

Modelo matemático fuzzy de classificação do BI-RADS de nódulos de mama

Ellen V. Silva¹ Graciele P. Silveira²
DFQM, CCTS – UFSCar, 18.052-780, Sorocaba/SP.

Resumo. O câncer é um grave problema de saúde pública mundial, cujas incidência e mortalidade têm aumentado de modo alarmante nos últimos anos. O câncer de mama constitui-se na neoplasia mais frequente entre as mulheres e um diagnóstico precoce é fundamental para o tratamento e a sobrevivência das pacientes. O propósito deste trabalho foi desenvolver um modelo matemático via teoria de conjuntos fuzzy, para classificar o BI-RADS (Breast Imaging Reporting Data System), levando em consideração o tempo de duplicação do tumor e outras características como formato, contorno e densidade. Pesquisas recentes indicam a importância do rastreamento e acompanhamento dos exames para identificar tais elementos. A metodologia usada são os sistemas baseados em regras fuzzy (SBRF), com inferência de Mamdani e defuzzificação pelo Centro de Gravidade. Simulações foram implementadas, supondo diferentes quadros clínicos de pacientes e os resultados apontaram que o modelo é útil, podendo contribuir com o diagnóstico e tratamento de tumores mamários malignos.

Palavras-chave: *Câncer de mama; Breast Imaging Reporting Data System; Conjuntos fuzzy.*

1. Introdução

Câncer é uma denominação que abrange um conjunto de mais de 100 doenças malignas, que possuem como característica comum o crescimento desordenado de células, podendo invadir tecidos adjacentes e outros órgãos (INCA, 2022). Estas células ao se dividirem rapidamente, tendem a ser muito agressivas

¹ellen.vilarinho@gmail.com

²graciele@ufscar.br

e incontroláveis, dando origem a formação de tumores, que podem se espalhar para as demais regiões do corpo (metástase).

Nos últimos anos, a incidência tem crescido consideravelmente, tanto em países desenvolvidos como naqueles em processo de desenvolvimento. De acordo com o Instituto Nacional do Câncer - INCA (2022), a doença atualmente constitui-se no principal problema de saúde pública, sendo responsável por 9,6 milhões de mortes (1/6 de todos os óbitos) e 18 milhões de novos casos no mundo em 2018.

O câncer mais incidente no globo foi o de pulmão, representando 2,1 milhões de casos, seguido pelo câncer de mama, que apresenta também 2,1 milhões de casos. Os tipos de câncer mais frequentes nos homens foram o câncer de pulmão e de próstata. Nas mulheres, o câncer de mama foi o mais incidente seguido pelo câncer de cólon e reto.

A Organização Pan-Americana da Saúde - OPAS (2020), prevê que a mortalidade por câncer nas Américas aumente para 2,1 milhões até 2030. Para o Brasil, a estimativa mais recente do INCA indica que para o triênio 2020-2022, ocorrerão 625 mil novos casos de câncer (INCA, 2019).

Não existe uma causa única para o câncer. As causas podem ser externas (fatores ambientais) ou internas como os hormônios, condições imunológicas e mutações genéticas. Entretanto, estudos mostram que entre 80% e 90% dos casos da doença envolvem fatores externos e cerca de um terço das mortes de câncer são ocasionadas por cinco principais riscos comportamentais e alimentares: alto índice de massa corporal, baixo consumo de frutas e vegetais, falta de atividade física e uso de álcool e tabaco (OPAS, 2020).

Pessoas idosas são mais suscetíveis à desenvolver a doença, já que foram expostas por mais tempo aos fatores de risco e o intrínseco processo de envelhecimento do ser humano traz mudanças às células. Essas considerações podem explicar parcialmente o motivo da doença ser mais frequente em pessoas mais velhas.

Uma das maiores dificuldades no tratamento do câncer é localizar as células malignas e destruí-las, sem comprometer as células saudáveis e os tecidos do paciente. Células tumorais que não forem eliminadas podem originar um novo tumor, chamado de *recidiva*.

Os tratamentos disponíveis para o câncer são cirurgia, quimioterapia, radioterapia, imunoterapia ou transplante de medula óssea. Em muitos casos é necessário combinar mais de uma modalidade de tratamento.

