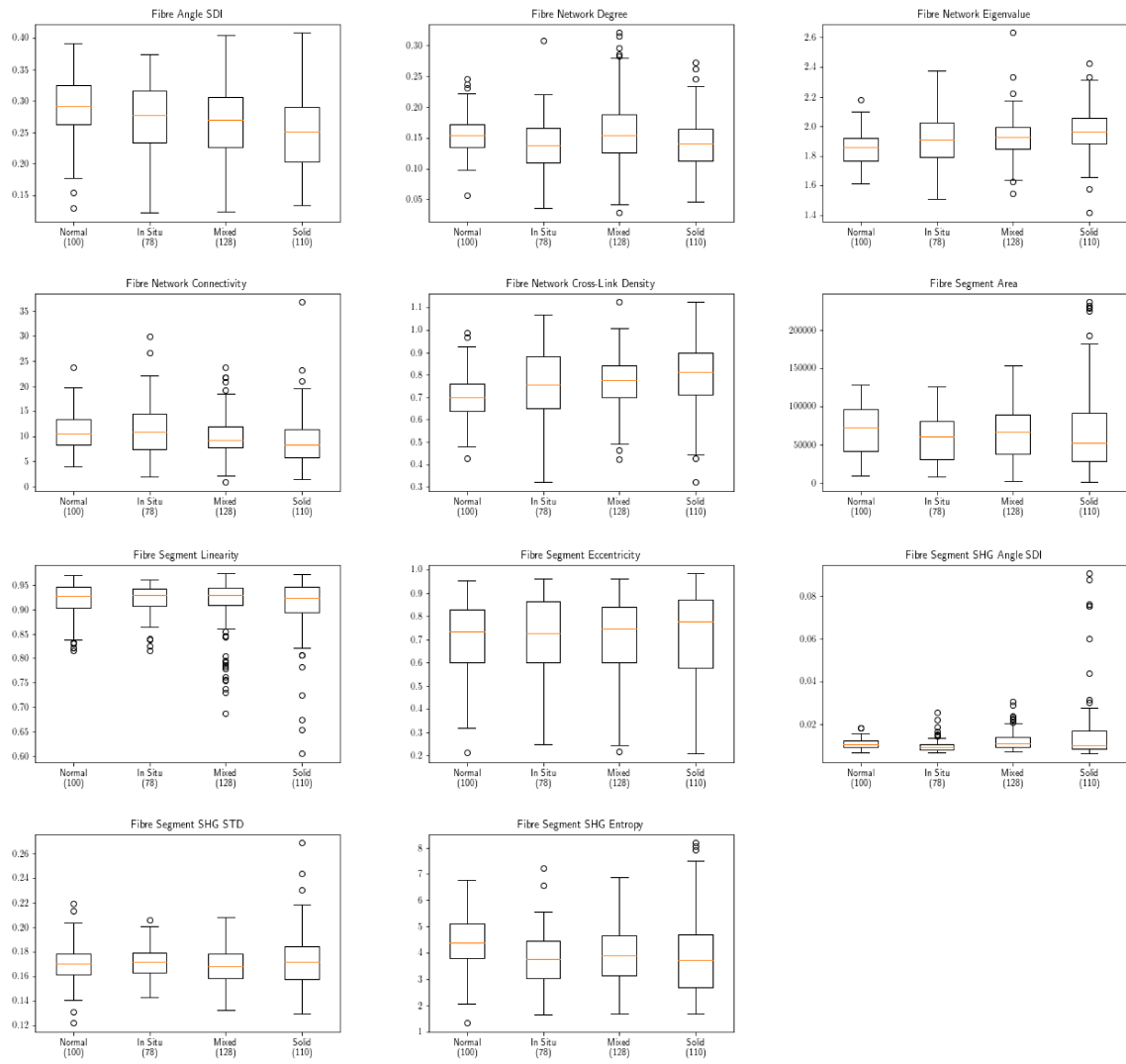


Figura C.2: Gráficos de caixa para as métricas relacionadas as fibras, com as métricas indicadas nos títulos de cada gráfico, para os diagnósticos de tecidos mamários normal e tumoral. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

