



INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Mestrado Profissional em Tecnologia das Radiações em Ciências da Saúde

**Desenvolvimento De Uma Metodologia Para Radiômica Em
Mamografia Digital**

ÉRICA MAURICIO POLICARPO

**Dissertação apresentada como parte
dos requisitos para obtenção do Grau de
Mestre Profissional em Tecnologia das
Radiações em Ciências da Saúde na Área de
Concentração
Processos de Radiação na Saúde**

Orientador:

Prof. Dr. Mário Olímpio de Menezes

Coorientador:

Prof. Dr. Orlando Rodrigues Jr.

**São Paulo
2024**

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Mestrado Profissional em Tecnologia das Radiações em Ciências da Saúde

**Desenvolvimento De Uma Metodologia Para Radiômica Em Mamografia
Digital**

ÉRICA MAURICIO POLICARPO

**Dissertação apresentada como parte dos
requisitos para obtenção do Grau de Mestre
Profissional em Tecnologia das Radiações em
Ciências da Saúde na Área de Concentração
Processos de Radiação na Saúde**

Orientador:

Prof. Dr. Mário Olímpio de Menezes

Coorientador:

Prof. Dr. Orlando Rodrigues Jr.

**São Paulo
2024**



Autorizo a reprodução e divulgação total ou parcial deste trabalho, para fins de estudo e pesquisa, desde que citada a fonte.

Como citar:

MAURICIO POLICARPO, E. ***Desenvolvimento De Uma Metodologia Para Radiômica Em Mamografia Digital***. 2024. 101 f. Dissertação (Mestrado Profissional em Tecnologia das Radiações em Ciências da Saúde), Instituto de Pesquisas Energéticas e Nucleares, IPEN-CNEN, São Paulo. Disponível em: <<http://repositorio.ipen.br/>> (data de consulta no formato: dd/mm/aaaa)

Ficha catalográfica elaborada pelo Sistema de geração automática da Biblioteca IPEN,
com os dados fornecidos pelo(a) autor(a).

Mauricio Policarpo, Erica
Desenvolvimento De Uma Metodologia Para Radiômica Em
Mamografia Digital / Erica Mauricio Policarpo; orientador Mário
Olimpio de Menezes; co-orientador Orlando Rodrigues Junior. --
São Paulo, 2024.
105 f.

Dissertação (Mestrado Profissional) - Programa de
Pós-Graduação em Tecnologia das Radiações em Ciências da Saúde
(Processos de Radiação na Saúde) -- Instituto de Pesquisas
Energéticas e Nucleares, São Paulo, 2024.

1. Radiomics. 2. Mamografia Digital. 3. Imagens
Médicas. 4. Processamento de Imagens. 5. Programas de
Computador. I. Olímpio de Menezes, Mário, orient. II. Rodrigues
Junior, Orlando, co-orient. III. Título.

FOLHA DE APROVAÇÃO

Autor: Érica Mauricio Policarpo

Título: Desenvolvimento De Uma Metodologia Para Radiômica Em Mamografia Digital

Dissertação apresentada como parte dos requisitos para obtenção do Grau de Mestre pelo programa de Pós Graduação Stricto-Sensu – Mestrado Profissional de Tecnologia das Radiações na Saúde ao IPEN-CNEN/SP na área de concentração de Processos de Radiação na Saúde

Data: ____ / ____ / ____

Banca Examinadora

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

AGRADECIMENTOS

Agradeço aos meus pais que sempre me ensinaram que o conhecimento é a chave para sermos seres humanos melhores, e sempre se dedicaram e me incentivaram ao estudo.

Ao meu querido companheiro Henrique, pela paciência e pelos finais de semana, feriados e férias sempre me deixando ao computador com a motivação de concluir o mestrado.

Aos meus queridos amigos do Mestrado Profissional, sempre fomos uma equipe forte, onde sempre houve tempo e disposição para ajudar e incentivar uns aos outros!

A professora dra. Kellen Adriana Curci Daros, por dividir todo seu conhecimento, por me ensinar como fazer pesquisa e me tornar uma profissional melhor, o meu eterno agradecimento.

A Universidade Federal de São Paulo, minha segunda casa, e onde foi possível desenvolver este trabalho.

Aos meus orientadores dr. Mário e dr. Orlando, que me apoiaram e acompanharam nesta jornada! O meu eterno agradecimento!

Ao IPEN-CNEN por promover os meios necessários para meu desenvolvimento intelectual e profissional, a todos os professores do Mestrado Profissional pela dedicação e ensinamentos, e as secretárias Andrea, Sabrina e Graziella sempre muito atenciosas, e com todas as respostas para nossas dúvidas e preocupações.

RESUMO

POLICARPO, Érica M. *Desenvolvimento de uma metodologia para radiômica em mamografia digital*. 2024. Dissertação (Mestrado Profissional em Tecnologia das Radiações em Ciências da Saúde) – Instituto de Pesquisa Energéticas e Nucleares - IPEN – CNEN/SP. São Paulo.

O câncer de mama é o tipo mais comum em mulheres, e é o responsável pela maioria das mortes por câncer. Seu prognóstico ainda depende de seu diagnóstico precoce. Em mamografia digital, apesar de ser considerado o método com maior sensibilidade e especificidade para o rastreamento do câncer de mama, detectar tumores em fase inicial ainda é uma tarefa difícil, dado ao fato das variações que o tecido mamário pode ter. A introdução de tecnologias de aquisição de imagens médicas no formato digital possibilitou que mais ferramentas, como a radiômica, auxiliem na análise e identificação de possíveis patologias. Neste trabalho foi desenvolvido uma estratégia para o planejamento de análises por radiômica e padronização da imagem mamográfica utilizando as ferramentas disponíveis para tratamento de imagens do software livre 3D Slicer, seguindo as recomendações do IBSI. Foram analisados os resultados da padronização de imagens de um simulador padrão que foram adquiridas em três equipamentos diferentes de mamografia, através do cálculo das características radiômicas dos test objects presentes no simulador. Das 108 características radiômicas extraídas dos test objects das imagens obtidas dos três equipamentos, 23 apresentaram coeficientes de variação menores de 15% para o test object fibra, 25 para o test object microcalcificação e massa. A mesma estratégia foi empregada em imagens de humanos e os valores calculados foram comparados com os resultados obtidos nos test objects, e 7 características apresentaram resultados semelhantes para fibra, 8 para microcalcificação e 7 para massa. Foi concluído que a estratégia empregada apresentou resultados satisfatórios para a padronização de imagens mamográficas para análises radiômicas.

Palavras chaves: Radiomics, mamografia digital, imagens médicas, processamento de imagens e programas de computador.

ABSTRACT

POLICARPO, Érica M. **Development of a methodology for radiomics in digital mammography**. 2024. Dissertação (Mestrado Profissional em Tecnologia das Radiações em Ciências da Saúde) – Instituto de Pesquisa Energéticas e Nucleares - IPEN – CNEN/SP. São Paulo.

Breast cancer is the most common type among women and accounts for the majority of cancer-related deaths. Its prognosis still relies on early diagnosis. In digital mammography, although it is considered the method with the highest sensitivity and specificity for breast cancer screening, detecting tumors at an early stage remains a challenging task due to the variations in breast tissue. The introduction of digital medical imaging acquisition technologies has enabled additional tools, such as radiomics, to assist in the analysis and identification of potential pathologies. This study developed a strategy for planning radiomic analyses and standardizing mammographic images using available image processing tools from the open-source software 3D Slicer, in accordance with the recommendations of the IBSI. The results of the standardization of images from a standard mammography phantom, acquired from three different digital mammography equipment, were analyzed through the calculation of radiomic features of the test objects present in the simulator. Out of 108 radiomic features extracted from the test objects in the images obtained from the three equipment, 23 showed coefficients of variation less than 15% for the fiber test object, and 25 for both the microcalcification and mass test objects. The same strategy was applied to human images, and the calculated values for the radiomic features were compared with the results obtained from the test objects, with 7 features showing similar results for fiber, 8 for microcalcifications, and 7 for mass. It was concluded that the employed strategy yielded satisfactory results for the standardization of mammographic images for radiomic analyses.

Key words: Radiomics, digital mammography, medical images, image processing and computer programming.

LISTA DE TABELAS

Tabela 1 – Características dos equipamentos de mamografia digital.....	41
Tabela 2 – Descrição dos test objects presentes no simulador ACR.....	42
Tabela 3 – Parâmetros de aquisição modo automático.....	43
Tabela 4 – CV das características radiômicas, onde: A – entre imagens do simulador sem processamento pós aquisição; B – entre imagens do simulador com reamostragem dos pixels; C – entre imagens do simulador com reamostragem dos pixels e correspondência do histograma.	56
Tabela 5 – Informações da imagem após processamento no domínio espacial. ...	58
Tabela 6 - Resultados do ICC para as medidas realizadas no test object fibra.....	62
Tabela 7 - Resultados do ICC para as medidas realizadas no test object microcalcificação.....	67
Tabela 8 - Resultados do ICC para as medidas realizadas no test object Massa. .	71
Tabela 9 – Valores das características do simulador ACR e do IBSI.	72
Tabela 10 – Características retiradas da análise por apresentarem correlação com outras.	82
Tabela 11 – Características elegíveis para análises por radiômica em mamografia digital.....	83

LISTA DE FIGURAS

Figura 1 – Ilustração de um tubo de raios X e equipamento dedicado ao exame de mamografia.....	23
Figura 2 – Ilustração da leitura à laser para sistemas CR.....	25
Figura 3 – Ilustração de receptor de radiografia digital.	26
Figura 4 - Geometria de varredura, posição dos colimadores e do detector do contador de fótons.	27
Figura 5 – Representação da função de coordenadas e das intensidades de níveis de cinza em imagem digital.	28
Figura 6 – Exemplo da perda da resolução espacial quando aplicadas operações de processamentos de imagens digitais, em A imagem com uma matriz de pixels de 4915x5355; em B modificada para 80x87.....	29
Figura 7 – Imagem do simulador em 16 bits em A, e depois convertida para 8 bits em B.	30
Figura 8 – Histograma da imagem do simulador utilizado nesta dissertação.	31
Figura 9 – Informações da imagem após reamostragem por interpolação linear..	32
Figura 10 – Tela do 3D Slicer do módulo de segmentação por limiarização (Thresholding).....	34
Figura 11 – Tela do sistema CAD evidenciando regiões com alterações sugestivas de patologia.....	36
Figura 12 – Sequência de eventos da análise Radiômica.	37
Figura 13 – Equipamento Contador de Fótons.....	40
Figura 14 – Equipamento de tomossíntese mamária.	41
Figura 15 – Imagem do simulador ACR.	42
Figura 16 – Imagens do CMMD, a esquerda com calcificações fibrosas, no centro microcalcificações agrupadas e a direita de uma massa tumoral.	44

Figura 17 - Imagem do manual de referência do IBSI, com a tabela e legenda original contendo os parâmetros de processamentos utilizados.....	45
Figura 18 – Simuladores utilizados pelo IBSI para a padronização das análises por radiômica. A direita, simulador digital e a esquerda simulador criado a partir de imagens de tomografia computadorizada de pulmão.....	46
Figura 19 – Tela do 3D Slicer módulo de visualização das informações da imagem.	50
Figura 20 – Tela do 3D Slicer módulo de redimensionamento Resample Scalar Volume.....	51
Figura 21 – Tela do 3D Slicer com as informações da imagem após o redimensionamento.	51
Figura 22 – Tela do 3D Slicer com as informação do histograma da imagem original.	52
Figura 23 – Tela do 3D Slicer módulo de correspondência de histograma, com as informações da imagem de referência.....	53
Figura 24 – Tela do 3D Slicer com a imagem resultante após aplicação das informações do histograma de referência.....	53
Figura 25 – Tela do 3D Slicer módulo de segmentação, com as ROIs delimitando as estruturas no simulador ACR.....	54
Figura 26 – Fluxo para a padronização da análise.	57
Figura 27 – Gráficos comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	59
Figura 28 – Gráfico comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	60
Figura 29 – Gráfico comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	60
Figura 30 – Gráfico comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	60

Figura 31 – Gráfico comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	60
Figura 32 – Gráficos comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	61
Figura 33 – Gráficos comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	61
Figura 34 – Gráfico comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	61
Figura 35 – Gráficos comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	61
Figura 36 – Gráficos comparando os valores do test object fibra entre ensaios no mesmo equipamento, e entre equipamentos.....	62
Figura 37 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	63
Figura 38 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	64
Figura 39 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	64
Figura 40 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	64
Figura 41 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	64
Figura 42 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	65
Figura 43 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	65
Figura 44 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	65

Figura 45 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	65
Figura 46 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	66
Figura 47 – Gráficos comparando os valores para o test object microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.	66
Figura 48 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	68
Figura 49 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	68
Figura 50 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	68
Figura 51 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	68
Figura 52 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	69
Figura 53 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	69
Figura 54 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	69
Figura 55 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	69
Figura 56 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	70
Figura 57 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	70
Figura 58 – Gráficos comparando os valores para o test object massa para ensaios no mesmo equipamento, e entre equipamentos.	70

Figura 59 – Resultados para o test object de fibra. Os pontos em vermelho representam os limiares de valores apresentados pelo IBSI.....	73
Figura 60 – Resultados para o test object de microcalcificação. Os pontos em vermelho representam os limiares de valores apresentados pelo IBSI.	73
Figura 61 – Resultados para o test object de massa. Os pontos em vermelho representam os limiares de valores apresentados pelo IBSI.....	74
Figura 62 – Análise comparativa para fibra entre o simulador ACR e o banco de imagens CMMD.	75
Figura 63 – Análise comparativa para segmentação de microcalcificação entre o simulador ACR e o banco de imagens CMMD.	76
Figura 64 – Análise comparativa para segmentação de massa entre o simulador ACR e o banco de imagens CMMD.....	77
Figura 65 – Matriz de correlação de características extraídas do test object Fibra.	79
Figura 66 – Matriz de correlação de características extraídas do test object Microcalcificação.....	80
Figura 67 – Matriz de correlação de características extraídas do test object Massa.	81

SUMÁRIO

1 INTRODUÇÃO	16
2 OBJETIVOS	21
2.1 Objetivo Geral.....	21
2.2 Objetivos Secundários	21
3 JUSTIFICATIVA.....	22
4 REVISÃO DA LITERATURA	23
4.1 Mamografia Digital.....	23
4.2 Imagem Digital.....	27
4.3 Simuladores.....	32
4.4 Segmentação de Imagem Digital	33
4.5 Biomarcadores de Imagem	35
5 MATERIAL E MÉTODOS	40
5.1 Equipamentos de Mamografia Digital	40
5.2 Simulador ACR.....	42
5.3 Parâmetros de Aquisição e Coleta das Imagens	43
5.4 Banco de Dados de Imagens Médicas	44
5.5 Resultados de Referência do IBSI	45
5.6 Pós Processamento, Análise das Imagens e Análise Estatística	46
5.7 Fases da Análise	47
5.8 Redimensionamento do Pixel.....	49
5.9 Correspondência ou Casamento do Histograma.....	52
5.10 Segmentação	54
5.11 Extração das Características Radiômicas.....	54
6 RESULTADOS E DISCUSSÃO	57
6.1 Análise da Estabilidade e Reprodutibilidade das Características Radiômicas .	58
6.1.1 Fibra.....	58
6.1.2 Microcalcificação	63
6.1.3 Massa	67
6.2 Comparativo dos valores calculados com o simulador ACR com dados IBSI .	71
6.3 Simulador ACR vs. Banco de Imagens CMMD	74
6.3.1 Fibra.....	75

6.3.2 Microcalcificação	76
6.3.3 Massa	77
6.4 Características Redundantes	78
7 CONCLUSÃO	85
REFERÊNCIAS BIBLIOGRÁFICAS	86
APÊNDICES	93

1 INTRODUÇÃO

A taxa de incidência do câncer de mama tem aumentado pelo mundo, tornando-se o tipo mais comum de câncer em mulheres. Em países desenvolvidos, ações de rastreamento da doença, de forma a detectá-la mais precocemente, tem proporcionado a diminuição na taxa de mortalidade pelo câncer de mama. Em contrapartida, países em desenvolvimento como o Brasil, tanto a taxa de incidência como a de mortalidade têm aumentado, dado ao fato de que a maioria dos casos de câncer de mama tem o seu diagnóstico muito tardio (ASSIS et al., 2020; CALEFFI et al., 2009). Dados do Instituto do Câncer (2024) mostram um aumento da taxa de mortalidade por câncer de mama, em 2022 o Brasil registrou mais de 19 mil óbitos por câncer de mama em mulheres, e foram estimados 73.610 novos casos para cada ano do triênio de 2023-2025 (CÂNCER (BRASIL), 2024; INSTITUTO NACIONAL DE CÂNCER - INCA, 2023).

O rastreamento do câncer de mama através da mamografia é uma prática considerada mundialmente como a de maior sensibilidade e especificidade. O objetivo do rastreamento por mamografia não é somente detectar a presença de câncer no tecido mamário, mas detectá-lo de forma precoce, e desta forma os tratamentos podem ser mais eficientes, diminuindo a taxa de mortalidade em decorrência da doença (DA SILVA et al., 2015; TOMAL et al., 2015).

A mamografia passou por várias modificações tecnológicas nos últimos anos, e esse constante aprimoramento que envolve a aquisição dessas imagens, é pelo fato do contraste entre os tecidos normal e patológico se apresentarem de formas muito semelhantes na imagem mamográfica (ALKHALIFAH et al., 2016).

Anteriormente adquirida de forma analógica, com receptores de imagem dedicados à imagem mamográfica, cassetes com uma tela intensificadora (écran), acondicionavam películas com uma emulsão gelatinosa de brometo de prata e iodeto de prata, onde após a interação da radiação com o tecido, forma-se uma imagem latente, e após passar por um processamento químico obtinha-se a imagem radiográfica (MINISTÉRIO DA SAÚDE; INSTITUTO NACIONAL DE CÂNCER JOSÉ

ALENCAR GOMES DA SILVA, 2019). As imagens mamográficas analógicas podiam apresentar uma alta resolução espacial, que permitia a visualização de estruturas muito pequenas, porém, problemas causados durante o processamento químico, armazenamento das películas, limpeza e temperatura dos químicos de revelação, podiam acarretar falhas na imagem adquirida (FREITAS et al., 2006).

Desenvolvido nos anos 80, o primeiro formato apresentado de sistema digital de aquisição de imagens radiográficas foi o conhecido como radiografia computadorizada ou CR (*Computed Radiography*), chamado de receptores indiretos, pois consistem em placas de fósforo fotoestimuláveis armazenadas em cassetes semelhantes aos utilizados nos sistemas analógicos, capazes de armazenar as informações provenientes da interação da radiação com o tecido (imagem latente), necessitando serem estimulados por um feixe de laser em uma leitora dedicada, que irá liberar a energia armazenada no fósforo em forma de luz, a intensidade da luz emitida é proporcional a quantidade de radiação absorvida pelo fósforo, a luz é capturada e intensificada por tubos fotomultiplicadores, convertidos em sinal eletrônico, para assim formar a imagem digital (COWEN et al., 2007; WRIGLEY, 2004).

Paralelamente, ocorria o desenvolvimento de um sistema em rede para o armazenamento, transferência e visualização de imagens digitais, impulsionado não somente pelo desenvolvimento dos sistemas CR, mas principalmente pela tomografia computadorizada e a ressonância magnética, que estavam ganhando um importante espaço no diagnóstico por imagem. O sistema PACS (*Picture Archive and Communication Systems*) foi a solução para o problema em se arquivar as películas que perdiam a qualidade com o tempo, como também diminuiu os custos envolvidos com a impressão, e principalmente a repetição de exames (STRICKLAND, 2000; WRIGLEY, 2004).

Nos anos 90, os dispositivos de aquisição de imagens tiveram outra adaptação tecnológica, a radiografia digital ou DR (*Digital Radiography*), que consiste em um dispositivo acoplado de carga (DCC) diretamente ao equipamento de mamografia. Estes detectores podem converter o sinal recebido em imagem digital de forma indireta ou direta (LANÇA; SILVA, 2009a, b).

Na forma indireta, cintiladores, sendo os de iodeto de céσιο ou oxissulfeto de gadolínio os mais utilizados, que após serem estimulados pela radiação incidente produzem uma leve cintilação, luz esta que será medida e convertida de forma semelhante ao sistema CR. Na forma direta, selênio amorfo é aplicado à uma placa de transistores de filme fino (por meio de evaporação), por ser um material fotocondutor, o selênio amorfo absorve a radiação, gera um elétron de alta energia, uma tensão polarizada é aplicada formando vários elétrons-buracos que se espalham pela placa, este sinal é diretamente intensificado e transformado em sinal eletrônico para a formação da imagem digital (BUSHONG, 2008; YAFFE et al., 2013).

Apesar do formato de imagem digital não alcançar a mesma resolução espacial entregue pelo formato analógico, pois nas imagens digitais a resolução espacial está diretamente associada ao tamanho do pixel, os recursos de pós-processamento disponíveis para a visualização da imagem, possibilidade de armazenamento por longo prazo, diminuir as reconvocações/repetições e desta forma a dose à qual a mama é exposta, são vantagens associadas a esta tecnologia (LANÇA; SILVA, 2009b; YAFFE et al., 2013).

Apesar da evolução pela qual a imagem mamográfica passou, sua análise requer tempo e habilidade técnica do médico, pois a interpretação de características benignas ou malignas na imagem não é simples. Com a introdução das imagens digitais, ferramentas de assistência para interpretação da imagem mamográfica foram desenvolvidas nas décadas de 90, como o CAD (*Computer-Aided Diagnosis or Detection*, do inglês diagnóstico ou detecção auxiliado por computador) software que a partir da análise das intensidades dos valores de pixel em regiões de interesse na imagem, apontam ao médico quais apresentam características com potencial patológico (ANTROPOVA et al., 2017; AVANZO et al., 2017; HUYNH et al., 2016).

A busca pelo aperfeiçoamento de ferramentas como o CAD, levaram ao desenvolvimento de uma ferramenta que elevasse o nível do diagnóstico por imagem, como também a ter um papel central principalmente no diagnóstico de doenças oncológicas. Em crescente desenvolvimento há mais de uma década, *Radiomics* é uma ferramenta que tem o potencial de extrair dados quantitativos de imagens médicas. O termo *Radiomics*, onde o prefixo “radio” refere-se às

modalidades radiológicas, e o sufixo “omics” que foi utilizado pela primeira vez quando foi realizado o mapeamento do genoma humano, a genômica, caracteriza a abrangência do sistema como um todo, é o correspondente à análise global de um sistema biológico, identificando, quantificando e caracterizando (AVANZO et al., 2017; PAREKH; JACOBS, 2016).

Embora seja um desenvolvimento derivado dos sistemas CAD, a análise dos dados é realizada de forma independente. Os dados extraídos formam uma base de dados que auxiliam na identificação de patologias oncológicas, como por exemplo o estágio da doença ou resposta ao tratamento. A ideia central é que os dados extraídos estão diretamente relacionados com padrões genéticos e proteicos encontrados nas células, encorajando *Radiomics* a se tornar uma ferramenta de suporte a medicina de precisão (AVANZO et al., 2017; PAREKH; JACOBS, 2016).

Diferentemente dos sistemas CAD a análise não se restringe somente aos valores de intensidade dos pixels/voxels, os dados extraídos envolvem descritores de distribuição de intensidade, relações espaciais entre os vários níveis de intensidade, padrões de heterogeneidade, de textura, descritores de forma e das relações das intensidades dos níveis de cinza, como por exemplo do tumor com os dos tecidos circundantes (AVANZO et al., 2017; GILLIES et al., 2016; LAMBIN, PHILIPPE et al., 2012; PAREKH; JACOBS, 2016; VAN TIMMEREN et al., 2020).

A base dos cálculos utilizados para a extração dos dados por radiômica fundamentam-se em técnicas já bem estabelecidas para imagens digitais. Desde 1948, por Claude Shannon, depois Haralick e Galloway, sendo os primeiros a incorporar medidas estatísticas às matrizes de tons de cinza, para desta forma ser possível descrever informações como textura, suavidade, rugosidade entre outras contidas nas imagens digitais de satélites (LAMBIN, PHILIPPE et al., 2012; PAREKH; JACOBS, 2016).

Esta capacidade que os dados extraídos têm em fornecer informações que podem ser correlacionadas como biomarcadores, fez de *Radiomics* um dos campos de pesquisa da atualidade (KUMAR et al., 2012; VALDORA et al., 2018).

Porém como toda tecnologia emergente, ainda possui obstáculos, como a robustez dos dados e a falta de uma padronização em toda a cadeia que envolve a geração desses dados.

Os estudos iniciais foram realizados em um mesmo centro de diagnóstico utilizando um único equipamento de Tomografia Computadorizada, porém ao se analisar imagens realizadas em equipamentos de Tomografia de fabricantes diferentes, foram encontradas divergências nos resultados causados por diferenças na espessura do corte e filtros de reconstrução dentre outros fatores (GER et al., 2018; MACKIN et al., 2015).

O volume de características quantificadoras extraídas, faz com que as escolhas dos dados a serem analisados sejam diferentes nos diversos estudos, diminuindo a reprodutibilidade dos ensaios. Assim como o método escolhido para a segmentação da área de interesse – ROI (*Region of Interest*), se automático, semiautomático ou manual que também geram variações nos valores das características. Pesquisas realizadas com programas *home made*, também dificultam a reprodução da técnica de extração realizada, pois a metodologia utilizada na construção do programa não é compartilhada (LAMBIN, PHILIPPE et al., 2017; VAN TIMMEREN et al., 2020; ZWANENBURG; VALLIÈRES; et al., 2020). E para se criar um modelo preditivo apropriado, requer-se uma amostra de pacientes muito alta, e em sua falta, a análise estatística desses dados podem gerar um *overfitting*, causado pela sobreposição entre os dados de validação e teste (GIGER, 2018; LIMKIN et al., 2017).

O *Image Biomarker Standardisation Initiative* (IBSI) foi recentemente formado com o intuito de criar uma padronização, e tentar tornar as pesquisas com *Radiomics* o mais reprodutíveis possível, porém ainda apresenta resultados preliminares (ZWANENBURG; VALLIÈRES; et al., 2020).

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver uma estratégia para a padronização e para o planejamento inicial de análises por radiômica em imagens de mamografia digital, segundo as recomendações do IBSI.

2.2 Objetivos Secundários

Viabilidade do uso de um simulador padrão de qualidade de imagem em mamografia para o processo de padronização das análises radiômicas.

Verificar a estabilidade e reprodutibilidade das características radiômicas extraídas das imagens do simulador que foram adquiridas em diferentes equipamentos de mamografia digital.

Comparar os valores das características calculadas nas imagens do simulador com os valores disponibilizados pelo grupo de trabalho responsável pela padronização operacional dos estudos de análise por radiômica.

Empregar a estratégia utilizada para extrair as características radiômicas nas imagens do simulador, em imagens de mamografia digital realizadas em humanos disponibilizadas publicamente em banco de dados digital, e comparar com os resultados obtidos no simulador.

3 JUSTIFICATIVA

O presente trabalho se justifica pela insuficiência de documentos exemplificando os processos e procedimentos, com os primeiros passos necessários para iniciar-se um projeto de análises utilizando a radiômica.

A minha jornada de aprendizado desta ferramenta foi uma motivação, pois sem encontrar artigos que demonstrassem os passos de forma a serem reproduzidos, onde os softwares e métodos de extração e análises dessas características radiômicas eram de propriedade dos pesquisadores, me motivou a criar um roteiro para que mais pessoas, que assim como eu, que se interessem por este campo de pesquisa, possam dar seus primeiros passos.

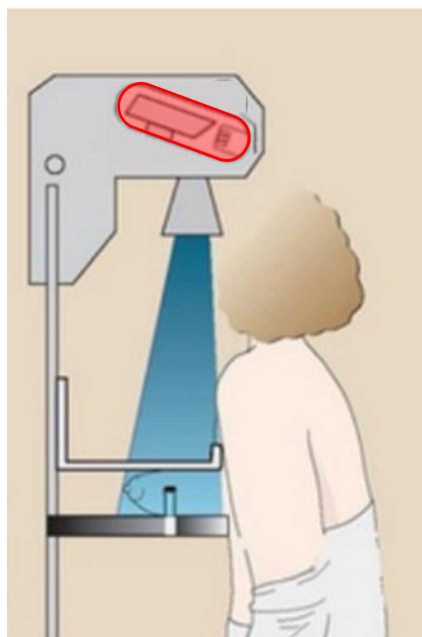
A possibilidade de usar ferramentas de fácil acesso, como softwares livres e um simulador de imagem padrão a qualquer serviço de mamografia, fazem com que esse trabalho seja facilmente reproduzido por outros, difundindo o conhecimento para que possa ser aprimorado no desenvolvimento de análises utilizando a radiômica de forma mais elaborada.

4 REVISÃO DA LITERATURA

4.1 Mamografia Digital

A mamografia é o exame que avalia o tecido mamário, a existência de uma modalidade de exame radiográfico dedicado a este fim, está ligado ao fato da densidade radiográfica das estruturas anatômicas da mama serem melhor representadas com um espectro de raios X diferentes daqueles utilizados nos exames de radiologia convencional, e principalmente por o câncer de mama ser o mais recorrente e a principal causa de morte em mulheres (AMORIM et al., 2022; MINISTÉRIO DA SAÚDE; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2019). Na figura 1 é apresentada uma ilustração simplificada de um exame de mamografia, onde pode-se observar a leve inclinação do tubo de raios X (destacado em vermelho na imagem), para um melhor uso benefício do efeito anódico e diminuição do ponto focal, e o sistema de compressão da mama, para espalhar o tecido, e desta forma diminuir as sobreposições de estruturas e diminuir a espessura da mama, resultando em uma dose de exposição à radiação mais baixa (BUSHONG, 2008).

Figura 1 – Ilustração de um tubo de raios X e equipamento dedicado ao exame de mamografia.



Fonte: (BUSHONG, 2008).

Os exames de mamografia são classificados como de rastreamento, que são realizados em mulheres assintomáticas, com o objetivo de diagnosticar precocemente alguma patologia na mama, e mamografia diagnóstica, exames realizados para confirmar se há a presença ou não de possíveis patologias em pacientes com sintomas ou algum sinal de doença na mama (MINISTÉRIO DA SAÚDE; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2019).

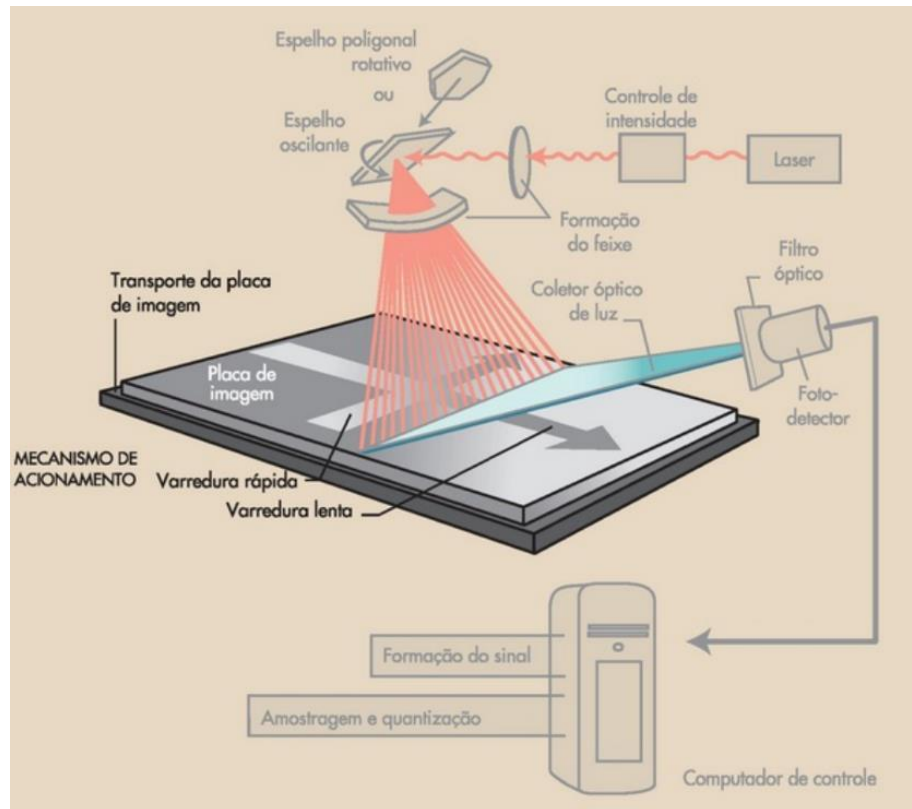
Os sistemas de mamografia evoluíram muito nas últimas décadas, nos anos 60, com a introdução de filtros e alvos de molibdênio nos tubos de raios X, nos anos 70 com sistema de grades difusoras, filmes e telas intensificadoras dedicadas para a produção da imagem mamográfica, e principalmente, pelos sistemas de compressão e controle automático de exposição (MINISTÉRIO DA SAÚDE; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA, 2019).

