

Técnicas de Inteligência Computacional Como Mecanismo Auxiliador na Triagem da Possibilidade de Ataque Cardíaco em Adultos

Roney Nogueira de Sousa¹, Rhyan Ximenes de Brito¹

¹Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Av. Tabeião Luiz Nogueira de Lima S/N – Tianguá – CE – Brasil

{nogueiraroney453, rxbrito}@gmail.com

Abstract. *Chronic non-communicable diseases are responsible for approximately 40 million deaths annually in the world. Based on this premise, this study aimed to verify the possibility of the presence or absence of heart attack, in order to assist in the classification of these individuals. The methodology used in this article was implemented based on 303 samples of individuals from a public database, with the application of the cross-validation techniques k-fold, z-score and the SMOTE technique. The results showed an average rate of 94.26% for RF accuracy, for SVM 90.65%, RBF 83.48% and 86.18% for MLP.*

Resumo. *As doenças crônicas não transmissíveis são responsáveis por aproximadamente 40 milhões de óbitos anuais no mundo. Com base nessa premissa, este trabalho teve como objetivo verificar a possibilidade da presença ou ausência de ataque cardíaco, de forma a auxiliar na classificação desse desses indivíduos. A metodologia trabalhada nesse artigo foi implementada com base em 303 amostras de indivíduos de um banco de dados público, com a aplicação das técnicas validação cruzada k-fold, z-score e a técnica SMOTE. Os resultados mostraram para a acurácia do RF uma taxa média de 94,26%, para a SVM 90,65%, RBF 83,48% e 86,18% para a MLP.*

1. Introdução

É estimado que 17,9 milhões de pessoas morreram de doenças relacionadas a problemas cardíacos em 2016, representando 31% de todas as mortes globais [Organization, W.H 2017]. Dentro deste grupo de causas, as maiores responsáveis por óbitos no Brasil em indivíduos acima de 40 anos de idade são as doenças isquêmicas do coração, responsáveis por 29,7% dos óbitos entre 2008 e 2012 [Souza et al. 2017].

Com base nessa premissa, este trabalho teve como objetivo principal verificar a possibilidade da presença ou ausência de ataque cardíaco em indivíduos adultos de forma a auxiliar no processo de classificação. Com base na implementação e análise dos classificadores RBF (*Radial Basis Function*), MLP (*Multilayer Perceptron*), RF (*Random Forest*) e SVM (*Support Vector Machine*), comparando os resultados baseados em procedimentos quantitativos, com ênfase no treinamento e teste.

A metodologia trabalhada nesse artigo foi desenvolvida tendo como base um grupo de indivíduos do banco de dados público “*Heart Disease Data Set*” obtido através do link, <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.

A motivação para este trabalho está relacionada à importância da aplicação de técnicas de Inteligência Artificial em situações que possam auxiliar profissionais da área de saúde dando suporte de forma direta ou indireta no diagnóstico de problemas cardíacos.

Este trabalho está dividido em seis seções: Seção 2 descreve alguns trabalhos relacionados, Seção 3 traz a fundamentação teórica sobre as doenças cardiovasculares. A Seção 4 apresenta os principais conceitos sobre os classificadores utilizados, na Seção 5 é descrita a metodologia. A Seção 6 apresenta os resultados e discussões e no fim a Seção 7 aborda as considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

Esta seção apresenta uma revisão bibliográfica de forma reflexiva sobre diferentes abordagens sobre o uso da Inteligência Computacional (IC) aplicada como ferramenta de auxílio no diagnóstico de doenças.

[Ambrósio 2002] desenvolveu um sistema computacional baseado na técnica de uma rede neural do tipo *Perceptron* Multicamadas, para auxílio ao médico radiologista na confirmação de diagnóstico das chamadas lesões intersticiais pulmonares. Os dados para treinamento da rede neural foram obtidos por protocolo gerado por especialistas, através da revisão de casos com diagnóstico confirmado para determinadas patologias.