C.2 Métricas: Tipos histológicos

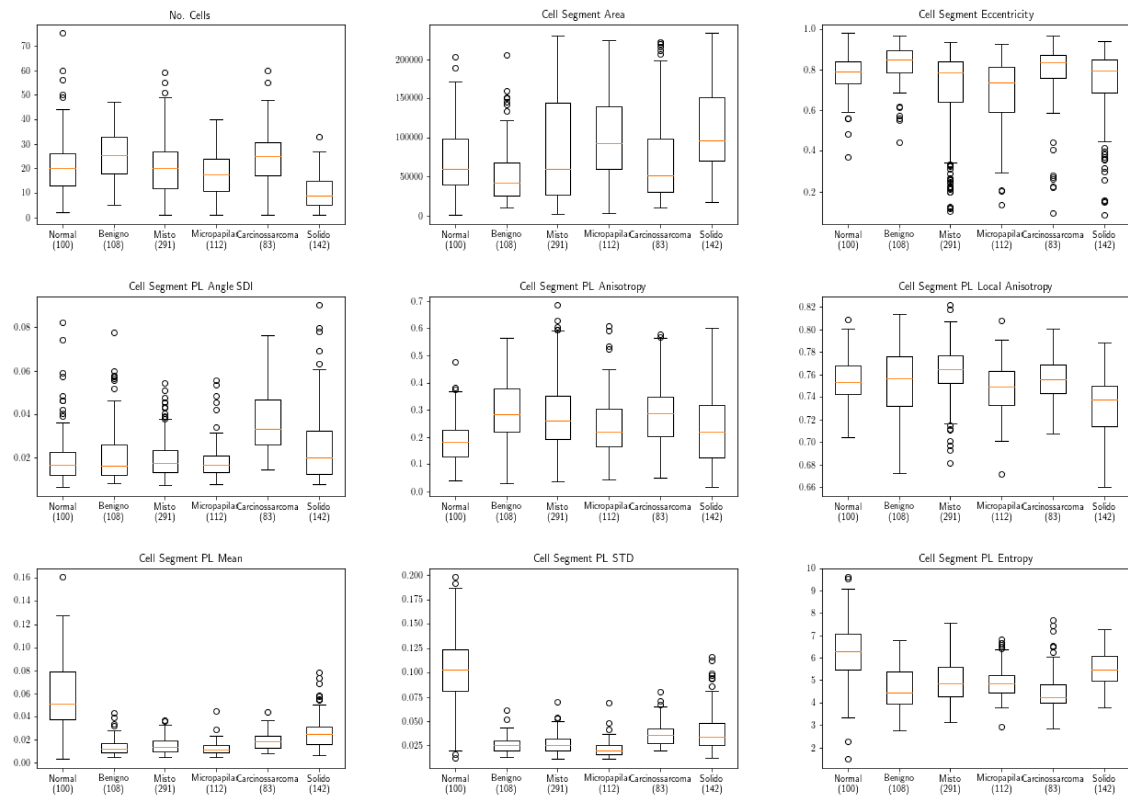


Figura C.3: Gráficos de caixa para as métricas relacionadas à região celular, com as métricas indicadas nos títulos de cada gráfico, para os diagnósticos de tecidos mamários normal e tumoral. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

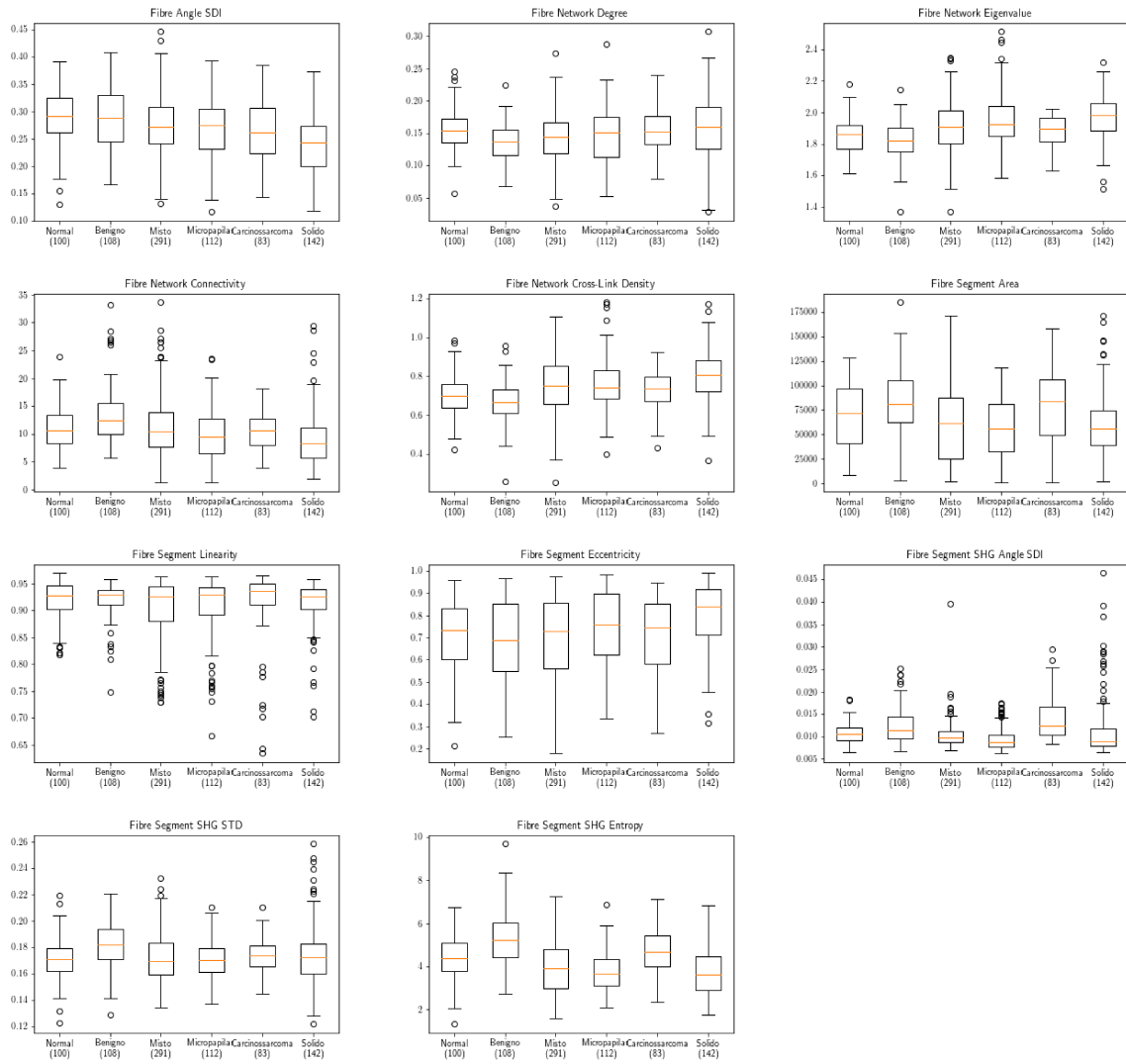


Figura C.4: Gráficos de caixa para as métricas relacionadas as fibras, com as métricas indicadas nos títulos de cada gráfico, para os diagnósticos de tecidos mamários normal e tumoral. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

