

PPM-UNet: Uma Rede Neural Convolutacional para a Segmentação de Rins em Imagens de TC

Marcus Vinicius Oliveira¹, Caio Eduardo Falcão Matos¹, Geraldo Braz Júnior¹,
Anselmo Cardoso de Paiva¹, João D. Sousa de Almeida¹, Gabriel Costa¹,
Matheus Levy¹, Mario Freitas¹

¹Núcleo de computação aplicada – Universidade Federal do Maranhão (UFMA)
65.080-805 – São Luís – MA – Brasil¹

Abstract. *Kidney cancer represents 3% of adult malignancies, and early diagnosis is an essential tool. Automatic segmentation aims to assist the physician in the diagnostic process. The objective of this work is to develop and evaluate a network to segment, in computed tomography images, the regions of the kidneys and tumor, if any. We evaluated the use of Pyramid Pooling Module replacing convolutional layers of the U-Net network to achieve the goal. The proposed methodology achieves as a result 0.91 of Iou for kidney regions and 0.88 of Iou for tumor regions.*

Resumo. *O câncer de rim representa 3% das doenças malignas do adulto, sendo o diagnóstico precoce uma ferramenta essencial. A segmentação automática tem como objetivo auxiliar o médico no processo de diagnóstico. O objetivo desse trabalho consiste em desenvolver e avaliação de uma rede neural para segmentar as imagens de tomografia computadorizada, das regiões dos rins e tumor, caso existam. Avaliamos a utilização de Pyramid Pooling Module em substituição de camadas convolucionais da rede U-Net para atingir o objetivo. A metodologia proposta atinge como resultado 0,91 de Iou para regiões do rim e 0,88 de Iou para regiões de tumor.*

1. Introdução

Os rins consistem em dois órgãos presentes no trato urinário localizado na parte posterior do abdômen sendo suas extremidades superiores localizadas na altura dos arcos costais, próximos a 10^a e 12^a costelas torácicas. Este órgão, é responsável pela eliminação de toxinas do sangue através de um sistema de filtragem, bem como, a regularização da pressão arterial e controle de líquidos do corpo [de Nefrologia 2022]. O câncer de rim, também conhecido como adenocarcinoma renal apresenta-se como o terceiro mais frequente do aparelho geniturinário e representa aproximadamente 3% das doenças malignas do adulto. Estatísticas estimam que, no Brasil, sua incidência é de 7 a 10 casos para cada 100 mil habitantes e anualmente surgem cerca de 51 mil novos casos da doença [Sírio-Libanês 2022, do Câncer 2022].

Estudos apontam alguns fatores que elevam o risco de desenvolver o câncer renal como obesidade, histórico familiar, tabagismo, hipertensão e doença de Von Hippel-Lindau¹ [Sírio-Libanês 2022]. Em relação a localidade dos tumores renais, por volta de

¹Doença genética e rara, caracterizada pela formação de tumores com muita vascularização em várias partes diferentes do corpo.

54% são encontrados confinados ao rim, 20% estão localmente avançados (acometendo gânglios regionais próximos ao rim) e 25% desenvolvem a metástase da doença espalhando-se principalmente para os pulmões, fígado e ossos [Sírio-Libanês 2022].

Desta forma, o diagnóstico precoce desta patologia torna-se ferramenta essencial para o prognóstico e tratamento da doença elevando assim, as chances de cura do paciente. Pacientes diagnosticados em estágio inicial da doença, ou seja, quando não há disseminação da doença, possuem cerca de 93% de taxa de sobrevivência relativa em cinco anos. Por outro lado, quando o diagnóstico é efetuado em estágio avançado (metástase da doença) esta taxa é reduzida para 12%, demonstrando assim, a importância deste diagnóstico precoce [Oncoguia 2022].

Aprendizagem profunda tem sido recentemente aplicado em problemas médicos com sucesso [Trevisan et al. 2021, Jatobá et al. 2020, Vogado et al. 2019], inclusive redes neurais convolucionais (CNN) provaram-se promissoras na segmentação de áreas dos rins [da Cruz et al. 2020]. Neste trabalho, o modelo atingiu um Dice de 96,33% e um Jaccard de 93,02%, para segmentação dos rins.

Do mesmo modo, surgiram outras redes que tentam superar a U-Net como [Isensee and Maier-Hein 2019] que fazem uso de uma rede U-Net padrão combinada com uma U-Net 3D para obter um dice do rim de 97,36%, e dice do Tumor de 85,73%. [Shahin et al. 2019] apresentou uma Rede Totalmente Convolucional (FCN) com o objetivo de superar os resultados da U-Net, no qual criaram uma rede inspirada na Pyramid Scene Parsing Network [Zhao et al. 2016]. O seu modelo alcançou um Dice de 90,3 e um Jaccard de 83,7.

