

Aplicação de Aprendizado de Máquina no Auxílio ao Diagnóstico de Doenças Renais Crônicas

João Pedro Stone Moreira¹, Thaiane Kreps Schönhofen²,
Diego Duarte Bottero¹, Giancarlo Lucca¹, Lizandro de Souza Oliveira¹

¹Universidade Católica de Pelotas (UCPEL)
Centro de Ciências Sociais e Tecnológicas (CCST)
Mestrado em Engenharia Eletrônica e Computação (MEEC)

²Universidade Católica de Pelotas (UCPEL)
Centro de Ciências da Saúde (CCS)

{joao.moreira,thaiane.schonhofen,diego.bottero}@sou.ucpel.edu.br

{giancarlo.lucca,lizandro.oliveira}@ucpel.edu.br

Abstract. Artificial Intelligence (AI) has revolutionized several fields, offering efficient solutions to complex problems. In healthcare, AI techniques have been employed to improve diagnoses and treatments. This paper compares algorithms for predicting Chronic Kidney Disease using machine learning. The K-Nearest Neighbors, Decision Tree, Random Forest, and FURIA algorithms were applied to a Chronic Kidney Disease dataset from the UCI Machine repository after appropriate preprocessing. The Random Forest model performed best, achieving 99% accuracy.

Resumo. A Inteligência Artificial (IA) tem revolucionado diversas áreas, oferecendo soluções eficientes para problemas complexos. Na área da saúde, técnicas de IA vêm sendo empregadas para aprimorar diagnósticos e tratamentos. Este trabalho aborda uma comparação entre algoritmos na predição da Doença Renal Crônica por meio de aprendizado de máquina. Foram aplicados os algoritmos K-Nearest Neighbors, Decision Tree, Random Forest e FURIA a um conjunto de dados Chronic Kidney Disease do repositório UCI Machine após pré-processamento adequado. O modelo Random Forest apresentou o melhor desempenho, destacando-se pela acurácia de 99%.

1. Introdução

O Aprendizado de Máquina (AM) é um ramo de pesquisa da Inteligência Artificial que busca desenvolver algoritmos capazes de aprender padrões a partir de dados e realizar previsões ou decisões sem programação explícita para cada tarefa específica. Essa abordagem tem revolucionado a forma como sistemas computacionais lidam com grandes volumes de dados, tornando-se uma ferramenta essencial para resolver problemas complexos em diferentes domínios científicos e industriais [Kumar et al. 2025].

Nos últimos anos, o AM tem sido amplamente adotado em áreas como finanças [Ouro Preto Investimentos 2025], segurança da informação [CashU 2025], alimentação [Magalhães et al. 2021] e saúde [Inforchannel 2024]. No setor financeiro,

técnicas de AM são empregadas para análise de risco, detecção de fraudes e automação de negociações [PhoenixNAP 2024]. Em segurança, o AM fortalece sistemas de proteção ao identificar padrões de ameaças e comportamentos anômalos em tempo real [Saxena 2025].

A Doença Renal Crônica (*Chronic Kidney Disease - CKD*) compreende um conjunto de distúrbios heterogêneos que comprometem tanto a estrutura quanto a função dos rins. As manifestações clínicas podem variar de acordo com a etiologia, o grau de comprometimento funcional e a velocidade de progressão da doença [Levin et al. 2024].

Na área da saúde, o AM tem impulsionado avanços em diagnósticos, prognósticos e personalização de tratamentos, além de otimizar a análise de imagens médicas e dados clínicos [Acropodium 2025, Kumar et al. 2025]. Esses exemplos evidenciam o impacto transversal do AM na sociedade contemporânea.

Nesse contexto, o uso de AM tem se mostrado promissor no auxílio à triagem e ao diagnóstico da CKD, uma vez que essas abordagens permitem explorar grandes volumes de dados clínicos e identificar padrões relevantes, contribuindo para a tomada de decisão médica [Metherall et al. 2024]. Neste trabalho, utilizaremos a sigla em inglês 'CKD' para Doença Renal Crônica, conforme o padrão adotado pela comunidade científica internacional e diretrizes clínicas reconhecidas [Group 2013].

Este trabalho tem como objetivo principal realizar um estudo comparativo entre algoritmos de AM para identificar os melhores que se adequam à realidade clínica e então propor uma abordagem capaz de auxiliar no processo de predição da Doença Renal Crônica, contribuindo para a detecção precoce e o suporte à tomada de decisão clínica.

