REPRESENTAÇÃO DO CONHECIMENTO RELACIONADO AO PROGNÓSTICO DE PACIENTES COM NEOPLASIA MALIGNA DE MAMA: UM ESTUDO DE CASO DO HOSPITAL BOM PASTOR

Erika Gonçalves de Assis
Programa de Pós-Graduação em Informática
PUC-Minas
dudabh@gmail.com

Cristiane Neri Nobre
Programa de Pós-Graduação em Informática
PUC-Minas
nobre@pucminas.br

RESUMO

O câncer de mama é o mais frequente em mulheres e a maior causa de mortes femininas. A representação do conhecimento relacionado ao prognóstico de pacientes com câncer de mama é um recurso que permite ao especialista em ciências da saúde encontrar fatores que impactam o tempo de vida das pacientes. A metodologia utilizada neste estudo consistiu inicialmente na pesquisa e avaliação do Registro Hospitalar de Câncer do Hospital Bom Pastor para sua posterior utilização no processo de KDD. Primeiramente foi feito um pré-processamento dos dados e depois a adequação dos dados a serem utilizados. Os modelos para representar o conhecimento foram gerados através de árvore de decisão e rede neural . Por fim, foi realizada a avaliação desses modelos. Os resultados mostram que os pacientes que estavam nos estágios iniciais do câncer de mama não morreram até a última data de acompanhamento. Pacientes que tiveram metástase e apresentavam estadiamento 3A, 3B e 3C tiveram menos de 2 anos de sobrevida. Pacientes com metástase morreram após 5 anos de sobrevida. E os pacientes que não tiveram metástases não morreram dentro de 5 anos de sobrevivência. Verificou-se que houve redução na probabilidade de sobrevivência de acordo com o aumento do estadiamento. Os resultados deste estudo vão ao encontro dos dados da literatura, apontando que o estadiamento do câncer de mama é uma importante variável que explica as disparidades de sobrevida entre mulheres com essa neoplasia.

Palavras-chave: Análise de sobrevida. Câncer de mama, Mineração de dados. Aprendizado de máquina, Representação do conhecimento

REPRESENTAÇÃO DO CONHECIMENTO RELACIONADO AO PROGNÓSTICO DE PACIENTES COM NEOPLASIA MALIGNA DE MAMA: UM ESTUDO DE CASO DO HOSPITAL BOM PASTOR

ABSTRACT

Breast cancer is the most common cancer in women and the most significant cause of female deaths. The representation of knowledge related to the prognosis of patients with breast cancer is a resource that allows the health science specialist to find factors that impact the lifespan of patients. The methodology used in this study initially consisted of research and evaluation of the

Hospital Cancer Registry of Hospital Bom Pastor for its subsequent use in the KDD process. Firstly, the data was pre-processed, and then the data to be used was adapted. The models to represent knowledge were generated using a decision tree and neural network. Finally, these models were evaluated. The results show that patients who were in the early stages of breast cancer did not die until the last follow-up date. Patients who had metastasis and had stages 3A, 3B, and 3C had less than two years of survival. Patients with metastasis died after five years of survival. Patients who did not have metastases did not die within five years of survival. It was found that there was a reduction in the probability of survival with increasing staging. The results of this study are in line with the literature data, pointing out that breast cancer staging is an important variable that explains survival disparities among women with this neoplasm.

Key words: Survival analysis. Breast cancer. Data mining. Machine Learning. Knowledge Representation

1. INTRODUÇÃO

Globalmente, o câncer de mama é a principal causa de mortalidade e o mais prevalente entre as mulheres (JBILOU J. ET AL., 2014) De acordo com a última estimativa mundial da Agência Internacional de Pesquisa sobre o Câncer (IARC) em 2022, houve cerca de 2,2 milhões de novos casos e aproximadamente 666 mil mortes por câncer de mama em mulheres em todo o mundo (GLOBOCAN, 2022). No Brasil, para cada ano do triênio 2023-2025, foram estimados 73.610 novos casos, representando uma taxa de incidência ajustada de 41,89 casos por 100.000 mulheres, excluindo os tumores de pele não melanoma. Este tipo de câncer também é o mais frequente entre as mulheres em todas as regiões do país, com taxas mais elevadas nas regiões Sul e Sudeste (INCA, 2023).