Em [Hafhouf et al. 2020] foi proposto uma modificação da U-Net, na qual se inseriu blocos de Pyramid pooling module (PPM) nos dois últimos skip connections. O modelo chegou a um dice de 89.6 e um Jaccard de 82.7 sendo melhor que a U-Net sem modificações.

Uma modificação de camadas de CNN, a Pyramid Pooling Module (PPM) [Zhao et al. 2016], foi usado em [Yang et al. 2018], onde se incorporou o PPM em uma 3D FCN, para se ter uma melhor campo de recepção de características. Os resultados chegam a um dice no rim de 0,931 e dice no tumor de 0.802.

O objetivo desse trabalho é avaliar uma versão modificada da rede U-Net através de blocos PPM para segmentação de rim e tumor em imagens de tomografia computadorizada. Os blocos Pyramid pooling module foram adicionados no lugar das convoluções da rede, com o objetivo de aumentar a quantidade de características obtidas das imagens e assim melhorar as taxas de assertividade.

O restante deste trabalho possui mais três seções. A Seção 2 descreve a metodologia usada na aplicação do modelo. A Seção 3 que serão apresentados os resultados e também serão discutidos. Por fim, na Seção 4 é destacada a conclusão e os trabalhos futuros.

2. Metodologia

A metodologia, apresentada na Figura 1, se divide em quatro etapas. A primeira sendo a aquisição das imagens usadas no trabalho, proveniente do desafio 2019 Kidney Tumor Segmentation Challenge ou Kits19 [Heller et al. 2019] que possui imagens de tomografia

computadorizada de 300 pacientes. A segunda etapa consiste num pré-processamento dos volumes. A terceira etapa se dá pela construção do modelo usando a PPM-UNet. A última etapa consiste na experimentação e avaliação do método proposto. Cada uma das etapas é detalhada nas seções seguintes.

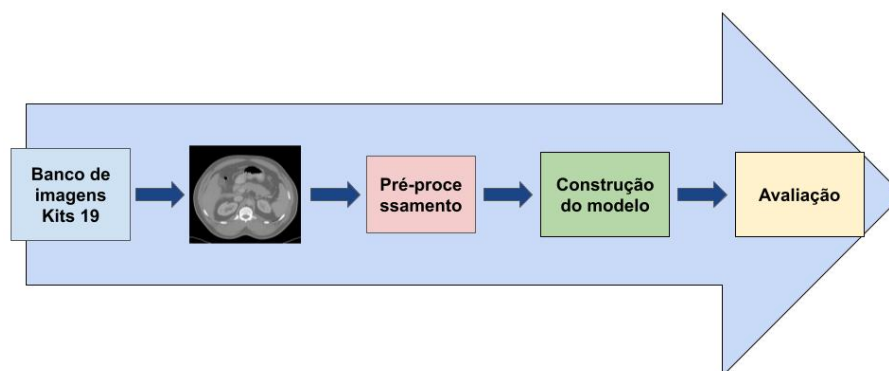


Figura 1. Etapas da metodologia proposta.

2.1. Aquisição de imagens

Para o desenvolvimento e avaliação do método, a base de imagens de tomografia computadorizada Kits19 [Heller et al. 2019] foi utilizada. Esta é composta por 300 pacientes, sendo 210 separados para treino e 90 para a avaliação do modelo. As imagens adquiridas vieram em escala de cinza e formato Neuroimaging Informatics Technology Initiative (NIfTI), como um tamanho de 512 x 512. Cada imagem passou por um pré-processamento e foram redimensionadas de 512x512 para 128x128 por conta de restrições de hardware.

2.2. Construção do modelo

O modelo foi baseado nas redes neurais convolucionais, usualmente utilizadas em visão computacional e classificação, essa rede é análogas as Redes Neurais Artificiais(ANN), sendo composta de neurônios que podem se otimizar a partir de aprendizagem. Cada neurônio recebendo uma entrada e fazendo uma operação, a base de muitas ANNs. Na CNN sua imagem de entrada até a sua saída final, vai expressar uma única função de pontuação perceptiva(peso), esses pesos podem ser treinados para se obter um melhor resultado, a ultima camada vai ter a função de loss associada as classes da mesma forma que a ANN [O'Shea and Nash 2015].