Com a introdução de tecnologias como da tomografia computadorizada, a possibilidade de geração e armazenamento de imagens médicas de forma digital, fizeram com que as demais modalidades de diagnósticos também se adaptassem aos novos formatos de imagens digitais (STRICKLAND, 2000)

A primeira tecnologia a ganhar popularidade foram os sistemas de radiografia computadorizada, CR (do inglês, *Computerized Radiography*), pois para a adequação ao sistema, eram feitas as trocas dos chassis de sistema tela-filme por chassis com placas de um fosforo fotoestimulável, o IP (do inglês *Image Plate*), que capta e armazena a energia incidente, que em uma unidade digitalizadora, um feixe de laser estimula (fotoestimulação) essa energia armazenada, e a libera em forma de luz, que é coletada por tubos fotomultiplicadores, que amplificam o sinal que é convertido em imagem digital. Esse sistema de conversão indireta, apresenta vantagens como a de poderem ser facilmente adaptados aos equipamentos de mamografia já instalado na clínica e por apresentarem uma resolução espacial de alto contraste próximos aos do sistema tela-filme, mas apesar de possuírem pixels pequenos, 50 μm , possuem ruído associado ao sistema que comprometem a resolução espacial, quando comparado com imagens geradas com mesma dose nos sistemas tela-filme (LANÇA; SILVA, 2009a). A figura 2 apresenta a sequência

do transporte da placa IP dentro da leitora, a fotoestimulação por laser e captação da luz resultante pelo fotodetector, para a digitalização da imagem em sistemas CR.

Figura 2 – Ilustração da leitura à laser para sistemas CR.

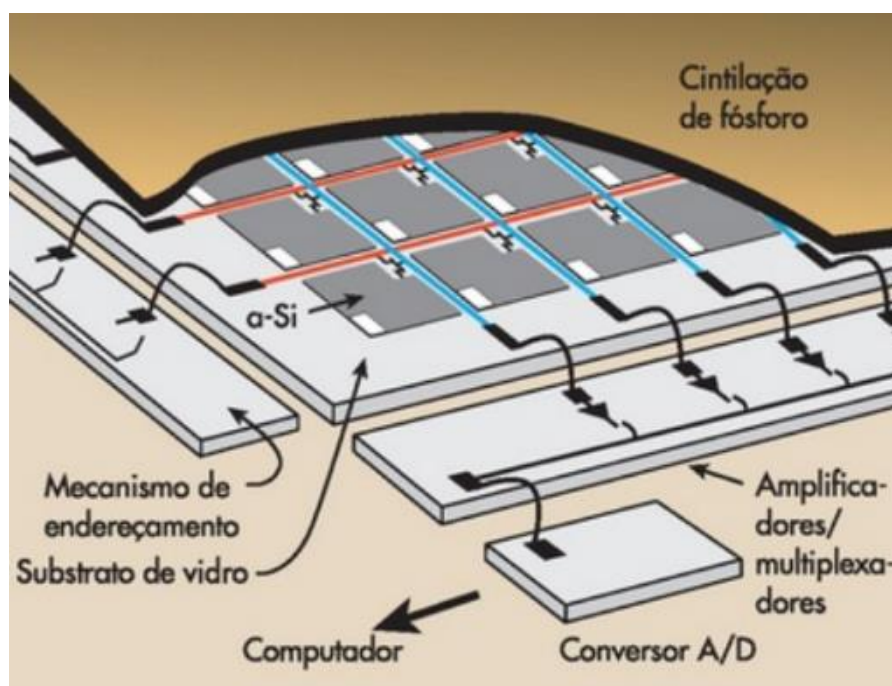


Fonte: (BUSHONG, 2008).

Já nos sistemas de radiografia digital, DR (do inglês, *Digital Radiography*) temos o digital indireto, onde a radiação resultante interage com os cristais do detetor que produzem luz, que é captada e convertida em sinal elétrico para formar a imagem digital. E os sistemas digitais direto, onde um material fotocondutor capta a radiação resultante e converte diretamente esses fótons em carga elétrica que irão formar a imagem (LANÇA; SILVA, 2009a). A figura 3 apresenta o esquema de confecção de uma placa DR de processo indireto, com material cintilador de Iodeto de Césio e o elemento de captura de silício amorfo. Neste trabalho foi utilizado um equipamento de tomossíntese que emprega este tipo de tecnologia de aquisição de imagem. Os equipamentos de tomossíntese realizam aquisições de imagens no plano 2D e 3D volumétricas (de forma independente). A tomossíntese foi desenvolvida com o objetivo de aprimorar o diagnóstico, pois em alguns casos o

tecido fibroglandular se sobrepõe a estruturas de interesse dificultando seu diagnóstico. As aquisições realizadas em 3D volumétricas fornecem múltiplas projeções, que são adquiridas em ângulos a partir de 0°; 11,7°; 33° (com variações de 8,8° e 14,6°) positivo e negativo, que são reconstruídas fornecendo secções paralelas de toda a altura da mama (VEDANTHAM et al., 2015).

Figura 3 – Ilustração de receptor de radiografia digital.

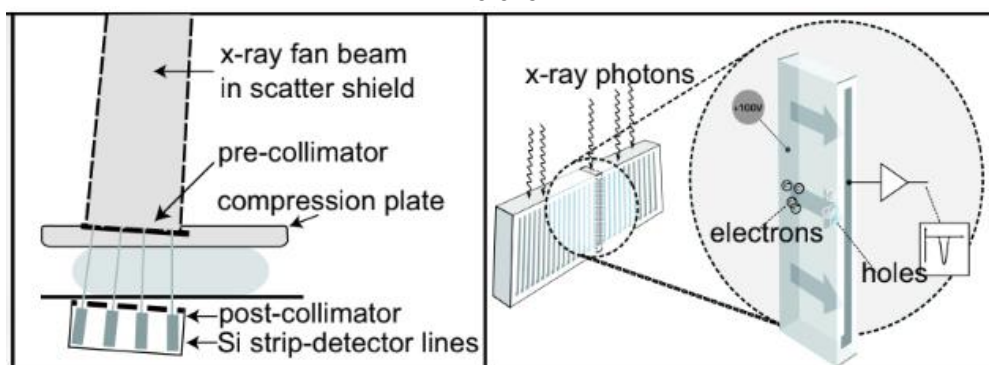


Fonte: (BUSHONG, 2008).

Outra tecnologia para aquisição de imagens empregada em equipamentos de mamografia digital são os contadores de fótons, do inglês *photon counting*, sua aquisição é em forma de varredura, a kilovoltagem e miliamperagem empregada para aquisição das imagens é menor do que aquelas utilizadas em equipamentos convencionais e com tomossíntese, pois possuem modulação de dose lateral ativa, que leva em consideração as densidades do tecido mamário conforme o feixe de raios X varre a mama durante a aquisição da imagem. Os equipamentos contadores de fótons utilizam dois sistemas de colimação do feixe de raios X, um pré colimador para que o feixe de radiação primária não interaja com o ar impedindo a produção de radiação secundária, e um pós colimador em coincidência com o primeiro, para impedir que a radiação secundária, desta vez proveniente da mama, chegue ao

detector, diminuindo desta forma sua interferência, a do ruído que ela gera e conseqüentemente diminuindo a dose de radiação para a aquisição da imagem. O pré colimador possui duas alturas, sendo o colimador baixo para mamas com espessuras de compressão abaixo de 60mm, e colimador alto para mamas com espessura de compressão acima de 60mm (ÅSLUND, 2007; WEIGEL et al., 2014). Neste trabalho foram utilizados dois equipamentos de mamografia digital com esta tecnologia. A figura 4 apresenta a geometria de varredura da aquisição e o posicionamento dos colimadores e do detector.

Figura 4 - Geometria de varredura, posição dos colimadores e do detector do contador de fótons.



Fonte: (ÅSLUND, 2007).

4.2 Imagem Digital

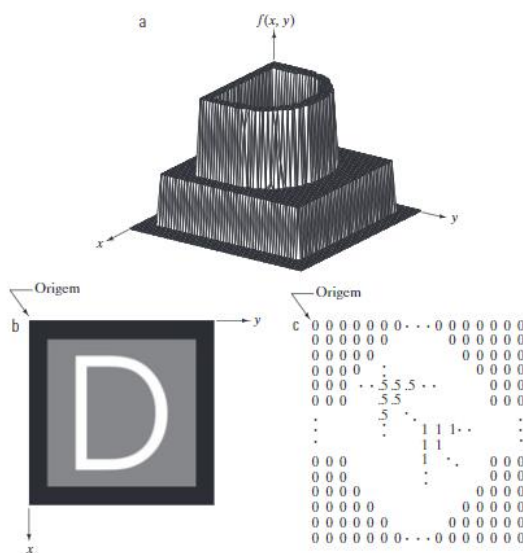
O uso de imagens radiográficas na medicina foi considerado como um grande passo para seu desenvolvimento, pois possibilitou a visualização do interior do corpo humano de uma forma não invasiva (PRINCE; LINKS, 2015).

Com o surgimento dos computadores e o rápido avanço das tecnologias associadas ao computador, o seu uso foi rapidamente aplicado à medicina, sendo o diagnóstico por imagem o maior beneficiado (BUSHONG, 2008).

Como explicado anteriormente, após a radiação ionizante, ou não ionizante, como no caso da ultrassonografia e da ressonância magnética, interagir com o objeto (paciente ou simulador) e sua resultante ser captada por um detector, inicia-se o processo da formação da imagem.

Vários são os processos envolvidos tanto no pré processamento quanto nos pós processamentos, para que uma imagem tenha uma apresentação visual satisfatória. A imagem digital é representada por uma função de coordenadas (x, y) , e cada ponto dessa coordenada é proporcional a um valor de intensidade – $f(x, y)$ (GONZALEZ; WOODS, 2010). No caso das imagens radiográficas essas intensidades estão associadas à uma escala de níveis de cinza (GONZALEZ; WOODS, 2010). A figura 5 apresenta o esquema das coordenadas espaciais da imagem, suas representações das amplitudes e dos valores dos tons de cinza.

Figura 5 – Representação da função de coordenadas e das intensidades de níveis de cinza em imagem digital.



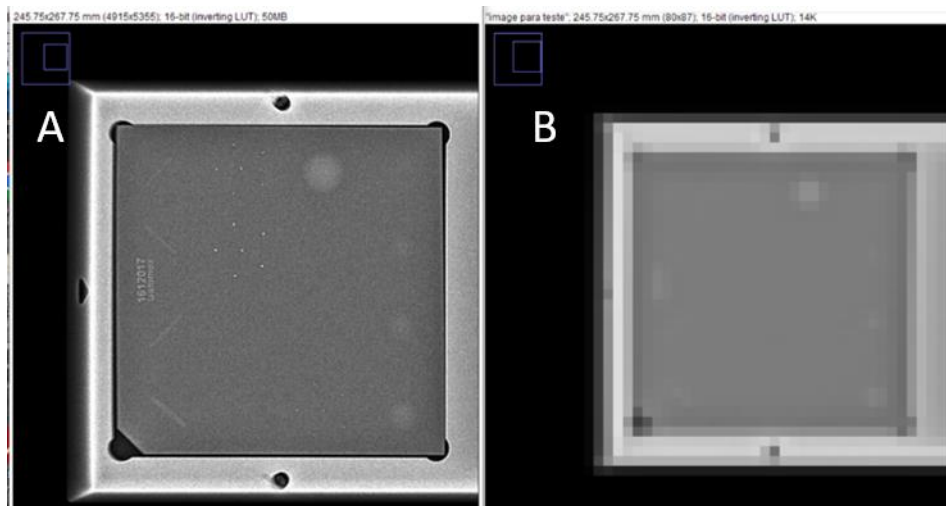
Fonte: (GONZALEZ; WOODS, 2010).

Esses dados obtidos passam por um processo de amostragem, que é a digitalização dos valores de coordenada, e pelo processo de quantização que é a digitalização dos valores de amplitude ou intensidade (GONZALEZ; WOODS, 2010).

Esse sistema de coordenadas irá formar uma matriz, que são as quantidades de linhas e colunas que será proporcional ao tamanho do detector de imagem, e cada um desses pontos do sistema de coordenadas recebe o nome de *pixel* (GONZALEZ; WOODS, 2010). Quanto mais pixels uma imagem possuir em sua matriz, maior será a sua resolução espacial, ou seja, maior será o detalhe dado a

está imagem, como por exemplo os detalhes de bordas e de estruturas pequenas contidas nesta imagem (GONZALEZ; WOODS, 2010). A figura 6 mostra a perda da resolução espacial de uma imagem do simulador utilizado nesta dissertação, quando aplicadas alterações em sua matriz de pixels. Originalmente a imagem foi adquirida com uma matriz de 4915 x 5355 pixels, e após tratamento em software dedicado ao processamento de imagens digitais *Image J*, esta foi alterada para 80 x 87 (ABRÀMOFF, 2004).

Figura 6 – Exemplo da perda da resolução espacial quando aplicadas operações de processamentos de imagens digitais, em A imagem com uma matriz de pixels de 4915x5355; em B modificada para 80x87.



Fonte: Autor da dissertação.

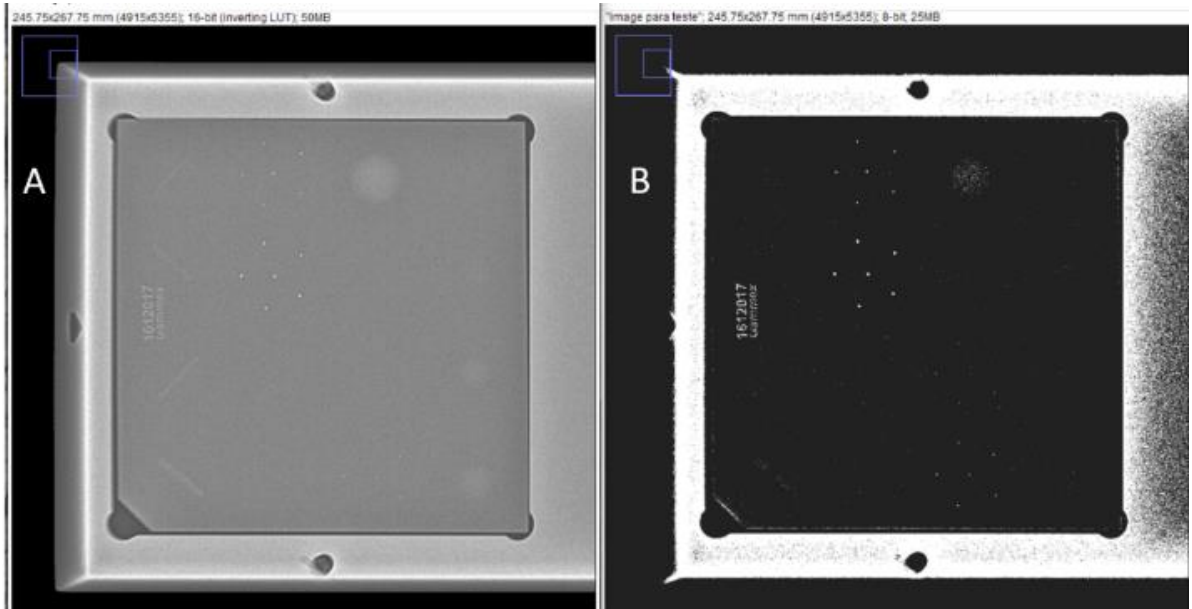
Outro fator importante para a qualidade de uma imagem digital é a quantidade de níveis discretos de tons de cinza que ela pode armazenar, por exemplo, uma imagem de 8 bits é capaz de armazenar 256 tons discretos de cinza (GONZALEZ; WOODS, 2010).

Portanto, imagens com uma matriz de 512 x 512 pixels e 256 níveis de cinza apresentarão um detalhe do contorno das estruturas e do contraste entre elas mais satisfatório, enquanto uma de 8, 4 ou 2 níveis de cinza e com a mesma matriz não apresentarão a mesma qualidade e poder de diferenciação de estruturas (GONZALEZ; WOODS, 2010).

Na figura 7 a imagem foi agora manipulada de 16 bits para 8 bits, nesta imagem a matriz de pixels foi mantida, por isso é possível observar alguns detalhes

menores das estruturas internas do simulador, porém os diferentes contrastes de tons de cinza que são observados em A não são bem diferenciados em B.

Figura 7 – Imagem do simulador em 16 bits em A, e depois convertida para 8 bits em B.



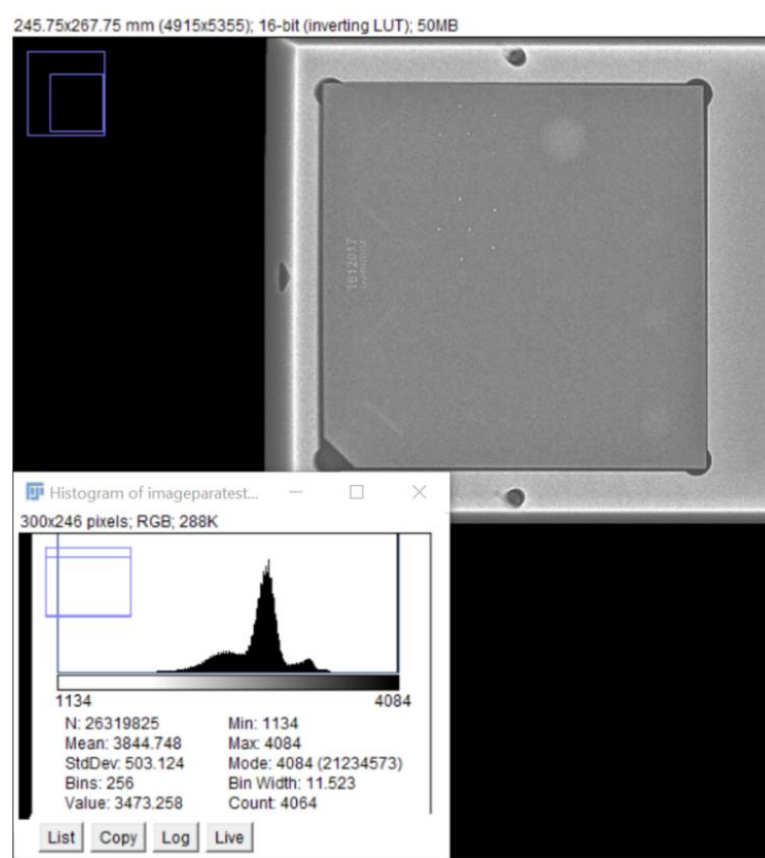
Fonte: Autor da dissertação.

Processamentos no domínio espacial são aplicados diretamente, de diversas formas, sobre a intensidade dos pixels de uma imagem, e podem alterar o contraste e dar realce em uma imagem (GONZALEZ; WOODS, 2010).

O histograma de uma imagem é a representação da quantidade de pixels com um determinado valor de intensidade. São utilizados nos processamentos no domínio espacial, para realçar as imagens, mas também são bastante utilizados para fornecer uma estatística da imagem, compressão e segmentação (GONZALEZ; WOODS, 2010).

A figura 8 apresenta o histograma com o gráfico da quantidade de pixels por distribuição de tons de cinza, e outras informações estatísticas da imagem do simulador.

Figura 8 – Histograma da imagem do simulador utilizado nesta dissertação.

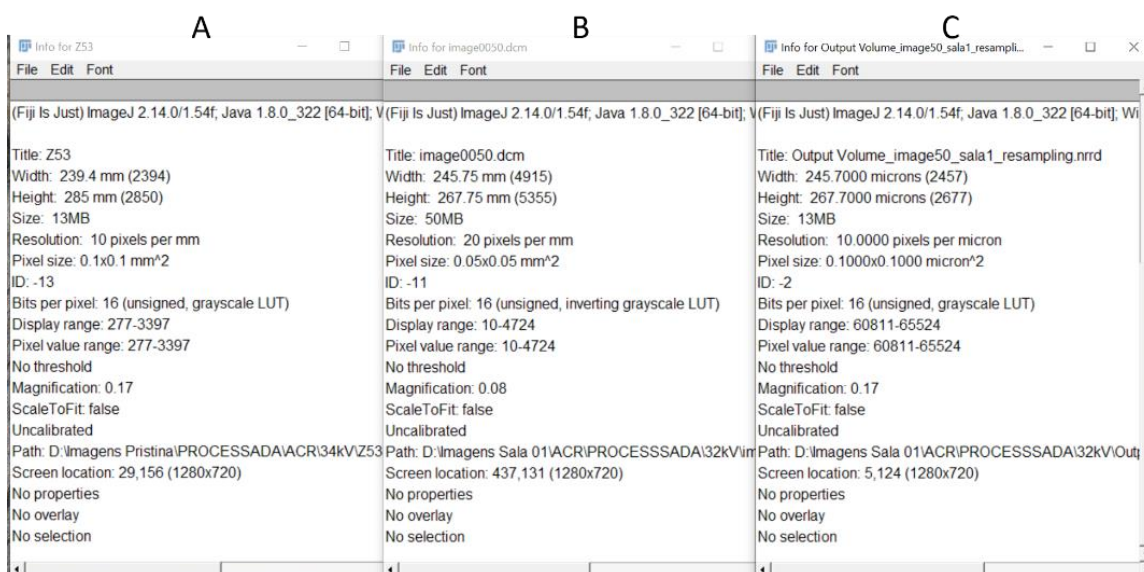


Fonte: Autor da dissertação.

Outro conceito importante a ser abordado são as transformações de registro das imagens digitais. Essas operações afetam diretamente os pixels de uma imagem (GONZALEZ; WOODS, 2010).

Neste trabalho, devido as características dos detectores utilizados nos equipamentos, foi necessária a realização de uma reamostragem de pixels para que apresentem o mesmo tamanho. O método utilizado realiza a reamostragem em coordenadas do espaço e não nos pixels. São realizadas interpolações nos valores de intensidade da imagem, as interpolações mais utilizadas são as lineares, as de vizinhos próximos e *bspline*. Para o registro das imagens utilizadas neste trabalho, foi utilizada a interpolação linear, para que desta forma as imagens apresentem o mesmo espaçamento entre os pixels. A figura 9 apresenta as informações da matriz da imagem, em A as informações originais da matriz do equipamento com pixel de tamanho 100 μm , em B as informações originais do equipamento com pixel de 50 μm e em C após a reamostragem dos pixels da imagem B, por interpolação linear.

Figura 9 – Informações da imagem após reamostragem por interpolação linear.



Fonte: Autor da dissertação.

4.3 Simuladores

Simuladores, do inglês *Phantom*, são amplamente utilizados no diagnóstico por imagem para estabelecer parâmetros de qualidade das imagens produzidas pelos sistemas de imagens, determinar as doses de exposição nos exames, entre outras aplicações.

Em mamografia, os simuladores de qualidade de imagens possuem características de uma mama média padrão, ou seja, aproximadamente 4,5 cm de tecido mamário comprimido e com densidade radiográfica semelhante a uma mama com 50% de tecido glandular e 50% de tecido adiposo (FLUKE BIOMEDICAL, 2023).

Possuem objetos de teste, ou *test objects*, com densidades e características semelhantes aos principais achados radiográficos em humanos, como calcificações fibrosas, microcalcificações agrupadas e massas tumorais. Estes *test objects* são dispostos em uma camada superficial de aproximadamente 7 mm de cera em uma base de polimetilmetacrilato, para que quando expostos aos raios X, os efeitos da atenuação dos raios X e da radiação secundária sejam sofridos pelos *test objects* da mesma forma que ocorrem em uma mama humana (FLUKE BIOMEDICAL, 2023).

O simulador mais utilizado é o do Colégio Americano de Radiologia (*American College of Radiology – ACR*), e que desde os anos 2000 acompanha todas as compras de equipamentos novos, inclusive para o Brasil.

Foi desenvolvido para integrar o Programa de Controle de Qualidade Mamográfica, que engloba o selo de acreditação do programa de qualidade para serviços de mamografia do colégio americano, e como parte dos requisitos do *Mammography Quality Standards Act – MQSA*, para que instalações americanas obtenham o certificado de funcionamento pelo FDA (*Food and Drugs Administration*), usado como parte da avaliação do desempenho do sistema de imagem de forma qualitativa, sendo possível verificar a capacidade do sistema em visualizar estruturas muito pequenas. O Brasil, através da Agência Nacional de Vigilância Sanitária – ANVISA, também adota os parâmetros estabelecidos pelo simulador do colégio americano, preconizado na Instrução Normativa 92, dispostos na RDC 611 de 2022, para verificação da qualidade da imagem mamográfica.

4.4 Segmentação de Imagem Digital

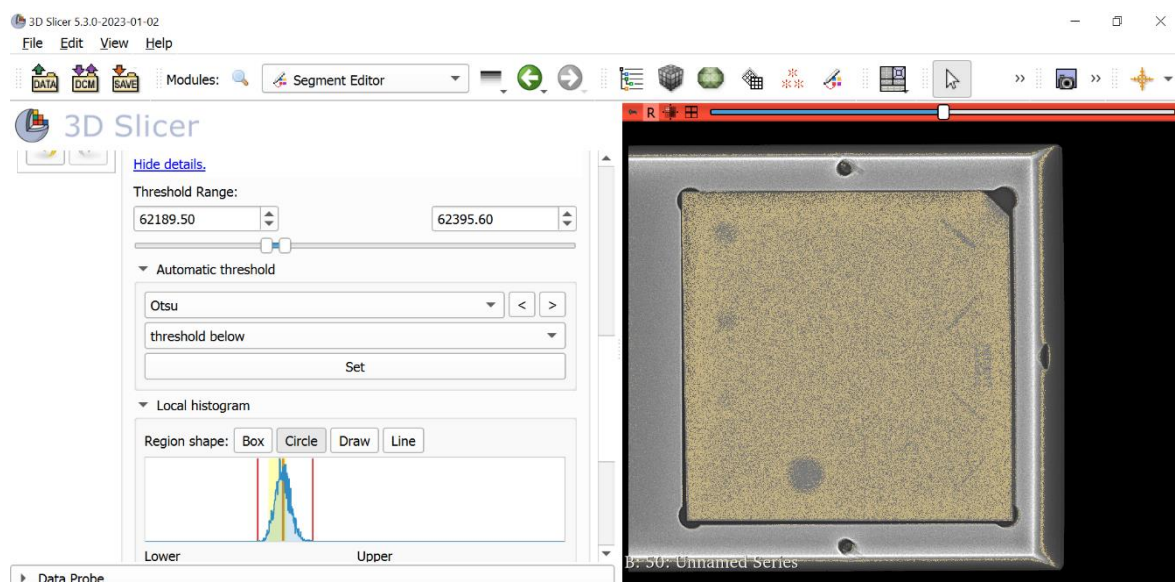
A segmentação de imagens digitais é o processo de delimitar regiões de interesse - ROI (*Region of Interest*). Na área médica esse processo pode ser aplicado para planejamento de uma região para tratamento radioterápico, para cirurgias guiadas, análises clínicas laboratoriais e entre outros (NOROUZI et al., 2014).

No radiodiagnóstico um dos primeiros sistemas a utilizar o processo de segmentação de imagens foram os sistemas CAD (*Computer-Aided Diagnosis ou Detection*), que interpreta possíveis padrões patológicos encontrados nas imagens a partir da segmentação de uma ROI (AVANZO et al., 2017).

Existem vários métodos para se segmentar uma imagem, os mais comuns são os baseados em detecção de bordas. Para se estabelecer as bordas em imagem digital, conceitos como o de conectividade são importantes, pois indicam se os pixels são adjacentes e se apresentam similaridades dentro de uma escala de tons de cinza (NUNES, 2006).

Um exemplo de segmentação é a por limiarização (*Thresholding*), que considera valores limiares dos pixels e divide a imagem em quantas partes forem necessárias, através da alteração dos valores dos pixels, deixando-os com menos quantidades de níveis de cinza e com isso separa os objetos, ou estruturas de uma imagem (GONZALEZ; WOODS, 2010; NUNES, 2006). A figura 10 apresenta a tela do módulo de segmentação por limiarização (*Thresholding*) do software de visualização e tratamento de imagens 3D Slicer, que será utilizado nesta dissertação, considerando os valores inferiores dos pixels (FEDOROV et al., 2012).

Figura 10 – Tela do 3D Slicer do módulo de segmentação por limiarização (Thresholding).



Fonte: Autor da dissertação.

A detecção de pontos isolados e de linhas são muito utilizados para remover por exemplo ruídos que possam existir na imagem. São utilizadas máscaras para que o valor do pixel central ou os pixels da vizinhança sejam realçados (NUNES, 2006). Outra forma bastante utilizada para segmentar imagens médicas é por rastreamento de nível, que define o contorno de borda e faz o preenchimento para a segmentação de uma estrutura da imagem que possuam os mesmos níveis de valores nos pixels.

4.5 Biomarcadores de Imagem

Como dito anteriormente o uso de imagens na medicina, permitiu que patologias fossem diagnosticadas com maior precisão e de forma precoce. Além de serem minimamente invasivas, podem oferecer um grande volume de informações da doença (ROGERS et al., 2020).

Por se tratar de imagens, é uma das modalidades dentro da medicina que se beneficiou com os avanços em hardware e software, tendo um significativo desenvolvimento tecnológico, e desta forma passando rapidamente de imagens analógicas para imagens digitais (BUSHONG, 2008).

A princípio as imagens médicas eram analisadas somente de forma qualitativa, e de acordo com a experiência do avaliador. Na década de 1950 iniciaram-se as pesquisas para desenvolvimento de métodos de análise das imagens médicas de forma quantitativa. Os sistemas de Diagnóstico Auxiliado por Computador (*CAD – Computer-Aided Diagnosis*) começaram a apresentar um melhor desempenho na década de 80 quando melhores recursos operacionais de informática possibilitaram a inclusão de algoritmos de reconhecimento de padrões às leituras das imagens radiológicas, sendo as principais aplicações no reconhecimento de patologias no pulmão e na mama (ROGERS et al., 2020; YANASE; TRIANTAPHYLLOU, 2019).

O uso dos sistemas CAD na avaliação de fenótipos encontrados nas imagens médicas auxiliam na tomada de decisões diagnósticas, como a indicação de biopsias, identificação de microcalcificações em mamografias ou pequenos nódulos pulmonares (ROGERS et al., 2020). Do mesmo modo, com as inovações tecnológicas, o volume de imagens geradas em um único exame passou a ser mais um obstáculo, e os sistemas CAD puderam oferecer maior precisão ao diagnóstico (JUNIOR; OLIVEIRA, 2017). A figura 11 mostra a análise do sistema CAD de uma mamografia.

Figura 11 – Tela do sistema CAD evidenciando regiões com alterações sugestivas de patologia.



Fonte: <https://www.siemens-healthineers.com/br/mammography/news/history-of-computer-aided-detection.html>, acessado em 03/09/2023 às 18:25.