Na quimioterapia são administrados medicamentos que possuem ação citotóxica, isto é, que causam danos às células. Pode ser utilizada para remover o tumor antes da cirurgia, mas também é empregada após o tratamento cirúrgico ou radioterápico.

A radioterapia é um procedimento que faz uso de radiação e tem como finalidade destruir as células doentes. É utilizado em tumores localizados, que não podem ser ressecados totalmente ou para tumores que costumam recidivar ou reaparecer. O uso desta terapia pode causar diversos efeitos colaterais localizados, então há constante preocupação em aumentar a precisão para não causar lesões nos tecidos vizinhos.

Existem algumas substâncias, chamadas de *marcadores tumorais*, cujas concentrações estão relacionadas à presença de células tumorais. Essas substâncias são detectadas e qualificadas no sangue ou em outros fluidos do corpo. Surgem com o aparecimento de neoplasias ou aumentam sua concentração na presença de células cancerígenas. Portanto, exames periódicos para medições das concentrações destes elementos podem auxiliar no diagnóstico da doença.

1.1. O câncer de mama

Segundo o INCA (2019), em 2018 o câncer de mama foi o segundo tipo de neoplasia maligna mais incidente no mundo, responsável por mais de 2,1 milhões de casos, o que representou 11,6% de todos os cânceres estimados. Entre as mulheres, este é o tumor de maior incidência, independentemente do Índice de Desenvolvimento Humano (IDH).

A World Health Organization (WHO, 2021), registra que em 2020 houve aproximadamente 2,3 milhões de mulheres no mundo diagnosticadas com câncer de mama e 685 mil mortes. Ao fim de 2020, existiam no mundo 7,8 milhões de mulheres vivas que foram diagnosticadas com câncer de mama nos últimos 5 anos. Devido à pandemia do novo Coronavírus, os dados referentes ao ano de 2020 podem estar subnotificados.

A estimativa mais recente apresentada pelo INCA (2019), mostra que no triênio 2020-2022 ocorrerão 66.280 mil novos casos de câncer de mama no Brasil. Esse valor corresponde a um risco estimado de 61,61 casos novos a cada 100 mil mulheres. Em 2017 ocorreram 16.724 óbitos por câncer de mama feminina, o que equivale a um risco de 16,16 por 100 mil.

Entre as décadas de 1930 e 1970 não aconteceram avanços ou aumento

na sobrevivência das pacientes. As taxas de mortalidade começaram a mudar a partir de 1980, com a detecção precoce e o uso de diferentes tratamentos para erradicar a doença invasiva (WHO, 2021).

O câncer de mama não é uma doença infecciosa ou transmissível. Como afirma o INCA (2019), não há um único fator de risco para o câncer de mama, contudo, a idade acima dos 50 anos é considerada o aspecto mais importante. Outros fatores que contribuem para o desenvolvimento da doença são as mutações dos genes BRCA1 e BRCA2. Fatores hereditários, menopausa tardia, obesidade, sedentarismo e exposições frequentes a radiações ionizantes também aumentam o risco de desenvolver a doença.

Tumores mamários frequentemente se apresentam como um nódulo indolor ou como espessamento na mama. É recomendável para as mulheres que ao perceberem um nódulo anormal na mama consultem um profissional da saúde rapidamente, entre 1 e 2 meses, havendo dor ou não. Alguns dos sintomas do câncer de mama são: alteração no tamanho, forma ou aparência da mama; ondulações, vermelhidão, depressões ou outras alterações na pele; mudança na aparência do mamilo ou alteração na pele ao redor do mesmo; secreção mamilar anormal.

A investigação inicial de tumores mamários ocorre através da ultrassonografia e da mamografia. Contudo, a maior parte dos cânceres de mama é descoberto pelas próprias mulheres, ao perceberem alterações na mama (INCA, 2019). Para detecção precoce, uma das estratégias é a mamografia de rastreamento em mulheres jovens (abaixo de 50 anos), possibilitando uma abordagem menos agressiva, já que o tumor será tratado nos estágios iniciais, o que eleva o potencial curativo e reduz a mortalidade por câncer de mama.