[Santos 2020] utilizou uma abordagem embasada em Máquina de Vetores de Suporte conhecidas como *Support Vector Machine* (SVM) para prever óbitos por ataques cardíacos, utilizando um *dataset* público intitulado como “*Echocardiogram Data Set*” a qual possuía 140 amostras. Através da ferramenta *WEKA* foi realizada a classificação das amostras chegando a uma acurácia de 95%.

[Pereira and Silveira 2015] utilizaram um sistema *fuzzy* a qual considerava como entradas alguns fatores de risco definidos pelo *National Cholesterol Education Program*, tais como a idade do paciente, hipertensão, diabetes e outros. Aplicaram o método de inferência *Mamdani* e a *defuzzificação* realizada foi via Método do Centro de Gravidade. Os resultados das simulações mostraram o grau de pertinência de um paciente com determinado quadro clínico, aos conjuntos *fuzzy* associados ao risco de ataque cardíaco.

[Saheki 2005] construiu um sistema especialista aplicado ao diagnóstico de doenças cardíacas, usando como ferramenta computacional redes *Bayesianas*. Os processos de modelagem e avaliação foram realizados com o auxílio de um especialista médico e de dados bibliográficos.

3. Doenças Cardíacas

As doenças cardiovasculares são as principais causas de mortes em todo o mundo anualmente se comparada a qualquer outra causa mortis [Organization, W.H 2017]. Pode-se afirmar que as principais causas de óbito por doenças do aparelho circulatório são as doenças isquêmicas do coração, as doenças cerebrovasculares e as doenças hipertensivas. Vale ressaltar que essas causas são evitáveis em sua maioria, pois há probabilidade de diminuição da ocorrência dessas mortes, se houver assistência ou prevenção oportunas [Ishitani et al. 2006].

Já o infarto agudo do miocárdio ou ataque cardíaco é a morte das células de uma região do músculo do coração por conta da formação de um coágulo que interrompe o

fluxo sanguíneo de forma súbita e intensa [Meneses et al. 2020]. Um ataque cardíaco pode acontecer quando plaquetas sanguíneas e outras substâncias formam um coágulo sanguíneo que bloqueia a maior parte ou todo o fluxo de sangue oxigenado para parte do músculo cardíaco.

Ao sofrer um ataque do cardíaco, o sujeito sente uma dor intensa no peito que pode irradiar pelos ombros e braços, além de empalidecer, sentir frio, náuseas e começar a suar. Com a redução da eficiência cardíaca há uma queda da pressão sanguínea que ocasiona tonturas [da Rocha 2017]. De acordo com [Organization, W.H 2017] cerca de 85% de todas as mortes causadas por doenças cardiovasculares são devido a ataques cardíacos e derrames. Os fatores que predis põem o ataque cardíaco estão relacionados à idade, colesterol alto, diabetes, tabagismo, obesidade e fatores hereditários [Meneses et al. 2020]

4. Classificadores

Esta seção apresenta os classificadores utilizados no trabalho como mecanismos de apoio a aprendizagem de máquina (AM).

4.1. *Multilayer Perceptron (MLP)*

[Bonifácio 2010] afirma que uma rede MLP consiste em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. É do tipo *feedforward*, ou seja, nenhuma saída de um neurônio de uma camada k será sinal de entrada para um neurônio de uma camada menor ou igual a k , e é completamente conectada, tal que cada neurônio fornece sua saída para cada unidade da camada seguinte.

4.2. *Support Vector Machine (SVM)*

O SVM baseia-se nos princípios da minimização do risco estrutural, proveniente da teoria do aprendizado estatístico, a qual está baseada no fato de que o erro do algoritmo de aprendizagem junto aos dados de validação é limitado pelo erro de treinamento mais um termo que depende da dimensão Vapnik e Chervonenkis [Semolini et al. 2002]. Vale destacar que as SVM foram primeiramente desenvolvidas para resolverem problemas de classificação, estendendo-se em seguida para problemas de regressão [Santos 2019].