A vigilância e o controle do câncer são de extrema importância, pois um aumento acelerado no número de novos casos pode sobrecarregar os recursos disponíveis para diagnóstico, tratamento e acompanhamento. Isso pode ter efeitos devastadores nos aspectos sociais e econômicos, representando um obstáculo significativo para o desenvolvimento socioeconômico, especialmente em países emergentes como o Brasil (INCA, 2023.

Diante desse cenário, a Mineração de Dados (MD) oferece métodos e instrumentos adequados para auxiliar na geração de conhecimentos mais robustos sobre o câncer, complementando os já existentes. A MD é o processo de descoberta automatizada de informações úteis em conjuntos massivos de dados, utilizando técnicas estatísticas e de inteligência artificial para identificar padrões, convergências e prognósticos.

Uma das vantagens desse estudo é a rapidez na identificação de padrões, o que permite definir prognósticos para os pacientes com base em algoritmos de mineração de dados. Isso resulta em uma tomada de decisão mais ágil em relação ao estado clínico e prognóstico dos pacientes (WITTEN et al, 2016).

A pesquisa foi conduzida utilizando dados fornecidos pela Fundação Hospitalar do Município de Varginha (Hospital Bom Pastor),

em conformidade com as diretrizes éticas estabelecidas, preservando o anonimato dos pacientes e respeitando os princípios éticos e legais aplicáveis à pesquisa médica. O Hospital Bom Pastor, localizado na região Sul de Minas Gerais, é uma instituição de saúde de atendimento geral que realiza diversos procedimentos anualmente e é o terceiro em atendimento para casos de câncer em Minas Gerais.

O estudo buscou identificar os fatores e combinações que influenciam na sobrevida de pacientes diagnosticados com câncer de mama maligno, analisando a evolução temporal dos padrões de morbidade e seus determinantes para embasar ações em saúde pública. Essas informações permitiram entender as características da população e compará-las em termos de gravidade da doença no momento do diagnóstico e distribuição demográfica.

Em saúde pública, é crucial estudar e melhorar as estruturas de atendimento para reduzir o tempo entre o início dos sintomas e o tratamento eficaz. O projeto teve um grande impacto ao fornecer informações mais precisas sobre a influência da qualidade da assistência na sobrevida, bem como o papel de outros determinantes potenciais.

2. DESENVOLVIMENTO

O processo de Descoberta de Conhecimento em Bases de Dados (KDD) é uma sequência de atividades contínuas que visam extrair e compartilhar conhecimento a partir de conjuntos de dados (WITTEN et al, 2016)..

O KDD é composto por cinco etapas que são descritas a seguir:

1. Seleção dos dados: preparação e escolha dos dados a serem utilizados;
2. Pré-processamento dos dados: remoção ou redução de ruídos presentes nos dados selecionados;
3. Transformação dos dados: aplicação de tratamentos e transformações para melhor adequação à extração de padrões;
4. Mineração de dados (Data Mining): busca e extração de padrões nos dados utilizando algoritmos;
5. Interpretação e avaliação dos resultados: análise da relevância e refinamento do conhecimento descoberto para o domínio em questão.

As técnicas de Data Mining desempenham tarefas como classificação, agrupamento e descoberta de regras de associação entre os dados. Entre essas técnicas, destacam-se árvores de decisão e redes neurais, que foram utilizadas neste estudo. Essas técnicas têm sido amplamente utilizadas na literatura para reconhecimento de padrões e classificação.

A classificação envolve a extração de padrões por meio de algoritmos de aprendizagem que buscam uma função capaz de classificar instâncias de dados em duas ou mais classes definidas a priori. O objetivo principal da classificação é construir um modelo que possa ser utilizado para fazer previsões (WITTEN et al, 2016).

2.1 Árvore de Decisão

A árvore de decisão é um método de classificação de dados que apresentam os resultados hierarquicamente, destacando o atributo mais importante no primeiro nó e os

atributos menos relevantes nos nós subsequentes. Quinlan (1992) desenvolveu a técnica de árvores de decisão, introduzindo o algoritmo ID3 e suas variantes, como ID4, ID6 e C4.5.