A confirmação da doença ocorre por meio da amostragem de tecido ou biópsia. A biópsia é uma técnica que consiste na retirada de um fragmento do nódulo através de punções, para ser analisado pelo patologista.

O câncer de mama pode sofrer metástase e desencadear outros sintomas. Um primeiro local de disseminação mais comum são os gânglios linfáticos, localizados embaixo do braço. As células tumorais podem se espalhar para os demais órgãos como pulmões, fígado, cérebro e ossos. Os sintomas relacionados ao câncer, quando as células tumorais atingem essas regiões incluem dor óssea ou dores de cabeça (WHO, 2021).

A figura 1 ilustra a progressão do câncer de mama até a ocorrência de metástase.

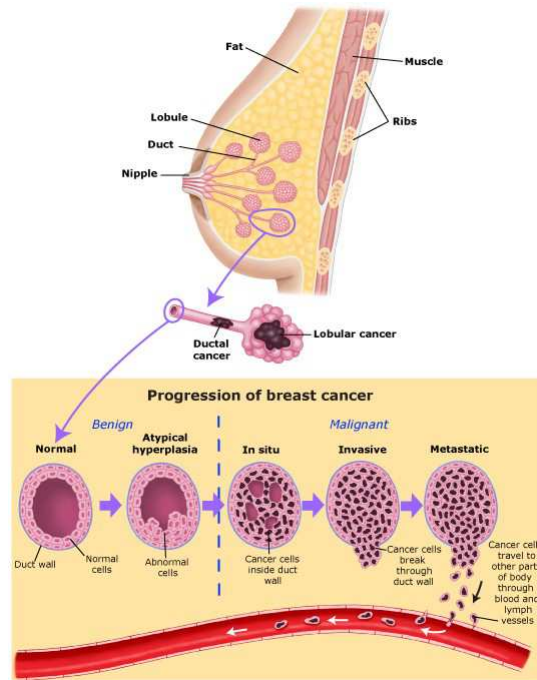


Figura 1: Ilustração do câncer de mama.

Fonte: Site do Saint John's Cancer Institute.

<https://www.saintjohnscancer.org/breast/breast-cancer/types-of-breast-cancer/>

1.2. Breast imaging reporting data system (BI-RADS)

Na década de 1980, com o aumento do uso da mamografia, problemas substanciais também surgiram, como disparidades nas práticas recomendadas ou nas doses de radiação prescritas, além de relatórios com indicações ambíguas (Burnside et al., 2009).

Em resposta, o Breast Imaging Reporting Data System (BI-RADS) de nódulos de mama foi desenvolvido no final da década de 1980, em colaboração por membros e comitês do ACR - *American College of Radiology*. Trata-se de um padrão de classificação, utilizado como um instrumento para guiar médicos e radiologistas acerca da lesão observada no exame de mamografia (Vieira e Toigo, 2002).

No laudo, o médico radiologista vai descrever o achado mamográfico, se

houver, em nódulos, cistos, linfonodos, calcificações etc. No fim do laudo, se encontra a classificação do BI-RADS, dada pelo médico radiologista após analisar o exame. Para caracterizar a lesão, os profissionais observam as categorias de avaliação do BI-RADS, que são numeradas de 0 a 5 (Miranda et al., 2009), onde:

0: indica que é necessária uma avaliação adicional da imagem, isso ocorre em situação de acompanhamento.

1: não há calcificações suspeitas ou presença de massas, caracterizando uma mamografia negativa.

2: apresenta achados benignos, sem indicação de malignidade.

3: apresenta alto grau de benignidade, porém menor do que o grau da categoria 2 e uma pequena probabilidade de malignidade (menor que 2%), recomendando-se o pedido de uma nova mamografia após seis meses.

4: há lesões com alta probabilidade de malignidade, porém elas não apresentam características morfológicas de câncer de mama, devendo ser realizado um estudo histológico da lesão.

5: apresenta lesões com alta probabilidade de ser câncer.

Por mais que a criação do padrão de classificação do BI-RADS tenha representado um avanço, a atribuição das categorias ocorre nos ambientes de análise e geração de diagnósticos, à partir de características visuais das imagens. Desse modo, existem incertezas e imprecisões envolvidas no processo, o que faz com que se busque diminuir a distância entre os resultados gerados e os resultados esperados pelos médicos especialistas.