Este trabalho está organizado da seguinte forma. A Seção 2, apresenta a fundamentação teórica. Após, na Seção 3, os trabalhos relacionados são discutidos. A metodologia e a avaliação são postas na Seção 4 e 5 respectivamente. Por fim, na Seção 6 é apresentado as conclusões e potenciais desdobramentos.

2. Fundamentação Teórica

A fundamentação teórica do presente trabalho aborda os conceitos necessários para a sua compreensão.

2.1. Doença Renal Crônica

A CKD é uma condição caracterizada pela perda progressiva e irreversível da função dos rins ao longo do tempo, frequentemente silenciosa por meses ou anos. Diferencia-se da lesão renal aguda, que envolve redução súbita e muitas vezes reversível da função renal. A CKD resulta de danos estruturais ou funcionais persistentes nos rins, comprometendo sua capacidade de filtrar o sangue adequadamente. Além disso, condições como hipertensão e diabetes podem contribuir para o desenvolvimento da CKD, que também é conhecida como insuficiência renal crônica [Levin et al. 2024].

2.2. Aprendizado de Máquina

O AM permite que os sistemas aprendam com padrões de dados sem a necessidade de programação explícita, fornecendo capacidade preditiva mesmo em contextos de dados complexos e de alta dimensão [Saravanan and Sujatha 2018].

No aprendizado supervisionado, algoritmos são treinados com dados rotulados, permitindo a construção de modelos capazes de realizar previsões precisas em novos conjuntos de dados [Kotsiantis 2007]. Os algoritmos utilizados neste trabalho incluem:

- **K-Nearest Neighbors (KNN):** método baseado em instâncias que realiza a predição de novas amostras a partir da maioria das classes dos k vizinhos mais próximos, calculados por uma métrica de distância, geralmente a Euclidiana, destacando-se pela simplicidade e ausência de modelo prévio;
- **Decision Tree (DT):** constrói uma estrutura hierárquica de decisões dividindo recursivamente os dados com base em atributos que maximizam critérios como ganho de informação, sendo um modelo interpretável e amplamente utilizado em saúde;
- **Random Forest (RF):** método de *ensemble learning* que combina múltiplas árvores de decisão construídas a partir de amostras aleatórias e subconjuntos de atributos, reduzindo o risco de sobreajuste e melhorando a precisão, especialmente eficaz em dados ruidosos e de alta dimensionalidade;
- **Fuzzy Unordered Rule Induction Algorithm (FURIA):** utiliza regras difusas para lidar com a incerteza e ambiguidade dos dados, superando limitações dos métodos tradicionais de indução de regras rígidas, gerando regras mais flexíveis e interpretáveis para problemas complexos.

3. Trabalhos Relacionados

O avanço do AM tem impulsionado soluções eficientes para a saúde, especialmente no diagnóstico de doenças crônicas como CKD. Diversos estudos aplicam algoritmos supervisionados para antecipar diagnósticos e melhorar a precisão clínica. Esta seção revisa trabalhos sobre o uso de AM na detecção da CKD, destacando métodos, algoritmos e resultados.

O estudo de [Kalpana et al. 2025] apresenta um modelo inovador para predição precoce da Doença Renal Crônica usando stacking com Explainable Boosting Classifier. O método combina alta precisão e interpretabilidade, destacando quais atributos clínicos influenciam as decisões do modelo. O sistema alcançou desempenho perfeito em métricas como acurácia e sensibilidade. Essa união entre desempenho e transparência facilita a confiança e adoção na saúde.

O trabalho de [Rafi et al. 2025] investigou vários algoritmos de aprendizado de máquina e profundo para detectar CKD, realizando rigoroso pré-processamento dos dados. Entre os modelos testados, o XGBoost se destacou, atingindo 99,1% de acurácia, superando outros métodos. O estudo evidenciou a força dos algoritmos de gradient boosting para tarefas médicas, além de reforçar a importância da seleção e tratamento adequado dos dados.

O estudo de [M and N 2025] propôs o uso do LightGBM para detectar precoceamente CKD com dados clínicos e demográficos. O modelo alcançou 96% de acurácia, mostrando-se mais eficiente e preciso comparado a KNN e Árvores de Decisão. O trabalho destaca o equilíbrio entre alta performance e eficiência computacional, além de identificar fatores de risco importantes para a progressão da doença.