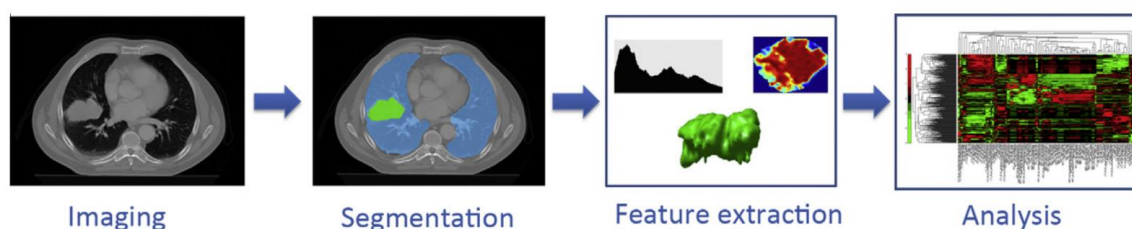
Contudo a medicina sempre evolui no sentido de buscar uma maior precisão nos diagnósticos, ter diagnósticos o mais precoce possível e tornar os tratamentos cada vez mais personalizados.

Evoluindo da ideia central dos sistemas CAD, do conhecimento da histologia das patologias, a alta qualidade, disponibilidade e volume de imagens geradas nos exames radiológicos, essas imagens passaram a ser interpretadas como biomarcadores de imagens quantitativos, através da extração de dados matemáticos e análise estatística (ABRAMSON et al., 2015; ZWANENBURG; VALLIÈRES; et al., 2020).

Os fenótipos radiográficos que antes apenas possuíam representações qualitativas, contando apenas com a acurácia visual para a sua avaliação, agora se aliam a uma análise estatística de dados de alto rendimento que podem ser explorados de informações contidas nas alterações das intensidades dos pixels/voxels, e que podem ser associadas a alterações morfológicas e metabólicas presentes em certas patologias (GILLIES et al., 2016; LAMBIN, PHILIPPE et al., 2017; SCAPICCHIO et al., 2021).

Dada a natureza da análise desses biomarcadores de imagens quantitativos ou características, como serão tratadas neste trabalho, as pesquisas relacionadas a essa aplicação que usa a ciência da computação para extrair dados de imagens médicas passou a ser conhecida como radiômica, do inglês *Radiomics*, que além de correlacionar essas variações nas intensidades dos pixels/voxels com patologias, também o correlaciona com a evolução e resposta da patologia ao tratamento (AVANZO et al., 2017; PAREKH; JACOBS, 2016). A figura 12 mostra a sequência de eventos das análises por radiômica.

Figura 12 – Sequência de eventos da análise Radiômica.



Fonte: (LAMBIN, PHILIPPE et al., 2012)

Para tanto, os processos envolvendo uma análise radiômica requerem uma padronização desde a aquisição das imagens, o tratamento no processamento pós aquisição e nas ferramentas envolvidas no processo de extração e cálculo dos dados quantitativos (MESEGUER et al., 2022; SCAPICCHIO et al., 2021).

Além disso, este é um tipo de processo que necessita de uma equipe multidisciplinar, cito, o tecnólogo especialista para a aquisição seguindo o protocolo padrão para esse tipo de imagem, o médico especialista para delimitação das ROIs, o especialista em tecnologia da informação e programação para a manipulação e extração dos valores computados e automação desse processo, e profissionais estatísticos para tratamento e interpretação do volume de dados gerados.

Dado o crescente interesse de pesquisas em radiômica e a falta de um consenso comum nas informações necessárias para a reprodução desses estudos, um grupo de pesquisadores internacionais e independentes criaram o *Image Biomarker Standardisation Initiative* – IBSI, que trabalha para validar uma diretriz e assim padronizar as análises por radiômica (ZWANENBURG; LEGER; et al., 2020a).

Até o momento o IBSI ainda conta com resultados preliminares, iniciados com um banco de imagens de câncer de pulmão em humanos, que incluem imagens adquiridas por Tomografia Computadorizada, Ressonância Magnética e Medicina Nuclear, e para um simulador digital com aquisições por Tomografia Computadorizada (IBSI, 2023; LAMBIN, P., 2021; ZWANENBURG; LEGER; et al., 2020a).

Uma questão que envolve as análises por radiômica é a forma como as características são calculadas, sendo importante utilizar ferramentas que já seguem algum tipo de padronização para o cálculo e extração dessas características. A biblioteca *PyRadiomics*, segue os padrões do IBSI e calcula e extrai 108 características radiômicas para cada área de segmentação selecionada (VAN GRIETHUYSEN et al., 2017), e são divididas em três categorias, sendo:

Morfológicas (*Shape*), analisa o formato, as irregularidades no formato da segmentação, mais utilizado em segmentações de tumores, pois fornece informações de volume, muito utilizado em imagens de Tomografia Computadorizada (TC), Ressonância Magnética (RM) e *Positron Emission Tomography* (PET), que são exames onde imagens de Medicina Nuclear são sobrepostas a imagens de TC, e em estudos associados com pacientes submetidos a radioterapia (PAREKH; JACOBS, 2016). Esta categoria engloba 14 características. Neste trabalho não abordamos os aspectos relacionados a estas características devido a utilização de imagens planas (2D).

Estatística de Primeira Ordem (*First Order*), análise por histograma das distribuições espaciais dos níveis das intensidades de tons de cinza dos pixels/voxels dentro da imagem (PAREKH; JACOBS, 2016). Mensuram, através dos desvios nos níveis de tons de cinza, a largura e os desvios da média desses histogramas, o grau dessas assimetrias no histograma, podem fornecer informações sobre a textura, como rugosidade (AGGARWAL; AGRAWAL, 2012; HARALICK et al., 1973). Esta categoria engloba 18 características.

Estatística de Alta Ordem, analisa as relações entre os pixels/voxels dentro da imagem, relaciona em diversas direções e diferentes matrizes de pixels/voxels (PAREKH; JACOBS, 2016). Aqui a avaliação não se restringe somente a como os tons do cinza estão distribuídos na ROI e o grau de assimetria do histograma, nesta

análise os pixels/voxels são analisados aos pares levando-se em consideração a orientação (em ângulos) e distância desses pixels dentro da imagem. Podendo ser relacionados com a textura e homogeneidade das estruturas, como por exemplo: regiões de necrose nas áreas adjacentes de um tumor; hipercaptação do meio de contraste pelo tumor; hipervascularização de um tumor (AGGARWAL; AGRAWAL, 2012; HARALICK et al., 1973; TAMAL, 2019). Esta categoria engloba 75 características, divididas em 5 classes:

- **Gray Level Co-occurrence Matrix – GLCM:** calcula por pares de pixels semelhantes, quantos e a sua distância do par de pixel localizado no centro da Região de Interesse. Essa classe engloba 24 características;
- **Gray Level Dependence Matrix – GLDM:** quantifica a dependência dos níveis de tons de cinza em uma imagem, que é dado pelo número e a distância de pixels conectados com o centro da Região de Interesse. Essa classe engloba 14 características;
- **Gray Level Run Length Matrix – GLRLM:** calcula quantos pixels consecutivos possuem o mesmo valor. Essa classe engloba 16 características;
- **Gray Level Size Zone Matrix – GLSZM:** calcula por zonas dentro da Região de Interesse quantos grupos de pixels possuem a mesma intensidade de tom de cinza. Essa classe engloba 16 características;
- **Neighbouring Gray Tone Difference Matrix – NGTDM:** calcula a média da intensidade do tom de cinza entre dois pixels vizinhos. Essa classe engloba 5 características.

As características que são calculadas dentro dessas 5 classes recebem seus nomes de acordo com a forma com que avaliam as relações dos tons de cinza, suas distribuições e suas matrizes. Normalmente na literatura os autores se restringem somente a informar a classe de características utilizadas, mas aqui exibimos todas as informações para que quando reproduzidos, haja uma maior compreensão sobre os dados que estão sendo coletados.

5 MATERIAL E MÉTODOS

Neste trabalho foi comparado o comportamento de características radiômicas após a aplicação de processamentos para a padronização de imagens médicas digitais de um simulador, que foram adquiridas em três equipamentos de mamografia diferentes, e quais dessas características que melhor representaram esse tipo de imagem. Foi verificada também a semelhança dos *test object* do simulador com imagens de humanos.

5.1 Equipamentos de Mamografia Digital

Foram utilizados três equipamentos de mamografia digital, sendo dois com tecnologia de contador de fótons (*Photon Counting*) e um com tecnologia tomossíntese (*Digital Breast Tomosynthesis*), nomeados de PC1, PC2 e DBT, respectivamente.

O equipamento PC1 está equipado com o colimador alto, e o equipamento PC2 equipado com o colimador baixo, suas indicações de uso já foram descritas no item 4.1 desta dissertação.

A figura 13 apresenta o equipamento PC1 com seu sistema de colimação (pré colimador) em modo de preparo para aquisição.

Figura 13 – Equipamento Contador de Fótons.



Fonte: Autor da dissertação.

A figura 14 apresenta uma imagem de um equipamento de tomossíntese semelhante ao utilizado neste trabalho.

Figura 14 – Equipamento de tomossíntese mamária.



Fonte: <https://www.gehealthcare.com/products/imaging>.

Como os equipamentos PC1 e PC2 geram somente imagens em 2D, neste trabalho, no equipamento de tomossíntese, foram adquiridas e utilizadas somente imagens em 2D.

As características dos detectores, matriz da imagem, tamanho do elemento detector (pixel) e a profundidade do pixel em bits alocados são apresentados na tabela 1.

Tabela 1 – Características dos equipamentos de mamografia digital.

Equipamento	Tipo de Detector	Matriz	Tam. Pixel (μm)	Bits
DBT	Iodeto de Césio – Cintilador	2394 x 2850	100 x 100	16
PC1	Silício Cristalino – Direto	5355 x 4915	50 x 50	16
PC2	Silício Cristalino – Direto	5355 x 4915	50 x 50	16

Fonte: Autor da Dissertação.

5.2 Simulador ACR

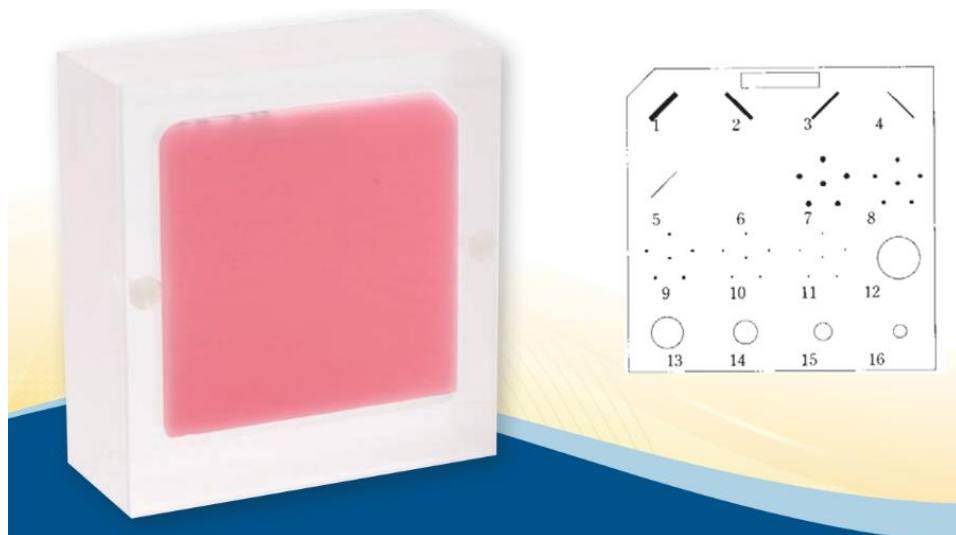
Para determinar quais características melhor representam imagens de mamografia digital, o simulador padrão do ACR foi utilizado na aquisição das imagens. Os *test objects* presentes no simulador foram segmentados e utilizados como corpo de prova para a quantificação das características radiômicas. Os materiais e tamanhos dos *test objects* são descritos conforme tabela 2. A figura 15 demonstra o aspecto externo e a disposição dos *test objects* no interior do simulador.

Tabela 2 – Descrição dos *test objects* presentes no simulador ACR.

Estrutura	Material	Tamanhos (mm)
Fibras	Nylon	1,56 / 1,12 / 0,89 / 0,75 / 0,54 / 0,40
Microcalcificação	Oxido de Alumínio	0,54 / 0,40 / 0,32 / 0,24 / 0,16
Massas	Espessuras de Cera	2,00 / 1,00 / 0,75 / 0,50

Fonte: Fluke Biomedical, disponível em www.flukebiomedical.com.

Figura 15 – Imagem do simulador ACR.



Fonte: Fluke Biomedical®, disponível em www.flukebiomedical.com.

5.3 Parâmetros de Aquisição e Coleta das Imagens

As imagens foram adquiridas no modo automático, com uso do controle automático de exposição, conforme o padrão adotado para aquisição das imagens na rotina clínica. Como as análises por radiômica tem como objetivo final reconhecer estágios de patologias em imagens radiológicas, foi optado por somente analisar imagens processadas para apresentação, que são as imagens clínicas, apresentadas para o diagnóstico. A tabela 3 apresenta os parâmetros técnicos utilizados por cada equipamento.

Tabela 3 – Parâmetros de aquisição modo automático.

Equipamento	kV	mAs	Alvo / Filtro
DBT	33	30	Ródio / Prata
PC1	32	13	Tungstênio / Alumínio
PC2	32	12	Tungstênio / Alumínio

Fonte: Autor da dissertação.

As aquisições das imagens foram divididas em dois ensaios, realizados da seguinte forma:

- 1º ensaio: coletado no período de 12/2022, com a aquisição de 5 imagens do simulador em cada equipamento, totalizando 15 imagens.
- 2º ensaio: coletado no período de 09/2023, com aquisição de 5 imagens em cada equipamento, totalizando 15 imagens.

Os ensaios geraram um total de 30 imagens para serem analisadas.

As imagens foram extraídas em formato “Processada para Apresentação”, ou seja, imagem clínica, que é apresentada com aplicação de filtros de processamento para melhorar os detalhes das estruturas e densidades presentes nas imagens mamográficas.

Destacando que os equipamentos utilizados neste trabalho estão sob rígido programa de controle de qualidade.

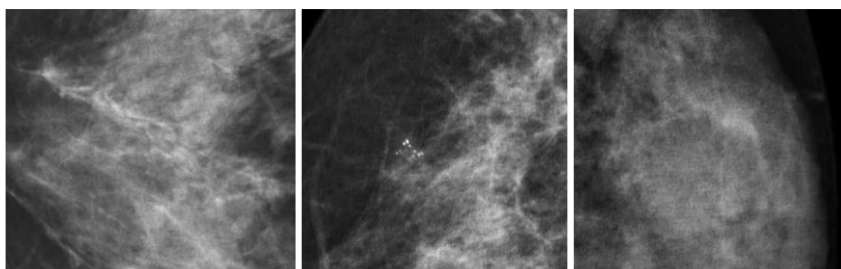
5.4 Banco de Dados de Imagens Médicas

Como forma de comparação dos valores obtidos com o simulador, foram utilizadas imagens de mamografia realizadas em humanos, disponibilizadas através do *Cancer Imaging Archive*, repositório público que viabiliza o acesso à imagens de várias modalidades diagnóstica (CLARK et al., 2013).

O banco de dados de mamografias chinês (CMMD) foi selecionado, e conta com 3.728 mamografias de pacientes chineses com doenças mamárias benignas e malignas que foram submetidos a mamografias no período de julho de 2012 a janeiro de 2016. Criado com o intuito de colaborar com os crescentes avanços na área de análise computacional, como o CAD e suas variações, e suprir a falta de grandes bancos de imagens geradas por mamografia para teste e avaliação dessas ferramentas de apoio ao diagnóstico. Neste banco, além das imagens, estão disponibilizados também o tipo de diagnóstico, confirmado por biopsia, assim como os subtipos moleculares, como luminal A, luminal B, HER2 positivo e triplo-negativo. Todas as imagens disponibilizadas, foram adquiridas em um equipamento GE Senographe DS, com detector de Iodeto de Césio e Silício amorfo e tecnologia radiologia digital indireta (CUI et al., 2021).

Neste trabalho foram selecionadas 2 imagens para cada achado radiológico, ou seja, calcificações fibrosas, microcalcificação e massas tumorais, que possuam distribuições semelhantes com os *test objects* do simulador, totalizando 6 imagens utilizadas do banco CMMD. Na figura 16 estão exemplificadas uma imagem de cada achado correspondente ao simulador que foi utilizado para a análise comparativa neste trabalho.

Figura 16 – Imagens do CMMD, a esquerda com calcificações fibrosas, no centro microcalcificações agrupadas e a direita de uma massa tumoral.



Fonte: (CUI et al., 2021)

5.5 Resultados de Referência do IBSI

Os valores do simulador também foram comparados com os valores preliminares publicados no manual do IBSI (ZWANENBURG; LEGER; et al., 2020a). De forma a criar uma relação com os valores obtidos neste trabalho, foram selecionados os resultados do IBSI provenientes de aquisições tomográficas, que apresentam segmentações de imagens 2D e com tamanhos de pixels semelhantes ou próximos aos das imagens do simulador ACR.

O grupo desenvolveu vários simuladores digitais e utilizando imagens de Tomografia Computadorizada de quatro pacientes com câncer de pulmão, com imagens em 2D e 3D, para testes de implementação e padronização de filtros de imagem e cálculos das características (DEPEURSINGE et al., 2020, 2024; LAMBIN, P., 2021). Neste estágio em que o grupo se encontra, apenas algumas características foram padronizadas e submetidas em seu manual de referência. A figura 17 é uma imagem da tabela com sua respectiva legenda, retirada do manual de referência do IBSI, nela contém as informações dos parâmetros de processamento pós aquisição das imagens tomográficas.

Figura 17 - Imagem do manual de referência do IBSI, com a tabela e legenda original contendo os parâmetros de processamentos utilizados.

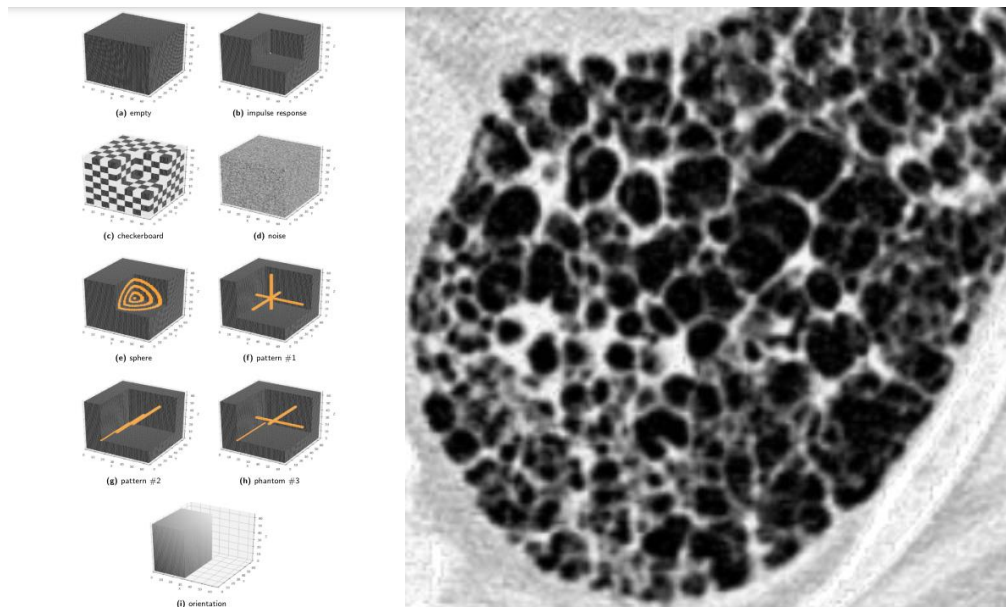
Parameter	Config. A	Config. B	Config. C	Config. D	Config. E
sample identifier	PAT1	PAT1	PAT1	PAT1	PAT1
ROI name	GTV-1	GTV-1	GTV-1	GTV-1	GTV-1
slice-wise or single volume (3D)	2D	2D	3D	3D	3D
interpolation	no	yes	yes	yes	yes
resampled voxel spacing (mm)		2 × 2 (axial)	2 × 2 × 2	2 × 2 × 2	2 × 2 × 2
interpolation method		bilinear	trilinear	trilinear	trilinear
intensity rounding		nearest integer	nearest integer	nearest integer	nearest integer
ROI interpolation method		bilinear	trilinear	trilinear	trilinear
ROI partial mask volume		0.5	0.5	0.5	0.5
re-segmentation					
range (HU)	[-500, 400]	[-500, 400]	[-1000, 400]	no	[-1000, 400]
outlier filtering	no	no	no	3σ	3σ
discretisation					
texture and IH	FBS: 25 HU	FBN: 32 bins	FBS: 25 HU	FBN: 32 bins	FBN: 32 bins
IVH	no	no	FBS: 2.5 HU	no	FBN: 1000 bins
texture parameters					
GLCM, NGTDM, NGLDM distance	1	1	1	1	1
GLSZM, GLDZM linkage distance	1	1	1	1	1
NGLDM coarseness	0.0	0.0	0.0	0.0	0.0

Table 5.1 — Different configurations for image processing. For details, refer to the corresponding sections in chapter 2. ROI: region of interest; HU: Hounsfield Unit; IH: intensity histogram; FBS: fixed bin size; FBN: fixed bin number; IVH: intensity-volume histogram; GLCM: grey level co-occurrence matrix; NGTDM: neighborhood grey tone difference matrix; NGLDM: neighbouring grey level dependence matrix; GLSZM: grey level size zone matrix; GLDZM: grey level distance zone matrix.

Fonte: (ZWANENBURG; LEGER; et al., 2020b)

Na figura 18 são apresentados os simuladores digitais e o produzido a partir de imagens de Tomografia Computadorizada de pulmão, utilizados para a implementação do processo de padronização das análises por radiômica pelo IBSI (IBSI, 2023; LAMBIN, P., 2021).

Figura 18 – Simuladores utilizados pelo IBSI para a padronização das análises por radiômica. A direita, simulador digital e a esquerda simulador criado a partir de imagens de tomografia computadorizada de pulmão.



Fonte: (DEPEURSINGE et al., 2024; IBSI, 2023; LAMBIN, P., 2021)

5.6 Pós Processamento, Análise das Imagens e Análise Estatística

Para visualização e processamentos pós aquisição das imagens, foi utilizado o software 3D Slicer, versão 5.5.0.2023-09-21, que possui uma plataforma amigável e código aberto e é disponibilizado gratuitamente. Com ferramentas que possibilitam segmentar e delimitar as regiões de interesse nas imagens, sendo também capaz de carregar o pacote de extensão *Pyradiomics*, versão 3.0.1, biblioteca do algoritmo *Radiomics* para o cálculo e extração das características radiômicas, que segue as diretrizes do IBSI para padronização do cálculo das características radiômicas (FEDOROV et al., 2012; VAN GRIETHUYSEN et al., 2017).

Foram segmentadas e calculadas as características radiômicas das estruturas fibra, microcalcificação e massa do simulador ACR, e de achados correspondentes nas imagens do CMMD.

Para as análises estatísticas das características foram utilizados os recursos e bibliotecas disponíveis no ambiente de aplicação web do Jupyter Notebook (EXECUTABLE BOOKS COMMUNITY, 2020; WASKOM, 2021).

5.7 Fases da Análise

A sequência dos eventos para o desenvolvimento da padronização teve as seguintes fases:

1. Imagens adquiridas em três equipamentos de mamografia digital. Sendo um equipamento de tomossíntese, identificado como DBT. E dois equipamentos com tecnologia *Photon Counting*, identificados como PC1 e PC2. E como objeto de teste para a aquisição das imagens, simulador padrão de qualidade de imagem mamográfica do ACR.
2. Aquisição das imagens do simulador para o primeiro ensaio, no período de 12/2022, nos três equipamentos de mamografia, gerando 5 imagens adquiridas em cada equipamento. Totalizando 15 imagens no primeiro ensaio.
3. Aquisição das imagens do simulador para o segundo ensaio, no período de 09/2023, nos três equipamentos de mamografia, gerando 5 imagens adquiridas em cada equipamento. Totalizando 15 imagens no segundo ensaio.
4. Após o término dos ensaios possuíamos um total de 30 imagens do simulador para serem analisadas.
5. O processo de padronização iniciou com a realização dos processamentos pós aquisição nas imagens, usando as ferramentas disponíveis no software 3D Slicer. Sendo o primeiro processamento o redimensionamento do tamanho do pixel, aplicado em todas as

imagens que foram adquiridas nos equipamentos PC1 e PC2 (1º ensaio e 2º ensaio). Para esse processamento foi tomado como referência o tamanho de pixel das imagens do DBT.

6. O próximo passo foi realizar a correspondência/casamento dos histogramas das imagens, e novamente foi aplicado em todas as imagens que foram adquiridas nos equipamentos PC1 e PC2 (1º ensaio e 2º ensaio), também tendo como referência as intensidades de cinza das imagens do DBT.
7. Seleção das ROIs e segmentação. As ROIs obedeceram a um tamanho e formato fixo para todas as imagens. Foram segmentadas três ROIs em cada imagem, que são referentes aos *test objects* presentes no simulador: fibra, microcalcificações e massa.
8. Cálculo e extração das características radiômicas. Para cada segmentação, são calculados os valores de 108 características radiômicas, e um relatório é gerado contendo o nome, e a qual classe a característica pertence.
9. Os relatórios referentes a cada imagem, foram agrupados da seguinte forma: por tipo de *test object*, de qual equipamento, e de qual ensaio.
10. Para cada agrupamento foi feito a média e desvio padrão das 108 características, para a avaliação da estabilidade das características radiômicas.
11. A média e o desvio padrão, foram usados para avaliar a reprodutibilidade entre os dois ensaios em um mesmo equipamento. E entre os equipamentos, calculando o coeficiente de variação das características. Este passo também serviu para selecionar as características radiômicas que melhor representaram as imagens mamográficas.
12. As características radiômicas que melhor representaram as imagens mamográficas foram as que tiveram um coeficiente de variação menor ou igual a 15%, determinado através das informações presentes no manual de referência do IBSI, que demonstra que as características

possuem um limiar de valores (SHUR et al., 2021; ZWANENBURG; LEGER; et al., 2020b).

13. Após as características serem selecionadas, a média dos valores encontrados com o simulador foram comparados com os valores informados pelo IBSI.
14. Com a finalidade de avaliar a resposta das características calculadas para os *test object* do simulador, foram utilizadas imagens de mamografias realizadas em humanos, disponibilizadas publicamente, onde os mesmos processamentos utilizados para a padronização das imagens e segmentação das ROIs no simulador foram aplicados nos achados radiológicos equivalentes nas mamografias realizadas em humanos. Os valores calculados para as características radiômicas do simulador e em humanos foram comparados, obedecendo os mesmos critérios dos 15% de coeficiente de variação.
15. E para uma análise mais completa, e conforme descrito na literatura, as características redundantes foram encontradas através de testes estatísticos de correlação entre pares. Deixando o conjunto de dados mais enxuto e pronto para a geração do banco de dados para aplicações que utilizarão aprendizado de máquina (*machine learn*).

5.8 Redimensionamento do Pixel

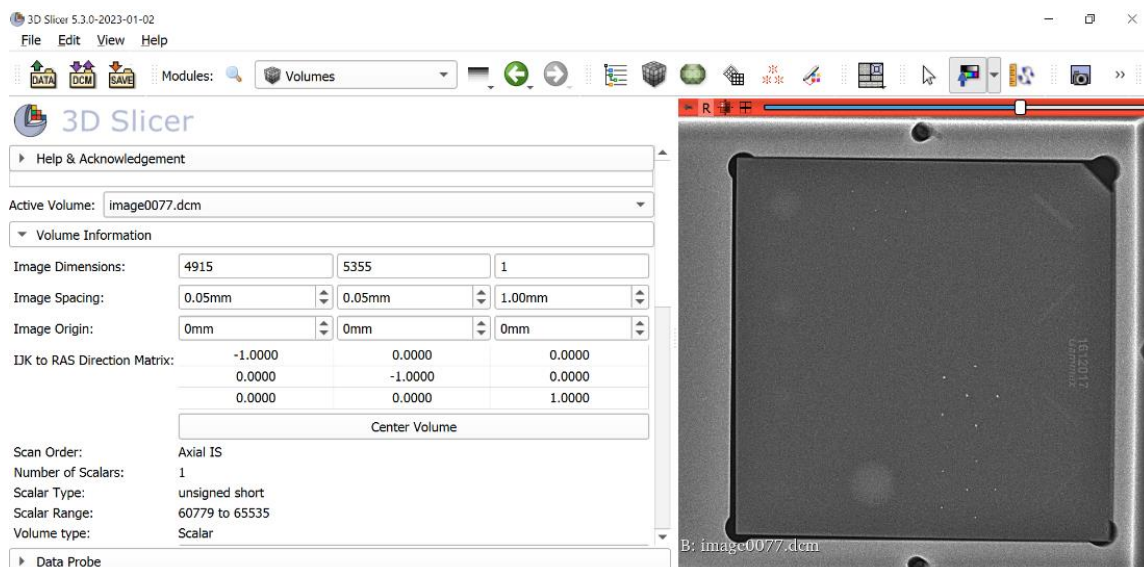
Os equipamentos de PC e DBT possuem distinções consideráveis em sua forma de aquisição e processamento das imagens, gerando diferenças intrínsecas a cada uma das tecnologias.

Os valores das características radiômicas se calculados em imagens sem processamento pós aquisição, apresentarão diferenças significativas para algumas características, pois serão utilizadas as matrizes de pixels para cálculos baseados nos valores de tons de cinza, variações em pixels vizinhos e nos baseados em histograma.

O processo de reamostragem ou redimensionamento do pixel foi realizado no domínio espacial, e nas imagens com menor espaçamento entre pixels, pois no

processo de interpolação linear, utilizado aqui para tal tarefa, os valores dos pixels vizinhos são calculados para então ser atribuído a um único pixel. A figura 19 mostra a tela do software 3D Slicer com as informações originais de dimensionamento da imagem.

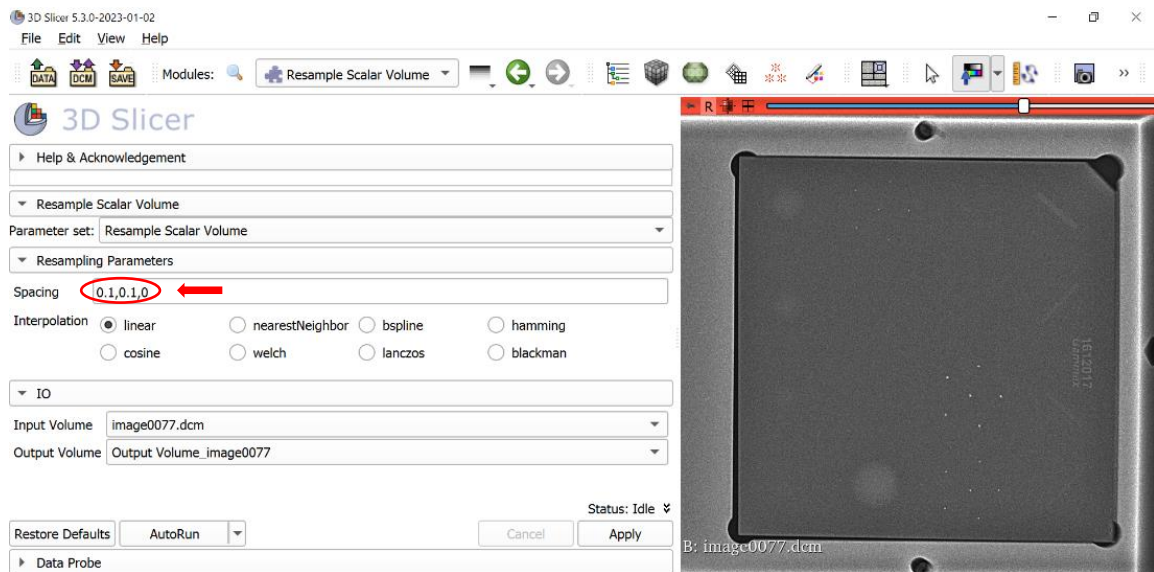
Figura 19 – Tela do 3D Slicer módulo de visualização das informações da imagem.



Fonte: Autor da dissertação.