Em contrapartida, Miranda et al. (2009) ressalta que o BI-RADS obedece a uma estrutura de linguagem natural. Por exemplo: “se o nódulo tem formato arredondado, margens bem definidas e densidade mediana, então sua classificação será: achado com alta probabilidade de benignidade”. Esse tipo de estrutura se ajusta perfeitamente à forma de abordagem dos conjuntos fuzzy e da lógica fuzzy, conforme veremos adiante.

2. Objetivos

Os objetivos desta pesquisa foram:

- Compreender o estudo quantitativo realizado por Pereira (2017), que versou sobre o câncer de mama.
- Desenvolver um sistema baseado em regras fuzzy (SBRF) para fornecer a classificação do Breast Imaging Reporting Data System (BI-RADS), considerando o tempo de duplicação do tumor como uma nova variável de entrada.
- Realizar simulações de cenários.

Os detalhes da modelagem matemática desenvolvida serão explicitados nas próximas seções.

3. Modelagem fuzzy para a classificação do BI-RADS

A modelagem matemática é uma ferramenta utilizada para descrever fenômenos do mundo real e suas tendências, no decorrer do tempo. Entretanto, as doenças, os sintomas, o diagnóstico e as terapias são envoltos de incertezas: categorizações linguísticas com limites vagos, medidas imprecisas, prontuários incompletos, entre outros. Surge então a Teoria de Conjuntos Fuzzy, capaz de lidar com este tipo de incerteza, a partir da matematização de termos linguísticos como “aproximadamente” (Barros e Bassanezi, 2015).

Um SBRF é constituído por quatro módulos principais: módulo de fuzzificação, que representa as variáveis linguísticas de entrada e saída do sistema, transformadas em conjuntos fuzzy; módulo de inferência; base de regras, composta por proposições fuzzy e o módulo de defuzzificação, que é responsável por transformar a saída dada em conjunto fuzzy para um valor numérico.

No estudo desenvolvido por Pereira (2017), três modelos matemáticos foram trabalhados, a saber: o primeiro deles para a dinâmica do câncer de mama, o segundo para o crescimento tumoral e por fim, um modelo para o Breast Imaging Reporting Data System (BI-RADS), usando a lógica fuzzy.

Para a modelagem matemática da classificação do BI-RADS, a referida autora se reportou a um outro estudo, feito por Miranda et al. (2009), que considerou os coeficientes de contorno, formato e densidade, como variáveis de entrada. Pereira (2017) então acrescentou uma variável de entrada a mais, que foi a velocidade de crescimento do tumor.

Na pesquisa que originou este artigo, ao estudar a modelagem matemática fuzzy feita por Pereira (2017), verificou-se que a variável linguística *Velocidade de Crescimento*, que havia sido incrementada, não apresentava valores para os intervalos nos seus subconjuntos. A autora considerou uma escala gradual entre 0 e 1 no domínio da variável, o que não permite saber quais valores seriam considerados baixo ou alto para a respectiva velocidade de crescimento. Dessa forma, a busca inicial foi por intervalos que pudessem ser escolhidos para representar a velocidade de crescimento do tumor da mama, no sistema baseado em regras fuzzy (SBRF).

Após um levantamento bibliográfico acerca do tema, foi encontrada a referência MacInnes et al. (2020), onde os autores estimam o crescimento de tumores em termos do tempo de duplicação das células, a partir de diferentes fatores.

De acordo com MacInnes et al. (2020), uma neoplasia mamária de grau histológico 1 tem seu tamanho duplicado, em média, após 294 dias. Com grau histológico 2, a duplicação do tamanho acontece após 164 dias, em média, enquanto que um tumor com grau histológico 3, leva em média, 128 dias para duplicação. Logo, quanto mais tempo uma neoplasia gasta para dobrar de tamanho, menos agressivo é, e se levar menos dias para duplicar, mais agressivo é o tumor. Utilizando essas informações como base, sugerimos considerar como variável de entrada o *Tempo de Duplicação (Dias)*, em substituição à variável velocidade de crescimento.