O trabalho de [S and VR 2025] abordou a predição de CKD usando Regressão Logística integrada a uma aplicação web prática. O modelo obteve acurácia de 97,5% e

destacou-se pela simplicidade e interpretabilidade, apesar de outros modelos terem desempenho um pouco melhor. A aplicação facilita que médicos e pacientes avaliem riscos instantaneamente, aproximando a inteligência artificial da rotina clínica.

Este trabalho apresenta um *benchmark* reproduzível com pré-processamento claro e divisão estratificada, além de demonstrar empiricamente que o *Random Forest* funciona como um *baseline* robusto no contexto estudado. Complementarmente, o método oferece uma camada extra de interpretabilidade por meio do uso do FURIA, que gera regras *fuzzy* compreensíveis, e apresenta uma lista enxuta dos preditores mais influentes, facilitando a orientação dos processos de coleta e triagem clínica.

4. Metodologia

A seção descreve detalhadamente os procedimentos e técnicas adotados para o desenvolvimento deste trabalho, desde a coleta e pré-processamento dos dados até a avaliação dos modelos.

4.1. Dataset

O trabalho utilizou o dataset *Chronic Kidney Disease* do repositório UCI [Rubini et al. 2015], composto por 400 instâncias e 24 atributos clínicos e laboratoriais, com a variável alvo indicando presença ou ausência da Doença Renal Crônica. A base é desbalanceada, com 62,5% dos casos classificados como CKD e 37,5% como NotCKD. A Tabela 1 resume os atributos (NdA), suas descrições (Desc) e intervalos de valores (IV).

Tabela 1. Descrição dos atributos do conjunto de dados

NdA	IV	Desc	NdA	IV	Desc
Age	2., 90	Idade	Sod	4.5,...,163	Sódio
BP	50, ..., 180	Pressão Arterial	Pot	2.5, ..., 47	Potássio
SG	1.005,...,1.025	Gravidade Específica	Hemo	3.1, ..., 17.8	Hemoglobina
Al	0....,5	Albumina	SC	0.4, ..., 76	Creatinina Sérica
Su	0....,5	Açúcar	Sod	4.5, ..., 163	Sódio
RBC	2.1,...,8	Glóbulos Vermelhos	Pot	2.5, ..., 47	Potássio
PC	Normal,Abnormal	Células Pus	Hemo	3.1,...,17.8	Hemoglobina
PCC	Present,Not Present	Agregados de Células Pus	PCV	9,...,54	Volume Corpúscular Empacotado
Ba	Present, Not Present	Bactérias	WBC	2200,...,26400	Contagem de Glóbulos Brancos
BGR	22, ..., 490	Glicose Sanguínea Aleatória	RBC	2.1,...,8	Contagem de Glóbulos Vermelhos
BU	1.5, ..., 391	Ureia Sanguínea	Htn	Yes,No	Hipertensão
SC	0.4, ..., 76	Creatinina Sérica	DM	Yes,No	Diabetes Mellitus
CAD	Yes,No	Doença Arterial Coronariana	Appet	Good,Poor	Apetite
PE	Yes,No	Edema de Pedal	Ane	Yes,No	Anemia
			Class	CKD,Not CKD	Classe (Doença Renal Crônica)

4.2. Pré-processamento

O pré-processamento confere as etapas fundamentais para a preparação dos dados brutos, garantindo sua qualidade e adequação para a aplicação dos algoritmos de aprendizado de máquina. Esse processo é essencial para assegurar que os modelos sejam treinados com informações consistentes e comparáveis, otimizando seu desempenho.

Inicialmente, valores ausentes foram tratados por meio de imputação, utilizando a média para atributos numéricos e a moda para atributos categóricos, assegurando a integridade das informações. Posteriormente, os atributos categóricos foram convertidos em valores numéricos por meio de codificação do tipo *label encoding*, tornando-os compatíveis

com os algoritmos de aprendizado de máquina. Por fim, aplicou-se a normalização utilizando a técnica de *Min-Max Scaling*, o que garantiu que todos os atributos numéricos estivessem em uma mesma faixa de valores.