Passo a passo o processo de redimensionamento da imagem do simulador ACR, a figura 20 mostra a tela do módulo *Resample Scalar Volume* do 3D Slicer, o círculo em vermelho mostra onde devem ser indicados os novos parâmetros do pixel nas três dimensões, neste caso como a direção **z** do pixel não foi alterada, informamos com o número **0**, para que desta forma ao ser aplicado o tipo de interpolação selecionado (seta vermelha na figura), esta dimensão se manterá como na imagem de origem.

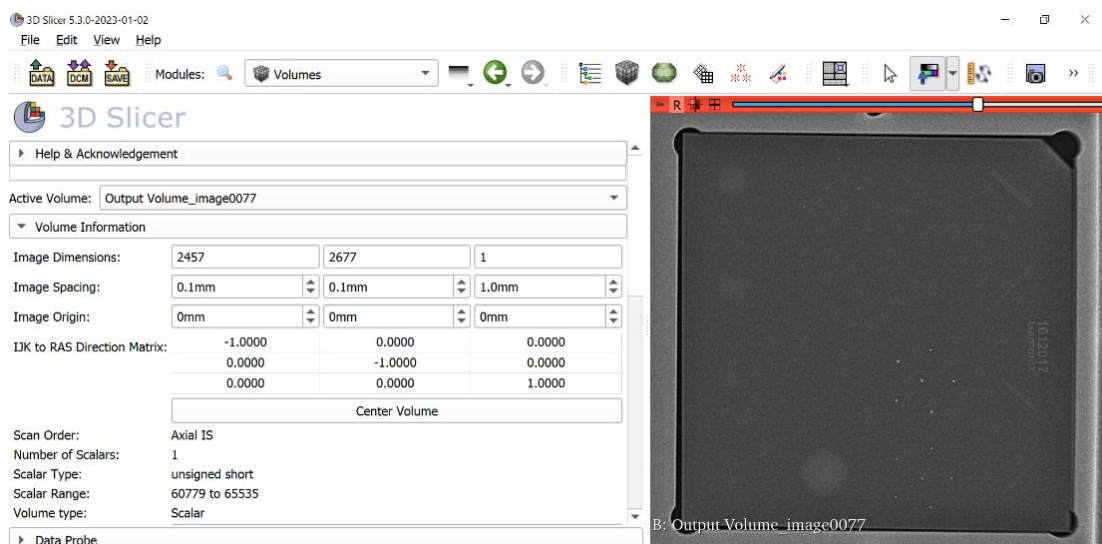
Figura 20 – Tela do 3D Slicer módulo de redimensionamento Resample Scalar Volume.



Fonte: Autor da dissertação.

Após o processamento das novas dimensões, uma nova imagem é gerada pelo software. A figura 21 mostra as alterações das informações de espaçamento entre pixels na imagem gerada. Este processo foi realizado em 20 imagens, referentes as aquisições do PC1 e PC2 feitas nos dois ensaios.

Figura 21 – Tela do 3D Slicer com as informações da imagem após o redimensionamento.



Fonte: Autor da dissertação.

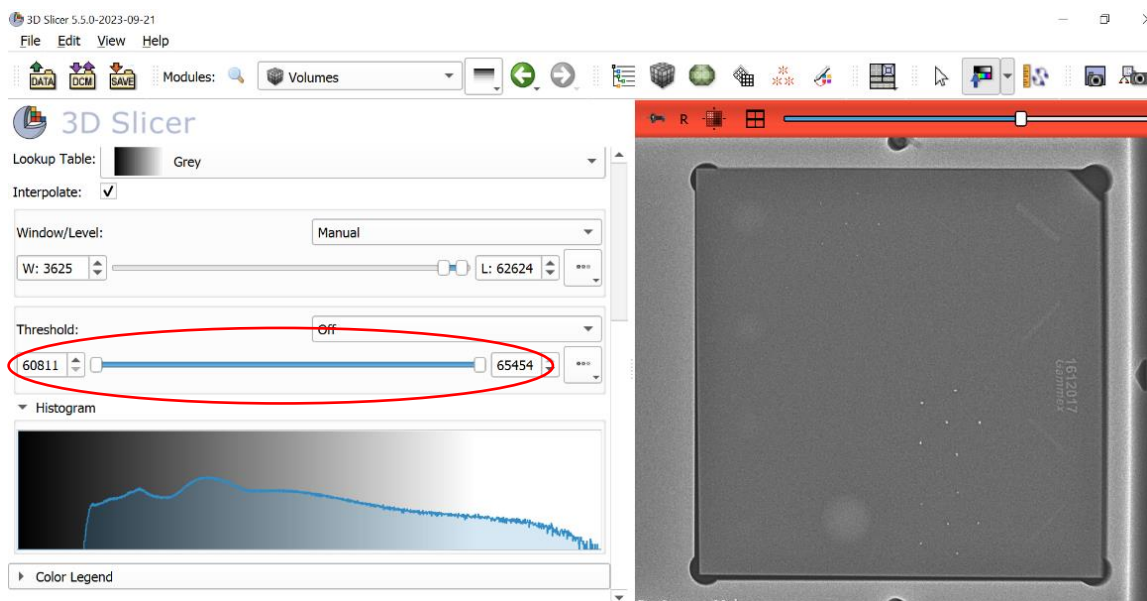
5.9 Correspondência ou Casamento do Histograma

Outro importante processo que envolve análises radiômicas, é a realização da correspondência ou casamento do histograma da imagem, quando as imagens são adquiridas em equipamentos com tecnologias e/ou protocolos diferentes.

Novamente foi aplicado um filtro no domínio espacial da imagem, onde a imagem com o menor intervalo de intensidades de tons de cinza foi usada como referência para a correspondência ou casamento do histograma nas demais imagens.

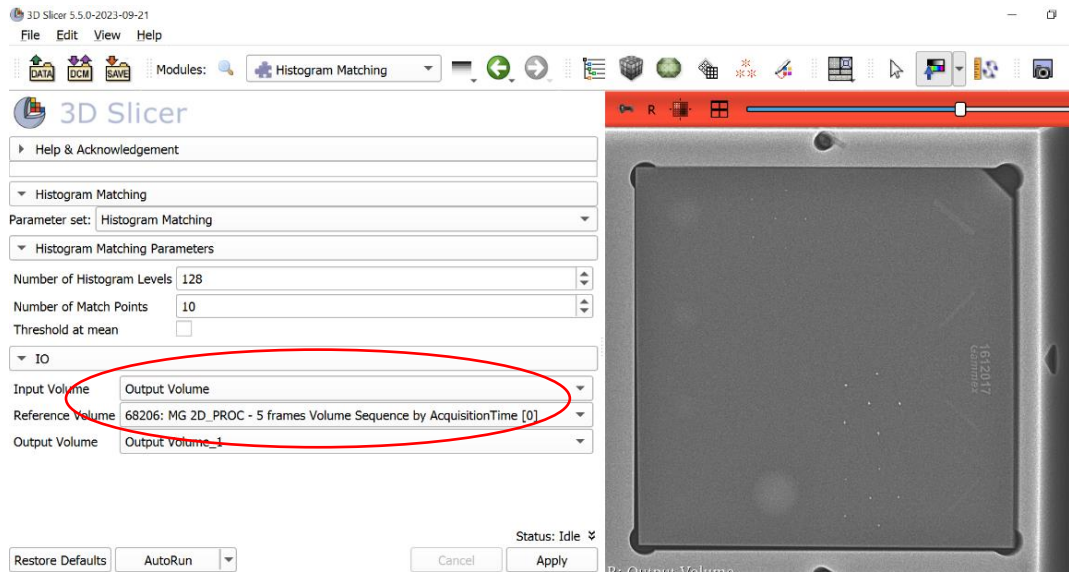
O módulo *Histogram Matching* do software 3D Slicer possibilita a aplicação de um filtro de correspondência nos pixels de uma imagem tomando como base uma imagem de referência. Nesse tipo de processamento é comum que valores de pixels de fundo sejam excluídos da imagem resultante, nas figuras 22, 23 e 24 os círculos em vermelho indicam na tela do 3D Slicer as informações da imagem original, de como o módulo de processamento solicita os dados para a correspondência e as informações da imagem resultante, respectivamente. Nas informações de histograma da imagem, nota-se como a faixa útil de tons de cinza da imagem foi estreitada na escala, demonstrando a exclusão dos valores de fundo da imagem.

Figura 22 – Tela do 3D Slicer com as informações do histograma da imagem original.



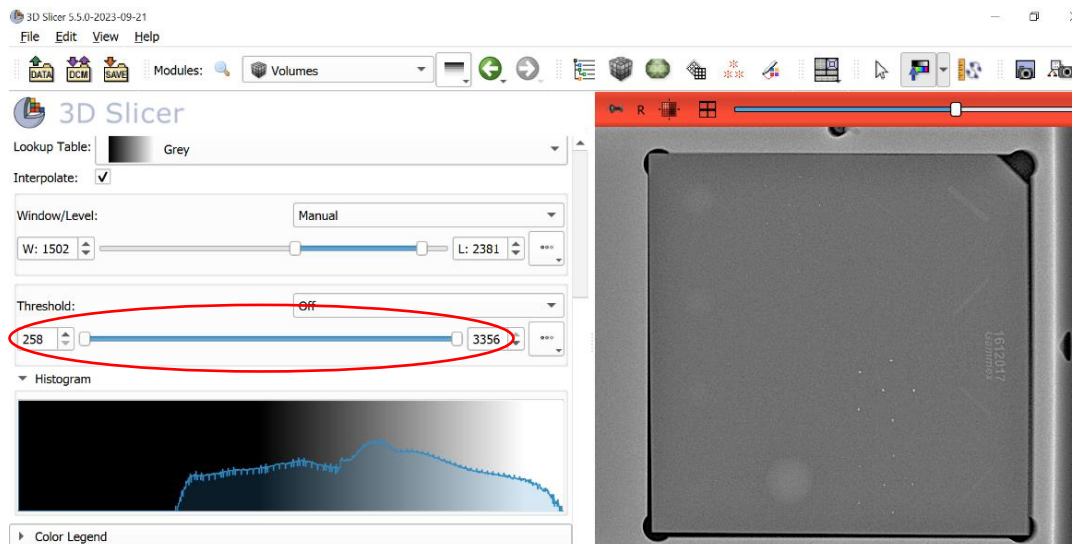
Fonte: Autor da dissertação.

Figura 23 – Tela do 3D Slicer módulo de correspondência de histograma, com as informações da imagem de referência.



Fonte: Autor da dissertação.

Figura 24 – Tela do 3D Slicer com a imagem resultante após aplicação das informações do histograma de referência.



Fonte: Autor da dissertação.

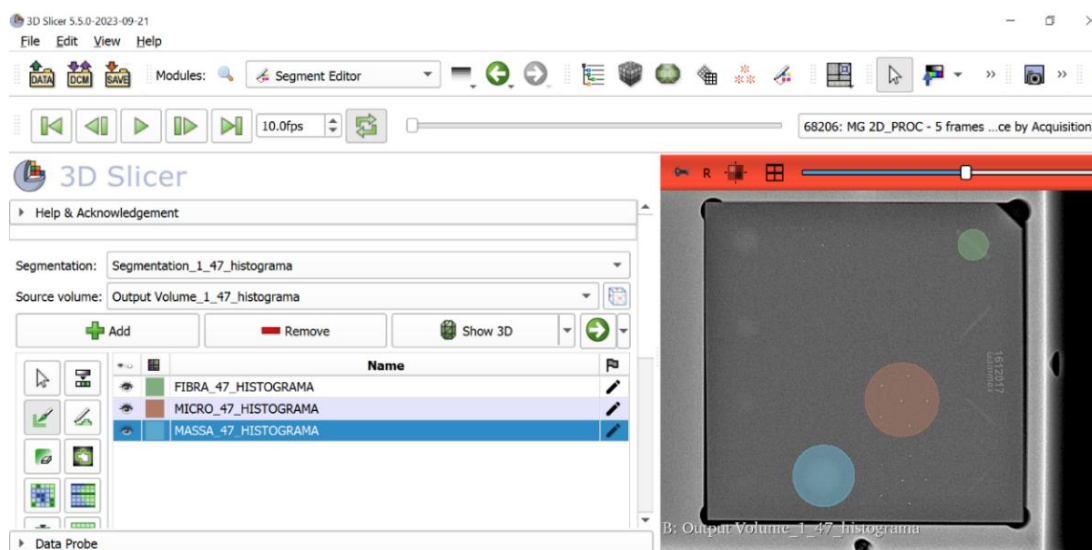
5.10 Segmentação

A segmentação da imagem é a delimitação de uma região de interesse ou ROI (*Region Of Interest*), essa segmentação pode ser manual, automática ou semi automática.

Para manter um padrão na segmentação, uma ROI em formato circular de raio definido foi utilizado para a delimitação das estruturas de interesse do simulador. Sendo o raio de 5 mm para fibra, 12 mm para as microcalcificações e 10 mm para a massa. Esse padrão de segmentação também foi seguido para as imagens do CMMD.

A figura 25 apresenta as ROIs delimitando as estruturas no simulador ACR.

Figura 25 – Tela do 3D Slicer módulo de segmentação, com as ROIs delimitando as estruturas no simulador ACR.



Fonte: Autor da dissertação.

5.11 Extração das Características Radiômicas

A literatura é muito abrangente no que se refere à seleção das características, os autores ou se baseiam em algoritmos de seleção ou em resultados de alterações para patologias específicas. O próprio IBSI também não caracteriza uma característica como sendo específica para um tipo de modalidade de imagem, apenas discrimina quais características são calculadas em informações

volumétricas (3D), e em informações planas (2D). Nos trabalhos realizados até o momento pelo IBSI, os valores de referência são de informações extraídas principalmente de imagens de Tomografia Computadorizada.

Para este trabalho, ficou definido que as características que apresentarem valores, que quando comparados entre os três equipamentos, apresentaram um Coeficiente de Variação (CV) de até 15%, serão selecionadas para as análises de reprodutibilidade e estabilidade, pois conforme resultados apresentados pelo IBSI, os valores extraídos possuem um limiar, e não um valor exato. Este valor de até 15% no CV, para definir quais características seriam selecionadas, foi determinado a partir do cálculo da média, desvio padrão e coeficiente de variação dos valores fornecidos pelo IBSI, no manual de referência de 2020, com as configurações de análises próximas as aplicadas neste trabalho, sendo que as características que puderam ser comparadas com o simulador ACR possuem valores que variaram entre 9% e 17% (SHUR et al., 2021; ZWANENBURG; LEGER; et al., 2020b).

A tabela 4 apresenta os CV de algumas características que foram analisadas neste trabalho, e destaca a importância dos processamentos pós aquisição. Na coluna **A** temos as análises com as imagens mantidas com seus parâmetros originais. Em **B** quando somente os espaçamentos entre pixel são processados para ficarem semelhantes em todas as imagens. E em **C** quando são aplicados todos os processamentos pós aquisição da imagem necessários para a padronização. Nesta tabela, através do CV, demonstra que as características extraídas apresentam valores aproximados, quando comparados entre os três equipamentos, conforme são aplicados os processamentos pós aquisição.

Tabela 4 – CV das características radiômicas, onde: **A** – entre imagens do simulador sem processamento pós aquisição; **B** – entre imagens do simulador com reamostragem dos pixels; **C** – entre imagens do simulador com reamostragem dos pixels e correspondência do histograma.

Classe	Característica	A (cv)	B (cv)	C (cv)
First Order	10 Percentile	81,84	81,84	1,04
	90 Percentile	81,75	81,75	0,54
	Energy	85,29	79,75	12,12
	Maximum	81,69	81,69	1,55
	Mean	81,79	81,80	0,24
	Median	58,21	58,21	0,18
	Root Mean Squared	81,79	81,80	0,24
GLCM	IMC2	17,99	13,60	7,21
	MCC	10,30	23,85	0,66
	Sum Entropy	35,41	33,84	13,97
GLDM	Dependence Entropy	15,73	12,43	3,93
	Small Dependence Low Gray Level Emphasis	84,01	98,10	13,81
GLRLM	Run Entropy	20,31	19,11	4,11
	Run Percentage	14,14	15,66	11,51
	Short Run Emphasis	9,78	11,0	8,20
GLSZM	Size Zone Non Uniformity Normalized	25,71	28,73	10,35
	Small Area Emphasis	12,85	14,51	5,25
	Zone Entropy	12,27	10,25	2,00

Fonte: Autor da dissertação.

6 RESULTADOS E DISCUSSÃO

O desenvolvimento da construção de soluções baseadas em Inteligência Artificial como a sugerida por *Radiomics*, requer primeiramente, muita pesquisa prévia, uma equipe multidisciplinar, e o desenho do fluxo de trabalho e quais as principais informações que serão utilizadas para a criação deste modelo, essa etapa inicial do desenvolvimento, descrita neste trabalho, é importante para “enxugar” a quantidade de informações geradas, para que posteriormente seja possível focar nas que realmente importam em uma análise que irá envolver um volume extraordinário de dados.

Na figura 26 é apresentado um fluxo de etapas para a padronização, com as sequências de processamentos aplicados nas imagens, e as análises realizadas neste trabalho.



Fonte: Autor da dissertação.

Na tabela 5 estão as informações da imagem no seu formato original e após o processamento pós aquisição. Lembrando que para este trabalho a imagem de referência foi a do DBT, pois se trata da imagem de maior tamanho de pixel, uma vez que no processo de reamostragem de pixels a chance de erros de informações são menores quando são realizadas interpolação de pixels do que na criação de novos elementos de imagem.

Tabela 5 – Informações da imagem após processamento no domínio espacial.

	Tamanho do Pixel (μm)	Matriz	Níveis de Intensidade de Tons de Cinza
Imagem de Referência (DBT)	100 x 100	2.394 x 2.850	258 – 3.356
Imagem Original (PC1 e PC2)	50 x 50	4.915 x 5.355	60.811 – 65.535
Processamento Pós Aquisição (PC1 e PC2)	100 x 100	2.457 x 2.677	258 – 3.356

Fonte: Autor da dissertação.

6.1 Análise da Estabilidade e Reprodutibilidade das Características Radiômicas

Reiterando, foram feitas aquisições em dois períodos de tempo, e em cada coleta dos ensaios foram adquiridas 5 imagens do simulador em cada equipamento, para que seja possível avaliar tanto a estabilidade das características, e a reprodutibilidade dessas características extraídas em um mesmo equipamento, entre os equipamentos, e de todo o processo aplicado para a extração das características. Para facilitar a observação do comportamento dos valores calculados em cada caso, as análises foram divididas em “1º ensaio” e “2º ensaio” e de acordo com o *test object* do simulador. As informações dos gráficos seguem o seguinte padrão: em azul para os ensaios realizados no DBT, em laranja no PC1 e em verde para o PC2. Os tons escuros e claros das barras azul, laranja e verde são para diferenciar os valores calculados no 1º ensaio e o 2º ensaio, respectivamente.

As características estão apresentadas em grupo de gráficos nas figuras a seguir, apenas por estarem dentro de uma mesma escala de valores, e não por estarem relacionadas entre si, apenas para uma melhor visualização.

6.1.1 Fibra

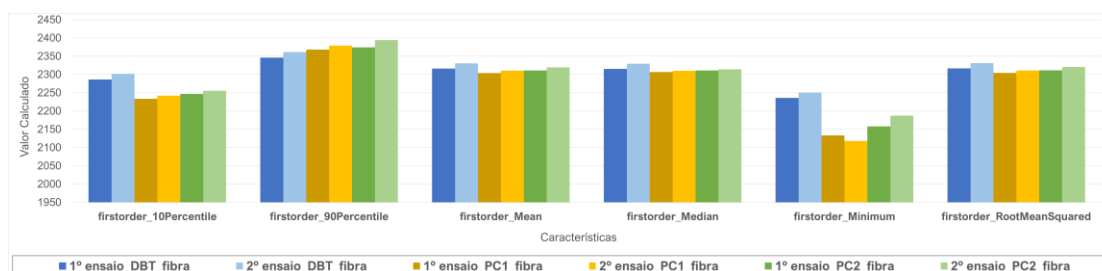
Na avaliação do *test object* fibra do simulador ACR, das 108 características calculadas pela biblioteca do *PyRadiomics*, 23 apresentaram um coeficiente de variação elegível para continuar nas análises ($\leq 15\%$), foi possível observar uma

flutuação dos valores calculados para um mesmo equipamento nos dois ensaios, alguns fatores que podem ocasionar tais variações, são, por exemplo, aqueles que afetam a qualidade da imagem no momento da aquisição, como o ruído eletrônico, desgaste natural do equipamento e outras que são as relacionadas a segmentação da imagem, que mesmo sendo realizada com formatos fixos e por um único observador (o autor da dissertação), podem sofrer variações devido ao seu posicionamento na imagem.

Já para as características extraídas em equipamentos diferentes, apesar da aplicação dos processamentos pós aquisição, ainda foi possível observar diferenças nos valores calculados, que também podem estar relacionados aos fatores que influenciam a aquisição de imagens digitais, como o ruído eletrônico, desgaste do tubo de raios X, temperatura e desgaste do detector, entre outros fatores que podem se alterar de formas diferentes nos três equipamentos. Mas apesar das flutuações as características apresentaram valores aproximados no 1º e 2º ensaios para um mesmo equipamento, como entre equipamentos nos dois ensaios.

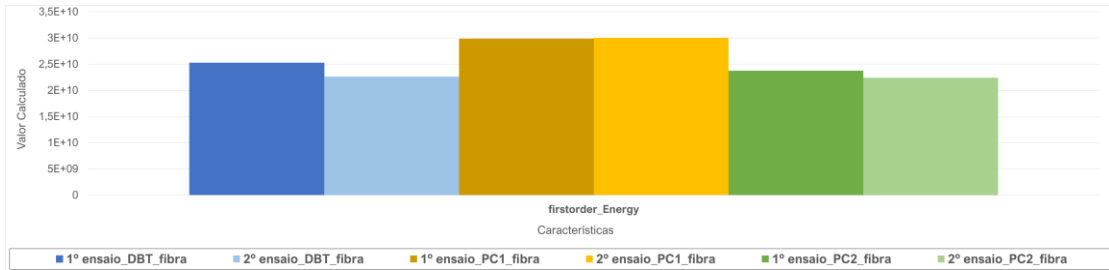
Os gráficos das figuras 27 a 36 apresentam os valores calculados das 23 características, com as comparações entre os ensaios para um mesmo equipamento, e entre equipamentos. Os gráficos seguem os critérios de cores e distribuições citados anteriormente.

Figura 27 – Gráficos comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



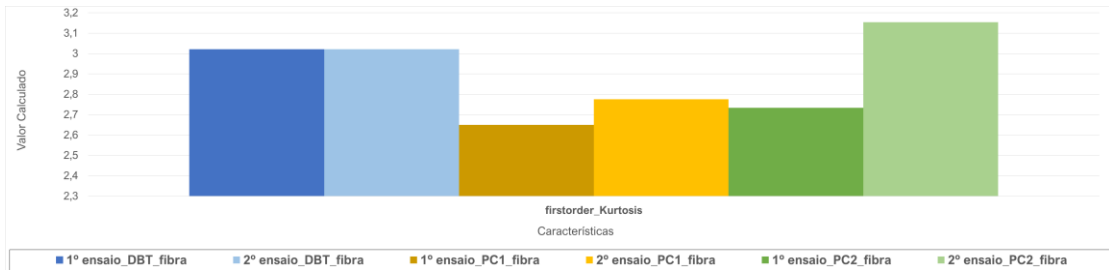
Fonte: Autor da dissertação.

Figura 28 – Gráfico comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



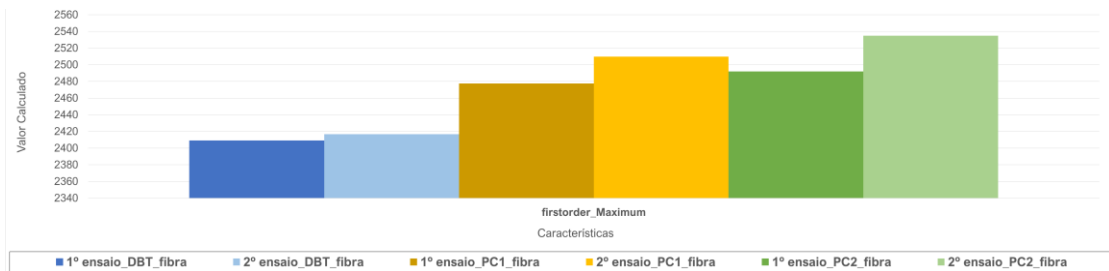
Fonte: Autor da dissertação.

Figura 29 – Gráfico comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



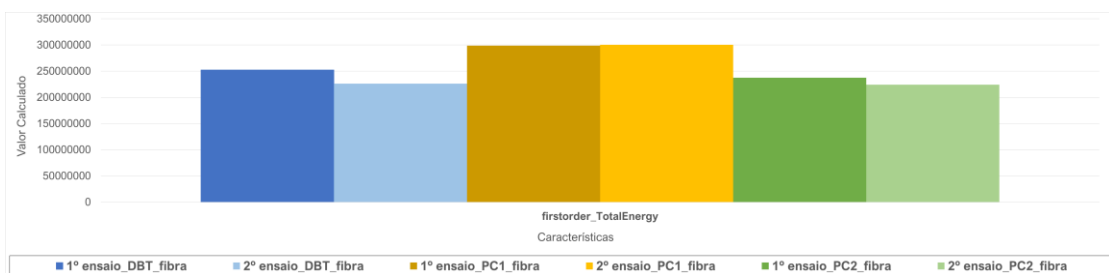
Fonte: Autor da dissertação.

Figura 30 – Gráfico comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



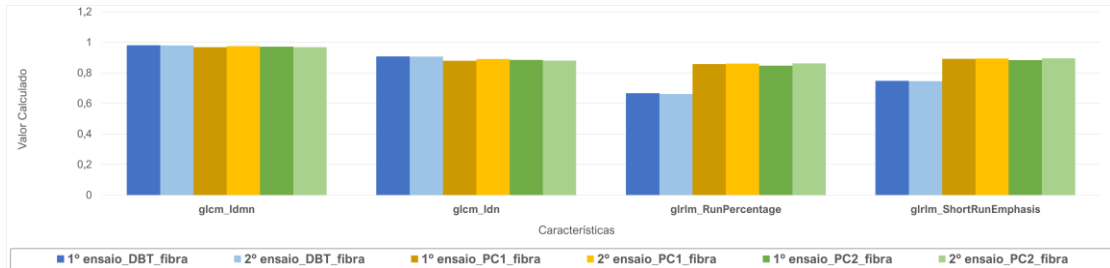
Fonte: Autor da dissertação.

Figura 31 – Gráfico comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



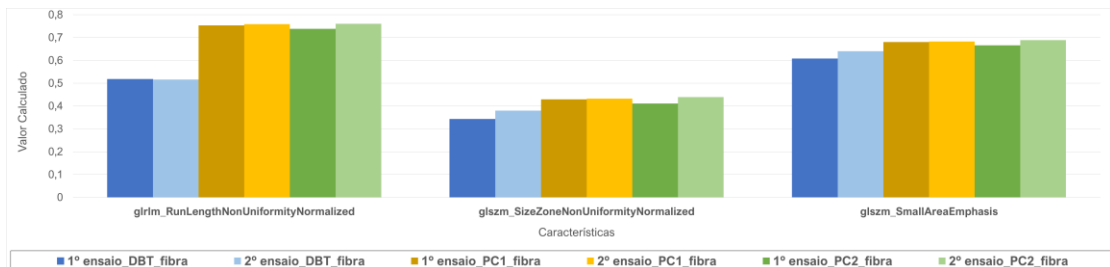
Fonte: Autor da dissertação.

Figura 32 – Gráficos comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



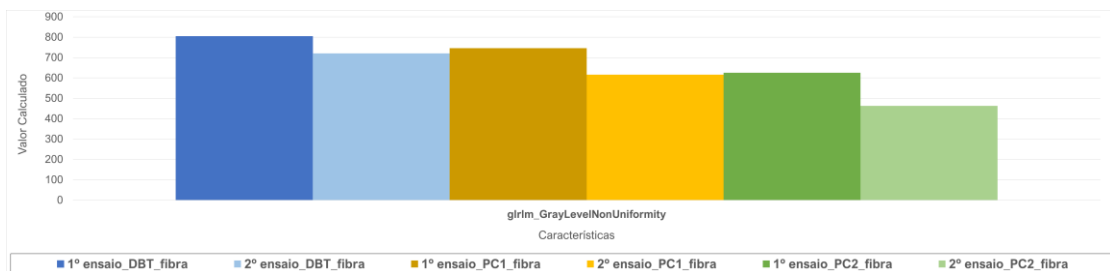
Fonte: Autor da dissertação.

Figura 33 – Gráficos comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



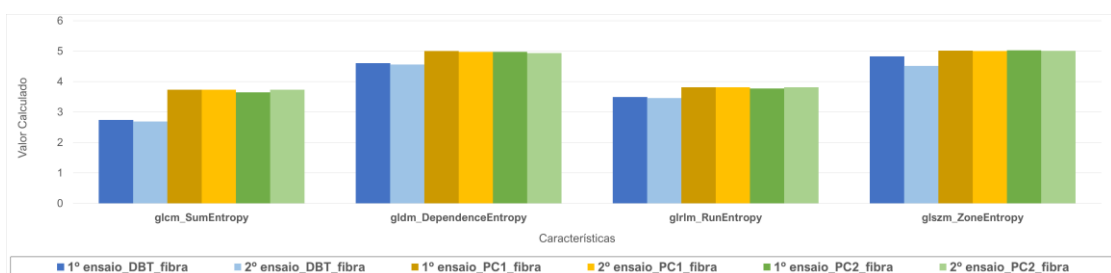
Fonte: Autor da dissertação.

Figura 34 – Gráfico comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



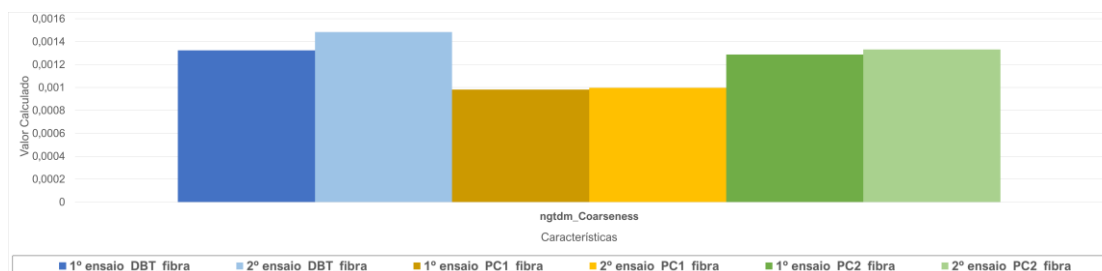
Fonte: Autor da dissertação.

Figura 35 – Gráficos comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



Fonte: Autor da dissertação.

Figura 36 – Gráficos comparando os valores do *test object* fibra entre ensaios no mesmo equipamento, e entre equipamentos.



Fonte: Autor da dissertação.

No sentido de compreender se a ocorrência de flutuações nos valores calculados nas 23 características, consideradas reprodutíveis e estáveis, estavam dentro de uma confiabilidade e concordância tanto entre os ensaios como entre os equipamentos, foi calculado o coeficiente de correlação intraclasse (para o mesmo equipamento) e interclasse (entre equipamentos) – (ICC).

As correlações são consideradas fortes quando o coeficiente está entre 0,75 a 1,0, significando que as medidas realizadas pelos avaliadores seguem um padrão.

Neste caso os equipamentos utilizados neste estudo (os avaliadores) apresentaram resultados confiáveis, ou homogêneos, para caracterizar uma mesma medida (BORTOLETTO LINO DOS SANTOS, 2018). Portanto as flutuações observadas nos valores calculados estão dentro de uma variação estatisticamente aceitável. A tabela 6 apresenta os resultados da análise de concordância ICC para o *test object* fibra.

Tabela 6 - Resultados do ICC para as medidas realizadas no *test object* fibra.