Portanto, o modelo matemático fuzzy aqui proposto possui como variáveis de entrada: *Coefficiente de Formato*, classificado nos termos linguísticos **Baixo**, **Médio** e **Alto**; *Coefficiente de Contorno*, nos conjuntos **Muito Baixo**, **Baixo**, **Médio** e **Alto**; *Coefficiente de Densidade*, classificado em **Baixo**, **Médio** e **Alto**; e, por fim, *Tempo de Duplicação (Dias)*, cujos termos linguísticos atribuídos foram **Curto**, **Médio** e **Longo**.

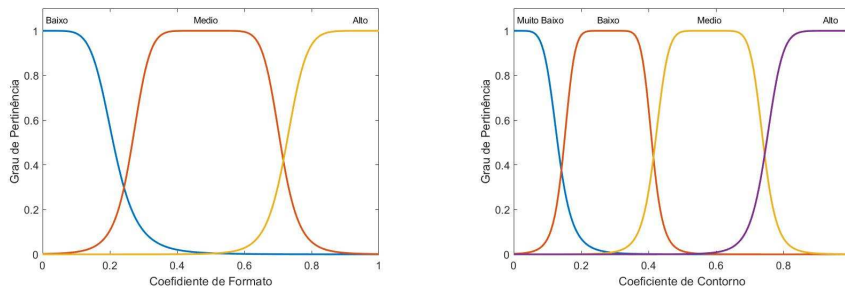
A variável de saída do sistema foi o padrão de classificação do BI-RADS, que foi dividido em quatro subconjuntos fuzzy, representativos das seguintes categorias do BI-RADS: 2, 3, 4 e 5.

Para os conjuntos fuzzy de todos os coeficientes considerados (formato, contorno e densidade), o domínio dos valores está entre 0 e 1, visto que conforme evidencia Miranda et al. (2009), quanto mais alto for o coeficiente, maior será a influência exercida sobre a malignidade da lesão. Já para os conjuntos fuzzy da variável *Tempo de Duplicação (Dias)*, as informações apresentadas

por MacInnes et al. (2020) auxiliaram nas escolhas, isto é, um tempo curto de duplicação do tumor foi considerado até 128 dias, um tempo médio para o tumor dobrar de tamanho entre 128 e 164 dias e um tempo de duplicação longo, acima de 164 dias.

De modo análogo ao que foi feito pelas referências consultadas, as funções de pertinências foram construídas em forma de sino, visando manter uma transição gradual entre os estados.

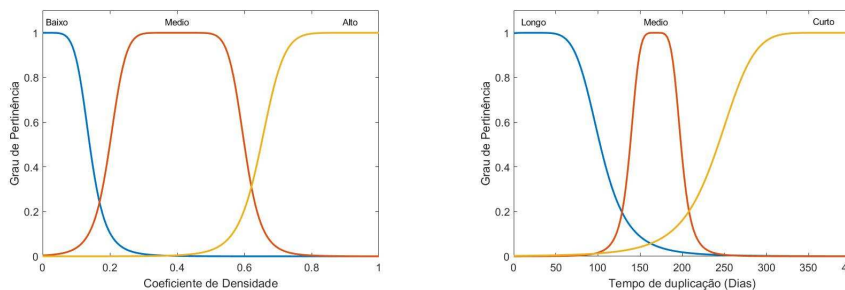
As figuras 2(a), 2(b), 3(a) e 3(b) apresentam as funções de pertinências para cada variável de entrada do SBRF.



(a) Funções de pertinências para a variável de entrada *Coeficiente de Formato*.

(b) Funções de pertinências para a variável de entrada *Coeficiente de Contorno*.

Figura 2: Funções de pertinências para a variável de entrada.



(a) Funções de pertinências para a variável de entrada *Coeficiente de Densidade*.

(b) Funções de pertinências para a variável de entrada *Tempo de Duplicação (Dias)*.

Figura 3: Funções de pertinências para a variável de entrada.

As funções de pertinências consideradas para a variável de saída do sis-

tema encontram-se na figura 4.