4.3. *Random Forest (RF)*

O algoritmo RF consiste em uma técnica de aprendizado de máquina que agrupa diversas árvores de decisão criadas a partir de uma base de treinamento, de modo que o modelo resultante consolida os resultados de todas as árvores para predições [da Costa Filho et al. 2019]. A criação das árvores é realizada por meio de seleção de atributos aleatória dos subconjuntos, sendo que a RF é formada por uma coleção de árvores de decisão [Silva et al. 2019].

4.4. *Radial Basis Function (RBF)*

A Rede RBF é similar à rede MLP pois também é constituída de várias camadas que utiliza aprendizado supervisionado. Sua principal característica é a utilização de funções de base radial em todos os nós da camada oculta que ao invés de utilizar como argumento de função o produto escalar entre os valores do registro de entrada e os valores do registro de pesos do neurônio, utiliza a distância entre os valores do registro de entrada e o seu centro [Bonifácio 2010].

5. Metodologia

Para o emprego dos algoritmos utilizou-se da linguagem *Python* em sua versão 3.7 e um banco de dados público intitulado *Heart Disease Data Set* obtido através do link: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.

O dataset é composto por 303 amostras e 75 atributos, que foram avaliadas com relação aos atributos: idade, sexo, tipo de dor no peito, colesterol sérico em mg/dl, concentração de açúcar no sangue, resultados eletrocardiográficos em repouso, frequência cardíaca máxima atingida, angina induzida por exercício, entre outras.

Ressalta-se ainda que foram adicionadas a base de dados 250 amostras sintéticas através do *SMOTE*, porém após o pré-processamento dos dados resultaram das 553, apenas 485 amostras. Quanto aos dados faltosos, ambos foram removidos, utilizou-se a normalização (*z-score*), quanto a variável desfecho (classe) a mesma foi trabalhada com relação à presença de doença cardíaca no paciente.

As classes foram codificadas com valores numéricos de 0 a 4 onde se concentraram em simplesmente tentar distinguir a presença (valores 1,2,3,4) da ausência (valor 0). Para os classificadores foram testadas vários hiperparâmetros de forma a ajustá-los através da estratégia *Randomized Search* para adquirir-se as melhores acurácias, sensibilidade e especificidade.

5.1. Treinamento e Teste com a Base de Dados

Para o treinamento e teste com o classificador MLP observou-se que com uma camada oculta com 10 neurônios, função de ativação tangente hiperbólica, *learning rate* de 0,2 e o algoritmo de treinamento *Resilient backpropagation* obteve-se os melhores resultados. Para o SVM obteve-se as melhores taxas de acurácias com o *kernel* polinomial de grau 5 juntamente com a constante de relaxamento de valor 2. Quanto ao algoritmo RF foi utilizado o algoritmo de entropia composto por 450 árvores de decisão com 4 *seeds*. Por fim observou-se que classificador RBF com uma camada oculta com 10 neurônios obteve os melhores resultados.

A base de dados foi normalizada (*z-score*) e o balanceamento foi realizado através do método *SMOTE* garantindo assim a proporcionalidade entre as amostras utilizadas. Vale salientar que a geração dos dados sintéticos foi relacionado a classe minoritária a partir de vizinhos. Além disso foi utilizado o método de validação cruzada *k-fold*, com o $k=10$ *folds*.

6. Resultados e Discussões

Os resultados obtidos e analisados tiveram como base os percentuais de acurácia adquiridos durante o treinamento e teste realizado com os classificadores, assim como nos resultados das sensibilidade e especificidade.

Como observado na Tabela 1, o melhor caso do classificador MLP atingiu uma acurácia de 89,06% de acertos e 10,94% de erros e o pior caso 82,67% de acertos e 17,33% de erros. A taxa média de acertos foi de 86,18% de acertos e 13,82% de erros. Para a especificidade a MLP atingiu 97,45% e 2,55% de falso negativo.