C.3 Métricas: Progressão Tumoral

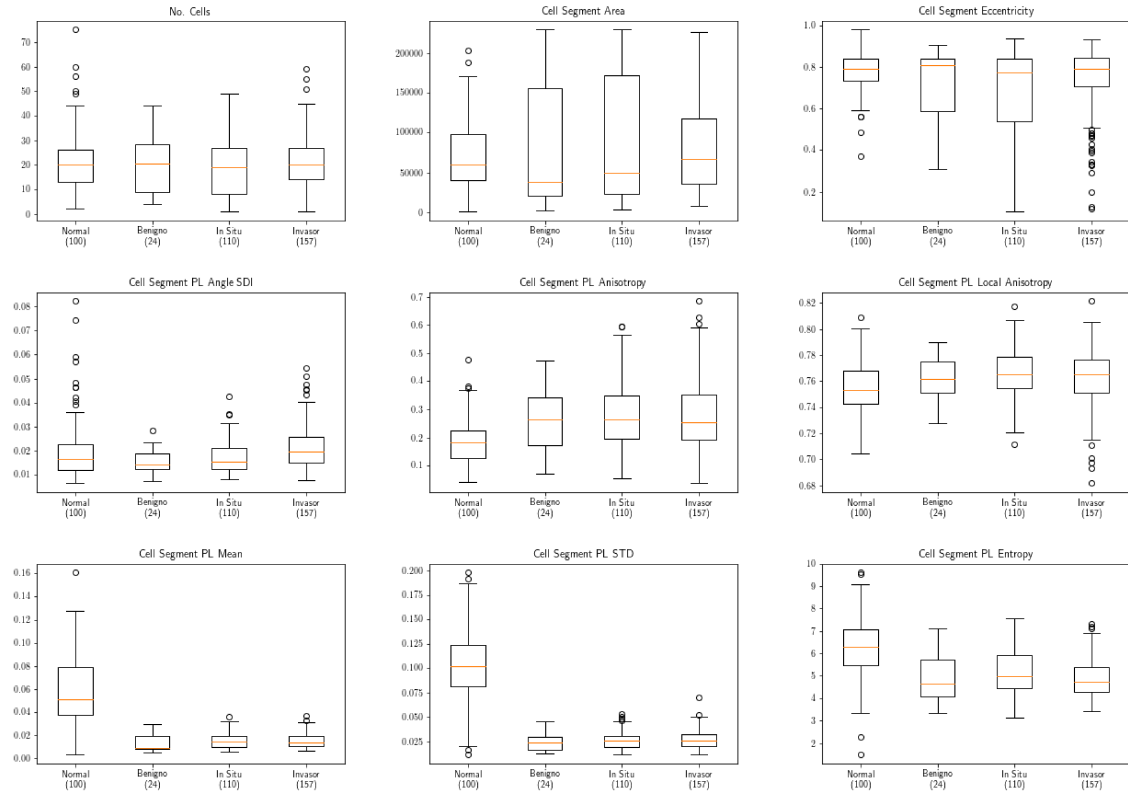


Figura C.5: Gráficos de caixa as métricas relacionadas às regiões celulares, com as métricas indicadas nos títulos dos gráficos, para os diagnósticos: Normal, regiões benignas de carcinoma em tumor misto, regiões *in situ* em carcinoma em tumor misto e regiões invasoras em carcinoma em tumor misto. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