A rede avaliada neste trabalho teve inspiração em uma rede U-Net [Ronneberger et al. 2015], sendo uma rede totalmente convolucional, esse tipo de rede tem como característica ser mais rápida, pois evita o uso de camadas mais densas e muitos parâmetros, sendo possível se utilizar para qualquer tamanho de imagem. Nessa rede ocorre uma redução de amostragem inicial, sendo utilizados convoluções, pooling e por

fim passa por um aumento de resolução, no caso da UNet feito por skip connection que ligam a etapa de aumento de resolução com a redução, se aproveitando dessa etapa para gerar uma melhor qualidade da imagem [Long et al. 2014].

A U-Net recebe essa nomenclatura por causa do seu formato, a rede se divide em 2 etapas, a primeira contrai a imagem para extrair suas características, a segunda é composta pelo aumento da resolução das imagens usando deconvoluções e os skip connections para no fim fim gerar a predição pixel a pixel.

Pensando na etapa de convolução onde se extrai características, foi proposto a inserção de camadas de Pyramid Pooling Module representada na Figura 2, para auxiliar na extração de características de tamanhos diferentes, e assim aumentar o nível de acerto da rede independente do tamanho ou formato do rim e tumor. Essa suposição leva em conta que no kits19 muitos pacientes sofreram com a perda de um rim ou perda parcial do mesmo.

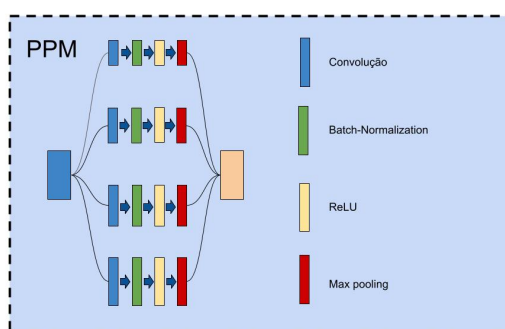


Figura 2. Representação do Pyramid Pooling Module

O Pyramid pooling module ou PPM tem como objetivo da rede enxergar a imagem em um contexto mais global, visando separar características locais e globais, e juntá-las para ter uma visão geral do que foi recebido. Com isso em vista, a aplicação do mesmo no meio da segmentação de rins se torna viável pela grande diferença de tamanho entre os rins de cada ser humano e o tamanho dos tumores que também variam ainda mais.

A aplicação do PPM se dá pela divisão de vários níveis, no qual se possui diferentes escalas, essas escalas devendo conter uma distância razoável entre elas, sendo elas medidas no núcleo da convolução. A quantidade de níveis e o tamanho de cada um é variável, dependendo do tamanho do mapa de características.

A representação de como fica uma rede U-Net com a presença do PPM está na Figura 3. A inserção dos blocos de PPM foram feitos de forma alternada com as camadas de convolução, sendo inseridas apenas na parte de contração da rede.

Na nossa arquitetura o PPM, apresentada na Figura 2, substitui 2 blocos de convoluções da U-Net na parte de redução da imagem. A PPM presente no trabalho se divide em 4 ramos, cada um deles com um tamanho de núcleo diferente, sendo eles 1,3,5 e 7, esses núcleos presentes nas convoluções também são acompanhados de camadas de Batch normalization, ReLU e max pooling no final, sendo as camadas as mesmas em cada galho a única diferença se dá no tamanho do núcleo. No final do PPM essas imagens de tamanhos diferentes são concatenadas todas juntas para passar para a outra camada.

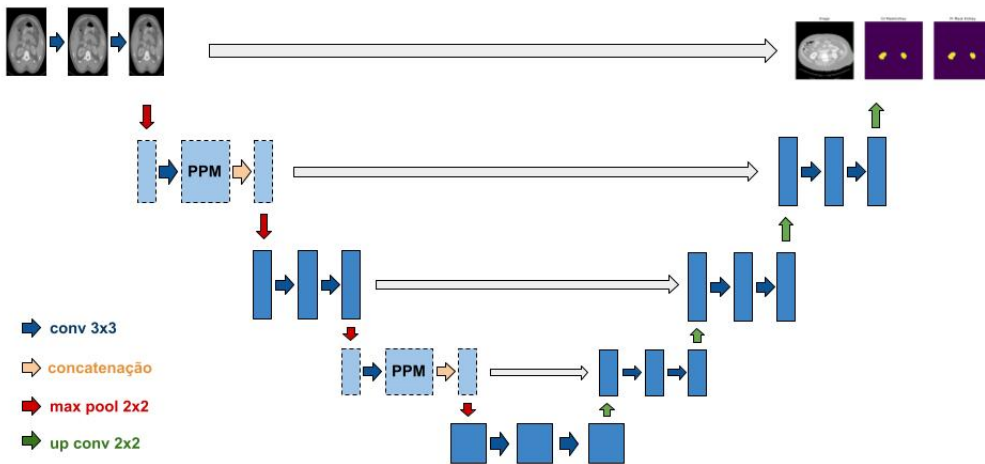


Figura 3. Modelo PPM-UNet proposto pelo método

2.3. Avaliação

Para a avaliação do modelo foi pensado em utilizar 50 épocas, se seguindo a sequência dos pacientes presente no Kits19. A cada 10 épocas, o resultado da rede é avaliado para verificação do aprendizado.