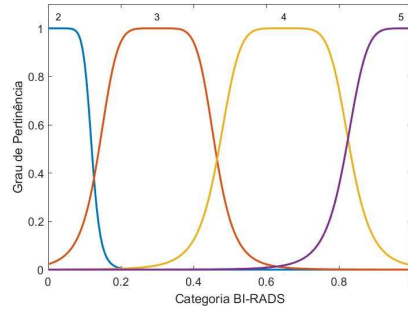


Figura 4: Funções de pertinências para a variável de saída *Classificação do BI-RADS*.

A figura 5 mostra o esquema geral do SBRF construído, considerando todas as variáveis de entrada e saída supracitadas. O método de inferência utilizado foi o de Mamdani e a defuzzificação realizada via método do Centro de Gravidade.

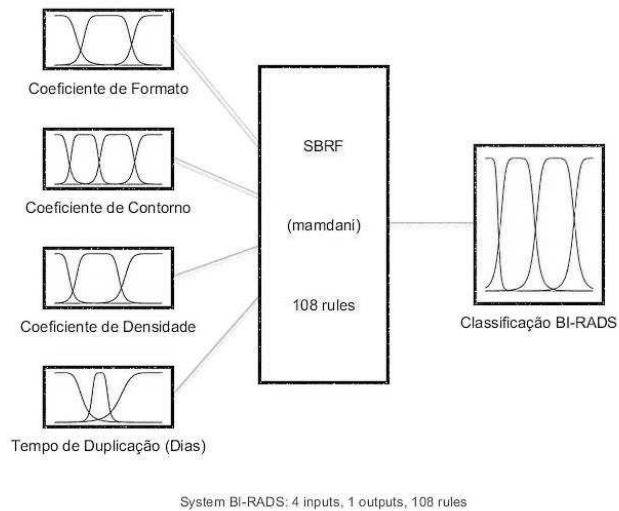


Figura 5: Esquema geral do SBRF construído.

A base de regras fuzzy foi elaborada tendo em vista os dados iniciais presentes em Pereira (2017) e Miranda et al. (2009), além das novas informações obtidas em MacInnes et al. (2020). Ao todo foram constituídas 108 regras fuzzy. Na tabela 1 encontram-se algumas regras da base de regras formada.

Tabela 1: Algumas das 108 regras da base de regras fuzzy. Na primeira coluna tem-se o número da regra e na última coluna o peso utilizado para ponderá-la.

N	Coef. de Form.	Coef. de Contor.	Coef. de Densid.	Tempo de Duplic.	BI-RADS	Peso
03	Baixo	Muito Baixo	Alto	Longo	4	1
21	Médio	Médio	Alto	Longo	5	1
38	Baixo	Muito Baixo	Médio	Médio	3	1
66	Alto	Baixo	Alto	Médio	4	1
84	Baixo	Alto	Alto	Curto	3	1

Ao acionar as condições de entrada para cada subconjunto fuzzy, uma determinada regra vai ser ativada. Ao acionar a regra n 38, por exemplo, tem-se:

R38 Se o Coeficiente de Formato é Baixo e o Coeficiente de Contorno é Muito Baixo e o Coeficiente de Densidade é Médio e o Tempo de Duplicação do tumor é Médio então a classificação do BI-RADS é 3 (1).

Todas as regras possuem a forma do exemplo anterior.

4. Simulações e Resultados

A modelagem matemática construída com o aporte da lógica fuzzy, após testes e ajustes a partir de dados reais, poderá vir a auxiliar médicos especialistas e radiologistas, se os mesmos dispuserem de dados pré-operatórios das pacientes. A ideia é que o modelo seja capaz de determinar com maior exatidão a classificação do BI-RADS, uma vez realizados os exames clínicos, como a mamografia, o que irá favorecer um tratamento mais efetivo da doença.

As simulações foram implementadas via MATLAB/Octave, tomando situações clínicas de pacientes hipotéticas. Após a defuzzificação, o sistema baseado em regras fuzzy fornece um número real e a partir deste foi calculada a pertinência à cada conjunto da saída do sistema.

Suponha uma paciente que em sua mamografia, o radiologista atribui os valores de Coeficiente de Formato equivalente a 0,8 (Alto), Coeficiente de Contorno igual a 0,7 (Médio) e Coeficiente de Densidade 0,6 (Médio). Se o tumor duplicar de tamanho em 100 dias (Curto), o sistema fuzzy fornece que a neoplasia possui pertinência de 31% para a categoria 4 do BI-RADS e 69% de pertinência para a categoria 5, indicando que o tumor contém lesões suspeitas e que a probabilidade de ser câncer maligno é maior em relação à chance de ser um tumor incerto.