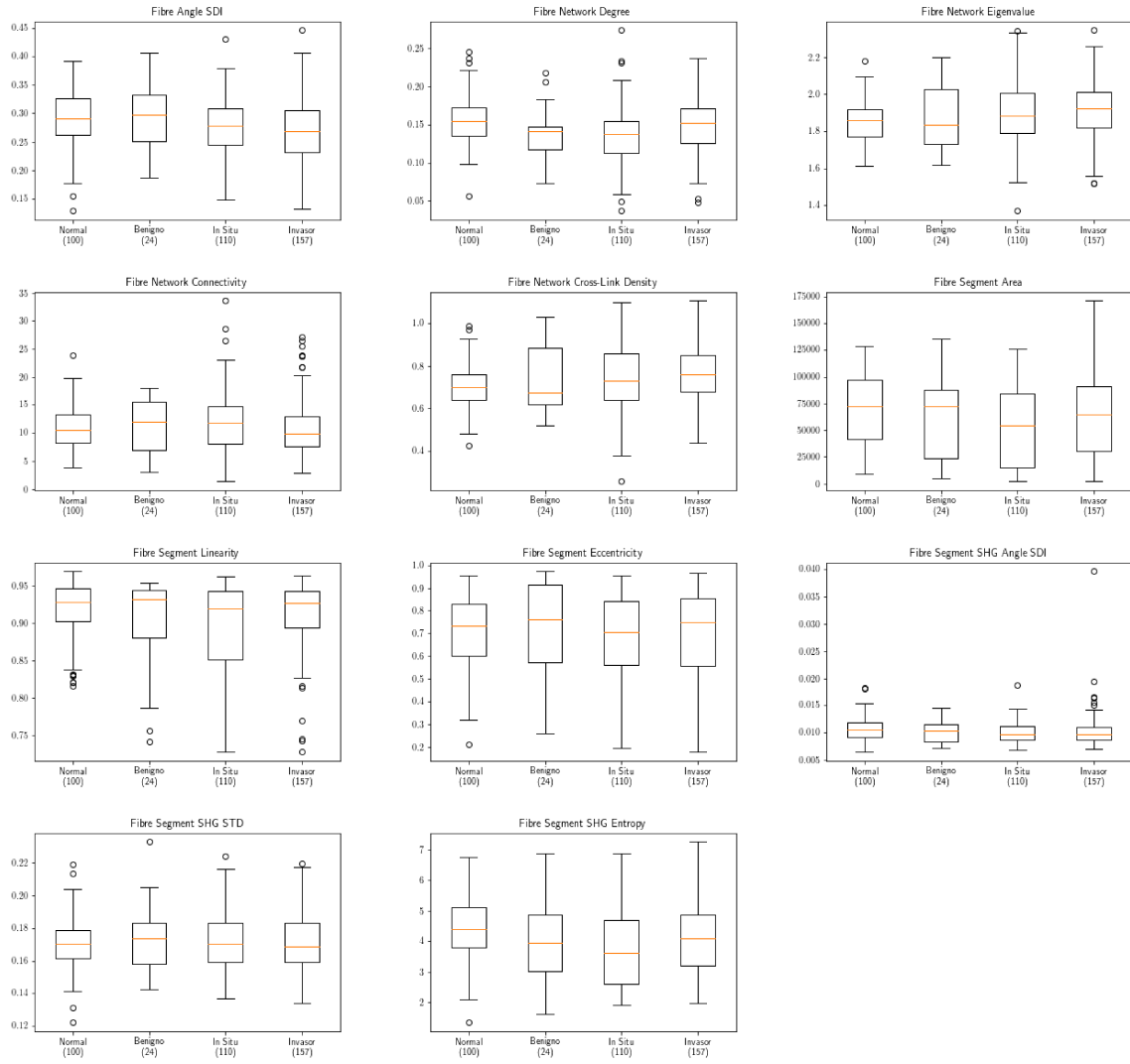


Figura C.6: Gráficos de caixa as métricas relacionadas as fibras, com as métricas indicadas nos títulos dos gráficos, para os diagnósticos: Normal, regiões benignas de carcinoma em tumor misto, regiões *in situ* em carcinoma em tumor misto e regiões invasoras em carcinoma em tumor misto. Os números entre parênteses são o número de imagens adquiridas para cada tipo histológico.

Apêndice D

Publicações

- **Luana A. Reis**, Ana P. V. Garcia, Egleidson F. A. Gomes, Francis G. J. Longford, Jeremy G. Frey, Geovanni D. Cassali, and Ana M. de Paula. *Canine mammary cancer diagnosis from quantitative properties of nonlinear optical images*. Biomed. Opt. Express 11, 6413-6427 (2020). <https://doi.org/10.1364/B0E.400871>
- Garcia, A.P.V., **Reis, L.A.**, Nunes, F.C. et al. *Canine mammary cancer tumour behaviour and patient survival time are associated with collagen fibre characteristics*. Sci Rep 11, 5668 (2021). <https://doi.org/10.1038/s41598-021-85104-w>

Referências Bibliográficas

- [1] Institute of cancer, Brazil, <https://www.inca.gov.br/numeros-de-cancer>. Accessed August 24th, 2021.
- [2] Mark Gray, James Meehan, Carlos Martínez-Pérez, Charlene Kay, Arran K Turnbull, Linda R Morrison, Lisa Y Pang, and David Argyle. Naturally-occurring canine mammary tumors as a translational model for human breast cancer. *Frontiers in Oncology*, 10, 2020.
- [3] Alessandra Estrela-Lima, Márcio SS Araújo, João M Costa-Neto, Andréa Teixeira-Carvalho, Stella M Barrouin-Melo, Sergio V Cardoso, Olindo A Martins-Filho, Rogéria Serakides, and Geovanni D Cassali. Immunophenotypic features of tumor infiltrating lymphocytes from mammary carcinomas in female dogs associated with prognostic factors and survival rates. *BMC Cancer*, 10(1):256, 2010.
- [4] W. Misdorp. *Tumors of the Mammary Gland*, chapter 12, pages 575–606. John Wiley & Sons, Ltd, 2008.
- [5] Paolo Uva, Luigi Aurisicchio, James Watters, Andrey Loboda, Amit Kulkarni, John Castle, Fabio Palombo, Valentina Viti, Giuseppe Mesiti, Valentina Zappulli, et al. Comparative expression pathway analysis of human and canine mammary tumors. *BMC Genomics*, 10(1):135, 2009.
- [6] P Rivera and H Von Euler. Molecular biological aspects on canine and human mammary tumors. *Veterinary Pathology*, 48(1):132–146, 2011.
- [7] Enni Markkanen. Know thy model: Charting molecular homology in stromal reprogramming between canine and human mammary tumors. *Frontiers in Cell and Developmental Biology*, 7:348, 2019.
- [8] Frédérique Nguyen, Laura Peña, Catherine Ibisch, Delphine Loussouarn, Adelina Gama, Natascha Rieder, Anton Belousov, Mario Campone, and Jérôme Abadie. Canine invasive mammary carcinomas as models of human breast cancer. part 1: natural history and prognostic factors. *Breast Cancer Research and Treatment*, 167(3):635–648, 2018.
- [9] Jérôme Abadie, Frédérique Nguyen, Delphine Loussouarn, Laura Peña, Adelina Gama, Natascha Rieder, Anton Belousov, Ingrid Bemelmans, Laëtitia Jaillardon,

- Catherine Ibisch, et al. Canine invasive mammary carcinomas as models of human breast cancer. part 2: immunophenotypes and prognostic significance. *Breast Cancer Research and Treatment*, 167(2):459–468, 2018.
- [10] I Agarwal and L Blanco. Who classification of breast tumors. PathologyOutlines.com website. <https://www.pathologyoutlines.com/topic/breastmalignantwhocclassification.html>. Accessed August 24th, 2021.
- [11] Geovanni D Cassali, Gleidice E Lavallo, Enio Ferreira, Alessandra Estrela-Lima, Andriago B De Nardi, Cristina Ghever, Renata A Sobral, Renne L Amorim, Luciana O Oliveira, Felipe AR Sueiro, et al. Consensus for the diagnosis, prognosis and treatment of canine mammary tumors-2013. *Brazilian Journal of Veterinary Pathology*, 10(3):87–99, 2014.
- [12] Xavier Catteau, Philippe Simon, Michel Jondet, Michel Vanhaeverbeek, and Jean-Christophe Noël. Quantification of stromal reaction in breast carcinoma and its correlation with tumor grade and free progression survival. *PLOS ONE*, 14(3):1–8, 03 2019.
- [13] Paolo P Provenzano, Kevin W Eliceiri, Jay M Campbell, David R Inman, John G White, and Patricia J Keely. Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Medicine*, 4(1):38, 2006.
- [14] Paolo P Provenzano, David R Inman, Kevin W Eliceiri, Justin G Knittel, Long Yan, Curtis T Rueden, John G White, and Patricia J Keely. Collagen density promotes mammary tumor initiation and progression. *BMC Medicine*, 6(1):11, 2008.
- [15] Anders Brabrand, Ian I Kariuki, Monica J Engstrøm, Olav A Haugen, Linda A Dyrnes, Bjørn O Åsvold, Magnus B Lilledahl, and Anna M Bofin. Alterations in collagen fibre patterns in breast cancer. a premise for tumour invasiveness? *APIMIS*, 123(1):1–8, 2015.
- [16] Mariano Bizzarri and Alessandra Cucina. Tumor and the microenvironment: a chance to reframe the paradigm of carcinogenesis? *BioMed Research International*, 2014, 2014.
- [17] Karin Wang, Fei Wu, Bo Ri Seo, Claudia Fischbach, Weisi Chen, Lauren Hsu, and Delphine Gourdon. Breast cancer cells alter the dynamics of stromal fibronectin-collagen interactions. *Matrix Biology*, 60–61:86–95, 2017.
- [18] Kathleen Burke, Ping Tang, and Edward B Brown. Second harmonic generation reveals matrix alterations during breast tumor progression. *Journal of Biomedical Optics*, 18(3):031106, 2012.