A técnica C4.5, uma dessas variantes, pode produzir tanto árvores de decisão quanto conjuntos de regras, sendo de fácil interpretação devido à clareza das regras derivadas. O J48, um indutor de árvores de classificação reimplementado no Weka, é baseado no algoritmo C4.5. Ele utiliza a entropia de Shannon para selecionar a melhor partição dos nós e como critério de parada (WITTEN et al, 2016)..

O algoritmo J48 realiza uma fase de pós-poda para converter sub-árvores sem ganhos de informação em folhas. Ele pode lidar com diferentes tipos de atributos e valores faltantes, e é capaz de classificar dados com precisão, gerando regras que representam a identidade dos dados. O objetivo é generalizar progressivamente a árvore de decisão para alcançar um equilíbrio entre flexibilidade e precisão (WITTEN et al, 2016).

2.2 Rede Neural

As Redes Neurais Artificiais são sistemas capazes de extrair informações não explicitamente fornecidas, realizando interpolação dos resultados. Inspiradas no cérebro humano, estas redes consistem em neurônios organizados em camadas interligadas por conexões sinápticas que possuem pesos associados, os quais armazenam conhecimento (GOODFELOW, BENGIO, COURVILLE, 2016).

Entre os diversos modelos de Redes Neurais, destacam-se as redes de múltiplas camadas (*MultiLayerPerceptron* - MLP), que possuem uma ou mais camadas intermediárias entre as camadas de entrada e de saída, responsáveis por processar os dados e gerar os resultados (GOODFELOW, BENGIO, COURVILLE, 2016).

As MLPs são treinadas utilizando o algoritmo de Retropropagação (*Backpropagation*), no qual cada camada desempenha uma função específica. A camada de saída recebe os estímulos da camada intermediária e produz a resposta final, enquanto as camadas intermediárias atuam como extratoras de características, codificando padrões de entrada e permitindo que a rede crie sua própria representação do problema (GOODFELOW, BENGIO, COURVILLE, 2016)..

O *Backpropagation* é um método comum de treinamento em redes neurais, pertencente à categoria de aprendizagem supervisionada, na qual os resultados desejados são conhecidos previamente e o objetivo do treinamento é fazer com que a rede aprenda a produzi-los (GOODFELOW, BENGIO, COURVILLE, 2016).

2.2 Materiais e Métodos

Os dados dos pacientes diagnosticados com câncer de mama no Hospital Bom Pastor (HBP) foram utilizados para análise. O banco de dados consiste em 17 atributos, abrangendo informações clínicas, demográficas e sociais, como faixa etária, cor/raça, nível de escolaridade, estágio do câncer, presença de metástases, entre outros. A classificação foi realizada para determinar se as pacientes estavam vivas ou mortas.

A metodologia empregada envolveu a pesquisa e avaliação do Registro Hospitalar de Câncer do HBP para uso no processo de KDD. Os dados foram pré-processados para aumentar ou reduzir seu nível de abstração, seguido pela adequação

para utilização. Dois modelos foram utilizados para representar o conhecimento gerado, sendo avaliados para determinar o mais adequado.

O banco de dados do HBP inclui 1666 casos de câncer de mama, diagnosticados entre 1989 e 2010, com a maioria dos pacientes sendo mulheres (98,6%). A amostra final consistiu em 872 registros de pacientes que faleceram de câncer de mama.

Para a análise, foram eliminados registros com informações incorretas ou ambíguas, resultando em uma amostra de 31 pacientes. A seleção de atributos foi realizada, excluindo aqueles que já continham informações classificatórias ou únicas para cada paciente. A transformação dos dados envolveu a representação de atributos contínuos e categóricos.

3. CONSIDERAÇÕES FINAIS

Os dados foram particionados em conjuntos de treinamento e teste, com 70% dos dados destinados ao treinamento e 30% aos testes. O algoritmo foi inicialmente treinado com uma parcela maior dos dados de treinamento e posteriormente avaliado com os dados de teste para validar seu desempenho. Somente após essa validação, o algoritmo foi colocado em produção, com a confiança de que ele é capaz de prever com precisão novos dados.