Após o pré-processamento, o conjunto de dados foi dividido em duas partes: 80% para treinamento e 20% para teste. Essa divisão foi realizada de maneira estratificada, preservando a proporção original das classes e, assim, evitando possíveis vieses na avaliação dos modelos. Com essa estratégia, foi possível avaliar de forma mais robusta a capacidade de generalização dos algoritmos, utilizando dados que não foram previamente vistos durante o treinamento.

4.3. Configuração da proposta

Os parâmetros de cada algoritmo se apresentam de forma padrão na implementação baseado na biblioteca de código *scikit-learn* da linguagem de programação Python. A combinação desses quatro algoritmos visa explorar diferentes estratégias de aprendizagem, permitindo uma análise comparativa abrangente em termos de desempenho preditivo e aplicabilidade no diagnóstico da CKD. O ambiente utilizado para o desenvolvimento e execução deste estudo foi o Google Colab, uma plataforma online gratuita que permite a execução da linguagem de programação Python diretamente no navegador, sem necessidade de configuração local. Os códigos com a implementação realizada estão disponíveis no repositório <https://github.com/pedromoreira49/weit2025-ckd-prediction-repository>.

5. Avaliação Realizada

Nesta seção, são apresentados os resultados obtidos a partir da avaliação dos algoritmos utilizados. Foram consideradas as métricas de avaliação em problemas de classificação: acurácia, precisão, *recall* e *f1-score*, além da análise das matrizes de confusão para compreender os padrões de acertos e erros cometidos pelos modelos. A seguir é apresentado a Tabela 2 que demonstra as métricas detalhadas de precisão, *f1-score* e *recall* para cada um dos algoritmos individualmente.

Tabela 2. Métricas por classe e algoritmo

Métrica	Classe	KNN	Decision Tree	Random Forest	FURIA
Precisão	CKD	0.82	0.96	0.96	0.98
	NotCKD	0.56	0.96	1.00	0.93
Recall	CKD	0.69	0.98	1.00	0.96
	NotCKD	0.71	0.93	0.93	0.96
F1-Score	CKD	0.75	0.97	0.98	0.97
	NotCKD	0.62	0.95	0.96	0.95

O algoritmo KNN apresentou desempenho moderado, com precisão e *recall* mais baixos, especialmente para a classe *NotCKD* (precisão 0,56 e F1-score 0,62), enquanto para CKD os valores foram melhores (precisão 0,82, recall 0,69 e F1-score 0,75), indicando dificuldades principalmente na identificação dos casos negativos. O DT demonstrou desempenho equilibrado e superior ao KNN, com altas métricas para ambas as classes (*F1-score* em torno de 0,95), mostrando-se eficaz na discriminação dos casos. O modelo RF destacou-se pelo melhor desempenho geral, alcançando precisão e *recall* de

1,00 para CKD e valores igualmente altos para *NotCKD* (*F1-score*: 0,98 para CKD e 0,96 para *NotCKD*), evidenciando grande robustez e acurácia. Por fim, o FURIA obteve resultados altos, embora ligeiramente inferiores ao RF, com precisão e *recall* por volta de 0,96–0,98 para CKD e 0,93–0,96 para *NotCKD* (*F1-score*: 0,97 e 0,95, respectivamente), confirmando sua confiabilidade e competitividade. A seguir é apresentado as matrizes de confusão gerada por cada algoritmo.