- [19] Matthew S Hall, Farid Alisafaei, Ehsan Ban, Xinzeng Feng, Chung-Yuen Hui, Vivek B Shenoy, and Mingming Wu. Fibrous nonlinear elasticity enables positive mechanical feedback between cells and ECMs. *Proceedings of the National Academy of Sciences*, 113(49):14043–14048, 2016.
- [20] Matthew W Conklin, Jens C Eickhoff, Kristin M Riching, Carolyn A Pehlke, Kevin W Eliceiri, Paolo P Provenzano, Andreas Friedl, and Patricia J Keely. Aligned collagen is a prognostic signature for survival in human breast carcinoma. *The American Journal of Pathology*, 178(3):1221–1232, 2011.
- [21] Aloísio M Garcia, Felipe L Magalhes, Jaqueline S Soares, Eduardo Paulino Junior, Mário F R de Lima, Marcelo Mamede, and Ana M de Paula. Second harmonic generation imaging of the collagen architecture in prostate cancer tissue. *Biomedical Physics & Engineering Express*, 4(2):025026, 2018.
- [22] Karissa Tilbury and Paul J Campagnola. Applications of second-harmonic generation imaging microscopy in ovarian and breast cancer. *Perspectives in Medicinal Chemistry*, 7:21–32, 2015.
- [23] Javier Adur, Vitor B Pelegati, Andre A de Thomaz, Mariana O Baratti, Liliana A L A Andrade, Hernandes F Carvalho, Fátima Bottcher-Luiz, and Carlos Lenz Cesar. Second harmonic generation microscopy as a powerful diagnostic imaging modality for human ovarian cancer. *Journal of Biophotonics*, 7(1-2):37–48, 2014.
- [24] Luana A Reis, Ana PV Garcia, Egleidson FA Gomes, Francis GJ Longford, Jeremy G Frey, Giovanni D Cassali, and Ana M de Paula. Canine mammary cancer diagnosis from quantitative properties of nonlinear optical images. *Biomedical Optics Express*, 11(11):6413–6427, 2020.
- [25] Adam E Tuer, Serguei Krouglov, Nicole Prent, Richard Cisek, Daaf Sandkuijl, Kazuhiro Yasufuku, Brian C Wilson, and Virginijus Barzda. Nonlinear optical properties of type I collagen fibers studied by polarization dependent second harmonic generation microscopy. *The Journal of Physical Chemistry B*, 115(44):12759–12769, 2011.
- [26] Paul J Campagnola and Leslie M Loew. Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. *Nature Biotechnology*, 21(11):1356–1360, 2003.
- [27] Paul Campagnola. Second harmonic generation imaging microscopy: applications to diseases diagnostics. *Analytical Chemistry*, 83(9):3224–3231, 2011.
- [28] Xiyi Chen, Oleg Nadiarynkh, Sergey Plotnikov, and Paul J Campagnola. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nature Protocols*, 7(4):654–669, 2012.

- [29] Zhi-Hua Zhou, Cheng-Dong Ji, Hua-Liang Xiao, Hai-Bin Zhao, You-Hong Cui, and Xiu-Wu Bian. Reorganized collagen in the tumor microenvironment of gastric cancer and its association with prognosis. *Journal of Cancer*, 8(8):1466, 2017.
- [30] Yuming Liu, Adib Keikhosravi, Guneet S Mehta, Cole R Drifka, and Kevin W Eliceiri. Methods for quantifying fibrillar collagen alignment. In *Fibrosis*, pages 429–451. Springer, 2017.
- [31] Ana PV Garcia, Luana A Reis, Fernanda C Nunes, Francis GJ Longford, Jeremy G Frey, Ana M de Paula, and Geovanni D Cassali. Canine mammary cancer tumour behaviour and patient survival time are associated with collagen fibre characteristics. *Scientific Reports*, 11(1):1–13, 2021.
- [32] Jeremy S Bredfeldt, Yuming Liu, Matthew W Conklin, Patricia J Keely, Thomas R Mackie, and Kevin W Eliceiri. Automated quantification of aligned collagen for human breast carcinoma prognosis. *Journal of Pathology Informatics*, 5, 2014.
- [33] Sara L. Best, Yuming Liu, Adib Keikhosravi, Cole R. Drifka, Kaitlin M. Woo, Guneet S. Mehta, Marie Altwegg, Terra N. Thimm, Matthew Houlihan, Jeremy S. Bredfeldt, E. Jason Abel, Wei Huang, and Kevin W. Eliceiri. Collagen organization of renal cell carcinoma differs between low and high grade tumors. *BMC Cancer*, 19(1):490, May 2019.
- [34] Matthew W Conklin, Ronald E Gangnon, Brian L Sprague, Lisa Van Gemert, John M Hampton, Kevin W Eliceiri, Jeremy S Bredfeldt, Yuming Liu, Nuntida Surachacharn, Polly A Newcomb, Andreas Friedl, Patricia J Keely, and Amy Trentham-Dietz. Collagen alignment as a predictor of recurrence after ductal carcinoma in situ. *Cancer Epidemiol Biomarkers Prev.*, 27(2):138–145, 2018.
- [35] Rodrigo de Andrade Natal, Geisilene R Paiva, Vitor B Pelegati, Ludwing Marengo, César A Alvarenga, Renato F Vargas, Sophie F Derchain, Luis O Sarian, Camille Franchet, Carlos L Cesar, et al. Exploring collagen parameters in pure special types of invasive breast cancer. *Scientific Reports*, 9(1):7715, 2019.
- [36] Rodrigo A Natal, José Vassallo, Geisilene R Paiva, Vitor B Pelegati, Guilherme O Barbosa, Guilherme R Mendonça, Caroline Bondarik, Sophie F Derchain, Hernandes F Carvalho, Carmen S Lima, et al. Collagen analysis by second-harmonic generation microscopy predicts outcome of luminal breast cancer. *Tumor Biology*, 40(4):1010428318770953, 2018.
- [37] P A Franken, Alan E Hill, C W Peters, and G Weinreich. Generation of optical harmonics. *Physical Review Letters*, 7(4):118, 1961.
- [38] Robert W Boyd. *Nonlinear Optics*. Academic Press, 2003.