Os resultados apresentados na Tabela 1 referem-se ao conjunto de teste. Foram utilizados os classificadores J48, que corresponde ao C45, redes neurais e regras JRip

No contexto das tarefas de classificação, a acurácia das instâncias corretamente classificadas é tão relevante quanto o coeficiente de Kappa, que indica o grau de concordância nas classificações. O Coeficiente Kappa é uma medida de associação usada para avaliar a confiabilidade e precisão na classificação, fornecendo uma indicação de quanto as observações divergem das esperadas.

Comparando os resultados dos dois métodos de classificação - árvore de decisão (J48) e rede neural - observa-se que a rede neural apresenta melhores indicadores, com 76,26% de instâncias corretamente classificadas e um coeficiente de Kappa de 48%. No entanto, ambos os métodos revelam uma baixa concordância entre as classificações.

Os resultados derivados da Árvore de Decisão no conjunto de dados estão expostos na Tabela 2.

A Taxa de Verdadeiros Positivos (TP) representa as instâncias corretamente classificadas em uma classe específica. No contexto deste estudo, 91% das pacientes foram corretamente classificadas como vivas, enquanto apenas 40% foram acertadamente identificadas como mortas.

Por outro lado, a Taxa de Falsos Positivos (FP) denota os casos erroneamente classificados como pertencentes a uma determinada classe. Neste estudo, 60% das pacientes foram incorretamente classificadas como vivas, enquanto apenas 9% foram indevidamente consideradas como mortas

.

Tabela 1 – Resultados Conjunto de Teste

	Árvore de decisão	Rede Neural

	n.	%	n.	%
Instâncias classificadas corretamente	442	72,5	665	76,26
Instâncias classificadas incorretamente	168	27,05	207	23.74
Kappa estatística	-	0,34	-	0,48
Erro absoluto médio	-	0,36	-	0,24
Erro quad. médio da raiz	-	0,43	-	0,45
Erro abs. relativo	-	79,87	-	53,54
Erro quad. relativo da raiz	-	90,81	-	94,91

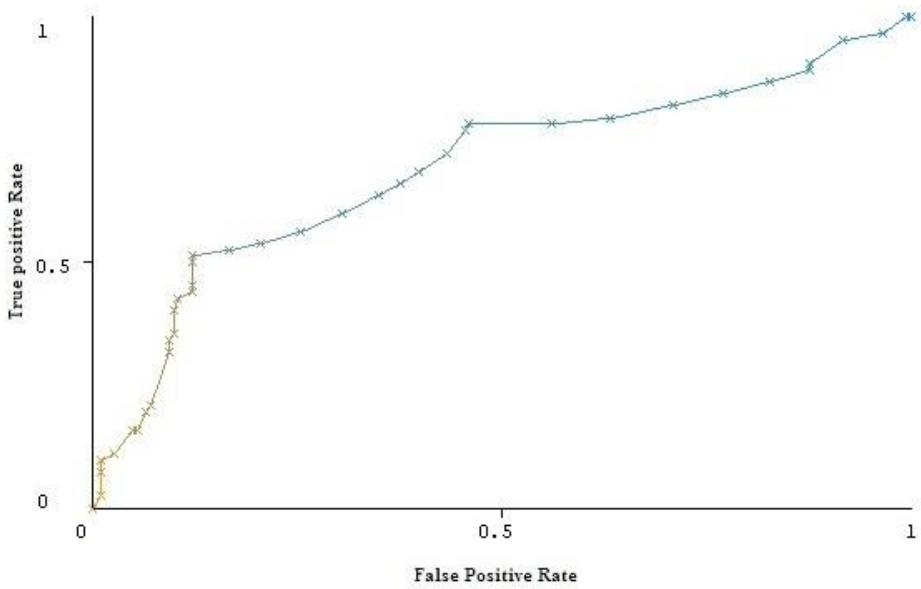
Fonte: Registro Hospitalar Hospital Bom Pastor