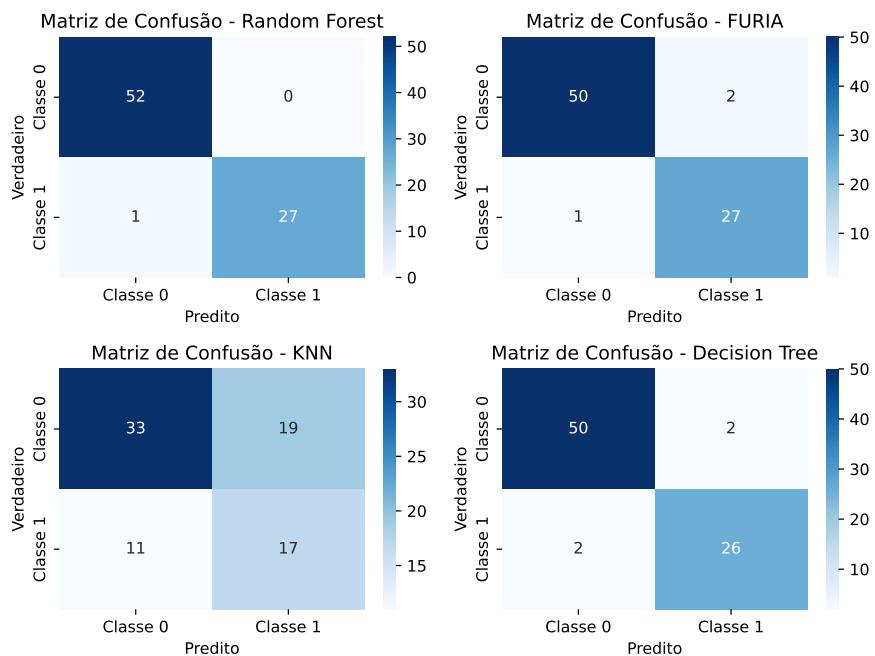


Figura 1. Matriz de Confusão 4 Modelos

Conforme é possível visualizar na Figura 1, cada uma das quatro matrizes apresentadas dos algoritmos utilizados ilustram a distribuição das classificações corretas e incorretas, permitindo observar o desempenho dos modelos na separação entre pacientes com e sem CKD.

5.1. Interpretação dos melhores modelos

Nesta seção, serão apresentados os processos de interpretação dos modelos que obtiveram as melhores métricas nos experimentos realizados. O objetivo é explorar as razões para o bom desempenho desses modelos, além de analisar quais fatores e variáveis mais influenciaram os resultados obtidos.

Inicialmente, considerando o *Random Forest*, é essencial analisar quais variáveis mais contribuíram para seu desempenho. Conforme visualizado na Figura 2, a gravidade específica foi o preditor de maior impacto. Clinicamente, a gravidade específica da urina é uma medida da concentração de solutos e reflete a capacidade dos rins de concentrar ou diluir a urina. Rins saudáveis ajustam a gravidade específica para manter a homeostase hídrica do corpo.

Em pacientes com CKD, a capacidade de concentração renal é frequentemente comprometida, resultando em uma gravidade específica baixa e fixa, tipicamente em torno de 1.010. Portanto, a alta relevância dessa variável no modelo está em total conformidade

com a fisiopatologia da CKD, tornando-a um indicador primário e não invasivo da função renal.

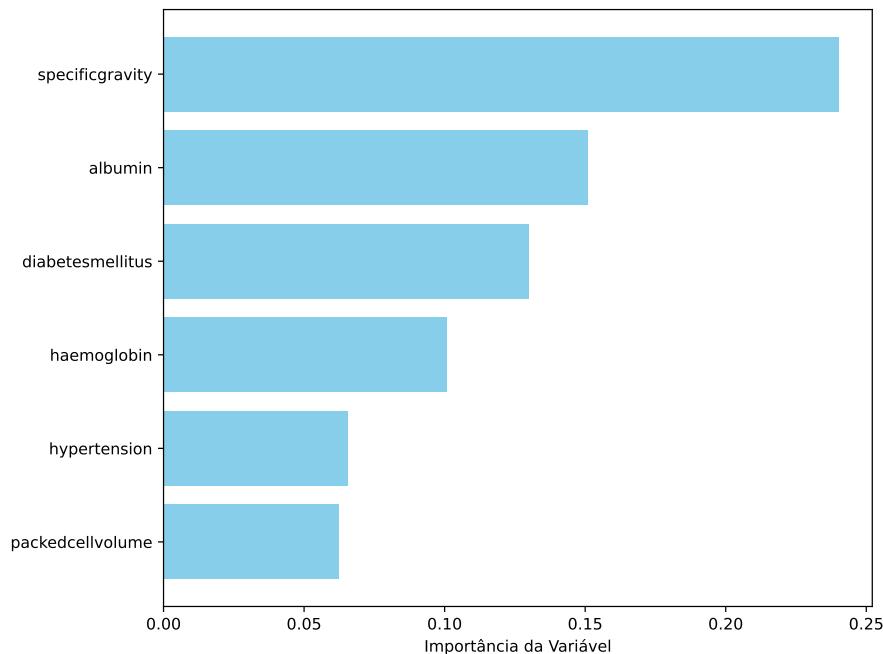


Figura 2. Relação de importância das variáveis no modelo Random Forrest.