Tipo	Descrição	ICC	CI95%
ICC1	Avaliadores absolutos únicos	0,981915682	[0.97 0.99]
ICC2	Avaliadores aleatórios únicos	0,981915732	[0.97 0.99]
ICC3	Avaliadores fixos únicos	0,981931873	[0.97 0.99]
ICC1k	Média dos avaliadores absolutos	0,996939829	[0.99 1.]
ICC2k	Média dos avaliadores aleatórios	0,996939838	[0.99 1.]
ICC3k	Média dos avaliadores fixos	0,996942611	[0.99 1.]

Fonte: Autor da dissertação.

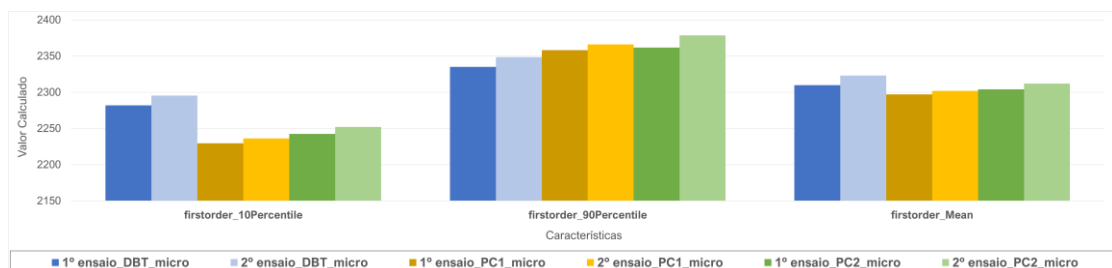
Os resultados ICC1k, ICC2k e ICC3k indicam a confiabilidade dos k avaliadores ao trabalhar como um grupo (enquanto ICC1, ICC2 e ICC3 representam a confiabilidade dos avaliadores como indivíduos). Os valores ICC1k, ICC2k e ICC3k sempre serão maiores porque vários avaliadores trabalhando juntos sempre darão um resultado estatisticamente mais confiável (KOO; LI, 2016).

6.1.2 Microcalcificação

Na avaliação do *test object* de microcalcificação, das 108 características calculadas pela biblioteca do *PyRadiomics*, 25 apresentaram coeficientes de variação dentro dos limites determinados. O mesmo tipo de flutuação nos valores calculados ocorreu para a avaliação do *test object* de microcalcificação do simulador, sendo que as características, apresentaram valores calculados aproximados tanto entre os ensaios para o mesmo equipamento como entre equipamentos.

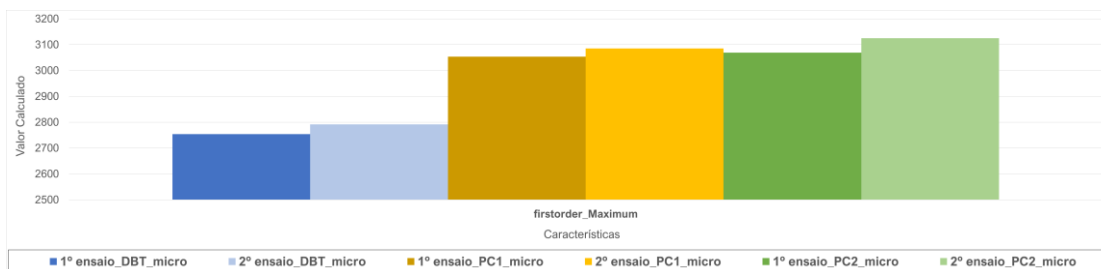
Os gráficos das figuras 37 a 47 apresentam os valores calculados das 25 características, com as comparações entre os ensaios para um mesmo equipamento, e entre os equipamentos. Os gráficos seguem os mesmos critérios de cores e distribuições citados anteriormente.

Figura 37 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



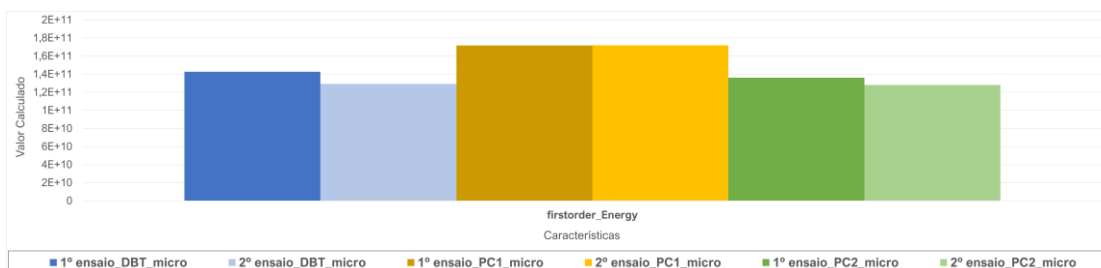
Fonte: Autor da dissertação.

Figura 38 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



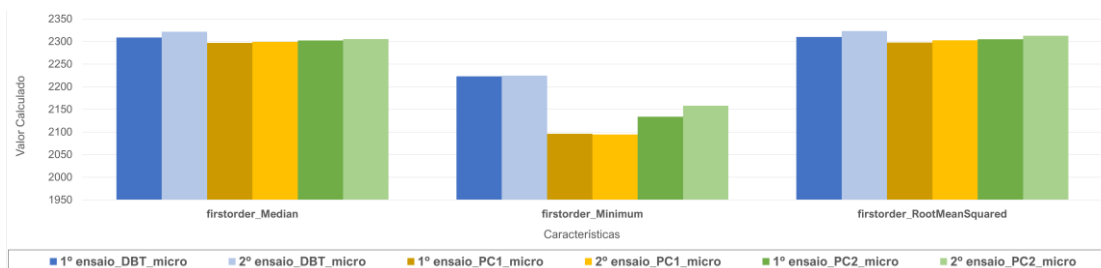
Fonte: Autor da dissertação.

Figura 39 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



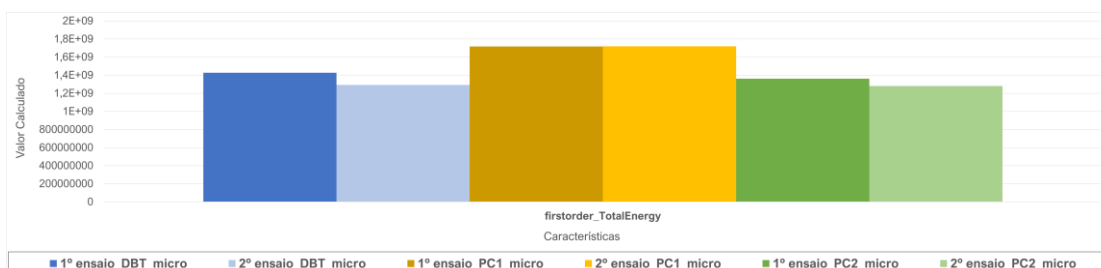
Fonte: Autor da dissertação.

Figura 40 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



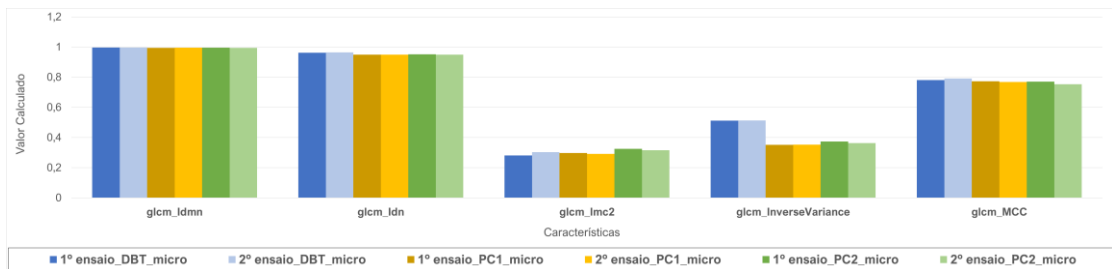
Fonte: Autor da dissertação.

Figura 41 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



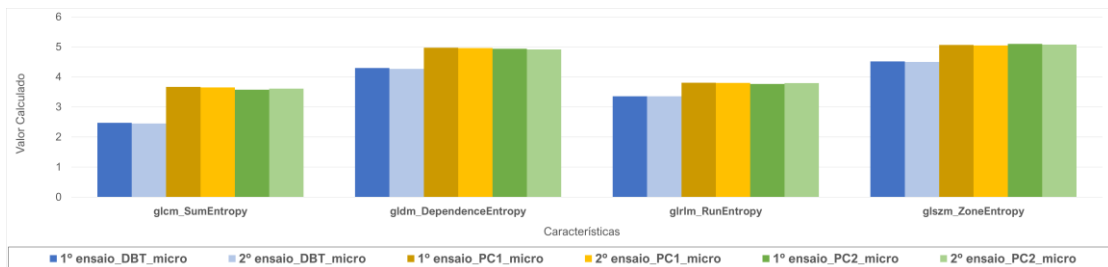
Fonte: Autor da dissertação.

Figura 42 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



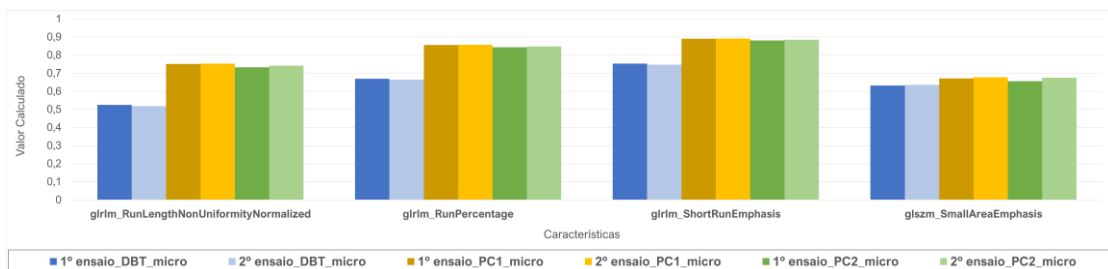
Fonte: Autor da dissertação.

Figura 43 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



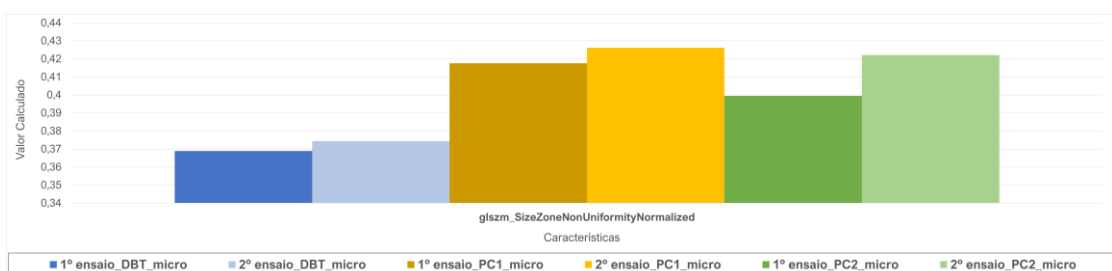
Fonte: Autor da dissertação.

Figura 44 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



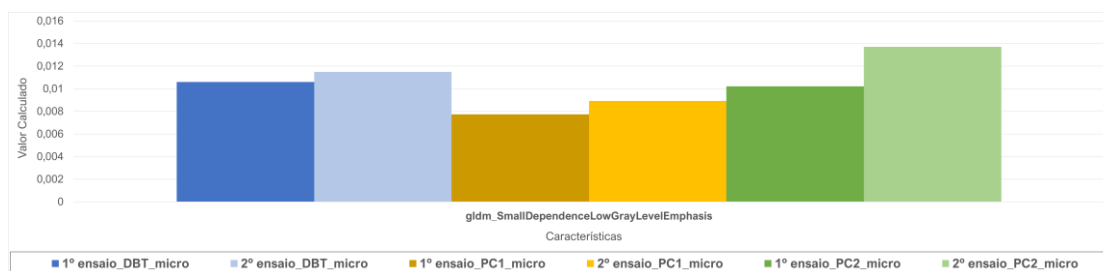
Fonte: Autor da dissertação.

Figura 45 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



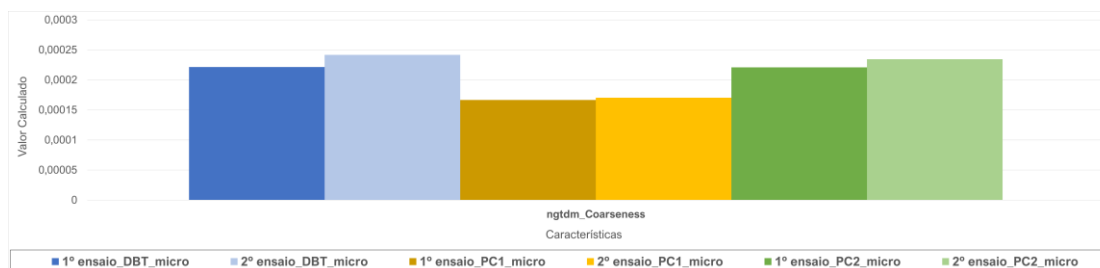
Fonte: Autor da dissertação.

Figura 46 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



Fonte: Autor da dissertação.

Figura 47 – Gráficos comparando os valores para o *test object* microcalcificação entre ensaios no mesmo equipamento, e entre equipamentos.



Fonte: Autor da dissertação.

No sentido de também compreender se as flutuações ocorridas nos valores calculados nas 25 características, consideradas reprodutíveis e estáveis, estavam dentro de uma confiabilidade e concordância tanto entre os ensaios como entre os equipamentos, também foi calculado o ICC das medidas realizadas no *test object* microcalcificação.

Os resultados do ICC indicaram que as medidas realizadas seguem um padrão para os avaliadores (equipamentos), que apresentaram resultados confiáveis, ou homogêneos, para caracterizar uma mesma medida. Portanto as flutuações observadas nos valores calculados estão dentro de uma variação estatisticamente aceitável. A tabela 7 apresenta os resultados da análise de concordância ICC para o *test object* microcalcificação.

Tabela 7 - Resultados do ICC para as medidas realizadas no *test object* microcalcificação.

Tipo	Descrição	ICC	CI95%
ICC1	Avaliadores absolutos únicos	0,981235961	[0.97 0.99]
ICC2	Avaliadores aleatórios únicos	0,98123601	[0.97 0.99]
ICC3	Avaliadores fixos únicos	0,981251352	[0.97 0.99]
ICC1k	Média dos avaliadores absolutos	0,996822982	[0.99 1.]
ICC2k	Média dos avaliadores aleatórios	0,996822991	[0.99 1.]
ICC3k	Média dos avaliadores fixos	0,996825629	[0.99 1.]

Fonte: Autor da dissertação.

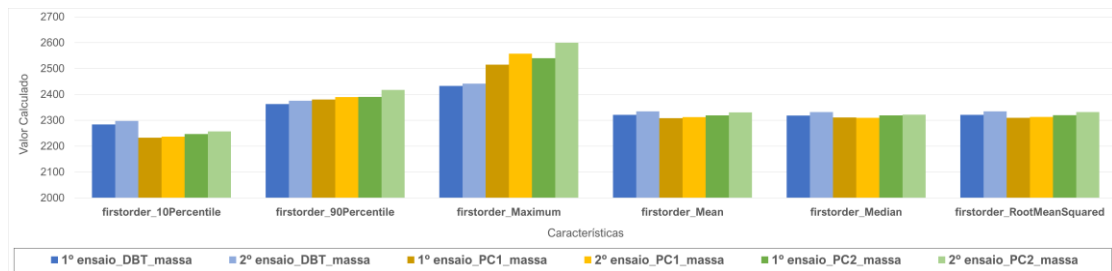
Os resultados ICC1k, ICC2k e ICC3k indicam a confiabilidade dos k avaliadores ao trabalhar como um grupo (enquanto ICC1, ICC2 e ICC3 representam a confiabilidade dos avaliadores como indivíduos). Os valores ICC1k, ICC2k e ICC3k sempre serão maiores porque vários avaliadores trabalhando juntos sempre darão um resultado estatisticamente mais confiável (KOO; LI, 2016).

6.1.3 Massa

Na avaliação do *test object* de massa das 108 características calculadas pela biblioteca do *PyRadiomics*, 25 apresentaram coeficientes de variação dentro dos limites determinados. Apresentaram também as mesmas flutuações observadas nos valores calculados nos *test object* de fibra e microcalcificação.

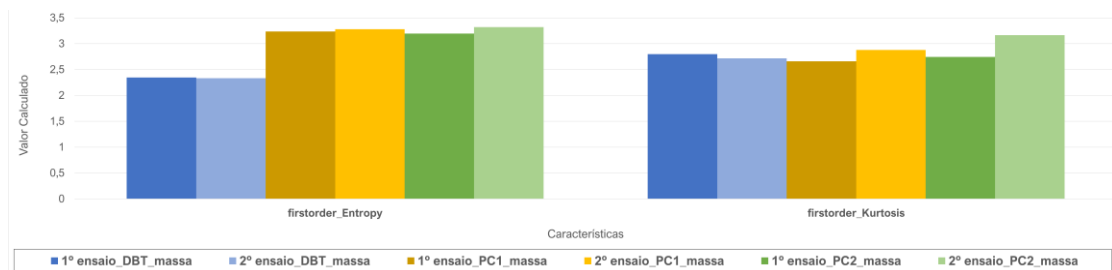
Os gráficos das figuras 48 a 58 apresentam os valores calculados das 25 características, com as comparações entre os ensaios para um mesmo equipamento, e entre os equipamentos. Os gráficos seguem os mesmos critérios de cores e distribuições citados anteriormente.

Figura 48 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



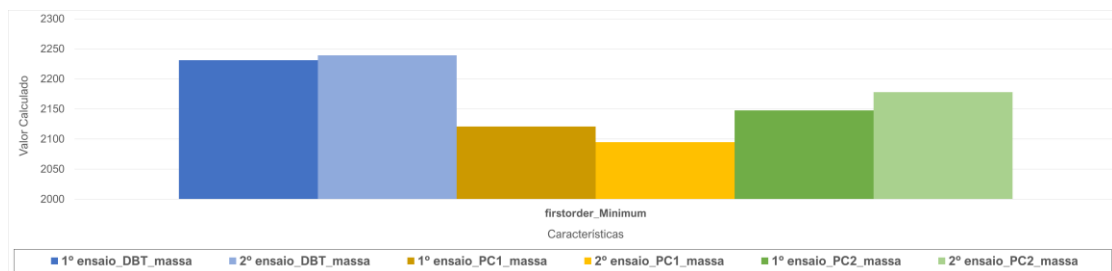
Fonte: Autor da dissertação.

Figura 49 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



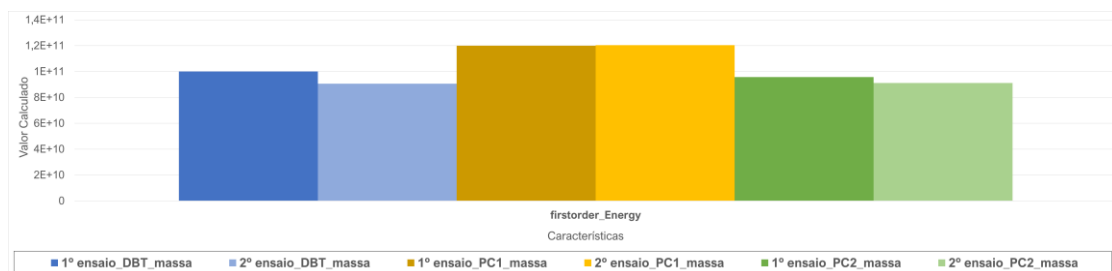
Fonte: Autor da dissertação.

Figura 50 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



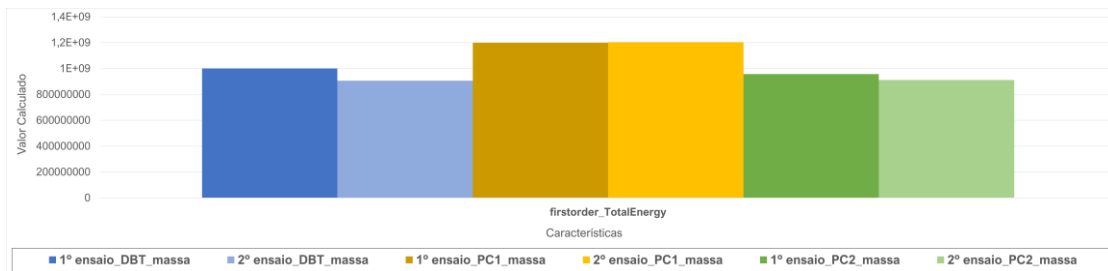
Fonte: Autor da dissertação.

Figura 51 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



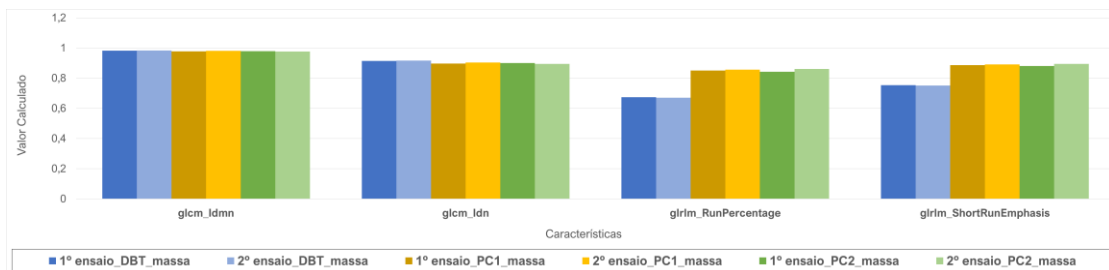
Fonte: Autor da dissertação.

Figura 52 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



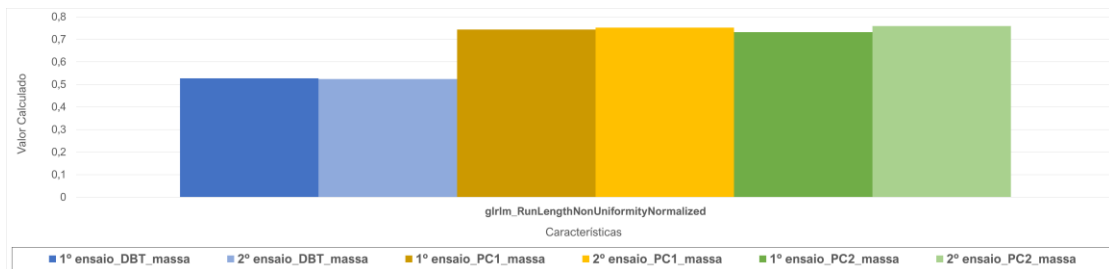
Fonte: Autor da dissertação.

Figura 53 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



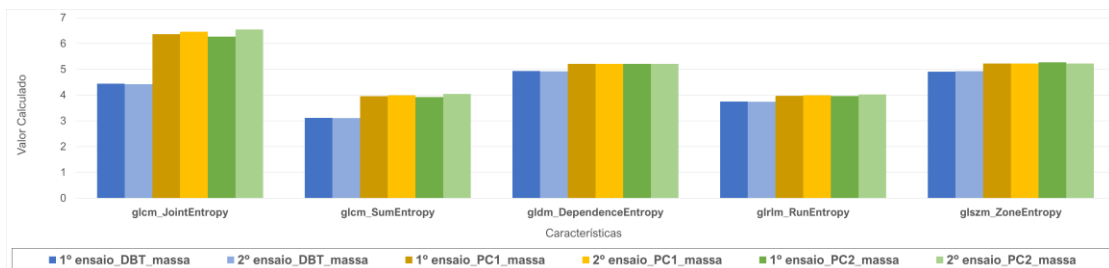
Fonte: Autor da dissertação.

Figura 54 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



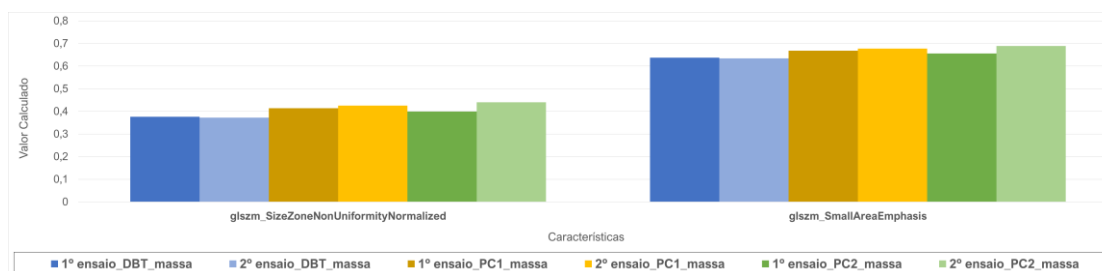
Fonte: Autor da dissertação.

Figura 55 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



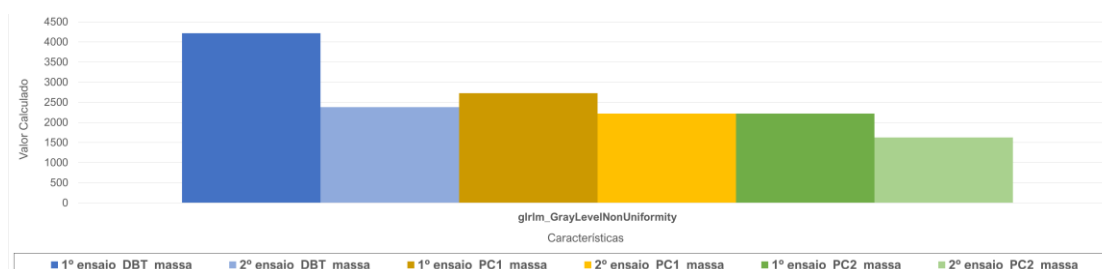
Fonte: Autor da dissertação.

Figura 56 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



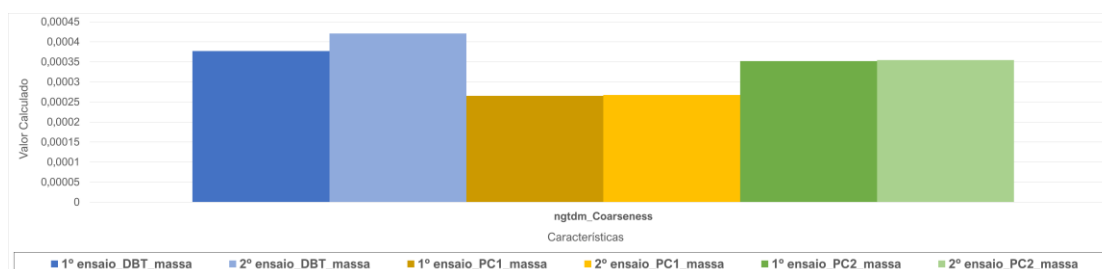
Fonte: Autor da dissertação.

Figura 57 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



Fonte: Autor da dissertação.

Figura 58 – Gráficos comparando os valores para o *test object* massa para ensaios no mesmo equipamento, e entre equipamentos.



Fonte: Autor da dissertação.

Para também compreender se a ocorrência das flutuações nos valores calculados nas 25 características, consideradas reprodutíveis e estáveis, estavam dentro de uma confiabilidade e concordância tanto entre os ensaios como entre os equipamentos, também foi calculado o ICC das medidas realizadas no *test object* massa.

Os resultados do ICC indicaram que as medidas realizadas seguem um padrão para os avaliadores (equipamentos), e apresentaram resultados confiáveis, ou homogêneos, para caracterizar uma mesma medida. Portanto as flutuações

observadas nos valores calculados estão dentro de uma variação estatisticamente aceitável. A tabela 8 apresenta os resultados da análise de concordância ICC para o *test object* massa.

Tabela 8 - Resultados do ICC para as medidas realizadas no test object Massa.

Tipo	Descrição	ICC	CI95%
ICC1	Avaliadores absolutos únicos	0,982428393	[0.97 0.99]
ICC2	Avaliadores aleatórios únicos	0,982428436	[0.97 0.99]
ICC3	Avaliadores fixos únicos	0,98244282	[0.97 0.99]
ICC1k	Média dos avaliadores absolutos	0,997027878	[0.99 1.]
ICC2k	Média dos avaliadores aleatórios	0,997027885	[0.99 1.]
ICC3k	Média dos avaliadores fixos	0,997030355	[0.99 1.]

Fonte: Autor da dissertação.

Os resultados ICC1k, ICC2k e ICC3k indicam a confiabilidade dos k avaliadores ao trabalhar como um grupo (enquanto ICC1, ICC2 e ICC3 representam a confiabilidade dos avaliadores como indivíduos). Os valores ICC1k, ICC2k e ICC3k sempre serão maiores porque vários avaliadores trabalhando juntos sempre darão um resultado estatisticamente mais confiável (KOO; LI, 2016).

6.2 Comparativo dos valores calculados com o simulador ACR com dados IBSI

O grupo de trabalho do IBSI apresentou dados preliminares de seus estudos das características radiômicas calculados em simuladores. O estudo disponibilizado pelo grupo, foi realizado em equipamentos de Tomografia Computadorizada e Medicina Nuclear, a descrição de toda a metodologia, incluindo as configurações de reconstrução e processamentos da imagem estão disponíveis nos materiais suplementares do trabalho de (ZWANENBURG; LEGER; et al., 2020a).

Para o presente trabalho apenas algumas características puderam ser comparadas com as do IBSI, dadas as diferenças intrínsecas de cada modalidade de imagem, portanto foram selecionados os resultados apresentados pelo IBSI de imagens com o espaçamento entre pixels próximas das utilizadas neste trabalho,

com segmentações de estruturas do simulador IBSI com densidades próximas ao do simulador ACR e de características que utilizaram as informações de imagens planas (2D).

Na tabela 9 são apresentados os limiares com os valores máximo e mínimo calculados com o simulador ACR e os apresentados pelo IBSI, para as características que puderam ser comparadas. É possível observar que os valores calculados neste trabalho se aproximaram dos calculados pelo IBSI, sugerindo que o software 3D Slicer e suas ferramentas de segmentação, juntamente com o plugin PyRadiomics para extração das características, foram capazes de reproduzir resultados semelhantes aos encontrados pelo principal grupo de trabalho para a padronização dos dados para pesquisas com radiômica.

Não foram realizados testes de significância para esta análise para saber o quanto os resultados encontrados são semelhantes, pois trata-se de uma amostra pequena, portanto os valores calculados foram somente confrontados se estavam dentro do limiar dos valores publicados pelo IBSI.

Tabela 9 – Valores das características do simulador ACR e do IBSI.

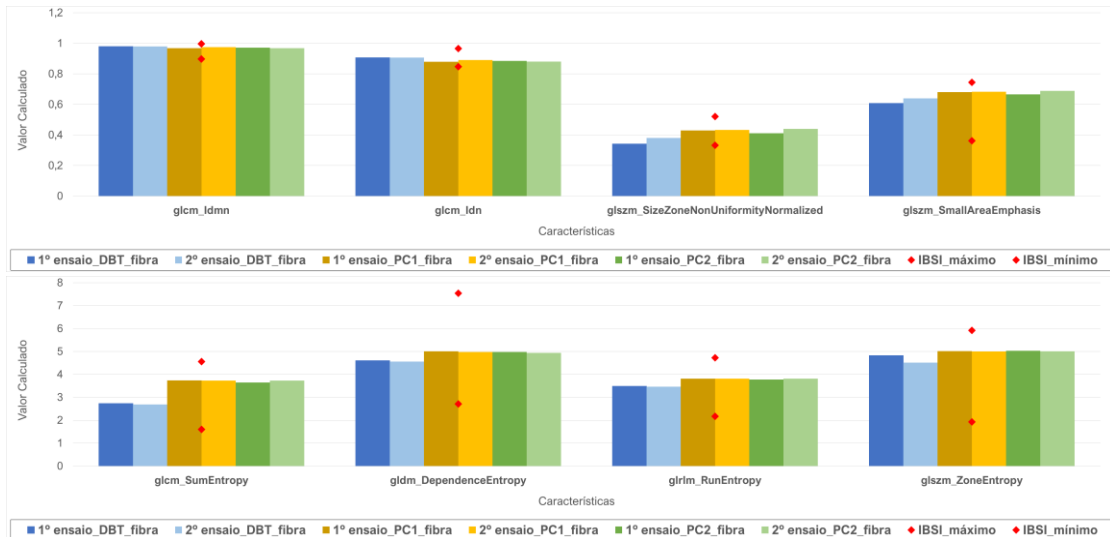
Classe	Característica	Limiar do Valor Calculado	
		ACR	IBSI
First Order	Entropy	2,05 – 3,02	1,27 – 3,73
GLCM	IDMN	0,96 – 0,99	0,89 – 0,99
GLCM	IDN	0,87 – 0,96	0,85 – 0,96
GLRLM	Run Entropy	3,48 – 3,81	2,17 – 5,35
GLRLM	Run Porcentage	0,66 – 0,84	0,55 – 0,70
GLSZM	Size Zone Non Uniformity Normalized	0,34 – 0,42	0,19 – 0,52
GLSZM	Zone Entropy	4,8 – 5,03	1,93 – 7,00

Fonte: Autor da dissertação.