- [39] Maria Göppert-Mayer. Über elementarakte mit zwei quantensprüngen. *Annalen der Physik*, 401(3):273–294, 1931.
- [40] W. Kaiser and C. G. B. Garrett. Two-photon Excitation in CaF_2 : Eu^{2+} . *Phys. Rev. Lett.*, 7:229–231, Sep 1961.
- [41] Winfried Denk, James H Strickler, and Watt W Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76, 1990.
- [42] Mariacristina Rumi and Joseph W Perry. Two-photon absorption: an overview of measurements and principles. *Advances in Optics and Photonics*, 2(4):451–518, 2010.
- [43] Peter T C So, Chen Y Dong, Barry R Masters, and Keith M Berland. Two-photon excitation fluorescence microscopy. *Annual Review of Biomedical Engineering*, 2(1):399–429, 2000.
- [44] Aloisio Miguel Garcia. Fluorescência por absorção de dois fótons em corantes com aplicações biológicas. Master’s thesis, Universidade Federal de Minas Gerais, 2013.
- [45] Michael Rubart. Two-photon microscopy of cells and tissue. *Circulation Research*, 95(12):1154–1166, 2004.
- [46] André Luiz Coelho Conceição. *Caracterização estrutural de tecidos mamários normais e neoplásicos através de espalhamento de raios X*. PhD thesis, Universidade de São Paulo, 2008.
- [47] M Bradley Calobrace, Bill G Kortesis, Gaurav Bharti, and Chet Mays. *Augmentation Mastopexy: Mastering the Art in the Management of the Ptotic Breast*. Springer Nature, 2020.
- [48] Daniel J Ledbetter and Paul RV Johnson. *Endocrine Surgery in Children*, volume 118. Springer, 2018.
- [49] Sonali Pandya and Richard G Moore. Breast development and anatomy. *Clinical Obstetrics and Gynecology*, 54(1):91–95, 2011.
- [50] Luiz CU JUNQUEIRA and José CARNEIRO. *Histologia Básica. Texto e Atlas*. 11ª edição. 2008.
- [51] Kerstin Lindblad-Toh, Claire M Wade, Tarjei S Mikkelsen, Elinor K Karlsson, David B Jaffe, Michael Kamal, Michele Clamp, Jean L Chang, Edward J Kulbokas, Michael C Zody, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069):803–819, 2005.

- [52] Geovanni Dantas Cassali, Angélica Cavalheiro Bertagnolli, Enio Ferreira, Karine Araújo Damasceno, Conrado de Oliveira Gamba, and Cecília Bonolo de Campos. Canine mammary mixed tumours: a review. *Veterinary Medicine International*, 2012:398–404, 2012.
- [53] Ana Paula Abdalla Goissis. *Análise estrutural do colágeno do tipo I-correlação estrutura: atividade biológica*. PhD thesis, Universidade de São Paulo, 2007.
- [54] Christian Frantz, Kathleen M Stewart, and Valerie M Weaver. The extracellular matrix at a glance. *Journal of Cell Science*, 123(24):4195–4200, 2010.
- [55] Jonathan N Ouellette, Cole R Drifka, Kelli B Pointer, Yuming Liu, Tyler J Lieberthal, W John Kao, John S Kuo, Agnes G Loeffler, and Kevin W Eliceiri. Navigating the collagen jungle: the biomedical potential of fiber organization in cancer. *Bioengineering*, 8(2):17, 2021.
- [56] Alfonso Gautieri, Simone Vesentini, Alberto Redaelli, and Markus J Buehler. Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up. *Nano Letters*, 11(2):757–766, 2011.
- [57] Jiexiang Lin, Yanping Shi, Yutao Men, Xin Wang, Jinduo Ye, and Chunqiu Zhang. Mechanical roles in formation of oriented collagen fibers. *Tissue Engineering Part B: Reviews*, 26(2):116–128, 2020.
- [58] Sylvie Ricard-Blum. The collagen family. *Cold Spring Harbor Perspectives in Biology*, 3(1):a004978, 2011.
- [59] Peter Fratzl. *Collagen: structure and mechanics, an introduction*. Springer, 2008.
- [60] Barbara Brodsky and John A M Ramshaw. The collagen triple-helix structure. *Matrix Biology*, 15(8-9):545–554, 1997.
- [61] Arnab Bhattacharjee and Manju Bansal. Collagen structure: the madras triple helix and the current scenario. *IUBMB Life*, 57(3):161–172, 2005.
- [62] Matthew D Shoulders and Ronald T Raines. Collagen structure and stability. *Annual Review of Biochemistry*, 78:929–958, 2009.
- [63] Markus J Buehler. Nature designs tough collagen: explaining the nanostructure of collagen fibrils. *Proceedings of the National Academy of Sciences*, 103(33):12285–12290, 2006.
- [64] Shengjie Ling, Wenshuai Chen, Yimin Fan, Ke Zheng, Kai Jin, Haipeng Yu, Markus J Buehler, and David L Kaplan. Biopolymer nanofibrils: Structure, modeling, preparation, and applications. *Progress in Polymer Science*, 85:1–56, 2018.