Por se tratar de um diferencial, a utilização de um modelo baseado em lógica fuzzy, e com uma boa performance, também é possível analisar regras fuzzy geradas pelo algoritmo FURIA. Essas regras são formuladas no formato “se-então”, permitindo inferências graduais que refletem melhor o raciocínio humano em situações complexas. A utilização dessa abordagem possibilita um modelo mais flexível e interpretável para a classificação no contexto da CKD. Cabe destacar que o antecedente das regras é formado pelos atributos apresentados na Tabela 1.

- Regra 1:** Se $SG \in (-\infty, 0.5, 0.75)$ e $PCV \in (-\infty, 0.086667, 0.091111)$, então classe = ckd (CF = 1.0).
- Regra 2:** Se $HTN \in [0, 1, \infty)$, então classe = ckd (CF = 0.99).
- Regra 3:** Se $AL \in [0, 0.2, \infty)$ e $SOD \in (-\infty, 0.084858, 0.085489)$ e $BU \in [0.060334, 0.062901, \infty)$, então classe = ckd (CF = 0.99).
- Regra 4:** Se $APPET \in [0, 1, \infty)$, então classe = ckd (CF = 0.99).
- Regra 5:** Se $DM \in [0, 1, \infty)$, então classe = ckd (CF = 0.99).
- Regra 6:** Se $PCV \in (-\infty, 0.064444, 0.093333)$ e $Hemo \in (-\infty, 0.063946, 0.067347)$ e $Hemo \in [0.054422, 0.055782, \infty)$, então classe = ckd (CF = 0.99).
- Regra 7:** Se $SG \in (-\infty, 0.25, 0.5)$ e $BU \in [0.012195, 0.039795, \infty)$, então classe = ckd (CF = 0.99).
- Regra 8:** Se $SG \in [0.5, 0.75, \infty)$ e $Htn \in (-\infty, 0, 1)$ e $Hemo \in [0.065306, 0.067347, \infty)$ e $SC \in (-\infty, 0.10582, 0.119048)$, então classe = notckd (CF = 0.99).
- Regra 9:** Se $SG \in [0.5, 0.75, \infty)$ e $Al \in (-\infty, 0, 0.2)$ e $DM \in (-\infty, 0, 1)$, então classe = notckd (CF = 0.93).

O algoritmo FURIA construiu nove regras para classificar a presença ou ausência de doença renal crônica (CKD). As regras utilizam combinações de variáveis clínicas, como gravidade específica, volume de células compactadas, hipertensão, albumina, sódio, ureia sanguínea, apetite, diabetes mellitus, hemoglobina e creatinina sérica.

A maioria das regras identifica a classe `ckd` quando há valores baixos ou intermediários de gravidade específica, baixos valores de volume de células compactadas, presença de hipertensão, diabetes mellitus, valores específicos de albumina, sódio, ureia sanguínea e hemoglobina. As regras para `not ckd` ativam-se quando a gravidade específica está em faixas intermediárias, ausência de hipertensão e diabetes mellitus, e valores específicos de albumina, hemoglobina e creatinina sérica.

Cada regra possui um fator de certeza (CF), indicando a confiança do algoritmo na classificação. O FURIA demonstrou capacidade de discernimento entre os casos, combinando múltiplos critérios clínicos para aumentar a precisão do diagnóstico.

6. Considerações Finais e Trabalhos Futuros

Este trabalho contribuiu para a predição da CKD por meio da aplicação de técnicas de aprendizado de máquina, com ênfase no uso do modelo RF e do algoritmo FURIA. O modelo RF demonstrou desempenho superior nas métricas de acurácia, precisão, *recall* e *F1-score*, confirmando sua eficácia na identificação de pacientes com CKD. Já o algoritmo FURIA evidenciou o potencial das abordagens *fuzzy* para a geração de modelos interpretáveis, facilitando a compreensão e a confiança dos profissionais de saúde nas decisões automatizadas.

Entretanto, algumas limitações devem ser destacadas: a definição manual das funções de pertinência no modelo *fuzzy* pode introduzir vieses; a base de dados utilizada, apesar de adequada, apresenta-se com desbalanceamento; e a integração entre o conhecimento médico e os modelos computacionais ainda exige maior aprofundamento para garantir maior robustez e aplicabilidade prática.

