

Detecção Automática do Risco de Doenças Renais Crônicas: Uma Abordagem Baseada em Visão Computacional

Hyan H. N. Batista¹

¹Departamento de computação – Universidade Federal Rural do Pernambuco (UFRPE)
Recife – PE – Brasil

hyan.batista@ufrpe.br

Abstract. *Chronic kidney diseases (CKD) cause millions of deaths worldwide every year. A widely used method to diagnose these diseases is renal scintigraphy. However, the way it is processed varies depending on the hospital and the physician. Many studies have addressed the issue. Thus, this work aims to use methods of computer vision and machine learning to automatically detect whether a patient presents a risk of CKD or not.*

Resumo. *Doenças renal crônicas (CKD) causam milhões de mortes ao redor do mundo todo ano. Um método bastante usado para diagnosticar essas doenças é a cintilografia renal. Entretanto, a maneira que é processada varia dependendo do hospital e do médico. Muitos trabalhos já dirigiram-se ao problema. Desse modo, este trabalho busca usar métodos de visão computacional e machine learning para detectar automaticamente se um paciente apresenta ou não risco de CKD.*

1. Introdução

Doenças renal crônicas causam milhões de mortes ao redor do mundo todo ano. Um método bastante usado para diagnosticar essas doenças é a cintilografia renal. Entretanto, a maneira que é processada varia dependendo do hospital e do médico. Muitos trabalhos já dirigiram-se ao problema.

[Rebouças Filho et al. 2019], por exemplo, tentou solucionar o problema usando métodos de visão computacional como o LBP para extrair características significadas de um conjunto de imagens cintilográficas e, então passar o vetor resultante para algoritmos de *machine learning* para, dessa forma, gerar um modelo preditivo. Esse estudo, entretanto, apenas apresenta as pontuações dos modelos no conjunto de treino, o que é uma falha metodológica.

Nesse contexto, este trabalho propõe a reprodução dos experimentos realizados em [Rebouças Filho et al. 2019] buscando aprimorar o seu processo metodológico com o objetivo de chegar a um modelo usando técnicas de visão computacional e *machine learning* que seja capaz de detectar automaticamente se um paciente precisa ou não de cuidados especializados pois corre risco de desenvolver CKD.

2. Metodologia

Esta seção tem como objetivo descrever cada uma das etapas dos experimentos realizados neste estudo, indo desde a coleta dos dados até o treinamento e validação dos modelos resultantes.

2.1. Dados

O conjunto de dados usado neste estudo foi o *Database of Dynamic Renal Scintigraphy*, uma coleção de estudos anonimizado de pacientes e dados simulados para auditorias clínicas, comparações inter-laboratoriais, desenvolvimento e testes de métodos analíticos e *software* clínico, assim como, para propósitos de ensino-aprendizagem e treinamento. Esse banco de dados inclui cinco conjuntos de dados diferentes: um de adultos, outro de crianças e três conjuntos de dados simulados. O conjunto de dados usado em [Rebouças Filho et al. 2019] foi o DRSPRG, uma série de 115 estudos cintilográficos em adultos. Nele, os estudos 3, 30, 33, 36, 43, 51, 103, 109 foram excluídos por estarem inadequados, restando então 107 estudos. Cada um desses estudos foi previamente classificado em cinco estágios de doenças renais crônicas (CKD) com base na estimação de GFR do sangue coletado.

No conjunto de dados DRSPRG, cada exame levou trinta minutos e o exame armazenado consiste de 180 imagens de tamanho 128x128 pixels que foram capturados a cada 10 segundos ao longo do procedimento do exame. Ambos, posterior e anterior foram capturados também. Notou-se que 5 exames de pacientes não tinham todas as 180 imagens, a saber, os pacientes enumerados como 25 (CKD 1), 55 (CKD 2), 65 (CKD 2), 81 (CKD 3) e, 106 (CKD 5); esses exames foram removidos dos experimentos com o objetivo de mantê-lo padronizado [Rebouças Filho et al. 2019].

Pacientes que se encontram no estágio 3 da doenças renais crônicas tem riscos de mortalidade maiores, entre outras comorbidades [Levey et al. 2003], como doenças nos ossos, além de terem maior sensibilidade aos efeitos colaterais de drogas [Harris and Stribling 2007]. Por essa razão, em [Rebouças Filho et al. 2019] optaram por trabalhar com apenas duas classes. A classe 0 que compreende os estágios 1 e 2, e a classe 1 que compreende os estágios 3, 4 e 5. Outro razão para isso, pode ter sido o desbalanceamento dos dados associado com o número reduzido de amostras por estágio. Na Figura 1, é possível notar a frequência em que cada estágio de CKD ao longo dos estudos.