- [65] Isaac Freund, Moshe Deutsch, and Aaron Sprecher. Connective tissue polarity. optical second-harmonic microscopy, crossed-beam summation, and small-angle scattering in rat-tail tendon. *Biophysical Journal*, 50(4):693–712, 1986.
- [66] Paul J Campagnola, Andrew C Millard, Mark Terasaki, Pamela E Hoppe, Christian J Malone, and William A Mohler. Three-dimensional high-resolution second-harmonic generation imaging of endogenous structural proteins in biological tissues. *Biophysical Journal*, 82(1):493–508, 2002.
- [67] K Burke, Marcel Smid, RP Dawes, MA Timmermans, P Salzman, CHM van Deurzen, David G Beer, JA Foekens, and E Brown. Using second harmonic generation to predict patient outcome in solid tumors. *BMC Cancer*, 15(1):929, 2015.
- [68] Javier Adur, Vitor B. Pelegati, Andre A. de Thomaz, Mariana O. Baratti, Diogo B. Almeida, L. A. L. A. Andrade, Fátima Bottcher-Luiz, Hernandez F. Carvalho, and Carlos L. Cesar. Optical biomarkers of serous and mucinous human ovarian tumor assessed with nonlinear optics microscopies. *Plos One*, 7(10):1–13, 10 2012.
- [69] Marc de Wergifosse, Jérôme de Ruyck, and Benoît Champagne.
- [70] Ariane Deniset-Besseau, Julien Duboisset, Emmanuel Benichou, François Hache, Pierre-François Brevet, and Marie-Claire Schanne-Klein. Measurement of the second-order hyperpolarizability of the collagen triple helix and determination of its physical origin. *The Journal of Physical Chemistry B*, 113(40):13437–13445, 2009.
- [71] Francis G.J. Longford. Pyfibre: Python fibrous image analysis toolkit. <https://github.com/franklongford/PyFibre>, Version 2, (2020).
- [72] Jeremy S Bredfeldt, Yuming Liu, Carolyn A Pehlke, Matthew W Conklin, Joseph M Szulczewski, David R Inman, Patricia J Keely, Robert D Nowak, Thomas R Mackie, and Kevin W Eliceiri. Computational segmentation of collagen fibers from second-harmonic generation images of breast cancer. *Journal of Biomedical Optics*, 19(1):016007, 2014.
- [73] Rana Rezakhaniha, Aristotelis Agianniotis, Jelle Tymen Christiaan Schrauwen, Alessandra Griffa, Daniel Sage, CVC vd Bouten, FN Van de Vosse, Michaël Unser, and Nikolaos Stergiopoulos. Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy. *Biomechanics and Modeling in Mechanobiology*, 11(3-4):461–473, 2012.
- [74] David Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pages 1177–1178, 2010.

- [75] Yoshinobu Sato, Shin Nakajima, Nobuyuki Shiraga, Hideki Atsumi, Shigeyuki Yoshida, Thomas Koller, Guido Gerig, and Ron Kikinis. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Medical Image Analysis*, 2(2):143–168, 1998.
- [76] Hiroki Sayama. *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks, 2015.
- [77] Andrew M Stein, David A Vader, Louise M Jawerth, David A Weitz, and Leonard M Sander. An algorithm for extracting the network geometry of three-dimensional collagen gels. *Journal of Microscopy*, 232(3):463–475, 2008.
- [78] Bernd Jahne. *Digital image processing*, volume 4. Springer Berlin, 2005.
- [79] Francisco J Ávila and Juan M Bueno. Analysis and quantification of collagen organization with the structure tensor in second harmonic microscopy images of ocular tissues. *Applied Optics*, 54(33):9848–9854, 2015.
- [80] Francisco J Ávila, Oscar del Barco, and Juan M Bueno. Polarization response of second-harmonic images for different collagen spatial distributions. *Journal of Biomedical Optics*, 21(6):066015, 2016.
- [81] Francisco J Ávila, Oscar Del Barco, and Juan M Bueno. Quantifying external and internal collagen organization from stokes-vector-based second harmonic generation imaging polarimetry. *Journal of Optics*, 19(10):105301, 2017.
- [82] Gabriela Barbosa Guedes Pereira, Lucas Pereira Fernandes, José Maurício Ramos de Souza Neto, Helon David de Macêdo Braz, and Leandro da Silva Sauer. A comparative study of linear discriminant analysis and an artificial neural network performances in breast cancer diagnosis. In *2020 IEEE ANDESCON*, pages 1–6. IEEE, 2020.
- [83] Mesut Toğaçar, Burhan Ergen, and Zafer Cömert. Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders. *Medical Hypotheses*, 135:109503, 2020.
- [84] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.
- [85] Aleix M Martínez and Avinash C Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.
- [86] Keinosuke Fukunaga. *Statistical pattern recognition*. World Scientific, 1993.