Pode ser notado na Tabela 2 que o melhor caso do classificador SVM atingiu uma acurácia de 94,53% de acertos e 5,47% de erros e o pior caso 88,18% de acertos e

Tabela 1. Resultados MLP

<i>Folds</i>	Taxa de Acerto(%)	Taxa de erro(%)	Situação (Resultado)
1	89,06	10,94	Melhor Caso
2	87,50	12,50	
3	84,37	15,63	
4	89,06	10,94	Melhor Caso
5	82,67	17,33	Pior Caso
6	83,46	16,54	
7	87,40	12,60	
8	86,61	13,09	
9	85,82	14,18	
10	85,82	14,18	
total	86,18	13,82	Caso Médio

11,82% de erros. A taxa média de acertos foi de 90,65% de acertos e 9,35% de erros. A especificidade para a SVM atingiu 98,25% e 1,75% de falso negativo.

Tabela 2. Resultados SVM

<i>Folds</i>	Taxa de Acerto(%)	Taxa de erro(%)	Situação (Resultado)
1	91,40	8,60	
2	91,40	8,60	
3	89,06	10,94	
4	94,53	5,47	Melhor Caso
5	92,12	7,88	
6	88,18	11,82	Pior Caso
7	91,33	8,67	
8	91,33	8,67	
9	88,97	11,03	
10	88,18	11,82	Pior Caso
total	90,65	9,35	Caso Médio

Pode ser notado na Tabela 3 que o melhor caso do classificador RF atingiu uma acurácia de 97,65% de acertos e 2,35% de erros e o pior caso 89,76% de acertos e 10,24% de erros. A taxa média de acertos foi de 94,26% de acertos e 5,74% de erros. A especificidade atingida pela RF foi 98,76% e 1,24% de falso negativo.

A Tabela 4 apresenta o melhor caso do classificador RBF, a qual atingiu uma acurácia de 95,27% de acertos e 4,73% de erros e o pior caso 48,03% de acertos e 51,97% de erros. A taxa média de acertos foi de 83,48% de acertos e 16,52% de erros. Para a especificidade a RBF atingiu 95,25% e 4,75% de falso negativo.

Conforme pode ser observado o classificador RF em comparação aos demais obteve a melhor taxa média de acerto atingindo ganhos de 10,78% comparado ao classificador RBF, de 8,08% com relação ao MLP, 3,61% comparado ao SVM. Dessa forma o ganho mais significativo foi quando relacionado ao classificador RBF.

Porém quando analisado os ganhos relacionados a especificidade o classificador

Tabela 3. Resultados RF

<i>Folds</i>	Taxa de Acerto(%)	Taxa de erro(%)	Situação (Resultado)
1	96,09	3,91	
2	97,65	2,35	Melhor Caso
3	95,31	4,69	
4	95,31	4,69	
5	93,70	6,30	
6	92,91	7,09	
7	95,27	4,73	
8	93,70	6,30	
9	92,91	7,09	
10	89,76	10,24	Pior Caso
total	94,26	5,74	Caso Médio

Tabela 4. Resultados RBF

<i>Folds</i>	Taxa de Acerto(%)	Taxa de erro(%)	Situação (Resultado)
1	94,53	5,47	
2	92,96	7,04	
3	93,75	6,25	
4	94,53	5,47	
5	86,61	13,39	
6	48,03	51,70	Pior Caso
7	49,60	50,40	
8	95,27	4,73	Melhor Caso
9	90,55	9,45	
10	88,97	11,03	
total	83,48	16,52	Caso Médio

RF atingiu a maior taxa de acerto com 98,76%, representando um ganho de 3,51% se comparado ao classificador RBF, 0,51% comparado ao SVM e 1,31% quando comparado ao resultado da MLP.