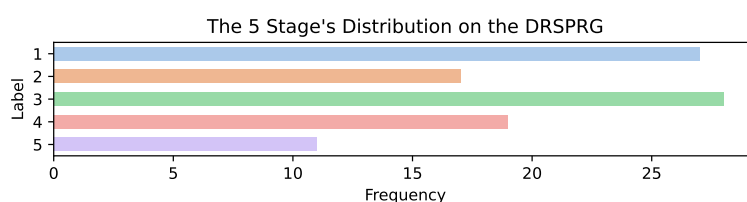


Figure 1. Frequência de cada estágio de CKD ao longo dos estudos cintilográficos.

Desse modo, o modelo final indicara se o paciente necessidade de monitoramento especializado por conta de possível progressão na doença, o que é indicado pela classe 1, ou não, representado pela classe 0. Portanto, como podemos ver na Figura 2, tem-se, no final do procedimento, uma distribuição de 44 amostras para a classe 0, e 58 para a classe 1.

2.2. Preparação dos Dados

O conjunto de dados descrito na Subseção 2.1 passou por uma série de processamentos para que pudesse ser utilizado como entrada para algoritmos de visão computacional e

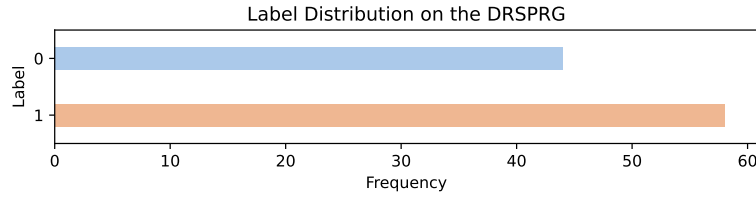


Figure 2. Distribuição das classes determinadas ao longo dos estudos cintilográficos.

machine learning que serão abordados nas subseções 2.3 e 2.4.

Os processamentos executados no conjunto de dados foram os mesmos usados em [Rebouças Filho et al. 2019]. Todas as 180 imagens de cada exame de paciente, portanto, passaram por um processo de *blurring* usando um filtro médio. Esse processo é realizado com o objetivo de remover o ruído presente e como resultado suaviza as bordas da imagem. Na Figura 3, pode-se ver uma amostra antes e após o *blurring*.

Depois disso, o conjunto de 180 imagens foi dividido em 15 subconjuntos onde cada um deles representa um *frame* de 2 minutos do exame, e contém 12 imagens cada. Segundo [Rebouças Filho et al. 2019], esse período de 2 minutos nada mais é do que uma prática comum na literatura, logo, o mesmo foi reproduzido neste estudo. Finalmente, após essa divisão, uma imagem média de todos os 15 subconjuntos foi gerada. A Fórmula 1 mostra como ela foi criada. Nela, n representa o total de imagens no subconjunto e x , uma instância de imagem. Na Figura 4 é possível observar uma das imagens resultados desse procedimento.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

2.3. Extração das Característica

Na etapa de preparação dos dados, detalhada na Subseção 2.2, os estudos foram separados em 15 subgrupos de 12 imagens, onde cada subgrupo foi reduzido à uma única imagem por meio de operação de média. Após isso, cada exame de paciente passou a ser representado por uma sequência de 15 imagens cintilográficas. Entretanto, para que elas estivessem em um estado adequado para um modelo de *machine learning*, ainda precisaria passar por uma etapa de extração de características.

A etapa extração de característica busca transformar os dados para que eles possam ser utilizados por algoritmos de aprendizagem. Neste estudo, utilizou-se o LBP (do inglês, *local binary patterns*) para extrair características significativas de cada uma das 15 imagens que representam o exame de cintilografia do paciente. Como esse algoritmo precisa que as imagens não estejam em RGB, deu-se então a execução de um processo de binarização.

Primeiramente, todas as imagens foram convertidas para cinza. Em seguida, deu-se o processo de binarização. O método usado para isso foi o *Otsu's Thresholding*, que consegue identificar automaticamente os limiares mínimo e máximo para binarização da

