Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Aprendizado de Máquina na Detecção de Câncer de Pâncreas

Beatriz Cogo Debs [1], Angela Leite Moreno [2]
UNIFAL-MG, Alfenas, MG
Reginaldo José da Silva [3]
FEIS/UNESP, Ilha Solteira, SP

O câncer de pâncreas (CP) é considerado o 3^o no índice de mortalidade por câncer, possuindo a menor taxa de sobrevivência [6]. Advindo da porção exócrina do pâncreas, a neoplasia apresenta sintomas comuns a outras doenças, como perda de peso, dor e incapacidade, o que torna o diagnóstico da doença mais difícil de ser detectado [4]. Embora haja tratamento, estes não são eficazes o suficiente para gerar sobrevida aos pacientes, principalmente devido ao diagnóstico tardio.

Existem alguns biomarcadores importantes como o CA19-9 e o REG1B para a detecção de CP. O CA19-9 é considerado o biomarcador de ouro para a doença, enquanto o indicador REG1B (Litostatina-1-beta) auxilia o CA19-9 a discriminar pacientes doentes e saudáveis, reduzindo o índice de falsos positivos e falsos negativos [2]. Técnicas de *machine learning* vêm sendo utilizadas como forma de auxiliar os médicos na tomada de decisão e detecção de doenças [5]. No presente trabalho, são apresentados os resultados utilizando o algoritmo *Random Forest* e a rede neural Euclidiana Autoexpansível baseada na Teoria da Ressonância Adaptativa (EAEART) [3], para classificação de pacientes com CP.

Para a realização da pesquisa utilizou-se o banco de dados [1]. Os dados coletados foram idade, sexo, diagnóstico da doença (1 = Controle, 2 = Benigno e 3 = Câncer), assim como resultados do teste ELISA para o Plasma CA19-9, Creatinina, e os bioindicadores receptor de hialuronano endotelial de vaso linfático 1 (LYVE1), Litostatina-1-beta (REG1B), Trefoil 1 (TFF1) e REG1A. Após análise os atributos "sexo" e REG1A foram excluídos das simulações.

Tabela 1: Resultados Obtidos.

		Sem CA19-9				Com CA19-9			
	Métricas	**Controle**	**Benigno**	**Câncer**	**Geral**	**Controle**	**Benigno**	**Câncer**	**Geral**
EAEART	ACU	–	–	–	0,4746	–	–	–	0,6476
	Se	0,5818	0,3871	0,4667	0,4785	0,6429	0,5938	0,6889	0,6418
	Sp	0,6311	0,6957	0,8889	0,7386	0,9351	0,6849	0,8500	0,8233
	F1	0,6055	0,4974	0,6120	0,5716	0,7619	0,6361	0,7610	0,7197
	AUC	–	–	–	0,6959	–	–	–	0,8046
RF	ACU	–	–	–	0,6441	–	–	–	0,7524
	Se	0,6182	0,4516	0,8667	0.6455	0,6071	0,7188	0,8667	0,7309
	Sp	0,8689	0,8174	0,7778	0.8213	0,9610	0,7671	0,9000	0,8761
	F1	0,7224	0,5818	0,8198	0.7080	0,7442	0,7421	0,8830	0,7898
	AUC	–	–	–	0,8348	–	–	–	0,8960

ACU: Acurácia, Se: Sensibilidade, Sp: Especificidade, AUC: AUC-ROC

[1] beatriz.debs@sou.unifal-mg.edu.br
[2] angela.moreno@unifal-mg.edu.br
[3] reginaldo.silva@unesp.br

As simulações foram realizadas em duas etapas, a primeira **com** o atributo CA19-9 e a segunda **sem** o CA19-9. Essa separação foi realizada com o objetivo de determinar como o CA19-9 atua na classificação. Para este processo, utilizou-se a separação 70/30 com o processo de validação cruzada *k-fold*, com $k = 5$, nos 70%. Na busca dos melhores parâmetros para configuração dos algoritmos, os parâmetros encontrados para a ART Euclidiana (EAEART) foram: $\beta = 0,3499$ e $\rho = 0,003$. Os resultados identificados foram melhores avaliados no algoritmo treinado e validado *Random Forest* (RF). Na Tabela 1 são apresentados os resultados obtidos no conjunto de teste.

Como pode ser observado, ao se utilizar o atributo CA19-9 juntamente com os outros indicadores o valor de AUC para o modelo é mais alto e preciso. Isso deve-se ao fato de que os modelos de aprendizado de máquina conseguem extrair melhor a informação quando outros marcadores são utilizados para treinamento e validação da rede. Ademais, é possível afirmar que a sensibilidade e a especificidade da rede também são incrementadas quando o modelo é treinado não somente com o CA19-9, mas sim em conjunto com outros bioindicadores importantes para a doença.

Além disso, os resultados ao se utilizar o atributo CA19-9 fornecem resultados superiores em relação às simulações que não o utilizam, como é possível examinar ao se analisar a AUC obtida em cada situação. O *Random Forest* obteve os melhores resultados para o problema, em relação a todas as métricas utilizadas. Outro ponto que merece destaque é que o uso do biomarcador favorece a identificação de pacientes doentes, influenciando tanto no grupo benigno quanto com câncer. Já o não uso deste biomarcador influencia a identificação de pacientes saudáveis. Desta forma, é permitido concluir que o biomarcador CA19-9 possui alta relevância na separação de classes, o que possui alta relevância no classificador do modelo.

Como trabalhos futuros, propõe-se o tratamento do problema de três classes para um problema binário, verificando como os algoritmos se comportam. Além de testar outros algoritmos de aprendizado de máquina.

Agradecimentos

Agradecemos ao apoio financeiro da FAPEMIG e UNIFAL-MG.

Referências

[1] S. Debernardi et al. “A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case–control study”. Em: **PLoS Medicine** 17.12 (2020), e1003489.

[2] S. Makawita et al. “Validation of four candidate pancreatic cancer serological biomarkers that improve the performance of CA19. 9”. Em: **BMC cancer** 13.1 (2013), pp. 1–11.

[3] A. L. Moreno. **Redes Neurais ART e ARTMAP com Treinamento Continuado**. Saarbrucken: Novas Edições acadêmicas, 2016.

[4] L. K. Ruth e W. Declan. “Symptoms of pancreatic cancer”. Em: **Journal of Pain and Symptom Management** 6.6 (1991), pp. 360–367. ISSN: 0885-3924. DOI: https://doi.org/10.1016/0885-3924(91)90027-2. URL: https://www.sciencedirect.com/science/article/pii/0885392491900272.

[5] C. Sammut e G. I. Webb. **Encyclopedia of machine learning**. Springer Science & Business Media, 2011.

[6] R. L. Siegel, K. D. Miller e A. Jemal. “Cancer statistics, 2019”. Em: **CA: a cancer journal for clinicians** 69.1 (2019), pp. 7–34.