Os gráficos nas figuras 59, 60 e 61 apresentam os valores calculados nos equipamentos PC1, PC2 e DBT com os limiares mínimo e máximo das características publicados pelo IBSI, para os *test objects* de fibra, microcalcificação e massa, respectivamente. Os valores encontrados no simulador ACR ficaram dentro do mesmo limiar dos apresentados pelo IBSI para 8 características

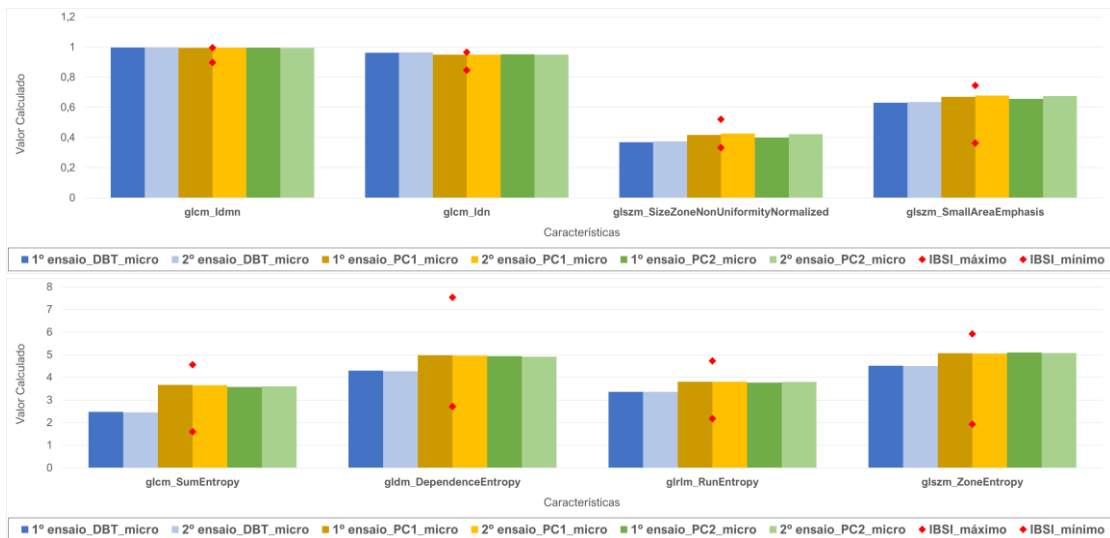
calculadas no *test object* fibra e microcalcificação, e 10 características no *test object* massa.

Figura 59 – Resultados para o *test object* de fibra. Os pontos em vermelho representam os limiares de valores apresentados pelo IBSI.



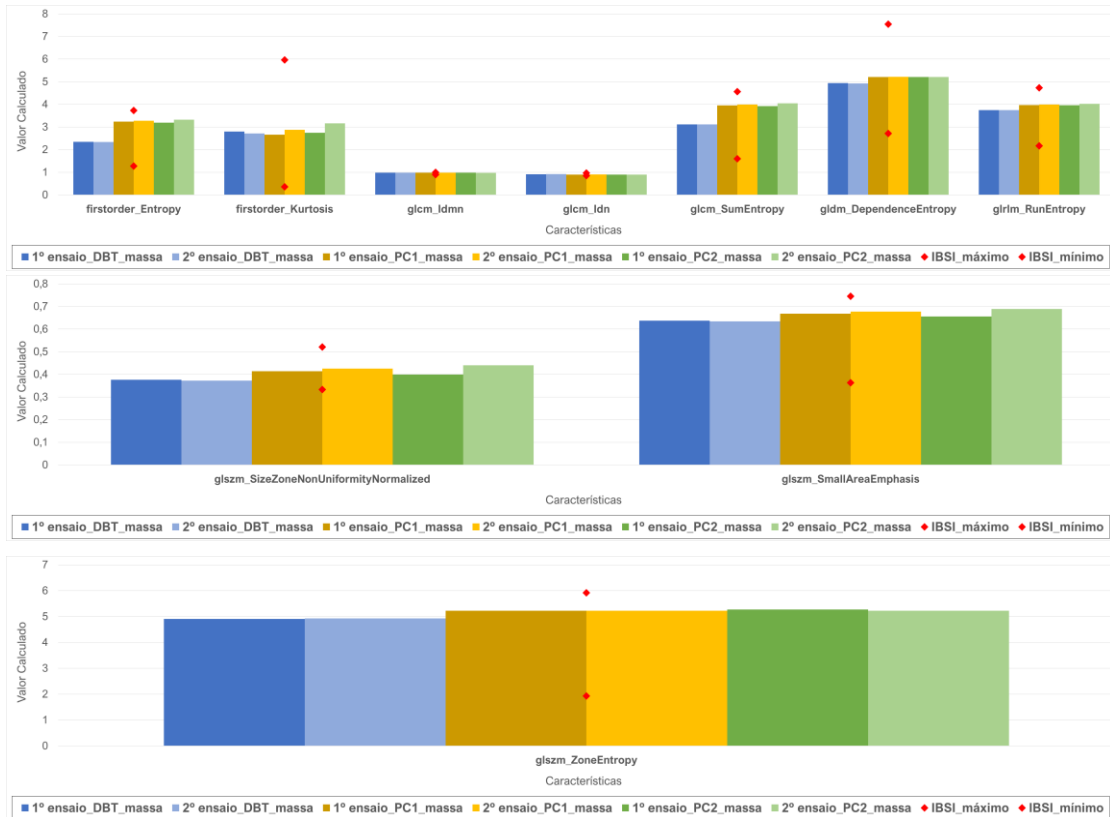
Fonte: Autor da dissertação.

Figura 60 – Resultados para o *test object* de microcalcificação. Os pontos em vermelho representam os limiares de valores apresentados pelo IBSI.



Fonte: Autor da dissertação.

Figura 61 – Resultados para o *test object* de massa. Os pontos em vermelho representam os limiares de valores apresentados pelo IBSI.



Fonte: Autor da dissertação.

6.3 Simulador ACR vs. Banco de Imagens CMMD

Para verificar a compatibilidade das estruturas dos *test objects* do simulador com os principais achados patológicos em humanos, os valores calculados das segmentações do simulador ACR foram comparados com valores calculados em segmentações de estruturas equivalentes encontradas nas imagens do banco CMMD.

As imagens do banco CMMD foram disponibilizadas no *Cancer Imaging Archive* em 8 bits, e como já mencionado anteriormente neste trabalho, imagens com 8 bits possuem menores níveis de intensidade de tons de cinza que aquelas com 16 bits. Como as imagens do simulador ACR originalmente são de 16 bits, estas foram submetidas a mais um processamento pós aquisição da imagem para convertê-las em 8 bits, além daquelas necessárias para a extração dos dados

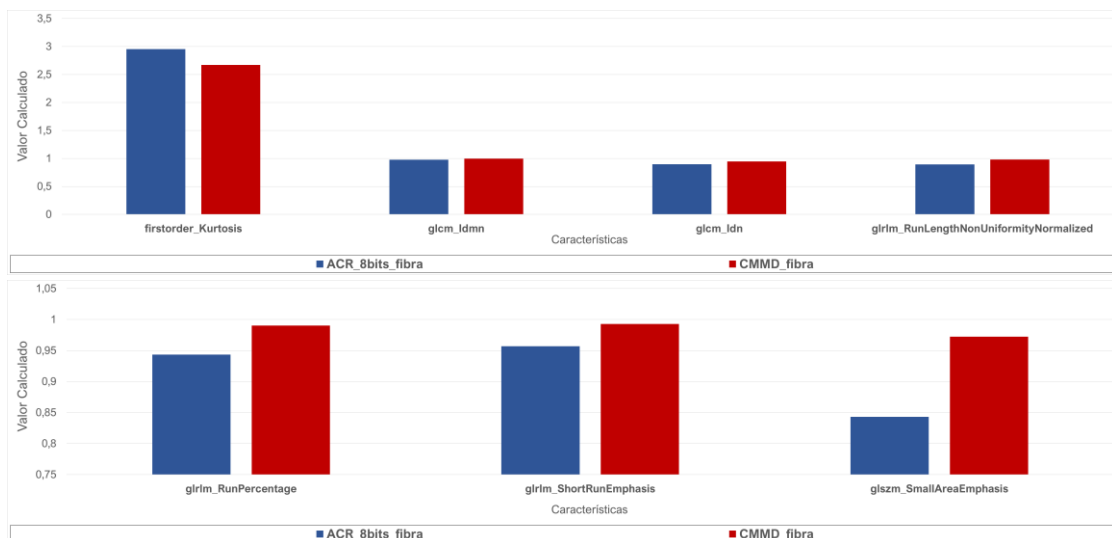
radiômicos. Vale destacar que para esta análise não foi possível comparar os valores obtidos com os reportados pelo IBSI, dada as diferenças entre os bits alocados em cada imagem.

Esse tipo de análise, além de possibilitar verificar se o simulador utilizado apresenta semelhanças com o tecido humano, pode servir também como linha de base para a seleção prévia das características que melhor representam a patologia ou alteração metabólica que se busca observar, é importante salientar, que apesar de existirem centenas de características radiômicas, as soluções criadas para a geração de um banco de dados para uma análise empregando machine learning utilizam apenas poucas características, e acabam sendo específicas para cada tipo de estudo, ou seja, são características selecionadas de acordo com a melhor resposta obtida para uma doença específica.

6.3.1 Fibra

Na análise da fibra, 7 características apresentaram valores calculados entre a segmentação no simulador e de achados patológicos nas imagens do CMMD, dentro do limite de 15% de CV estabelecido neste trabalho, os valores calculados para as duas situações são apresentados nos gráficos da figura 62.

Figura 62 – Análise comparativa para fibra entre o simulador ACR e o banco de imagens CMMD.

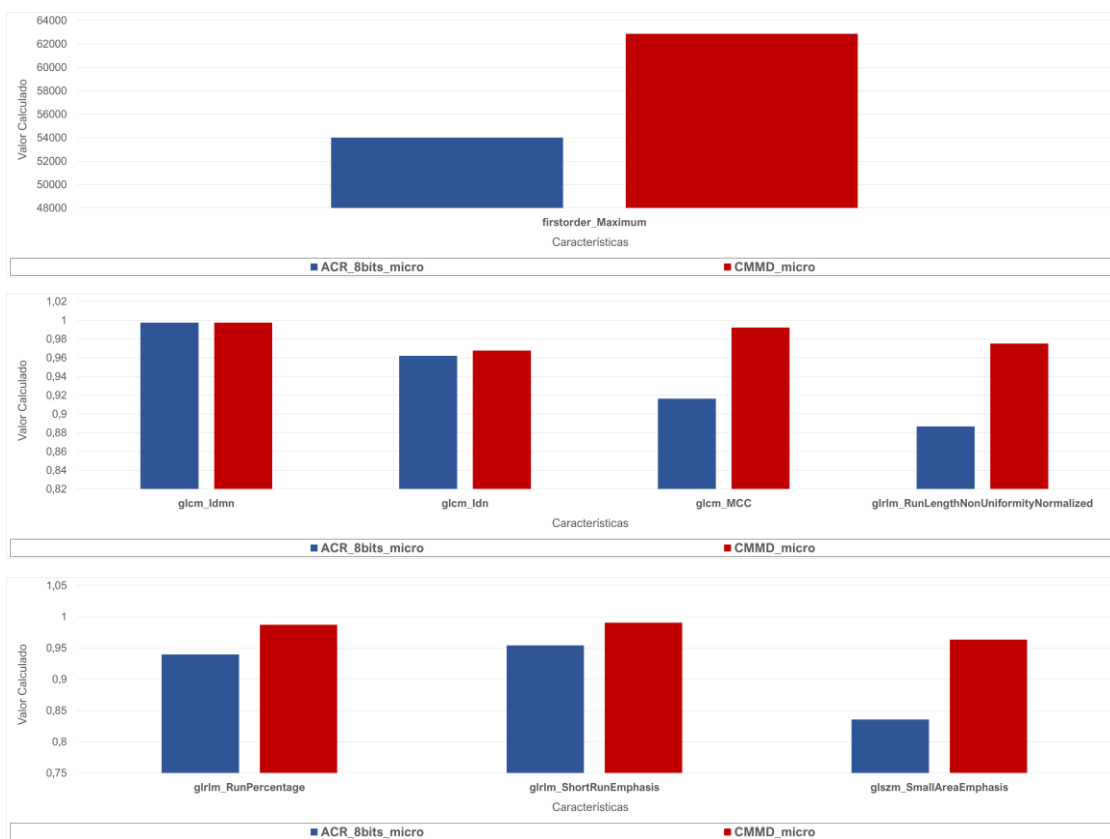


Fonte: Autor da dissertação.

6.3.2 Microcalcificação

Na análise de microcalcificações, 8 características apresentaram valores calculados dentro do limite de 15% de CV estabelecido neste trabalho, para a segmentação no simulador e dos achados patológicos nas imagens do CMMD, os valores calculados para as duas situações estão plotados na figura 63.

Figura 63 – Análise comparativa para segmentação de microcalcificação entre o simulador ACR e o banco de imagens CMMD.

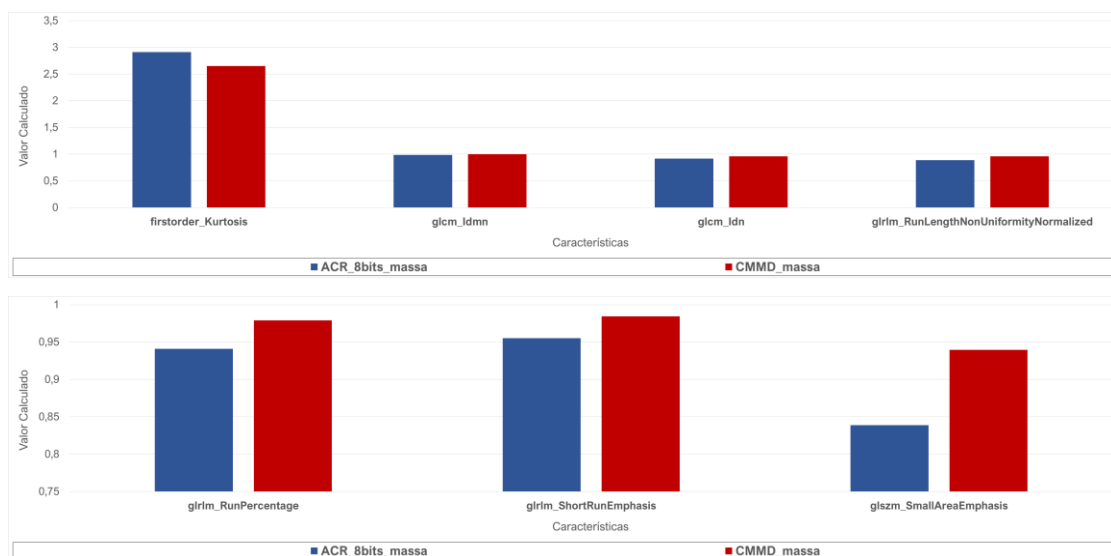


Fonte: Autor da dissertação.

6.3.3 Massa

Na análise da massa, 7 características apresentaram valores calculados para a segmentação no simulador e de achados patológicos nas imagens do CMMD, dentro do limite de 15% de CV estabelecido neste trabalho, os valores calculados para as duas situações são apresentados nos gráficos da figura 64.

Figura 64 – Análise comparativa para segmentação de massa entre o simulador ACR e o banco de imagens CMMD.



Fonte: Autor da dissertação.

As flutuações observadas nas análises comparativas dos *test objects* com os achados nas imagens do CMMD, podem ser ocasionadas pelas variações que o tecido mamário apresenta, as diferentes espessuras de compressão e por consequência os fatores técnicos aplicados para a aquisição da imagem como kilovtagem, miliamperage e combinação de filtro e alvo do mamógrafo. Enquanto que as imagens adquiridas com o simulador, apesar das diferenças tecnológicas dos equipamentos utilizados, a estrutura homogênea e a invariabilidade da espessura fazem com que ocorram variações mais sutis nos valores calculados das características.

6.4 Características Redundantes

Algumas características podem apresentar comportamento semelhantes entre si, sejam elas pertencentes a uma mesma classe ou não. Isto se dá ao fato de que algumas equações que dimensionam as relações entre pixels ou voxels utilizadas para calcular os valores das características radiômicas, realizam cálculos muito parecidos, muitas vezes diferenciando-se apenas em suas grandezas, como por exemplo na quantidade de elementos de imagens que estão sendo usados pela equação para realizar esses cálculos.

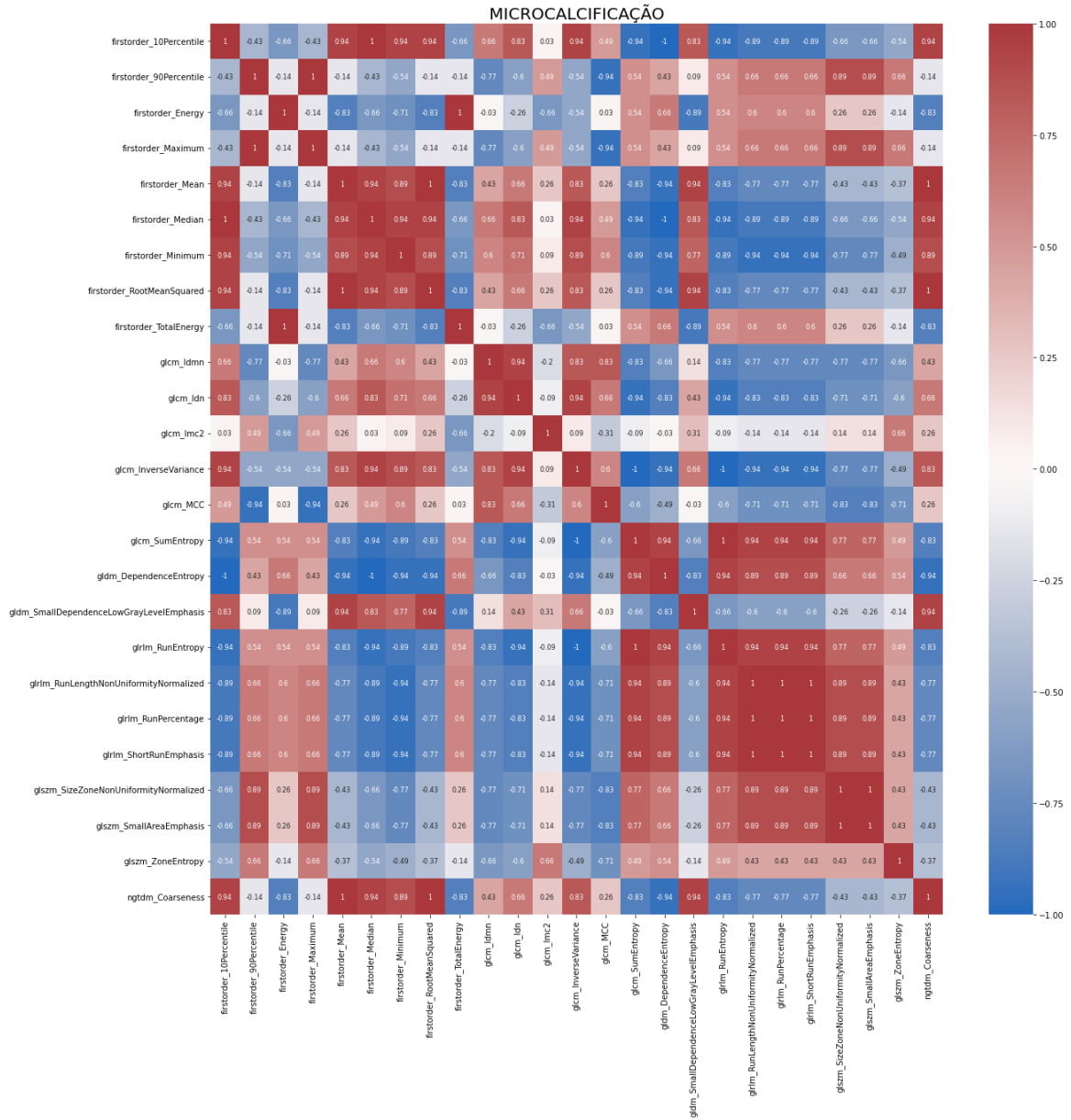
Este passo reduz o número de características a serem analisadas, para agilizar o processamento utilizado por soluções computacionais aplicadas para análise de imagens clínicas, pois irá envolver um volume extraordinário de dados.

A literatura é muito abrangente quanto aos métodos, e existem vários, que podem ser utilizados para verificar tais redundâncias, e desta forma reduzir a dimensionalidade de características a serem analisadas. Para este trabalho será realizada uma correlação entre pares como sugerido no estudo de (DORAN et al., 2021; SHUR et al., 2021), por se tratar de uma amostra pequena e este trabalho não tratar de análises com aprendizado de máquina, apenas a avaliação do comportamento das características radiômicas quando um mesmo simulador foi utilizado para adquirir imagens em equipamentos diferentes.

Como os dados a serem correlacionados não obedeceram a uma distribuição normal pelo Teste de Normalidade de Shapiro-Wilk (valor de $p < 0,05$) o teste utilizado para determinar a correlação entre as características foi o Teste de Correlação de Spearman, e onde o valor de $p < 0,05$ demonstrou que é estatisticamente significativa essa correlação entre as características, e o coeficiente de correlação de Spearman (ρ), que mensura se essa correlação é diretamente proporcional (+1), se é inversamente proporcional (-1) ou se é fraca ou moderada (quando é 0 ou próxima de 0).

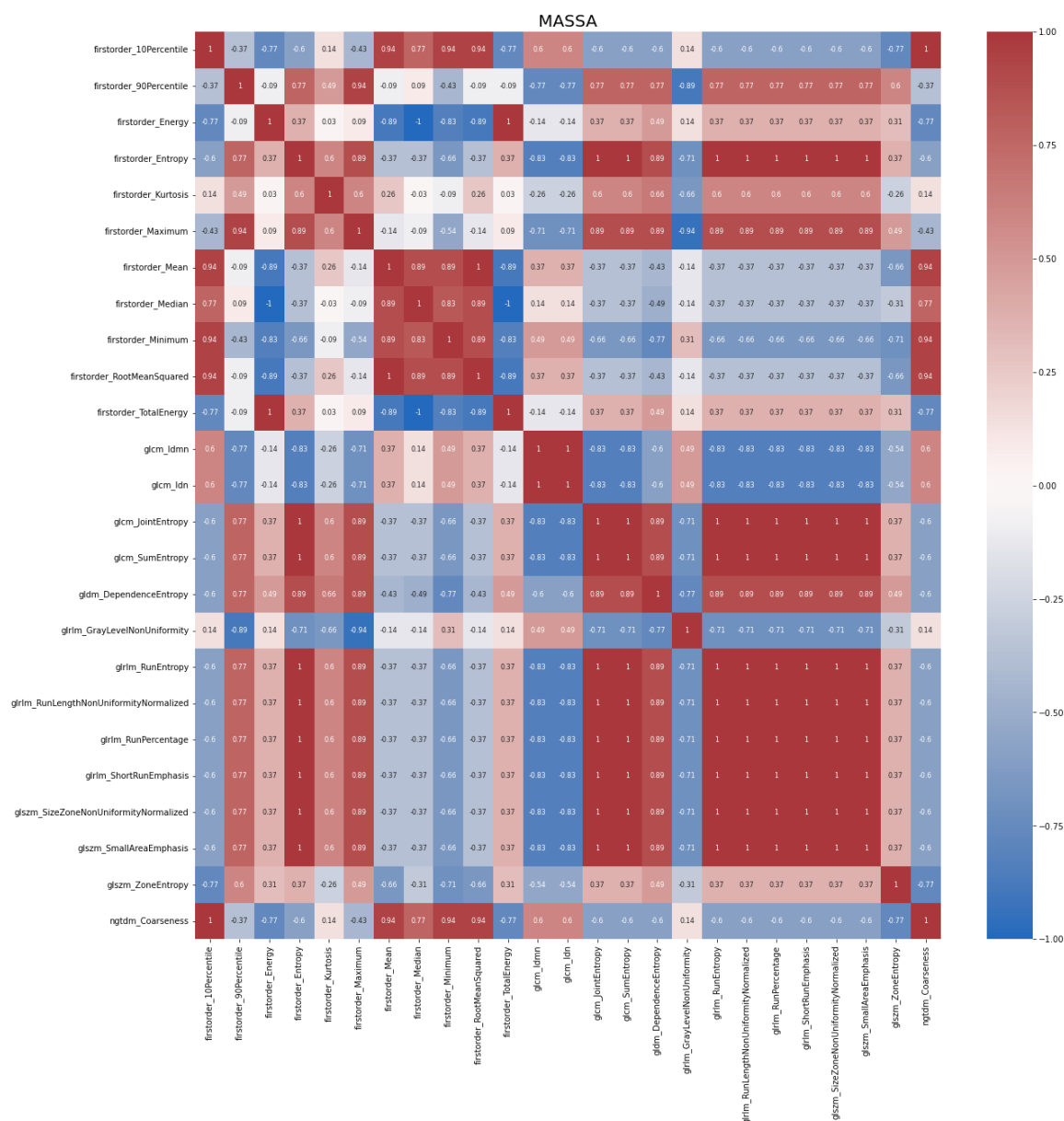
A correlação de Spearman foi realizada com o uso das bibliotecas e recursos para análises estatísticas disponíveis para a linguagem Python, dentro do ambiente do Jupyter Notebook. Os notebooks com as análises aplicadas neste trabalho encontram-se nos apêndices A, B e C. Heatmaps com as matrizes de correlação

Figura 66 – Matriz de correlação de características extraídas do *test object* Microcalcificação.



Fonte: Autor da dissertação.

Figura 67 – Matriz de correlação de características extraídas do test object Massa.



Fonte: Autor da dissertação.

Após a análise de correlação, no dataframe foi aplicado código com comandos para localizar as interações em qualquer ponto na matriz de correlação os pares de características que possuíam um coeficiente de correlação maior que 0,9. A partir da lista gerada com os pares de características, foi escrito no código que a segunda característica ou características do par seria retirada da análise, gerando uma nova lista com os nomes dessas características que foram retiradas.

A tabela 10 apresenta listados os nomes das características que foram retiradas para a análise dos *Test Objects* fibra, microcalcificação e massa, por

apresentarem uma correlação ($\rho=0,9$) com outras características, sendo estas portanto consideradas redundantes.

Tabela 10 – Características retiradas da análise por apresentarem correlação com outras.

Fibra	
firstorder_RootMeanSquared	glcm_Idn
firstorder_Median	glszm_SizeZoneNonUniformityNormalized
firstorder_Minimum	glszm_SmallAreaEmphasis"glrlm_RunPercentage
firstorder_TotalEnergy	glrlm_RunEntropy
firstorder_Maximum	glrlm_ShortRunEmphasis
firstorder_Mean	ngtdm_Coarseness
gldm_DependenceEntropy	
Microcalcificação	
firstorder_Minimum	gldm_SmallDependenceLowGrayLevelEmphasis
firstorder_Maximum	gldm_DependenceEntropy
firstorder_TotalEnergy	glszm_SmallAreaEmphasis
firstorder_RootMeanSquared	glrlm_RunEntropy
firstorder_Median	glrlm_RunPercentage
firstorder_Mean	glrlm_RunLengthNonUniformityNormalized
glcm_Idn	glrlm_ShortRunEmphasis
glcm_InverseVariance	ngtdm_Coarseness
Massa	
firstorder_Maximum	glszm_SmallAreaEmphasis
firstorder_RootMeanSquared	glszm_SizeZoneNonUniformityNormalized
firstorder_Minimum	glrlm_ShortRunEmphasis
firstorder_TotalEnergy	glrlm_RunEntropy
firstorder_Mean	glrlm_RunLengthNonUniformityNormalized
glcm_Idn	glrlm_RunPercentage
glcm_SumEntropy	ngtdm_Coarseness
glcm_JointEntropy	

Fonte: Autor da dissertação.

Para concluir, estão listadas na tabela 11 as características consideradas elegíveis a serem utilizadas nas análises por radiômica em imagens de mamografia digital utilizando o simulador ACR, de acordo com a estratégia para padronização sugerida neste trabalho. Os nomes das características que aparecem em negrito,

são para destacar as características que apresentaram similaridades na resposta para o simulador e no seu correspondente no tecido humano.

Tabela 11 – Características elegíveis para análises por radiômica em mamografia digital.

Fibra	
firstorder_Energy	glcm_Idmn
firstorder_10Percentile	glszm_ZoneEntropy
firstorder_90Percentile	glrlm_RunLengthNonUniformityNormalized
firstorder_Kurtosis	glrlm_GrayLevelNonUniformity
glcm_SumEntropy	
Microcalcificação	
firstorder_90Percentile	glcm_Idmn
firstorder_10Percentile	glcm_Imc2
firstorder_Energy	glszm_ZoneEntropy
glcm_MCC	glszm_SizeZoneNonUniformityNormalized
glcm_SumEntropy	
Massa	
firstorder_Entropy	firstorder_90Percentile
firstorder_10Percentile	gldm_DependenceEntropy
firstorder_Kurtosis	glcm_Idmn
firstorder_Energy	glszm_ZoneEntropy
firstorder_Median	glrlm_GrayLevelNonUniformity

Fonte: Autor da dissertação.

Esta redução da dimensionalidade de características promoverá modelos mais eficientes de aprendizado de máquina, por gerar um menor custo computacional, quando comparado com o volume inicial de dados analisados neste trabalho.

A estratégia sugerida que iniciou com um total de 108 características para cada *test object*, ao final selecionou 9 características para o *test object* Fibra e Microcalcificação e 10 para Massa, que poderão prosseguir como parte de análises mais robustas. Este passo finaliza a estratégia proposta neste trabalho para a padronização e desenvolvimento de análises por radiômica em imagens de mamografia digital.

Este tipo de análise por radiômica irá incluir um volume extraordinário de dados de imagens, históricos clínicos e resultados anatomopatológicos, para o desenvolvimento de soluções diagnósticas com aprendizado de máquina.

Vale destacar que as características podem apresentar comportamentos diferentes, podendo uma característica melhor representar um tipo de patologia que outra, mesmo para imagens de uma mesma modalidade diagnóstica, sendo o tipo de tumor, hormônio ou a vascularização que a patologia apresenta as principais causas para essas variações de comportamento das características, pois interferem nas representações das intensidades de tons de cinza da imagem. Por esse motivo essas análises necessitam de um grande volume de imagens (dados) e de todas as informações disponíveis sobre a patologia a qual se quer criar um banco de dados para análises por radiômica, como por exemplo: resultados do anatomopatológico, estadiamento, dentre outros.

7 CONCLUSÃO

A análise por radiômica em imagens de mamografia digital ainda não possui uma padronização, nem valores estabelecidos pelo IBSI. O fato de estudos com imagens mamográficas apresentarem seus próprios modelos de software para processamento, segmentação e cálculo das características, dificulta a reprodução e entendimento das metodologias utilizadas, assim como a não publicação dos valores calculados, tanto para os realizados com objetos simuladores, como em humanos.

A estratégia proposta neste trabalho apresentou resultados satisfatórios para a padronização de imagens mamográficas adquiridas em equipamentos de diferentes fabricantes, seguindo as diretrizes do IBSI.