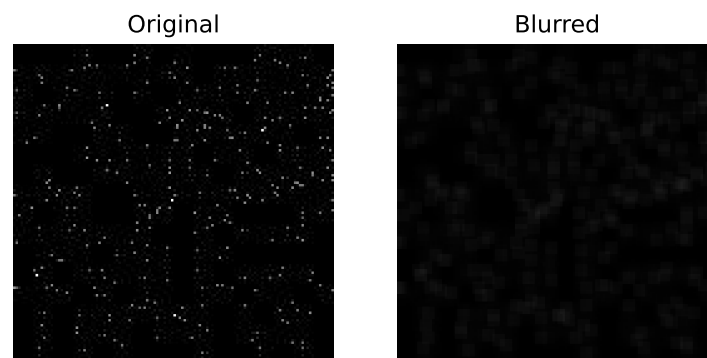


Figure 3. Imagem cintilográfica antes e depois do processo de *blurring*

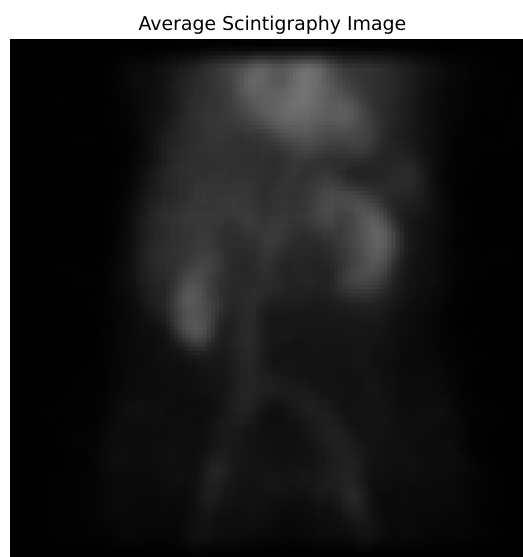


Figure 4. Imagem cintilográfica resultante do procedimento de *averaging*

imagem. Na Figura 5 é possível observar o resultado final disso para uma das imagens. Feito este procedimento, a imagem estava pronta para o algoritmo de LBP.

Para cada imagem foi, então, gerado um vetor de 108 inteiros caracterizando a textura da imagem usando LBP. Os vetores resultados foram, em seguida, concatenados. Desse modo, para cada exame gerou-se um vetor de 1620 inteiros que veriam a ser pas-



Figure 5. Imagem cintilográfica média binarizada.

sados como entrada para os algoritmos de aprendizagem para gerar o modelo preditivo.

Os valores de raio e *sampling pixels* para o algoritmo LBP foram 1 e 106, respectivamente. Estes valores foram usados para tentar se aproximar do número de características que foram extraídas segundo [Rebouças Filho et al. 2019] em seus experimentos, haja que os autores não deixaram claro qual foi a configuração utilizada.

2.4. Seleção dos Modelos

Uma vez que os dados passaram pelos processos de preparação e extração de características, os quais foram detalhados nas subseções 2.2 e 2.3, o experimento entrou na fase de seleção de modelos. Nesta etapa busca-se desenvolver modelos preditivos usando técnicas de *machine learning*.

Dessa forma, o primeiro passo foi separar o conjunto de dados (2.1) em treino e teste. Em [Rebouças Filho et al. 2019], os autores não deixaram claro quais parâmetros usaram para realizar esse procedimento nem se este o foi feito. Desse modo, neste estudo optou-se por estabelecer uma divisão de 80% das amostras para o conjunto de treino e o restante, 20%, para o de teste.

Os experimentos, em seguida, foram executados com as seguintes combinações de algoritmos: LBP e SVM. O primeiro para extrair os vetores de características que descrevessem significativamente as imagens cintilográficas e o segundo para modelar os dados. Para mais, é necessário destacar que duas versões do SVM foram empregadas: o RBF (do inglês, *radial basis function*) e o *linear SVM*.

No que se refere aos parâmetros usados no SVM em cada um dos experimentos, eles foram selecionados usando o algoritmo *random search* em configurações de hiper-parâmetro, buscando obter a melhor performance em cada um deles. O número de iterações foi definido como 40 e o número de *folders* como 10. Depois da otimização de hiper-parâmetros, os melhores foram usados em um processo de *10-fold cross-validation*.

Além disso, vale ressaltar que todos os experimentos em [Rebouças Filho et al. 2019] foram executados no conjunto de treino. Essa não é a maneira mais adequada de validar um modelo preditivo, visto que não é possível detectar inúmeros problemas, como por exemplo *overffiting*. Pode-se supor então que o *holdout set* não foi produzido. Desse modo, neste estudo, decidiu-se mostrar os resultados do melhor modelo resultante em cima do conjunto de teste, buscando aproximar o modelo de um cenário mais realístico e detectar problemas como o *overfitting*.