- [87] Sebastian Raschka. Linear discriminant analysis – bit by bit. https://sebastianraschka.com/Articles/2014_python_lda.html#\T1\textendash-bit-by-bit, (2014).
- [88] Srishti Sawla. Linear discriminant analysis. <https://medium.com/@srishtisawla/linear-discriminant-analysis-d38decf48105>, (2018).
- [89] Clewton Luís Ferreira da Fonseca et al. A técnica de análise do discriminante linear (lda) no estudo da composição química dos raios cósmicos primários em uma parametrização fenomenológica do perfil longitudinal de chuueiros atmosféricos extensos. Master’s thesis, Intituto de Física Gleg Wataghin - UNICAMP.
- [90] Daniel Pelliccia. Classification of nir spectra by linear discriminant analysis in python. <https://nirpyresearch.com/classification-nir-spectra-linear-discriminant-analysis-python/>, (2018).
- [91] Scikit-learn documentation. https://scikit-learn.org/stable/modules/cross_validation.html.
- [92] Scikit-learn documentation. https://scikit-learn.org/stable/modules/feature_selection.html#rfe.
- [93] G Falzon, S Pearson, and R Murison. Analysis of collagen fibre shape changes in breast cancer. *Physics in Medicine & Biology*, 53(23):6641, 2008.
- [94] Tord Hompland, Arne Erikson, Mikael Lindgren, Tore Lindmo, and Catharina de Lange Davies. Second-harmonic generation in collagen as a potential cancer diagnostic parameter. *Journal of Biomedical Optics*, 13(5):054050, 2008.
- [95] Craig E Barcus, Kathleen A O’Leary, Jennifer L Brockman, Debra E Rugowski, Yuming Liu, Nancy Garcia, Menggang Yu, Patricia J Keely, Kevin W Eliceiri, and Linda A Schuler. Elevated collagen-i augments tumor progressive signals, intravasation and metastasis of prolactin-induced estrogen receptor alpha positive mammary tumor cells. *Breast Cancer Research*, 19(1):1–13, 2017.
- [96] Ashley Case, Becky K. Brisson, Amy C. Durham, Suzanne Rosen, James Monslow, Elizabeth Buza, Pascale Salah, Julie Gillem, Gordon Ruthel, Sridhar Veluvolu, Veronica Kristiansen, Ellen Puré, Dorothy C. Brown, Karin U. Sørenmo, and Susan W. Volk. Identification of prognostic collagen signatures and potential therapeutic stromal targets in canine mammary gland carcinoma. *PLOS ONE*, 12(7):1–19, 07 2017.
- [97] Xiaoxing Han, Ryan M Burke, Martha L Zettel, Ping Tang, and Edward B Brown. Second harmonic properties of tumor collagen: determining the structural relationship between reactive stroma and healthy stroma. *Optics Express*, 16(3):1846–1859, 2008.

- [98] Ara Ghazaryan, Halley F Tsai, Gor Hayrapetyan, Wei-Liang Chen, Yang-Fang Chen, Myung-Yung Jeong, Chang-Seok Kim, Shean-Jen Chen, and Chen-Yuan Dong. Analysis of collagen fiber domain organization by fourier second harmonic generation microscopy. *Journal of Biomedical Optics*, 18(3):031105, 2012.
- [99] Yuankai K Tao, Dejun Shen, Yuri Sheikine, Osman O Ahsen, Helen H Wang, Daniel B Schmolze, Nicole B Johnson, Jeffrey S Brooker, Alex E Cable, James L Connolly, et al. Assessment of breast pathologies using nonlinear microscopy. *Proceedings of the National Academy of Sciences*, 111(43):15304–15309, 2014.
- [100] Douglas C Montgomery and George C Runger. *Estatística Aplicada e Probabilidade para Engenheiros*. 2^a. Ed. Rio de Janeiro: Editora LTC, 2003.
- [101] Geovanni D. Cassali, Karine A. Damasceno, Angélica C. Bertagnolli, Alessandra Estrela-Lima, Gleidice E. Lavallo, Giovana W. Di Santis, Andriago B. De Nardi, Cristina G Fernandes, Bruno Cogliati, Renata Sobral, Fernanda V. Amorim da Costa, Enio Ferreira, Breno S. Salgado, Cecilia B. Campos, Mario J. M. H. D’Assis, Laís P. Silva, Marília C. A. Machado, Bruna F. Firmo, Fernanda C. Nunes, and Karen Y. R. Nakagaki. Consensus regarding the diagnosis, prognosis and treatment of canine mammary tumors: benign mixed tumors, carcinomas in mixed tumors and carcinosarcomas. *Brazilian Journal of Veterinary Pathology*, 7(2):38–69, 2017.
- [102] Geovanni Dantas Cassali, Angélica Cavaleiro Bertagnolli, Enio Ferreira, Karine Araújo Damasceno, Conrado de Oliveira Gamba, and Cecília Bonolo de Campos. Canine mammary mixed tumours: a review. *Veterinary Medicine International*, 2012:1–7.
- [103] KA Damasceno, E Ferreira, A Estrela Lima, Y Bosco, LP Silva, ALB Barros, AC Bertagnolli, and Geovanni Cassali Dantas. Relationship between the expression of versican and egfr, her-2, her-3 and cd44 in matrix-producing tumours in the canine mammary gland. *Histology and Histopathology: Cellular and Molecular Biology*, 31(6):675–688, 2016.
- [104] Kun Zhang, Callie A Corsa, Suzanne M Ponik, Julie L Prior, David Piwnica-Worms, Kevin W Eliceiri, Patricia J Keely, and Gregory D Longmore. The collagen receptor discoidin domain receptor 2 stabilizes snail1 to facilitate breast cancer metastasis. *Nature Cell Biology*, 15(6):677–687, 2013.
- [105] Peter Friedl and E-B Bröcker. The biology of cell locomotion within three-dimensional extracellular matrix. *Cellular and Molecular Life Sciences CMLS*, 57(1):41–64, 2000.
- [106] Billy Samuel Hill, Annachiara Sarnella, Giuliana D’Avino, and Antonella Zannetti. Recruitment of stromal cells into tumour microenvironment promote the metastatic

- spread of breast cancer. In *Seminars in Cancer Biology*, volume 60, pages 202–213. Elsevier, 2020.
- [107] Hendrik Ungefroren, Susanne Sebens, Daniel Seidl, Hendrik Lehnert, and Ralf Hass. Interaction of tumor cells with the microenvironment. *Cell Communication and Signaling*, 9(1):1–8, 2011.
- [108] Sen Guo and Chu-Xia Deng. Effect of stromal cells in tumor microenvironment on metastasis initiation. *International Journal of Biological Sciences*, 14(14):2083, 2018.