Tabela 2 – Resultados Árvore de Decisão e Rede Neural

	Árvore de Decisão			Rede Neural		
	Sim	Não	Méd. Pond.	Sim	Não	Méd. Pond.
TP Rate	0,40	0,91	0,73	0,66	0,82	0,76
FP Rate	0,09	0,60	0,42	0,18	0,34	0,28
Precisão	0,71	0,73	0,72	0,67	0,81	0,76
Recall	0,40	0,91	0,70	0,66	0,82	0,76
F-Measure	0,51	0,81	0,70	0,66	0,82	0,76
MCC	0,37	0,37	0,37	0,48	0,48	0,48
ROC Area	0,69	0,69	0,69	0,82	0,82	0,82
PRC Area	0,59	0,76	0,70	0,72	0,87	

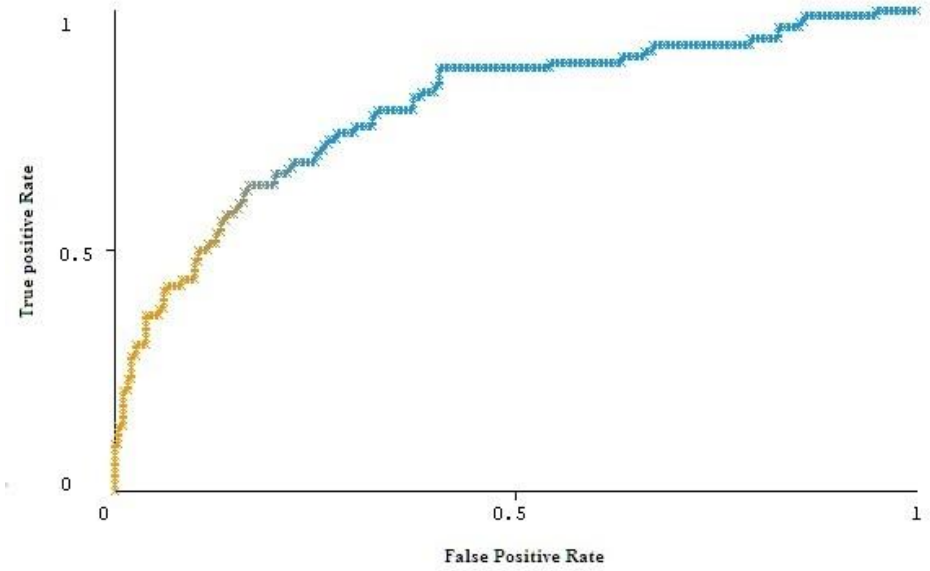
Fonte: Registro Hospitalar Hospital Bom Pastor

Gráfico 1 – Árvore de Decisão – Curva Roc



Fonte: Fonte: Registro Hospitalar Hospital Bom Pastor

Gráfico 2 – Rede Neural – Curva Roc



Fonte: Fonte: Registro Hospitalar Hospital Bom Pastor

2.1 Árvore de Decisão

A precisão é a proporção de instâncias corretamente classificadas em relação ao total de instâncias classificadas naquela classe. Para a classe "sim" (pacientes falecidos), a precisão foi de 71%, enquanto para a classe "não" (pacientes vivos), foi de 73%.

A sensibilidade (*recall*) é a proporção de instâncias corretamente classificadas como pertencentes a uma determinada classe em relação

ao total de instâncias verdadeiramente pertencentes a essa classe. A sensibilidade para a classe "sim" (pacientes falecidos) foi muito baixa, indicando um alto número de falsos negativos. Consequentemente, utilizando o classificador de árvore de decisão, o modelo apresentou baixa sensibilidade para a classe "sim" e alta sensibilidade para a classe "não".

Inferimos que pacientes em estágios iniciais do câncer de mama não vieram a óbito até o último acompanhamento. Por outro lado, pacientes com metástase e estágios 3A, 3B e 3C apresentaram menos de 2 anos de sobrevida. A presença de metástase esteve associada a um óbito em até 5 anos de sobrevida, enquanto pacientes sem metástase não faleceram no mesmo período. A sobrevida também se correlacionou com a graduação histológica do tumor, onde graus mais elevados de indiferenciação histológica estiveram relacionados a um pior prognóstico e menor sobrevida. Dessa forma, os pacientes nos estágios I e II apresentaram a maior sobrevida (BATSCHAUER, 2009); (GUERRA, 2007); (SCHNEIDER, 2009) e (ROSA, 2012).