A utilização de ferramentas de fácil acesso como o simulador ACR, o software livre de análise de imagens 3D Slicer e das bibliotecas disponíveis no Jupyter Notebook para a análise estatística, foi possível avaliar a estabilidade e reprodutibilidade das características radiômicas de acordo com a estratégia adotada para a padronização.

Algumas características radiômicas calculadas com o simulador ACR apresentaram resultados dentro do limiar de valores publicados pelo IBSI.

Os métodos utilizados para o cálculo e extração de características radiômicas do simulador e de imagens de humanos apresentaram valores dentro de um limiar de 15% de variação.

Para análises por radiômica mais robustas, que envolvem várias imagens médicas e diferentes informações diagnósticas, como anatomopatológico, evolução clínica e outras informações, exigirá a dedicação de uma equipe multiprofissional no seu desenvolvimento.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRÀMOFF, D. M. D. Image Processing with ImageJ. 2004.

ABRAMSON, R. G.; BURTON, K. R.; YU, J.-P. J.; SCALZETTI, E. M.; YANKEELOV, T. E.; ROSENKRANTZ, A. B.; MENDIRATTA-LALA, M.; BARTHOLMAI, B. J.; GANESHAN, D.; LENCHIK, L.; SUBRAMANIAM, R. M. Methods and Challenges in Quantitative Imaging Biomarker Development. *Academic radiology*, v. 22, n. 1, p. 25–32, jan. 2015. '

AGGARWAL, N.; AGRAWAL, R. K. First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images. v. 2012, , 30 maio. 2012. Disponível em: <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=19553>. Acesso em: 16 jun. 2024.

ALKHALIFAH, K. H.; BRINDHABAN, A.; ASBEUTAH, A. M. Comparison between image quality in electronic zoom and geometric magnification in digital mammography. *Journal of X-Ray Science and Technology*, v. 24, n. 5, p. 681–689, 6 out. 2016.

AMORIM, P. S.; BARRETO, I. L. D. S.; SOUZA, D. do N. Controle de qualidade em unidades móveis de mamografia. *Revista Brasileira de Física Médica*, v. 16, p. 699–699, 31 dez. 2022.

ANTROPOVA, N.; HUYNH, B. Q.; GIGER, M. L. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Medical Physics*, v. 44, n. 10, p. 5162–5171, out. 2017.

ÅSLUND, M. *Digital mammography with a photon counting detector in a scanned multislit geometry*. Stockholm: Fysik, Kungliga Tekniska högskolan, 2007. 41 p.

ASSIS, M. D.; SANTOS, R. O. M. Dos; MIGOWSKI, A. Breast cancer early detection in the Brazilian media during the Breast Awareness Month. *Physis: Revista de Saúde Coletiva*, v. 30, , 2020.

AVANZO, M.; STANCANELLO, J.; NAQA, I. E. Beyond imaging: The promise of radiomics. *Physica Medica: European Journal of Medical Physics*, v. 38, p. 122–139, 1 jun. 2017.

BORTOLETTO LINO DOS SANTOS, T. *Tamanho De Amostra Para O Teste-Reteste Na Determinação Do Coeficiente De Correlação Intraclasse*. 2018. Universidade Federal de Uberlândia Faculdade de Matemática, 2018. Disponível em: <https://repositorio.ufu.br/bitstream/123456789/22300/1/TamanhoAmostraTesteReteste.pdf>. Acesso em: 22 out. 2024.

BUSHONG, S. C. *Radiologic science for technologists: physics, biology, and protection*. 9th ed, St. Louis, Mo: Mosby/Elsevier, 2008. 685 p.

CALEFFI, M.; RIBEIRO, R. A.; FILHO, D. L. D.; ASHTON-PROLLA, P.; BEDIN, A. J.; SKONIESKI, G. P.; ZIGNANI, J. M.; GIACOMAZZI, J.; FRANCO, L. R.; GRAUDENZ, M.; POHLMANN, P.; FERNANDES, J. G.; KIVITZ, P.; WEBER, B. A model to optimize public health care and downstage breast cancer in limited-resource populations in southern Brazil. (Porto Alegre Breast Health Intervention Cohort). *BMC Public Health*, v. 9, n. 1, p. 83, dez. 2009.

CÂNCER (BRASIL), I. N. De. *Controle do câncer de mama no Brasil*. Rio de Janeiro, RJ: Instituto Nacional De Câncer, 2024.

CLARK, K.; VENDT, B.; SMITH, K.; FREYMAN, J.; KIRBY, J.; KOPPEL, P.; MOORE, S.; PHILLIPS, S.; MAFFITT, D.; PRINGLE, M.; TARBOX, L.; PRIOR, F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*, v. 26, n. 6, p. 1045–1057, dez. 2013.

COWEN, A. R.; DAVIES, A. G.; KENGYELICS, S. M. Advances in computed radiography systems and their physical imaging characteristics. *Clinical Radiology*, v. 62, n. 12, p. 1132–1141, dez. 2007.

CUI, C.; LI, L.; CAI, H.; FAN, Z.; ZHANG, L.; DAN, T.; LI, J.; WANG, J. , 2021. . [S.l.]: The Cancer Imaging Archive. Disponível em: <https://wiki.cancerimagingarchive.net/x/7KEvB>. Acesso em: 10 out. 2023.

DA SILVA, S. D.; JOANA, G. S.; OLIVEIRA, B. B.; DE OLIVEIRA, M. A.; LEYTON, F.; NOGUEIRA, M. D. S. Dosimetry and image quality in digital mammography facilities in the State of Minas Gerais, Brazil. *Radiation Physics and Chemistry*, v. 116, p. 292–299, nov. 2015.

DEPEURSINGE, A.; ANDREARCZYK, V.; WHYBRA, P.; VAN GRIETHUYSEN, J.; MÜLLER, H.; SCHAER, R.; VALLIÈRES, M.; ZWANENBURG, A. Standardised convolutional filtering for Radiomics. jun. 2020.

DEPEURSINGE, A.; ANDREARCZYK, V.; WHYBRA, P.; VAN GRIETHUYSEN, J.; MÜLLER, H.; SCHAER, R.; VALLIÈRES, M.; ZWANENBURG, A. , 7 fev. 2024. . [S.l.]: arXiv. Disponível em: <http://arxiv.org/abs/2006.05470>. Acesso em: 21 abr. 2024.

DORAN, S. J.; KUMAR, S.; ORTON, M.; D'ARCY, J.; KWAKS, F.; O'FLYNN, E.; AHMED, Z.; DOWNEY, K.; DOWSETT, M.; TURNER, N.; MESSIOU, C.; KOH, D.-M. “Real-world”

radiomics from multi-vendor MRI: an original retrospective study on the prediction of nodal status and disease survival in breast cancer, as an exemplar to promote discussion of the wider issues. **Cancer Imaging**, v. 21, n. 1, p. 37, 20 maio. 2021.

EXECUTABLE BOOKS COMMUNITY. , 12 fev. 2020. . [S.l.]: [object Object]. Disponível em: <https://zenodo.org/record/4539666>. Acesso em: 06 maio 2024.

FEDOROV, A.; BEICHEL, R.; KALPATHY-CRAMER, J.; FINET, J.; FILLION-ROBIN, J.-C.; PUJOL, S.; BAUER, C.; JENNINGS, D.; FENNESSY, F.; SONKA, M.; BUATTI, J.; AYLWARD, S.; MILLER, J. V.; PIEPER, S.; KIKINIS, R. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. **Magnetic resonance imaging**, v. 30, n. 9, p. 1323–1341, nov. 2012.

FLUKE BIOMEDICAL. **Mammographic Accreditation Phantom**. Disponível em: www.flukebiomedical.com. Acesso em: 22 jul. 2023.

FREITAS, A. G. D.; KEMP, C.; LOUVEIRA, M. H.; FUJIWARA, S. M.; CAMPOS, L. F. Mamografia digital: perspectiva atual e aplicações futuras. **Radiologia Brasileira**, v. 39, n. 4, p. 287–296, ago. 2006.

GER, R. B.; ZHOU, S.; CHI, P.-C. M.; LEE, H. J.; LAYMAN, R. R.; JONES, A. K.; GOFF, D. L.; FULLER, C. D.; HOWELL, R. M.; LI, H.; STAFFORD, R. J.; COURT, L. E.; MACKIN, D. S. Comprehensive Investigation on Controlling for CT Imaging Variabilities in Radiomics Studies. **Scientific Reports**, v. 8, n. 1, p. 13047, 29 ago. 2018.

GIGER, M. L. Machine Learning in Medical Imaging. **Journal of the American College of Radiology**, v. 15, n. 3, p. 512–520, mar. 2018.

GILLIES, R. J.; KINAHAN, P. E.; HRICAK, H. Radiomics: Images Are More than Pictures, They Are Data. **Radiology**, v. 278, n. 2, p. 563–577, fev. 2016.

GONZALEZ, R. C.; WOODS, R. E. **Processamento Digital de Imagens**. 3º Edição, São Paulo: Pearson, 2010.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural Features for Image Classification. **Institute of Electrical and Electronics Engineers Transactions on systems, man, and cybernetics**, v. SMC-3, n. 6, p. 610–621, 1973.

HUYNH, B. Q.; LI, H.; GIGER, M. L. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. **Journal of Medical Imaging**, v. 3, n. 3, p. 034501, 22 ago. 2016.

IBSI. **IBSI – Image Biomarker Standardisation Initiative**. Disponível em: <https://theibsi.github.io/>. Acesso em: 03 dez. 2023.

INSTITUTO NACIONAL DE CÂNCER - INCA. **Conceito e Magnitude**. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/gestor-e-profissional-de-saude/controlado-cancer-de-mama/conceito-e-magnitude>. Acesso em: 30 jul. 2023.

JUNIOR, J. R. F.; OLIVEIRA, M. C. Unidade de processamento gráfico no diagnóstico auxiliado por computador. **Journal of Health Informatics**, v. 9, n. 3, , 16 ago. 2017. Disponível em: <https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/469>. Acesso em: 01 set. 2023.

KOO, T. K.; LI, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. **Journal of Chiropractic Medicine**, v. 15, n. 2, p. 155–163, 1 jun. 2016.

KUMAR, V.; GU, Y.; BASU, S.; BERGLUND, A.; ESCHRICH, S. A.; SCHABATH, M. B.; FORSTER, K.; AERTS, H. J. W. L.; DEKKER, A.; FENSTERMACHER, D.; GOLDFOF, D. B.; HALL, L. O.; LAMBIN, P.; BALAGURUNATHAN, Y.; GATENBY, R. A.; GILLIES, R. J. Radiomics: the process and the challenges. **Magnetic Resonance Imaging**, v. 30, n. 9, p. 1234–1248, nov. 2012.

LAMBIN, P. , 2021. . [S.l.]: DOI 10.17195/candat.2016.08.1

LAMBIN, Philippe; LEIJENAAR, R. T. H.; DEIST, T. M.; PEERLINGS, J.; DE JONG, E. E. C.; VAN TIMMEREN, J.; SANDULEANU, S.; LARUE, R. T. H. M.; EVEN, A. J. G.; JOCHEMS, A.; VAN WIJK, Y.; WOODRUFF, H.; VAN SOEST, J.; LUSTBERG, T.; ROELOFS, E.; VAN ELMPT, W.; DEKKER, A.; MOTTAGHY, F. M.; WILDBERGER, J. E.; WALSH, S. Radiomics: the bridge between medical imaging and personalized medicine. **Nature Reviews Clinical Oncology**, v. 14, n. 12, p. 749–762, dez. 2017.

LAMBIN, Philippe; RIOS-VELAZQUEZ, E.; LEIJENAAR, R.; CARVALHO, S.; VAN STIPHOUT, R. G. P. M.; GRANTON, P.; ZEGERS, C. M. L.; GILLIES, R.; BOELLARD, R.; DEKKER, A.; AERTS, H. J. W. L. Radiomics: Extracting more information from medical images using advanced feature analysis. **European Journal of Cancer**, v. 48, n. 4, p. 441–446, mar. 2012.

LANÇA, L.; SILVA, A. Digital radiography detectors – A technical overview: Part 1. **Radiography**, v. 15, n. 1, p. 58–62, fev. 2009. a.

LANÇA, L.; SILVA, A. Digital radiography detectors – A technical overview: Part 2. **Radiography**, v. 15, n. 2, p. 134–138, maio. 2009. b.

- LIMKIN, E. J.; SUN, R.; DERCLE, L.; ZACHARAKI, E. I.; ROBERT, C.; REUZÉ, S.; SCHERNBERG, A.; PARAGIOS, N.; DEUTSCH, E.; FERTÉ, C. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. ***Annals of Oncology***, v. 28, n. 6, p. 1191–1206, 1 jun. 2017.
- MACKIN, D.; FAVE, X.; ZHANG, L.; FRIED, D.; YANG, J.; TAYLOR, B.; RODRIGUEZ-RIVERA, E.; DODGE, C.; JONES, A. K.; COURT, L. Measuring Computed Tomography Scanner Variability of Radiomics Features: ***Investigative Radiology***, v. 50, n. 11, p. 757–765, nov. 2015.
- MESEGUER, E.; BARBERÁ-TOMÁS, D.; BENITO-AMAT, C.; DÍAZ-FAES, A. A.; MARTÍ-BONMATÍ, L. What do biomarkers add: Mapping quantitative imaging biomarkers research. ***European Journal of Radiology***, v. 146, p. 110052, jan. 2022.
- MINISTÉRIO DA SAÚDE; INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. ***Atualização em Mamografia para Técnicos em Radiologia***. [S.l.]: Coordenação de Ensino, 2019.
- NOROUZI, A.; RAHIM, M. S. M.; ALTAMEEM, A.; SABA, T.; RAD, A. E.; REHMAN, A.; UDDIN, M. Medical Image Segmentation Methods, Algorithms, and Applications. ***IETE Technical Review***, v. 31, n. 3, p. 199–213, 4 maio. 2014.
- NUNES, F. L. S. ***Livro das Jornadas de Atualizações em Informática***. [S.l.: s.n.], 2006. 73–126 p.
- PAREKH, V.; JACOBS, M. A. Radiomics: a new application from established techniques. ***Expert Review of Precision Medicine and Drug Development***, v. 1, n. 2, p. 207–226, 3 mar. 2016.
- PRINCE, J. L.; LINKS, J. M. ***Medical imaging signals and systems***. 2. ed., Boston: Pearson, 2015. 519 p.
- ROGERS, W.; THULASI SEETHA, S.; REFAEE, T. A. G.; LIEVERSE, R. I. Y.; GRANZIER, R. W. Y.; IBRAHIM, A.; KEEK, S. A.; SANDULEANU, S.; PRIMAKOV, S. P.; BEUQUE, M. P. L.; MARCUS, D.; VAN DER WIEL, A. M. A.; ZERKA, F.; OBERIJE, C. J. G.; VAN TIMMEREN, J. E.; WOODRUFF, H. C.; LAMBIN, P. Radiomics: from qualitative to quantitative imaging. ***The British Journal of Radiology***, v. 93, n. 1108, p. 20190948, abr. 2020.
- SCAPICCHIO, C.; GABELLONI, M.; BARUCCI, A.; CIONI, D.; SABA, L.; NERI, E. A deep look into radiomics. ***La Radiologia Medica***, v. 126, n. 10, p. 1296–1311, 2021.

SHUR, J. D.; DORAN, S. J.; KUMAR, S.; AP DAFYDD, D.; DOWNEY, K.; O'CONNOR, J. P. B.; PAPANIKOLAOU, N.; MESSIOU, C.; KOH, D.-M.; ORTON, M. R. Radiomics in Oncology: A Practical Guide. *RadioGraphics*, v. 41, n. 6, p. 1717–1732, out. 2021.

STRICKLAND, N. H. Current topic: PACS (picture archiving and communication systems): filmless radiology. *Archives of Disease in Childhood*, v. 83, n. 1, p. 82–86, 1 jul. 2000.

TAMAL, M. Grey Level Co-occurrence Matrix (GLCM) as a Radiomics Feature for Artificial Intelligence (AI) Assisted Positron Emission Tomography (PET) Images Analysis. *IOP Conference Series: Materials Science and Engineering*, v. 646, n. 1, p. 012047, 1 out. 2019.

TOMAL, A.; PEREZ, A. M. M. M.; SILVA, M. C.; POLETTI, M. E. Experimental evaluation of the image quality and dose in digital mammography: Influence of x-ray spectrum. *Radiation Physics and Chemistry*, v. 116, p. 282–286, nov. 2015.

VALDORA, F.; HOUSSAMI, N.; ROSSI, F.; CALABRESE, M.; TAGLIAFICO, A. S. Rapid review: radiomics and breast cancer. *Breast Cancer Research and Treatment*, v. 169, n. 2, p. 217–229, jun. 2018.

VAN GRIETHUYSEN, J. J. M.; FEDOROV, A.; PARMAR, C.; HOSNY, A.; AUCCOIN, N.; NARAYAN, V.; BEETS-TAN, R. G. H.; FILLION-ROBIN, J.-C.; PIEPER, S.; AERTS, H. J. W. L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, v. 77, n. 21, p. e104–e107, 1 nov. 2017.

VAN TIMMEREN, J. E.; CESTER, D.; TANADINI-LANG, S.; ALKADHI, H.; BAESSLER, B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into Imaging*, v. 11, n. 1, p. 91, dez. 2020.

VEDANTHAM, S.; KARELLAS, A.; VIJAYARAGHAVAN, G. R.; KOPANS, D. B. Digital Breast Tomosynthesis: State of the Art. *Radiology*, v. 277, n. 3, p. 663–684, dez. 2015.

WASKOM, M. seaborn: statistical data visualization. *Journal of Open Source Software*, v. 6, n. 60, p. 3021, 6 abr. 2021.

WEIGEL, S.; BERKEMEYER, S.; GIRNUS, R.; SOMMER, A.; LENZEN, H.; HEINDEL, W. Digital Mammography Screening with Photon-counting Technique: Can a High Diagnostic Performance Be Realized at Low Mean Glandular Dose? *Radiology*, v. 271, n. 2, p. 345–355, maio. 2014.

WRIGLEY, R. H. Computed Radiology. *Clinical Techniques in Equine Practice*, v. 3, n. 4, p. 341–351, dez. 2004.

YAFFE, M. J.; BLOOMQUIST, A. K.; HUNTER, D. M.; MAWDSLEY, G. E.; CHIARELLI, A. M.; MURADALI, D.; MAINPRIZE, J. G. Comparative performance of modern digital mammography systems in a large breast screening program: Comparative performance of CR and DR mammography systems. *Medical Physics*, v. 40, n. 12, p. 121915, 27 nov. 2013.

YANASE, J.; TRIANTAPHYLLOU, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, v. 138, p. 112821, 30 dez. 2019.

ZWANENBURG, A.; LEGER, S.; VALLIÈRES, M.; LÖCK, S. Image biomarker standardisation initiative. *Radiology*, arXiv:1612.07003 [cs, eess], v. 295, n. 2, p. 328–338, maio. 2020. a.

ZWANENBURG, A.; LEGER, S.; VALLIÈRES, M.; LÖCK, S. Image biomarker standardisation initiative. *Radiology*, arXiv:1612.07003 [cs, eess], v. 295, n. 2, p. 328–338, maio. 2020. b.

ZWANENBURG, A.; VALLIÈRES, M.; ABDALAH, M. A.; AERTS, H. J. W. L.; ANDREARCZYK, V.; APTE, A.; ASHRAFINIA, S.; BAKAS, S.; BEUKINGA, R. J.; BOELLAARD, R.; BOGOWICZ, M.; BOLDRINI, L.; BUVAT, I.; COOK, G. J. R.; DAVATZIKOS, C.; DEPEURSINGE, A.; DESSEROIT, M.-C.; DINAPOLI, N.; DINH, C. V.; ECHEGARAY, S.; EL NAQA, I.; FEDOROV, A. Y.; GATTA, R.; GILLIES, R. J.; GOH, V.; GÖTZ, M.; GUCKENBERGER, M.; HA, S. M.; HATT, M.; ISENSEE, F.; LAMBIN, P.; LEGER, S.; LEIJENAAR, R. T. H.; LENKOWICZ, J.; LIPPERT, F.; LOSNEGÅRD, A.; MAIER-HEIN, K. H.; MORIN, O.; MÜLLER, H.; NAPEL, S.; NIOCHE, C.; ORLHAC, F.; PATI, S.; PFAEHLER, E. A. G.; RAHMIM, A.; RAO, A. U. K.; SCHERER, J.; SIDDIQUE, M. M.; SIJTSEMA, N. M.; SOCARRAS FERNANDEZ, J.; SPEZI, E.; STEENBAKKERS, R. J. H. M.; TANADINI-LANG, S.; THORWARTH, D.; TROOST, E. G. C.; UPADHAYA, T.; VALENTINI, V.; VAN DIJK, L. V.; VAN GRIETHUYSEN, J.; VAN VELDEN, F. H. P.; WHYBRA, P.; RICHTER, C.; LÖCK, S. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*, v. 295, n. 2, p. 328–338, maio. 2020.

APÊNDICES

Apêndice A – Documentação do Jupyter Notebook utilizado neste trabalho para a análise do *test object* FIBRA.

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline

from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report

import warnings

fibra = pd.read_excel(r'C:\Users\erica\OneDrive\MPTRCS\planilhas_quarta_analise_BOAS')

fibra.head()

```

	firstorder_10Percentile	firstorder_90Percentile	firstorder_Energy	firstorder_Kurtosis	firstorder_Max
0	2286.00	2346.00	2.523424e+10	3.020177	2.
1	2302.00	2361.40	2.262975e+10	3.021159	2.
2	2233.60	2368.40	2.987070e+10	2.649508	2.
3	2242.35	2379.25	3.004947e+10	2.775414	2.
4	2247.00	2374.40	2.376400e+10	2.733361	2.

5 rows × 23 columns

```

corr = fibra.corr('spearman')
corr

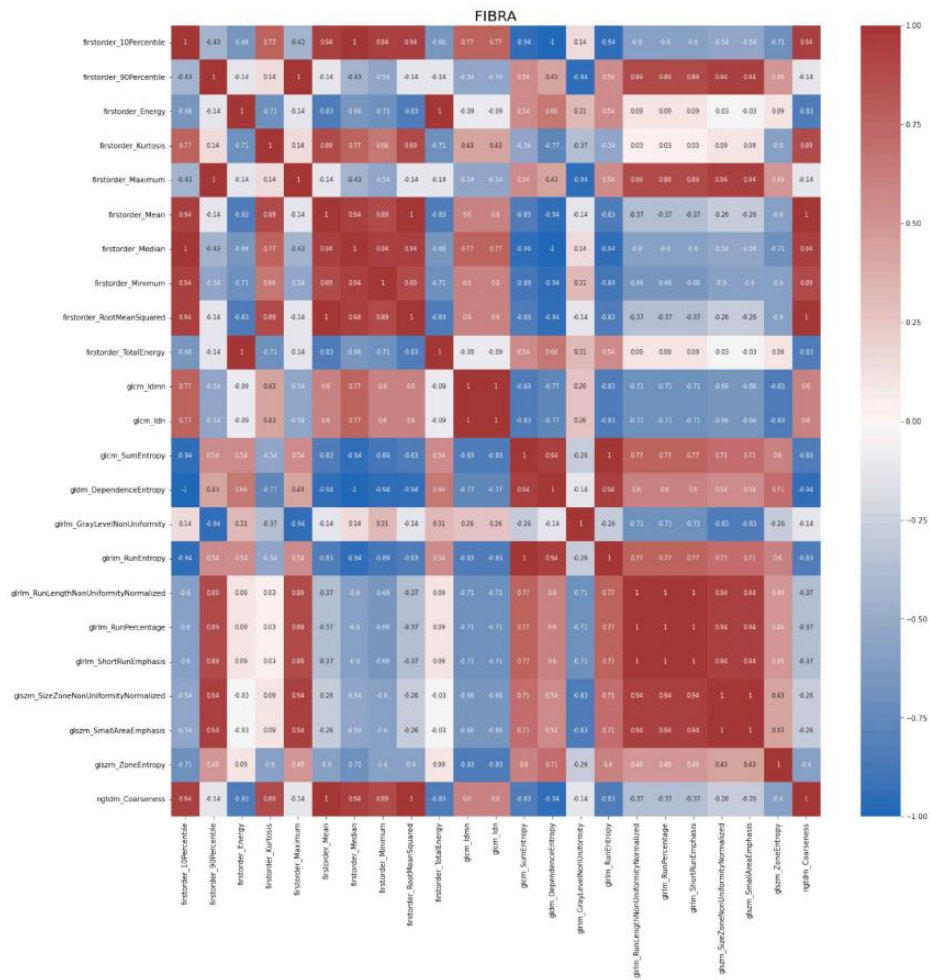
```

	firstorder_10Percentile	firstorder_90Percentile	firstorder_E
firstorder_10Percentile	1.000000	-0.428571	-0.61
firstorder_90Percentile	-0.428571	1.000000	-0.14
firstorder_Energy	-0.657143	-0.142857	1.00
firstorder_Kurtosis	0.771429	0.142857	-0.77
firstorder_Maximum	-0.428571	1.000000	-0.14
firstorder_Mean	0.942857	-0.142857	-0.81
firstorder_Median	1.000000	-0.428571	-0.61
firstorder_Minimum	0.942857	-0.542857	-0.77
firstorder_RootMeanSquared	0.942857	-0.142857	-0.81
firstorder_TotalEnergy	-0.657143	-0.142857	1.00
gldm_Idmn	0.771429	-0.542857	-0.01
gldm_Idn	0.771429	-0.542857	-0.01
gldm_SumEntropy	-0.942857	0.542857	0.54
gldm_DependenceEntropy	-1.000000	0.428571	0.61

```

corr = fibra.corr('spearman')
plt.figure(figsize=(20,20))
plot = sns.heatmap(fibra.corr('spearman').round(2), annot = True, annot_kws={'fontsize': 8},
cmap='vlag', center=0, vmin=-1, vmax=1, xticklabels = 1, yticklabels = 1)
plot.set_title('FIBRA', fontsize = 20)

```



```

corr_pairs=[]
for i in range(len(corr.columns)):
    for j in range(i):
        if corr.iloc[i,j]>0.9:
            corr_pairs.append((corr.columns[i],corr.columns[j],corr.iloc[i,j]))

corr_pairs

```

```

[('firstorder_Maximum', 'firstorder_90Percentile', 1.0),
('firstorder_Mean', 'firstorder_10Percentile', 0.9428571428571428),
('firstorder_Median', 'firstorder_10Percentile', 1.0),
('firstorder_Median', 'firstorder_Mean', 0.9428571428571428),
('firstorder_Minimum', 'firstorder_10Percentile', 0.9428571428571428),
('firstorder_Minimum', 'firstorder_Median', 0.9428571428571428),
('firstorder_RootMeanSquared', 'firstorder_10Percentile', 0.9428571428571428),
('firstorder_RootMeanSquared', 'firstorder_Mean', 1.0),
('firstorder_RootMeanSquared', 'firstorder_Median', 0.9428571428571428),
('firstorder_TotalEnergy', 'firstorder_Energy', 1.0),
('glcm_Idn', 'glcm_Idmn', 1.0),
('gldm_DependenceEntropy', 'glcm_SumEntropy', 0.9428571428571428),
('gldm_RunEntropy', 'glcm_SumEntropy', 1.0),
('gldm_RunEntropy', 'gldm_DependenceEntropy', 0.9428571428571428),

```

```
( 'glrlm_RunPercentage', 'glrlm_RunLengthNonUniformityNormalized', 1.0),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunLengthNonUniformityNormalized', 1.0),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunPercentage', 1.0),
( 'glszm_SizeZoneNonUniformityNormalized',
  'firstorder_90Percentile',
  0.9428571428571428),
( 'glszm_SizeZoneNonUniformityNormalized',
  'firstorder_Maximum',
  0.9428571428571428),
( 'glszm_SizeZoneNonUniformityNormalized',
  'glrlm_RunLengthNonUniformityNormalized',
  0.9428571428571428),
( 'glszm_SizeZoneNonUniformityNormalized',
  'glrlm_RunPercentage',
  0.9428571428571428),
( 'glszm_SizeZoneNonUniformityNormalized',
  'glrlm_ShortRunEmphasis',
  0.9428571428571428),
( 'glszm_SmallAreaEmphasis', 'firstorder_90Percentile', 0.9428571428571428),
( 'glszm_SmallAreaEmphasis', 'firstorder_Maximum', 0.9428571428571428),
( 'glszm_SmallAreaEmphasis',
  'glrlm_RunLengthNonUniformityNormalized',
  0.9428571428571428),
( 'glszm_SmallAreaEmphasis', 'glrlm_RunPercentage', 0.9428571428571428),
( 'glszm_SmallAreaEmphasis', 'glrlm_ShortRunEmphasis', 0.9428571428571428),
( 'glszm_SmallAreaEmphasis', 'glszm_SizeZoneNonUniformityNormalized', 1.0),
( 'ngtdm_Coarseness', 'firstorder_10Percentile', 0.9428571428571428),
( 'ngtdm_Coarseness', 'firstorder_Mean', 1.0),
( 'ngtdm_Coarseness', 'firstorder_Median', 0.9428571428571428),
( 'ngtdm_Coarseness', 'firstorder_RootMeanSquared', 1.0) ]
```

```
caracteristicas_a_tirar = []
for i, j, k in corr_pairs:
    caracteristicas_a_tirar.append(i)

caracteristicas_a_tirar = list(set(caracteristicas_a_tirar))
caracteristicas_a_tirar
```

```
[ 'firstorder_Median',
  'ngtdm_Coarseness',
  'firstorder_Minimum',
  'glrlm_RunEntropy',
  'glszm_SmallAreaEmphasis',
  'gldm_DependenceEntropy',
  'glcm_Idn',
  'firstorder_Maximum',
  'glrlm_ShortRunEmphasis',
  'glszm_SizeZoneNonUniformityNormalized',
  'firstorder_Mean',
  'firstorder_RootMeanSquared',
  'glrlm_RunPercentage',
  'firstorder_TotalEnergy' ]
```

```
caracteristicas_a_tirar = list(set(caracteristicas_a_tirar))
print (caracteristicas_a_tirar)
```

```
[ 'firstorder_Median', 'ngtdm_Coarseness', 'glszm_SmallAreaEmphasis', 'firstorder_Minimum',
  'glrlm_RunEntropy', 'glcm_Idn', 'firstorder_Maximum', 'glrlm_ShortRunEmphasis',
  'glszm_SizeZoneNonUniformityNormalized', 'glrlm_RunPercentage', 'firstorder_Mean',
  'gldm_DependenceEntropy', 'firstorder_RootMeanSquared', 'firstorder_TotalEnergy' ]
```

```
caracteristicas_seleccionadas = fibra.drop(caracteristicas_a_tirar, axis=1)
caracteristicas_seleccionadas = list(set(caracteristicas_seleccionadas))
caracteristicas_seleccionadas
```

```
[ 'firstorder_Energy',
  'firstorder_Kurtosis',
  'glcm_SumEntropy',
  'firstorder_10Percentile',
  'firstorder_90Percentile',
  'glszm_ZoneEntropy',
  'glrlm_RunLengthNonUniformityNormalized',
  'glrlm_GrayLevelNonUniformity',
  'glcm_Idmn' ]
```

```
[1]: import pingouin as pg
import pandas as pd
```

```
[2]: import numpy as np
```

```
[3]: fibra = pd.read_excel(r'C:\Users\erica\OneDrive\MPTRCS\planilhas_quarta_analise_BOAS\fibra_ICC_2.xlsx')
fibra.head()
```

```
[3]:
```

	caracteristica	equipamento	valor
0	firstorder_10Percentile	1º ensaio_DBT_fibra	2286.00
1	firstorder_10Percentile	2º ensaio_DBT_fibra	2302.00
2	firstorder_10Percentile	1º ensaio_PC1_fibra	2233.60
3	firstorder_10Percentile	2º ensaio_PC1_fibra	2242.35
4	firstorder_10Percentile	1º ensaio_PC2_fibra	2247.00