Em um segundo caso, considere uma paciente que possui Coeficiente de Formato igual a 0,5 (Médio), Coeficiente de Contorno equivalente a 0,8 (Alto), Coeficiente de Densidade igual a 0,3 (Médio) e que transcorridos 80 dias (Longo) foi constatado que o tumor dobrou de tamanho. Logo, para essa paciente o sistema fuzzy forneceu 98% para a categoria 3 do BI-RADS e 2% para a categoria 4, apontando que o tumor apresenta um nível de malignidade que não pode ser desprezado e deve ser observado.

5. Conclusões

O câncer de mama é o mais incidente entre as mulheres, sendo responsável por 2,1 milhões de casos e 685 mil mortes em 2020, segundo dados divulgados pela World Health Organization (WHO, 2021).

O rastreamento, a detecção precoce e a combinação de tratamentos são cruciais para aprimorar a sobrevivência das pacientes. Não há uma única causa para a doença; a idade acima de 50 anos, fatores hereditários e fatores ambientais podem estar relacionados ao aparecimento de tumores mamários.

Em razão da urgência que o tema traz, os objetivos deste trabalho foram inicialmente estudar e compreender o modelo matemático fuzzy proposto por Pereira (2017), cujo intuito foi obter a classificação do BI-RADS - Breast Imaging Reporting Data System, a partir de dados pré-operatórios, por meio de um sistema baseado em regras fuzzy. Em seguida, com o avanço da pesquisa, optou-se por considerar uma nova variável de entrada, a saber, o tempo de duplicação do tumor.

As simulações executadas supondo pacientes hipotéticas, demonstraram que o modelo tem potencial para fornecer uma classificação mais acurada para o BI-RADS, o que pode auxiliar os médicos especialistas e radiologistas no processo de tomada de decisão, com relação ao tumor de mama. Em trabalhos

futuros, pretende-se fazer ajustes contando com a participação de especialistas, a fim de aperfeiçoar os resultados alcançados.

Agradecimentos

A primeira autora agradece ao CNPq e a UFSCar pela bolsa de Iniciação Científica do PIBIC.

Referências

- Barros, L. C. e Bassanezi, R. C. (2015). *Tópicos de Lógica Fuzzy e Biomatemática*, volume 3 of *Coleção Textos Didáticos*. IMECC–UNICAMP.
- Burnside, E. S., Sickles, E. A., e Bassett, L. W. (2009). The acr bi-rads experience: Learning from history. *Journal of the American College of Radiology*, 6(12):851–860.
- INCA (2019). Estimativa 2020: incidência de câncer no brasil. Disponível em: <https://www.inca.gov.br/publicacoes/livros/estimativa-2020-incidencia-de-cancer-no-brasil>. Acesso em: 25/07/2022.
- INCA (2022). O que é câncer? Disponível em: <https://www.inca.gov.br/o-que-e-cancer>. Acesso em: 25/07/2022.
- MacInnes, E. G., Duffy, S. W., e Simpson, J. A. (2020). Radiological audit of interval breast cancers: Estimation of tumour growth rates. *The Breast*, 51:114–119.
- Miranda, G. H. B., Marques, P. M. A., e Felipe, J. C. (2009). Aplicação dos conceitos da lógica nebulosa à classificação bi-rads de nódulos de mama. *J. Health Inform.*, 1(1):07–16.
- OPAS (2020). Folha informativa atualizada em outubro de 2020. Disponível em: <https://www.paho.org/pt/topicos/cancer>. Acesso em: 25/07/2022.
- Pereira, N. K. M. (2017). Um estudo quantitativo sobre o câncer de mama. Dissertação de Mestrado, IMECC–Unicamp, Campinas/SP.
- Vieira, A. V. e Toigo, F. P. (2002). Classificação bi-rads: categorização de 4968 mamografias. *Radiol. Bras.*, 35(4):205–208.

WHO (2021). Breast cancer. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. Acesso em: 25/07/2022.