Verificou-se uma redução na probabilidade de sobrevida à medida que o estadiamento do câncer de mama avançava (HOFELMANN, 2014).

2.2 Rede Neural

Os resultados encontrados através da Rede neural no conjunto de dados estão apresentados na tabela 2.

A Taxa de Verdadeiros Positivos (TP) refere-se às instâncias corretamente classificadas em uma classe específica. Neste estudo, observou-se que 82% das pacientes foram devidamente classificadas como vivas, enquanto 66% foram acertadamente identificadas como mortas.

A Taxa de Falsos Positivos (FP), por sua vez, indica os casos erroneamente classificados como pertencentes a uma determinada classe. No contexto deste estudo, constatou-se que 34% das pacientes foram incorretamente classificadas como vivas, e 18% foram indevidamente consideradas como mortas.

A precisão representa a proporção de instâncias corretamente classificadas em relação ao total de instâncias classificadas naquela classe específica. Para a classe "sim" (pacientes que faleceram), a precisão foi de 67%, enquanto para a classe "não" (pacientes que sobreviveram), foi de 81%.

A sensibilidade (*recall*) denota a proporção de instâncias corretamente classificadas como pertencentes a uma determinada classe em relação ao total de instâncias verdadeiramente pertencentes a essa classe. Observou-se que a sensibilidade para a classe "sim" (pacientes falecidos) foi de 66%, enquanto para a classe "não" (pacientes vivos) foi de 82%.

2.3 Curva ROC

O diagrama ROC (*Receiver Operating Characteristic*) constitui um método gráfico empregado na avaliação, organização e seleção de sistemas de diagnóstico e/ou prognóstico. Notavelmente, a curva ROC emerge como uma ferramenta poderosa para a análise de modelos de classificação, especialmente em cenários onde ocorre uma significativa disparidade entre as classes, ou quando se faz necessário ponderar diferentes custos e benefícios associados aos

distintos erros e acertos de classificação (FAWCETT, 206).

Essas curvas oferecem uma métrica quantitativa da precisão de um classificador, sendo esta diretamente proporcional à área sob a curva ROC. Quanto mais próxima a curva estiver do canto superior esquerdo do diagrama, maior será a área sob a curva e, por conseguinte, maior será a precisão do classificador.

No presente estudo, a área sob a curva ROC foi de 0,69 para o modelo baseado em árvore de decisão. Portanto, constata-se que este modelo de classificação não apresentou uma eficácia tão expressiva em comparação com a rede neural, a qual obteve uma área sob a curva ROC de 0,82.

Utilizamos o algoritmo JRip (*Repeated Incremental Pruning to Produce Error Reduction*) para a extração de regras de classificação (WITTEN, 2016).

Este algoritmo, fundamentado em regras de decisão do tipo "Se" e "Então", opera em duas fases distintas. Na primeira fase, são geradas um conjunto inicial de regras para classificação. Na segunda etapa, ocorre a otimização desse conjunto inicial, visando reduzir erros e tornar o processo de classificação mais seletivo. Ambas as fases são repetidas iterativamente para refinar o conjunto de regras.

O conjunto resultante é composto por 5 regras, conforme apresentado abaixo:

- R1: (Metástase = Sim) e (Estadiamento = 3B) => Óbito = Sim
- R2: (Sobrevida > 5) => Óbito = Sim
- R3: (Sobrevida < 1) => Óbito = Sim
- R4: (Metástase = Sim) e (Raça/Cor = Preta) => Óbito = Sim
- R5: (Sobrevida = 5) e (Faixa-etária = 50-59) => Óbito = Sim

A regra R2 apresenta a melhor precisão, alcançando 85\%, com uma cobertura de 27\%. Indica que pacientes com sobrevida superior a cinco anos vieram a óbito. Contudo, não encontramos associação na literatura especializada entre uma maior sobrevida e um aumento no risco de morte.