Como desdobramentos futuros, recomenda-se aprimorar as funções de pertinência por meio de métodos automatizados e validar os modelos em bases externas e mais amplas. Outra direção promissora é o desenvolvimento de *architectures* híbridas que integrem técnicas *fuzzy* e de aprendizado profundo, bem como uma colaboração mais estreita entre especialistas clínicos e cientistas de dados, para criar sistemas de decisão médica fundamentados em evidências e passíveis de adoção em ambientes reais.

7. Agradecimentos

Este trabalho foi financiado pelas seguintes agências de fomento: CAPES, FAPERGS (24/2551-0001396-2) e FAPERGS/CNPq (23/2551-0000126-8).

Referências

- Acropolium (2025). Machine learning in healthcare: [7 real use cases included]. Acesso em: 13 jul. 2025.
- CashU (2025). Segurança de dados: você está protegido? veja o que fazer. Reportagem sobre cibercrime e segurança da informação no Brasil, destacando o uso de IA e aprendizado de máquina para detecção de fraudes e proteção de dados.

- Group, K. D. I. G. O. K. C. W. (2013). Kdigo 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney International Supplements*, 3(1):1–150.
- Inforchannel (2024). Especialista destaca ferramentas para análise de dados em 2025. Discussão sobre o uso de análise de dados e inteligência artificial, incluindo aprendizado de máquina, em setores como saúde.
- Kalpana, C., Jagadeesh, K., Vennila, A., and Radhika, R. (2025). A stacking based explainable boosting model for early prediction of chronic kidney disease. In *2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 1584–1589.
- Kotsiantis, S. (2007). Supervised machine learning: A review of classification techniques. *Informatica (Slovenia)*, 31:249–268.
- Kumar, A. et al. (2025). Machine learning applications in healthcare clinical practice and research. *Frontiers in Medicine*. Acesso em: 13 jul. 2025.
- Levin, A., Ahmed, S. B., Carrero, J. J., Foster, B., Francis, A., Hall, R. K., Herrington, W. G., Hill, G., Inker, L. A., Kazancioğlu, R., Lamb, E., Lin, P., Madero, M., McIntyre, N., Morrow, K., Roberts, G., Sabanayagam, D., Schaeffner, E., Shlipak, M., Shroff, R., Tangri, N., Thanachayanont, T., Ulasi, I., Wong, G., Yang, C.-W., Zhang, L., Robinson, K. A., Wilson, L., Wilson, R. F., Kasiske, B. L., Cheung, M., Earley, A., and Stevens, P. E. (2024). Executive summary of the kdigo 2024 clinical practice guideline for the evaluation and management of chronic kidney disease: known knowns and known unknowns. *Kidney International*, 105(4):684–701.
- M, K. and N, S. (2025). Early chronic kidney disease detection using machine learning techniques. In *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, pages 151–155.
- Magalhães, V., Carmona, J., Lucca, G., Santos, H., and Borges, E. (2021). Uma classificação de vinhos baseada em regras fuzzy utilizando o algoritmo farc-hd. In *Anais do VI Workshop-Escola de Informática Teórica*, pages 148–155, Porto Alegre, RS, Brasil. SBC.
- Metherall, B., Berryman, A., and Brennan, G. (2024). Machine learning for classifying chronic kidney disease and predicting creatinine levels using at-home measurements.
- Ouro Preto Investimentos (2025). Inteligência artificial no mercado de capitais. Acesso em 15 jul. 2025.
- PhoenixNAP (2024). Machine learning in finance: Examples and benefits. Acesso em: 13 jul. 2025.
- Rafi, S., Revanth, N., Raghava Reddy, K. V., Babu, K. M., Kumar, Y. L. P., and Kumar, N. V. (2025). Chronic kidney disease prediction using machine learning and deep learning models. In *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, volume 3, pages 1–6.
- Rubini, L., Soundarapandian, P., and Eswaran, P. (2015). Chronic Kidney Disease. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5G020>.

S, K. and VR, G. (2025). Machine learning-based prediction of chronic kidney disease with logistic regression. In *2025 8th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 330–334.

Saravanan, R. and Sujatha, P. (2018). A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 945–949.

Saxena, V. (2025). The role of machine learning in strengthening application security. Acesso em: 13 jul. 2025.