2.5. Avaliação dos Resultados

Em [Rebouças Filho et al. 2019], os autores adotaram uma abordagem baseada em classificação binária para solucionar o problema da detecção de doenças renais crônicas. Ao trabalhar-se com classificação binária majoritariamente são empregadas as seguintes medidas: acurácia, precisão, *recall* e *F₁ score* [Kowsari et al. 2019]. A acurácia, segundo [Hossin and Sulaiman 2015], mensura a razão entre as classificações positivas e negativas corretas e o total de amostras avaliadas. A precisão é uma métrica estatística usada para indicar a porcentagem de amostras corretamente classificadas como positivas considerando todas as que foram assim classificadas. O *recall* é mostra a porcentagem das amostras positivas classificadas corretamente. Por fim, o *F₁ score* é a média harmônica entre essas duas últimas métricas, passando uma ideia geral do desempenho do modelo. As fórmulas 3 e 4 mostram como foram calculadas a precisão, cobertura e *F₁ score*, respectivamente.

$$Acurácia = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Precisão = \frac{TP}{TP + FP} \quad Cobertura = \frac{TP}{TP + FN} \quad (3)$$

$$F\text{-measure} = \frac{2 \times precisão \times cobertura}{precisão + cobertura} \quad (4)$$

3. Resultados

Os experimentos realizados neste estudo tinham dois objetivos distintos. O primeiro era descobrir se era possível usar técnicas de visão computacional e *machine learning* para detectar automaticamente se um paciente precisa de monitoramento especializado por conta de risco de progressão de CKD. O segundo, era comparar o desempenho de diferentes métodos nessa tarefa. Nesta seção, portanto, os resultados alcançados.

Os algoritmos descritos na Seção 2, SVM-RBF e o *Linear SVM*, foram implementados, treinados e testados usando o LBP como método de extração de características. A Tabela 1 mostra as pontuações obtidas pelo SVM-RBF com LBP. É possível notar que ele apresentou um bom *F₁ score* para a classe 1, 0,81%, ou seja, consegue detectar com bastante precisão quando um paciente precisa de monitoramento especializado. Contudo, alcançou apenas 0,67% de *F₁ score* para a classe 0. Olhando para o sua precisão e *recall* (0,83% e 0,56%, respectivamente), pode-se a entender que o modelo embora seja preciso, não consegue recuperar todos os pacientes que estão nessa classe.

A Tabela 2, por sua vez, apresenta as pontuações obtidas pelo *Linear SVM* com LBP. O modelo apresentou o mesmo *F₁ score* para cada uma das classes. Entretanto, ao

	Precision	Recall	F-Score	Support
0	0,83	0,56	0,67	9
1	0,73	0,92	0,81	12
Accuracy			0,76	

Table 1. Pontuações do SVM-RBF no conjunto de teste.

se observar suas pontuações para precisão e *recall* para cada uma das classes, pode-se entender que há uma espécie de *trade-off* em cada uma delas. Ele tem uma boa precisão para a classe 1, 0,78%, ele tem um *recall* ruim. O inverso acontece para a classe 0.

	Precision	Recall	F-Score	Support
0	0,58	0,78	0,67	9
1	0,78	0,58	0,67	12
Accuracy			0,67	

Table 2. Pontuações do *Linear* SVM no conjunto de teste.

Ao comparar-se os resultados de ambos os modelos, o SVM-RBF e o *Linear* SVM, pode-se chegar a conclusão que o SVM-RBF obteve mais sucesso em modelar os dados que o *Linear* SVM. Quando se observa as acurácias dos respectivos modelos, isso torna-se ainda mais evidente. Tendo isso em mente, não é exagerado concluir-se que, não só o SVM-RBF é o melhor modelo dentre os analisados, como também que é possível detectar automaticamente a presença de CKD em pacientes com certa acurácia.

4. Conclusão

Este trabalho teve como objetivo o desenvolvimento de para detecção automática de doenças renais crônicas (CKD). Estudo experimentais são apresentados para mostrar como a abordagem proposta tem sucesso em classificar de forma automática os pacientes em dois grupos, um para pacientes sem risco e outro para pacientes que tem risco de desenvolver CKD e por isso necessitam de acompanhamento especializado. Os resultados mostram que, embora o estudo realizado em [Rebouças Filho et al. 2019] não tenha apresentado as pontuações dos modelos no conjunto de teste, o método proposto tem validade e consegue solucionar o problema com acurácia de 0,76%. Para trabalhos futuros, pretende-se aumentar o conjunto de dados utilizado acrescentando dados sintéticos e reavaliar o método proposto.

References

Harris, K. and Stribling, B. (2007). Automated estimated gfr reporting: A new tool to promote safer prescribing in patients with chronic kidney disease? *Therapeutics and clinical risk management*, 3(5):969–972.

- Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2):1.
- Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., and Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4):150.
- Levey, A. S., Coresh, J., Balk, E., Kausz, A. T., Levin, A., Steffes, M. W., Hogg, R. J., Perrone, R. D., Lau, J., and Eknoyan, G. (2003). National kidney foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Annals of internal medicine*, 139(2):137–147.
- Rebouças Filho, P. P., da Silva, S. P. P., Almeida, J. S., Ohata, E. F., Alves, S. S. A., and Silva, F. d. S. H. (2019). An approach to classify chronic kidney diseases using scintigraphy images. In *Anais Estendidos do XXXII Conference on Graphics, Patterns and Images*, pages 156–159. SBC.