Por outro lado, a regra R1, com precisão de 48\% e cobertura de 15\%, revela que pacientes com metástase e estadiamento 3B faleceram. Esse resultado está em consonância com estudos anteriores, os quais indicam que mulheres com estadiamento avançado e doença metastática apresentam maior risco de morte (BRITO, 2004) (FAYER, 2016) e (BHOO-PATHY, 2012).

A regra R3, apesar de sua precisão de 47\%, possui uma cobertura de apenas 5\%, indicando que pacientes com menor sobrevida vieram a óbito. Contudo, esta regra é uma tautologia, pois pacientes com menor sobrevida inevitavelmente faleceram.

A regra R4, com precisão de 17\% e cobertura de 6\%, associa mulheres entre 50 e 59 anos com uma sobrevida de 5 anos ao óbito. Entretanto, a relação entre idade e prognóstico é complexa, com fatores como a falta de consenso sobre definições de juventude e velhice em relação à idade cronológica (FAYER, 2016) Mulheres com idade inferior a 35 anos e acima de 70 anos apresentam pior sobrevivência, influenciada por fatores biológicos e comorbidades (SILVA, 2011).

A regra R5, com precisão de 9\% e cobertura de 2\%, sugere que mulheres negras com metástase faleceram. Estudos realizados nos EUA indicam que mulheres negras têm maior probabilidade de

morrer de câncer de mama, mesmo com tumores de tamanho pequeno (JAVAID, 2015).

2.4 Conclusão

A Fundação Hospitalar do Município de Varginha (Hospital Bom Pastor) tem desempenhado um papel crucial como centro de referência para o tratamento de neoplasias malignas na região Sul de Minas Gerais. Neste estudo, empregamos os dados do Registro Hospitalar de Câncer do referido hospital, cuja relevância e contribuição para a pesquisa acadêmica são incontestáveis.

O objetivo primordial deste trabalho foi empregar algoritmos de indução e regras de decisão para a representação do conhecimento relacionado ao prognóstico de pacientes diagnosticadas com neoplasia de mama. A análise do prognóstico visa aprofundar a compreensão do perfil do paciente, sua história clínica e os fatores associados a um prognóstico favorável.

Embora a Árvore de Decisão não se destaque como o método de classificação mais eficaz quando comparada à Rede Neural, suas vantagens são notáveis. A principal delas reside na capacidade de tomar decisões baseadas nos atributos mais relevantes, sendo ainda compreensível para a maioria dos usuários e visualmente didática. Ao ordenar e apresentar os atributos por ordem de importância, as Árvores de Decisão possibilitam aos usuários identificar quais fatores exercem maior influência em suas decisões.

Quanto às regras de decisão, algumas se mostraram relevantes quando comparadas à literatura especializada, corroborando a importância do diagnóstico precoce na detecção do câncer em estágios iniciais, o que, por sua vez, favorece melhores respostas ao tratamento e contribui para um aumento na sobrevida.

Os resultados deste estudo estão alinhados com os achados da literatura, evidenciando que o estadiamento do câncer de mama figura como uma variável crucial que explica as disparidades na sobrevivência entre as mulheres afetadas por essa neoplasia.

REFERÊNCIAS

Batschauer, A. P. B. Avaliação hemostática e molecular em mulheres com câncer de mama receptor hormonal negativo. 2009. Tese (Doutorado em Ciências Farmacêuticas). Faculdade de Farmácia, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

Bhoo-Pathy N, Hartman M, Cheng-Har Y, Saxena N, Taib NA, Siew-Eng L, et al. Ethnic Differences in Survival after Breast Cancer in South East Asia. PLoS One 2012; v.7, n.2, pp. 1-6. DOI: 10.1371/journal.pone.0030995

Brito, C. Avaliação do tratamento a paciente com câncer de cama nas unidades oncológicas do Sistema Único de Saúde no Estado do Rio de Janeiro [dissertação]. Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz; 2004. 141p

Fawcett T., An introduction to ROC analysis, Pattern, *Recognition Letters*, vol. 27, n.8, 2006, pp. 861-874. Disponível em: https://doi.org/10.1016/j.patrec.2005.10.010. Acesso: 10 de maio de 2024.