7. Considerações Finais e Trabalhos Futuros

O artigo relatou a utilização de classificadores como ferramentas de aprendizagem de máquina supervisionada abordando doenças cardíacas como um problema de classificação de dados com base em um banco de dados composto por 485 amostras de indivíduos e na aplicação de técnicas como *z-score*, *SMOTE* e *k-fold*.

Foi constatado que o classificador *Random Forest* atingiu os melhores resultados com a média de 94,26% de acertos, assim como 98,76% de especificidade e 1,24% de taxa para o falso positivo enquanto o classificador com os piores resultados foi o RBF com 83,48% de taxa média, 95,25% de especificidade e 4,75% de falso positivo.

Como trabalhos futuros sugere-se a implementação da base de dados com a técnica de aprendizagem profundo (*deep learning*) analisando os resultados com vista na melhoria das acurácias alcançadas.

8. Agradecimentos

Os autores agradecem ao IFCE pelo apoio, através do projeto PEVPI.

Referências

- Ambrósio, P. E. (2002). *Redes neurais artificiais no apoio ao diagnóstico diferencial de lesões intersticiais pulmonares*. PhD thesis, Universidade de São Paulo.
- Bonifácio, F. N. (2010). Comparação entre as redes neurais artificiais mlp, rbf e lvq na classificação de dados. *Paraná: Universidade Estadual do Oeste do Paraná*.
- da Costa Filho, S. V. S., Arce, J. E., Montañó, R. A. N. R., and Pelissari, A. L. (2019). Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. *Ciência Florestal*, 29(4):1501–1515.
- da Rocha, A. C. S. (2017). Ataque cardíaco em jovens no Brasil: Conscientização e prevenção. *Universidade do Vale do Paraíba - UNIVAP. (Trabalho de Conclusão de Curso)*.
- Ishitani, L. H., Franco, G. d. C., Perpétuo, I. H. O., and França, E. (2006). Desigualdade social e mortalidade precoce por doenças cardiovasculares no brasil. *Revista de Saúde Pública*, 40:684–691.
- Meneses, L. S. L., Caxias, A. M., Franco, A. M., Dantas, A. S. F., de Oliveira, A. K. C., Leal, E. G., Lisboa, J. H. V., Pereira, K. A. C., Tavares, L. J. C., dias Santos, R., et al. (2020). Assistência de enfermagem ao paciente com infarto agudo do miocárdio em uma urgência e emergência: relato de experiência. *Brazilian Journal of Development*, 6(9):69116–69121.
- Organization, W.H (2017). Doenças cardiovasculares. [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). acessado em 23/03/2021.
- Pereira, N. K. M. and Silveira, G. P. (2015). Aplicação da lógica fuzzy no estudo do risco de ataque cardíaco. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 3(1).
- Saheki, A. H. (2005). *Construção de uma rede Bayesiana aplicada ao diagnóstico de doenças cardíacas*. PhD thesis, Universidade de São Paulo.
- Santos, E. J. R. d. (2019). Previsão de precipitação usando máquinas de vetores de suporte visando sua implementação em sistemas embarcados. *Universidade de Brasília*.
- Santos, J. P. M. S. (2020). Metodologia baseada em máquina de vetores de suporte para previsão de óbitos por ataque cardíaco (Monografia). *UNIFAP*.
- Semolini, R. et al. (2002). Support vector machines, inferência transdutiva e o problema de classificação.
- Silva, R. J. H., Pazoti, M. A., da Silva, F. A., Pereira, D. R., and de Almeida, L. L. (2019). Autenticação biométrica para sistemas por meio da dinâmica da digitação. In *Colloquium Exactarum*. ISSN: 2178-8332, volume 11, pages 26–33.

Souza, H. P., Werneck, G. L., de Medeiros, L. T., Sabroza, P. C., and dos Santos, J. P. C. (2017). Análise espacial de causas de morte por doenças cardiovasculares e fatores associados em uma área de implantação de um grande empreendimento. *Hygeia-Revista Brasileira de Geografia Médica e da Saúde*, 13(24):199–214.