```
[4]: print(pd.pivot_table(fibra, index='equipamento', columns='caracteristica').T)
```

```
equipamento          1º ensaio_DBT_fibra \
caracteristica
valor firstorder_10Percentile      2.286000e+03
      firstorder_90Percentile      2.346000e+03
      firstorder_Energy            2.523424e+10
      firstorder_Kurtosis          3.020177e+00
      firstorder_Maximum          2.409000e+03
      firstorder_Mean              2.316049e+03
      firstorder_Median            2.315000e+03
      firstorder_Minimum          2.235600e+03
      firstorder_RootMeanSquared  2.316174e+03
      firstorder_TotalEnergy       2.523424e+08
      glcm_Idmn                    9.791312e-01
      glcm_Idn                      9.068674e-01
      glcm_SumEntropy              2.733769e+00
      glcm_DependenceEntropy       4.604411e+00
      glrlm_GrayLevelNonUniformity 8.042178e+02
      glrlm_RunEntropy             3.489846e+00
```

```
[10]: results_fibra = pg.intraclass_corr(data=fibra, targets='caracteristica', raters='equipamento', ratings='valor')

pd.set_option('display.max_columns', 8)
pd.set_option('display.width', 200)

print(results_fibra)
df = pd.DataFrame.from_dict(results_fibra)
df.to_excel("results_fibra.xlsx", index=False)
```

	Type	Description	ICC	F	df1	df2	pval	CI95%
0	ICC1	Single raters absolute	0.981916	326.779175	22	115	4.969179e-93	[0.97, 0.99]
1	ICC2	Single random raters	0.981916	327.076487	22	110	9.192604e-90	[0.97, 0.99]
2	ICC3	Single fixed raters	0.981932	327.076487	22	110	9.192604e-90	[0.97, 0.99]
3	ICC1k	Average raters absolute	0.996940	326.779175	22	115	4.969179e-93	[0.99, 1.0]
4	ICC2k	Average random raters	0.996940	327.076487	22	110	9.192604e-90	[0.99, 1.0]
5	ICC3k	Average fixed raters	0.996943	327.076487	22	110	9.192604e-90	[0.99, 1.0]

```
[8]: results_fibra = results_fibra.set_index('Description')
icc = results_fibra.loc['Single random raters', 'ICC']

print(icc.round(3))
```

```
0.982
```

```
[9]: lower_ci = results_fibra.loc['Single random raters', 'CI95%'][0]
upper_ci = results_fibra.loc['Single random raters', 'CI95%'][1]
```

```
print(f'ICC = {icc:.3f}, 95% CI [{lower_ci}, {upper_ci}]')
```

```
ICC = 0.982, 95% CI [0.97, 0.99]
```

Apêndice B – Documentação do Jupyter Notebook utilizado neste trabalho para a análise do *test object* MICROCALCIFICAÇÃO.

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline

from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report

import warnings

micro = pd.read_excel(r'C:\Users\erica\OneDrive\MPTRCS\planilhas_quarta_analise_BOAS
micro.head()

```

	firstorder_10Percentile	firstorder_90Percentile	firstorder_Energy	firstorder_Maximum	firstorder_M
0	2282.00	2335.00	1.426813e+11	2754.00	2309.734
1	2295.40	2348.60	1.291862e+11	2791.80	2322.932
2	2229.40	2358.20	1.717862e+11	3054.20	2297.135
3	2236.25	2366.00	1.718655e+11	3085.25	2301.984
4	2242.40	2361.68	1.360773e+11	3069.60	2304.127

5 rows x 25 columns

```

corr = micro.corr('spearman')
corr

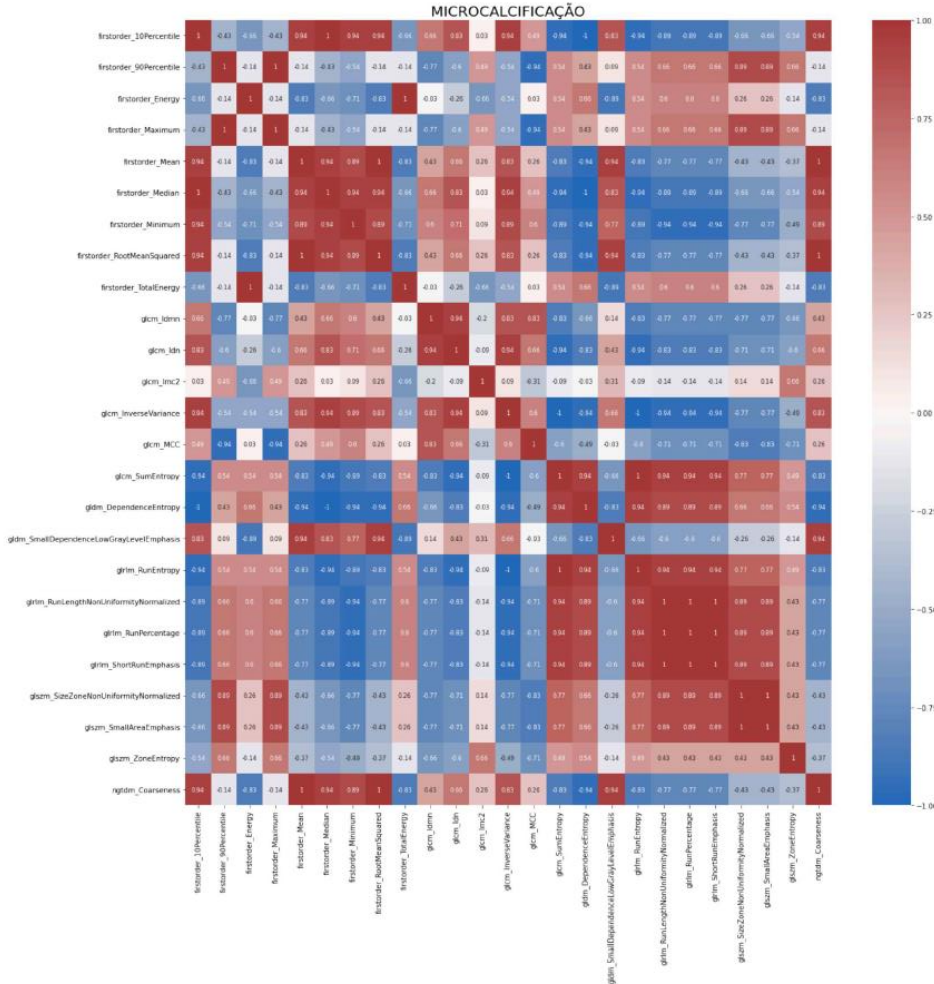
```

	firstorder_10Percentile	firstorder_90Percentile	firstorde
firstorder 10Percentile	1.000000	-0.428571	-
firstorder 90Percentile	-0.428571	1.000000	-
firstorder_Energy	-0.657143	-0.142857	-
firstorder Maximum	-0.428571	1.000000	-
firstorder Mean	0.942857	-0.142857	-
firstorder Median	1.000000	-0.428571	-
firstorder Minimum	0.942857	-0.542857	-
firstorder RootMeanSquared	0.942857	-0.142857	-
firstorder TotalEnergy	-0.657143	-0.142857	-
glcm_Idmn	0.657143	-0.771429	-
glcm_Idn	0.828571	-0.600000	-
glcm_Imc2	0.028571	0.485714	-
glcm_InverseVariance	0.942857	-0.542857	-
glcm_MCC	0.485714	-0.942857	-

```

corr = micro.corr('spearman')
plt.figure(figsize=(20,20))
plot = sns.heatmap(micro.corr('spearman').round(2), annot = True, annot_kws={'fontsize': 8},
cmap='vlag', center=0, vmin=-1, vmax=1, xticklabels = 1, yticklabels = 1)
plot.set_title('MICROCALCIFICAÇÃO', fontsize = 20)

```



```

corr_pairs=[]
for i in range(len(corr.columns)):
    for j in range(i):
        if corr.iloc[i,j]>0.9:
            corr_pairs.append((corr.columns[i],corr.columns[j],corr.iloc[i,j]))
corr_pairs

```

```

[('firstorder_Maximum', 'firstorder_90Percentile', 1.0),
 ('firstorder_Mean', 'firstorder_10Percentile', 0.9428571428571428),
 ('firstorder_Median', 'firstorder_10Percentile', 1.0),
 ('firstorder_Median', 'firstorder_Mean', 0.9428571428571428),
 ('firstorder_Minimum', 'firstorder_10Percentile', 0.9428571428571428),
 ('firstorder_Minimum', 'firstorder_Median', 0.9428571428571428),
 ('firstorder_RootMeanSquared', 'firstorder_10Percentile', 0.9428571428571428),
 ('firstorder_RootMeanSquared', 'firstorder_Mean', 1.0),
 ('firstorder_RootMeanSquared', 'firstorder_Median', 0.9428571428571428),
 ('firstorder_TotalEnergy', 'firstorder_Energy', 1.0),
 ('glcm_Idn', 'glcm_Idmn', 0.9428571428571428),
 ('glcm_InverseVariance', 'firstorder_10Percentile', 0.9428571428571428),
 ('glcm_InverseVariance', 'firstorder_Median', 0.9428571428571428),
 ('glcm_InverseVariance', 'glcm_Idn', 0.9428571428571428),

```

```
( 'gldm_DependenceEntropy', 'glcm_SumEntropy', 0.9428571428571428),
( 'gldm_SmallDependenceLowGrayLevelEmphasis',
  'firstorder_Mean',
  0.9428571428571428),
( 'gldm_SmallDependenceLowGrayLevelEmphasis',
  'firstorder_RootMeanSquared',
  0.9428571428571428),
( 'glrlm_RunEntropy', 'glcm_SumEntropy', 1.0),
( 'glrlm_RunEntropy', 'gldm_DependenceEntropy', 0.9428571428571428),
( 'glrlm_RunLengthNonUniformityNormalized',
  'glcm_SumEntropy',
  0.9428571428571428),
( 'glrlm_RunLengthNonUniformityNormalized',
  'glrlm_RunEntropy',
  0.9428571428571428),
( 'glrlm_RunPercentage', 'glcm_SumEntropy', 0.9428571428571428),
( 'glrlm_RunPercentage', 'glrlm_RunEntropy', 0.9428571428571428),
( 'glrlm_RunPercentage', 'glrlm_RunLengthNonUniformityNormalized', 1.0),
( 'glrlm_ShortRunEmphasis', 'glcm_SumEntropy', 0.9428571428571428),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunEntropy', 0.9428571428571428),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunLengthNonUniformityNormalized', 1.0),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunPercentage', 1.0),
( 'glszm_SmallAreaEmphasis', 'glszm_SizeZoneNonUniformityNormalized', 1.0),
( 'ngtdm_Coarseness', 'firstorder_10Percentile', 0.9428571428571428),
( 'ngtdm_Coarseness', 'firstorder_Mean', 1.0),
( 'ngtdm_Coarseness', 'firstorder_Median', 0.9428571428571428),
( 'ngtdm_Coarseness', 'firstorder_RootMeanSquared', 1.0),
( 'ngtdm_Coarseness',
  'gldm_SmallDependenceLowGrayLevelEmphasis',
  0.9428571428571428) ]
```

```
i]:
caracteristicas_a_tirar = []
for i, j, k in corr_pairs:
    caracteristicas_a_tirar.append(i)

caracteristicas_a_tirar = list(set(caracteristicas_a_tirar))
caracteristicas_a_tirar
```

```
i]: [ 'glcm_Idn',
      'glrlm_ShortRunEmphasis',
      'firstorder_Mean',
      'firstorder_Median',
      'gldm_DependenceEntropy',
      'glcm_InverseVariance',
      'gldm_SmallDependenceLowGrayLevelEmphasis',
      'firstorder_RootMeanSquared',
      'glrlm_RunLengthNonUniformityNormalized',
      'glrlm_RunPercentage',
      'firstorder_Maximum',
      'firstorder_Minimum',
      'glszm_SmallAreaEmphasis',
      'glrlm_RunEntropy',
      'ngtdm_Coarseness',
      'firstorder_TotalEnergy' ]
```

```
i]:
caracteristicas_a_tirar = list(set(caracteristicas_a_tirar))
print (caracteristicas_a_tirar)
```

```
['glcm_Idn', 'firstorder_Mean', 'firstorder_Median', 'gldm_DependenceEntropy', 'glcm_InverseVariance', 'gldm_SmallDependenceLowGrayLevelEmphasis', 'firstorder_Minimum', 'firstorder_RootMeanSquared', 'glrlm_RunLengthNonUniformityNormalized', 'glrlm_RunPercentage', 'firstorder_Maximum', 'glrlm_ShortRunEmphasis', 'glzsm_SmallAreaEmphasis', 'glrlm_RunEntropy', 'ngtdm_Coarseness', 'firstorder_TotalEnergy']
```

```
caracteristicas_selecionadas = micro.drop(caracteristicas_a_tirar, axis=1)
caracteristicas_selecionadas = list(set(caracteristicas_selecionadas))
caracteristicas_selecionadas
```

```
['glcm_MCC',
'glzsm_ZoneEntropy',
'glcm_SumEntropy',
'glcm_Idmn',
'firstorder_90Percentile',
'firstorder_10Percentile',
'glcm_Imc2',
'glzsm_SizeZoneNonUniformityNormalized',
'firstorder_Energy']
```

```
[1]: import pingouin as pg
import pandas as pd
import numpy as np
```

```
[2]: micro = pd.read_excel(r'C:\Users\erica\OneDrive\MPTRCS\planilhas_quarta_analise_BOAS\micro ICC.xlsx')
micro.head()
```

```
[2]:
```

	caracteristica	equipamento	valor
0	firstorder_10Percentile	1º ensaio_DBT_micro	2282.00
1	firstorder_10Percentile	2º ensaio_DBT_micro	2295.40
2	firstorder_10Percentile	1º ensaio_PC1_micro	2229.40
3	firstorder_10Percentile	2º ensaio_PC1_micro	2236.25
4	firstorder_10Percentile	1º ensaio_PC2_micro	2242.40

```
[3]: print(pd.pivot_table(micro, index='equipamento', columns='caracteristica').T)
```

```
equipamento
caracteristica      1º ensaio_DBT_micro \
valor firstorder_10Percentile      2.282000e+03
      firstorder_90Percentile      2.335000e+03
      firstorder_Energy            1.426813e+11
      firstorder_Maximum            2.754000e+03
      firstorder_Mean                2.309734e+03
      firstorder_Median              2.309900e+03
      firstorder_Minimum            2.223000e+03
      firstorder_RootMeanSquared    2.309908e+03
      firstorder_TotalEnergy        1.426813e+09
      glcm_Idmn                      9.969510e-01
      glcm_Idn                       9.627589e-01
      glcm_Imc2                      2.815550e-01
      glcm_InverseVariance           5.114210e-01
      glcm_MCC                       7.805844e-01
      glcm_SumEntropy                2.471865e+00
      gldm_DependenceEntropy         4.295965e+00
```

```
[4]: results_micro = pg.intraclass_corr(data=micro, targets='caracteristica', raters='equipamento', ratings='valor')
```

```
pd.set_option('display.max_columns', 8)
pd.set_option('display.width', 200)
```

```
print(results_micro)
df = pd.DataFrame.from_dict(results_micro)
df.to_excel("results_micro.xlsx", index=False)
```

	Type	Description	ICC	F	df1	df2	pval	CI95%
0	ICC1	Single raters absolute	0.981236	314.760580	24	125	5.415283e-100	[0.97, 0.99]
1	ICC2	Single random raters	0.981236	315.023073	24	120	9.155284e-97	[0.97, 0.99]
2	ICC3	Single fixed raters	0.981251	315.023073	24	120	9.155284e-97	[0.97, 0.99]
3	ICC1k	Average raters absolute	0.996823	314.760580	24	125	5.415283e-100	[0.99, 1.0]
4	ICC2k	Average random raters	0.996823	315.023073	24	120	9.155284e-97	[0.99, 1.0]
5	ICC3k	Average fixed raters	0.996826	315.023073	24	120	9.155284e-97	[0.99, 1.0]

```
[5]: results_micro = results_micro.set_index('Description')
icc = results_micro.loc['Single random raters', 'ICC']
```

```
print(icc.round(3))
0.981
```

```
[6]: lower_ci = results_micro.loc['Single random raters', 'CI95%'][0]
upper_ci = results_micro.loc['Single random raters', 'CI95%'][1]
```

```
print(f'ICC = {icc:.3f}, 95% CI [{lower_ci}, {upper_ci}]')
ICC = 0.981, 95% CI [0.97, 0.99]
```

Apêndice C – Documentação do Jupyter Notebook utilizado neste trabalho para a análise do *test object* MASSA.

```

[]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline

from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report

import warnings

massa = pd.read_excel(r'C:\Users\erica\OneDrive\MPTRCS\planilhas_quarta_analise_BOAS
massa.head()

```

```

[]:

```

	firstorder_10Percentile	firstorder_90Percentile	firstorder_Energy	firstorder_Entropy	firstorder_Kurt
0	2284.00	2363.0	1.000638e+11	2.345149	2.799
1	2297.20	2375.8	9.059263e+10	2.333631	2.716
2	2232.80	2380.4	1.199668e+11	3.235314	2.659
3	2236.75	2390.0	1.203341e+11	3.278062	2.877
4	2246.80	2390.6	9.572042e+10	3.194683	2.745

5 rows × 25 columns

```

[]:

```

```

corr = massa.corr('spearman')
corr

```

```

[]:

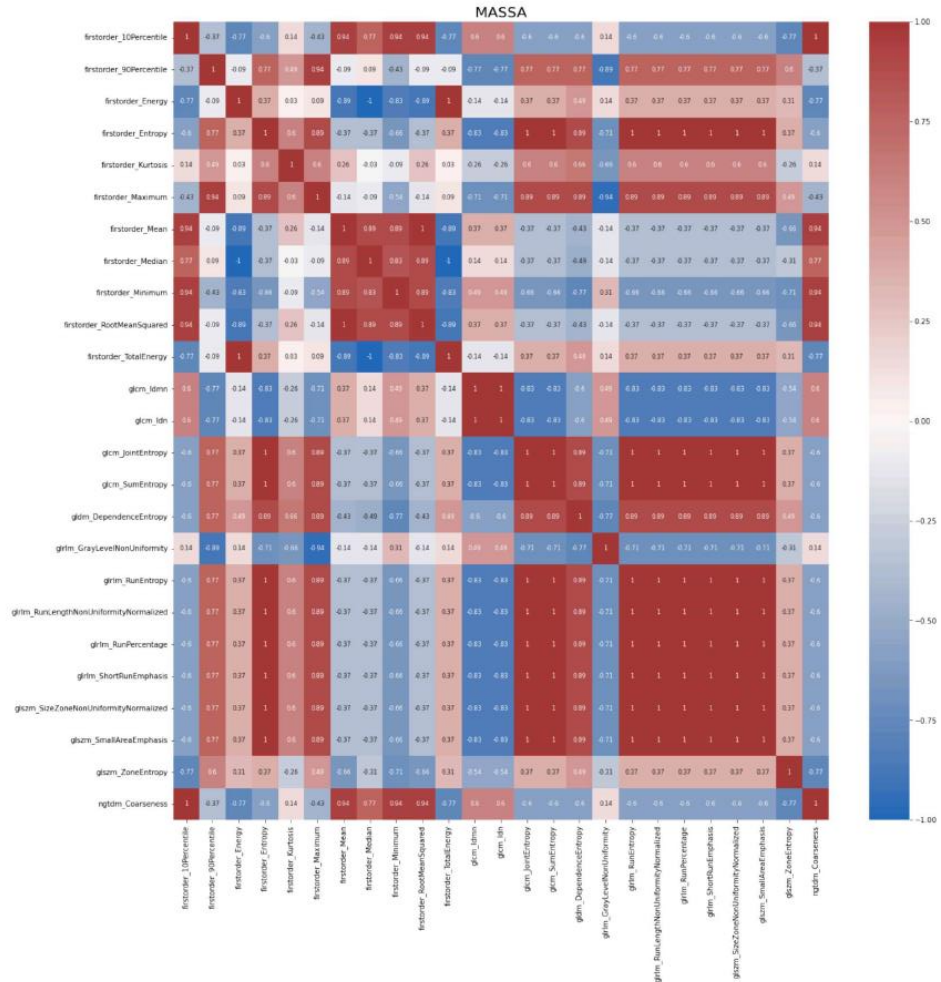
```

	firstorder_10Percentile	firstorder_90Percentile	firstorder_E
firstorder 10Percentile	1.000000	-0.371429	-0.771429
firstorder 90Percentile	-0.371429	1.000000	-0.085714
firstorder_Energy	-0.771429	-0.085714	1.000000
firstorder Entropy	-0.600000	0.771429	0.371429
firstorder Kurtosis	0.142857	0.485714	0.000000
firstorder Maximum	-0.428571	0.942857	0.000000
firstorder Mean	0.942857	-0.085714	-0.857143
firstorder Median	0.771429	0.085714	-1.000000
firstorder Minimum	0.942857	-0.428571	-0.857143
firstorder RootMeanSquared	0.942857	-0.085714	-0.857143
firstorder TotalEnergy	-0.771429	-0.085714	1.000000
gpcm_Idmn	0.600000	-0.771429	-0.142857
gpcm_Idn	0.600000	-0.771429	-0.142857
gpcm_JointEntropy	-0.600000	0.771429	0.371429

```

corr = massa.corr('spearman')
plt.figure(figsize=(20,20))
plot = sns.heatmap(massa.corr('spearman').round(2), annot = True, annot_kws={'fontsize': 8},
cmap='vlag', center=0, vmin=-1, vmax=1, xticklabels = 1, yticklabels = 1)
plot.set_title('MASSA', fontsize = 20)

```



```

corr_pairs=[]
for i in range(len(corr.columns)):
    for j in range(i):
        if corr.iloc[i,j]>0.9:
            corr_pairs.append((corr.columns[i],corr.columns[j],corr.iloc[i,j]))
corr_pairs

```

```

[('firstorder_Maximum', 'firstorder_90Percentile', 0.9428571428571428),
 ('firstorder_Mean', 'firstorder_10Percentile', 0.9428571428571428),
 ('firstorder_Minimum', 'firstorder_10Percentile', 0.9428571428571428),
 ('firstorder_RootMeanSquared', 'firstorder_10Percentile', 0.9428571428571428),
 ('firstorder_RootMeanSquared', 'firstorder_Mean', 1.0),
 ('firstorder_TotalEnergy', 'firstorder_Energy', 1.0),
 ('glcm_Idn', 'glcm_Idmn', 1.0),
 ('glcm_JointEntropy', 'firstorder_Entropy', 1.0),
 ('glcm_SumEntropy', 'firstorder_Entropy', 1.0),
 ('glcm_SumEntropy', 'glcm_JointEntropy', 1.0),
 ('glrlm_RunEntropy', 'firstorder_Entropy', 1.0),
 ('glrlm_RunEntropy', 'glcm_JointEntropy', 1.0),
 ('glrlm_RunEntropy', 'glcm_SumEntropy', 1.0),
 ('glrlm_RunLengthNonUniformityNormalized', 'firstorder_Entropy', 1.0),

```

```
( 'glrlm_RunLengthNonUniformityNormalized', 'glcm_JointEntropy', 1.0),
( 'glrlm_RunLengthNonUniformityNormalized', 'glcm_SumEntropy', 1.0),
( 'glrlm_RunLengthNonUniformityNormalized', 'glrlm_RunEntropy', 1.0),
( 'glrlm_RunPercentage', 'firstorder_Entropy', 1.0),
( 'glrlm_RunPercentage', 'glcm_JointEntropy', 1.0),
( 'glrlm_RunPercentage', 'glcm_SumEntropy', 1.0),
( 'glrlm_RunPercentage', 'glrlm_RunEntropy', 1.0),
( 'glrlm_RunPercentage', 'glrlm_RunLengthNonUniformityNormalized', 1.0),
( 'glrlm_ShortRunEmphasis', 'firstorder_Entropy', 1.0),
( 'glrlm_ShortRunEmphasis', 'glcm_JointEntropy', 1.0),
( 'glrlm_ShortRunEmphasis', 'glcm_SumEntropy', 1.0),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunEntropy', 1.0),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunLengthNonUniformityNormalized', 1.0),
( 'glrlm_ShortRunEmphasis', 'glrlm_RunPercentage', 1.0),
( 'glszm_SizeZoneNonUniformityNormalized', 'firstorder_Entropy', 1.0),
( 'glszm_SizeZoneNonUniformityNormalized', 'glcm_JointEntropy', 1.0),
( 'glszm_SizeZoneNonUniformityNormalized', 'glcm_SumEntropy', 1.0),
( 'glszm_SizeZoneNonUniformityNormalized', 'glrlm_RunEntropy', 1.0),
( 'glszm_SizeZoneNonUniformityNormalized',
  'glrlm_RunLengthNonUniformityNormalized',
  1.0),
( 'glszm_SizeZoneNonUniformityNormalized', 'glrlm_RunPercentage', 1.0),
( 'glszm_SizeZoneNonUniformityNormalized', 'glrlm_ShortRunEmphasis', 1.0),
( 'glszm_SmallAreaEmphasis', 'firstorder_Entropy', 1.0),
( 'glszm_SmallAreaEmphasis', 'glcm_JointEntropy', 1.0),
( 'glszm_SmallAreaEmphasis', 'glcm_SumEntropy', 1.0),
( 'glszm_SmallAreaEmphasis', 'glrlm_RunEntropy', 1.0),
( 'glszm_SmallAreaEmphasis', 'glrlm_RunLengthNonUniformityNormalized', 1.0),
( 'glszm_SmallAreaEmphasis', 'glrlm_RunPercentage', 1.0),
( 'glszm_SmallAreaEmphasis', 'glrlm_ShortRunEmphasis', 1.0),
( 'glszm_SmallAreaEmphasis', 'glszm_SizeZoneNonUniformityNormalized', 1.0),
( 'ngtdm_Coarseness', 'firstorder_10Percentile', 1.0),
( 'ngtdm_Coarseness', 'firstorder_Mean', 0.9428571428571428),
( 'ngtdm_Coarseness', 'firstorder_Minimum', 0.9428571428571428),
( 'ngtdm_Coarseness', 'firstorder_RootMeanSquared', 0.9428571428571428)]
```

```
] :
caracteristicas_a_tirar = []
for i, j, k in corr_pairs:
    caracteristicas_a_tirar.append(i)

caracteristicas_a_tirar = list(set(caracteristicas_a_tirar))
caracteristicas_a_tirar
```

```
] : ['glszm_SmallAreaEmphasis',
'glrlm_RunEntropy',
'glrlm_ShortRunEmphasis',
'ngtdm_Coarseness',
'firstorder_Maximum',
'firstorder_TotalEnergy',
'glcm_Idn',
'firstorder_Minimum',
'glcm_JointEntropy',
'glcm_SumEntropy',
'firstorder_RootMeanSquared',
'glrlm_RunLengthNonUniformityNormalized',
'glrlm_RunPercentage',
'glszm_SizeZoneNonUniformityNormalized',
'firstorder_Mean']
```

```
] :
caracteristicas_a_tirar = list(set(caracteristicas_a_tirar))
print (caracteristicas_a_tirar)
```

```
['glszm_SmallAreaEmphasis', 'glrlm_RunEntropy', 'glrlm_ShortRunEmphasis', 'ngtdm_Coar
seness', 'firstorder_Maximum', 'firstorder_TotalEnergy', 'glcm_Idn', 'firstorder_Min
imum', 'glcm_JointEntropy', 'glcm_SumEntropy', 'firstorder_RootMeanSquared', 'glrlm_Ru
nLengthNonUniformityNormalized', 'glrlm_RunPercentage', 'glszm_SizeZoneNonUniformityN
ormalized', 'firstorder_Mean']
```

```
1: caracteristicas_selecionadas = massa.drop(caracteristicas_a_tirar, axis=1)
caracteristicas_selecionadas = list(set(caracteristicas_selecionadas))
caracteristicas_selecionadas
```

```
1: ['firstorder_Entropy',
'gldm_DependenceEntropy',
'glszm_ZoneEntropy',
'glcm_Idmn',
'firstorder_10Percentile',
'glrlm_GrayLevelNonUniformity',
'firstorder_Kurtosis',
'firstorder_Energy',
'firstorder_Median',
'firstorder_90Percentile']
```

```
[1]: import pinguin as pg
import pandas as pd
import numpy as np
```

```
[2]: massa = pd.read_excel(r'C:\Users\erica\OneDrive\MPTRCS\planilhas_quarta_analise_BOAS\massa ICC.xlsx')
massa.head()
```

	caracteristica	equipamento	valor
0	firstorder_10Percentile	1º ensaio_DBT_massa	2284.00
1	firstorder_10Percentile	2º ensaio_DBT_massa	2297.20
2	firstorder_10Percentile	1º ensaio_PC1_massa	2232.80
3	firstorder_10Percentile	2º ensaio_PC1_massa	2236.75
4	firstorder_10Percentile	1º ensaio_PC2_massa	2246.80

```
[3]: print(pd.pivot_table(massa, index='equipamento', columns='caracteristica').T)
```

equipamento	1º ensaio_DBT_massa \
caracteristica	
valor firstorder_10Percentile	2.284000e+03
firstorder_90Percentile	2.363000e+03
firstorder_Energy	1.000638e+11
firstorder_Entropy	2.345149e+00
firstorder_Kurtosis	2.799404e+00
firstorder_Maximum	2.433000e+03
firstorder_Mean	2.320921e+03
firstorder_Median	2.318000e+03
firstorder_Minimum	2.231000e+03
firstorder_RootMeanSquared	2.321119e+03
firstorder_TotalEnergy	1.000638e+09
glcm_Idmn	9.826573e-01
glcm_Idn	9.137934e-01
glcm_JointEntropy	4.444602e+00
glcm_SumEntropy	3.116289e+00
gldm_DependenceEntropy	4.936064e+00

```
[5]: results_massa = pg.intraClass_corr(data=massa, targets='caracteristica', raters='equipamento', ratings='valor')
pd.set_option('display.max_columns', 8)
pd.set_option('display.width', 200)
print(results_massa)
df = pd.DataFrame.from_dict(results_massa)
df.to_excel("results_massa.xlsx", index=False)
```

	Type	Description	ICC	F	df1	df2	pval	CI95%
0	ICC1	Single raters absolute	0.982428	336.459943	24	125	9.067845e-102	[0.97, 0.99]
1	ICC2	Single random raters	0.982428	336.740532	24	120	1.800925e-98	[0.97, 0.99]
2	ICC3	Single fixed raters	0.982443	336.740532	24	120	1.800925e-98	[0.97, 0.99]
3	ICC1k	Average raters absolute	0.997028	336.459943	24	125	9.067845e-102	[0.99, 1.0]
4	ICC2k	Average random raters	0.997028	336.740532	24	120	1.800925e-98	[0.99, 1.0]
5	ICC3k	Average fixed raters	0.997030	336.740532	24	120	1.800925e-98	[0.99, 1.0]

```
[6]: results_massa = results_massa.set_index('Description')
icc = results_massa.loc['Single random raters', 'ICC']
print(icc.round(3))
0.982
```

```
[7]: lower_ci = results_massa.loc['Single random raters', 'CI95%'][0]
upper_ci = results_massa.loc['Single random raters', 'CI95%'][1]
print(f'ICC = {icc:.3f}, 95% CI [{lower_ci}, {upper_ci}]')
ICC = 0.982, 95% CI [0.97, 0.99]
```

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Diretoria de Pesquisa, Desenvolvimento e Ensino
Av. Prof. Lineu Prestes, 2242 – Cidade Universitária CEP: 05508-000
Fone (11) 2810-1570 ou (11) 2810-1572
SÃO PAULO – São Paulo – Brasil
<http://mprofissional.ipen.br>

O Instituto de Pesquisas Energéticas e Nucleares (IPEN) é uma Autarquia vinculada à Secretaria de Desenvolvimento Econômico do Governo do Estado de São Paulo e gerida técnica e administrativamente pela Comissão Nacional de Energia Nuclear (CNEN), órgão do Ministério da Ciência, Tecnologia e Inovações (MCTI) do Governo Federal.