Fayer, Vivian Assis et al. Sobrevida de dez anos e fatores prognósticos para o câncer de mama na região Sudeste do Brasil. Rev. Bras. epidemiol., São Paulo , v. 19, n. 4, p. 766-778, Dec. 2016. Disponível em: https://doi.org/10.1590/1980-5497201600040007. Acesso: 01 de maio de 2024.

Globocan 2022: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012, 2022. Disponível em: https://gco.iarc.who.in. Acesso em: 09 de maio de 2024.

Goodfelow I, Bengio Y., Courville A. Deep Learning. Cambridge: The Mit Press, 2016. 775 p

Guerra, M. R. Sobrevida e fatores prognósticos para o câncer de mama em Juiz de Fora, Minas Gerais, na coorte diagnosticada entre 1998 e 2000. Tese (Doutorado em Saúde Coletiva). Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2007

Hofelmann, D. A.; Anjos, J. C; Ayala, A. L. Sobrevida em dez anos e fatores prognósticos em mulheres com

câncer de mama em Joinville, Santa Catarina, Brasil. Cienc. Saúde coletiva, Rio de Janeiro, v. 19, n. 6, p. 1813-1824, Jun 2014. Disponível em: https://doi.org/10.1590/1413-81232014196.03062013. Acesso: 01 de maio de 2024.

Hospital Bom Pastor. Disponível em <http://www.fhomuv.com.br/>. Acesso em: 10 fev. 2024.

Instituto Nacional de Câncer. Dados e números sobre câncer de mama - Relatório anual 2023. Instituto Nacional de Câncer José Alencar Gomes da Silva – Rio de Janeiro: INCA, 2023.

Javaid Iqbal, MD et al. Differences in Breast Cancer Stage at Diagnosis and Cancer-Specific Survival by Race and Ethnicity in the United States.JAMA. 2015; 313(2):165-173

Jbilou J. et al. Medical genetic counseling for breast cancer in primary care: a synthesis of major determinants of physicians' practices in primary care settings. Public health genomics, v. 17, n. 4, p. 190-208, 2013. DOI: 10.1159/000362358

Quinlan, J.R.C4.5 Programs for Machine Learning, San Mateo, CA: Morgan Kaufmann, 1992.

Rosa, Luciana Martins da; RADUNZ, Vera. Taxa de sobrevida na mulher com câncer de mama: estudo de revisão. Texto contexto - enferma, Florianópolis, v. 21, n. 4, p. 980-989, Dec. 2012. Disponível em: https://doi.org/10.1590/S0104-07072012000400031 Acesso: 01 de maio de 2024.

Steiner, M. T. A., Névoa, J. C., Soma, N. Y., Shimizu, T., Steiner Neto, P. J. Extração de regras de classificação a partir de redes neurais para auxílio a tomada de decisão na concessão de crédito bancário. Pesquisa Operacional, v. 27, n. 3, p. 407-426, 2007. Disponível em: https://doi.org/10.1590/S0101-74382007000300002. Acesso: 01 de maio de 2024.

Schneider, Ione Gaye Ceola; D'ORSI, Eleonora. Sobrevida em cinco anos e fatores prognósticos em mulheres com câncer de mama em Santa Catarina, Brasil. Cad. Saúde pública, Rio de Janeiro, v. 25, n. 6, p. 1285-1296, jun. 2009. Disponível em: Sobrevida em cinco anos e fatores prognósticos em mulheres com câncer de mama em Santa Catarina, Brasil. Acesso: 01 de maio de 2024.

Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. Stat Med 2000; 19: 541-61

Silva, Gulnar Azevedo e et al. Cancer mortality trends in Brazilian state capitals and other municipalities between 1980 and 2006. Rev. Saúde pública, São Paulo, v. 45, n. 6, p. 1009-1018, Dec. 2011.

Witten I, Eibe E., Hall M., Pal C.J. Data Mining: Practical Machine Learning Tools and Technique. Burlington: Morgan Kaufmann, 2016. 664 p.