As métricas utilizadas na avaliação do modelo foram as métricas de Iou (Intersection over Union) que serve para medir o acerto de um detector de objeto em um base de dados, usando bounding boxes. Sendo calculado usando a área de interesse sobre a área de união como mostra na Equação 1.

$$IoU = \frac{P \cap Gt}{P \cup Gt} \quad (1)$$

Outra métrica utilizada foi o Dice mostrado na equação 4, que é obtido por um calculo entre o recall e o precision, levando em consideração o verdadeiro positivo(TP) e ambos os falsos positivos(FP) e os falsos negativos(FN), como mostra na equação 2 e 3.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Dice = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \quad (4)$$

3. Resultados

Nessa seção, é apresentada os resultados dos experimentos realizados. O método foi aplicado usando a biblioteca Python. Utilizou-se a biblioteca Keras, com o Tensorflow

GPU de back-end. A placa de vídeo utilizada foi a NVIDIA GeForce RTX 3060.

As imagens pré-processadas foram divididas pelo número identificador dos pacientes, sendo que todas as imagens de um determinado paciente fará parte de apenas um dos conjuntos. Foram formados três conjuntos: treino, validação e teste. Um total de 70% das imagens foram alocadas para treino, 20% para teste e 10% para validação. O modelo foi treinado com o conjunto de treino, e avaliado a cada época com o conjunto de validação. A cada 10 épocas, o modelo foi avaliado com a base de teste, onde no momento comparado a rede U-Net padrão, treinada com a mesma base.

Quando a rede foi executada a cada 10 épocas, o crescimento dela a cada época teve um ritmo muito similar a U-Net padrão como apresentado na Figura 4,. No entanto no resultado final a nossa rede teve um resultado superior a U-Net no final das 50 épocas, como mostra na Tabela 1.

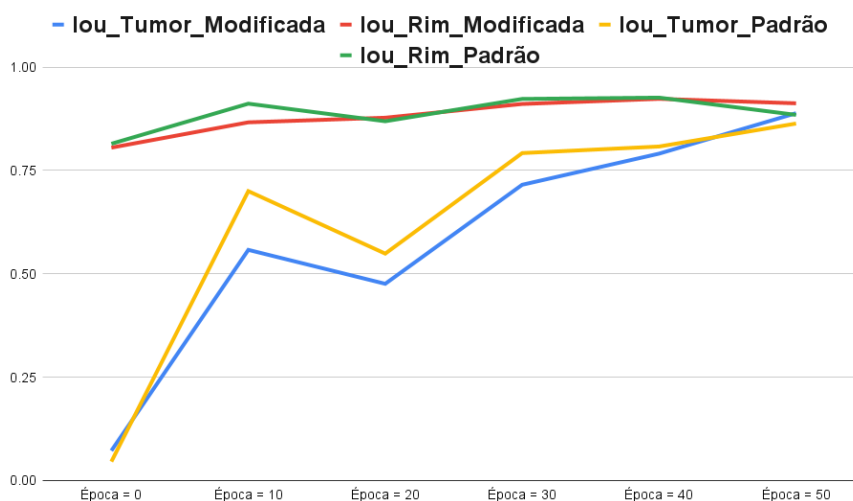


Figura 4. Comparação a cada 10 épocas

Tabela 1. Experimento a cada 10 épocas

Rede	Iou-Score	Iou-Tumor	Iou-Rim	Dice
PPM-UNet	0,93	0,88	0,91	0,94
U-Net	0,92	0,86	0,88	0,93

Tanto a Tabela 1 quanto a Figura 4 demonstram que a rede proposta é capaz de convergir mais rápido, e ainda aprender com mais eficiência dentro do problema.

O resultado final, após às 50 épocas para a PPM-UNet é apresentado na Tabela 2. Verifica-se que o IoU obtido para segmentação do RIM chegou a 0,91, e 0,88 para a segmentação das áreas de tumor.

Tabela 2. Experimento na sequência do Kits19

Rede	Iou-Score	Iou-Tumor	Iou-Rim	Dice
PPM-UNet	0,93	0,88	0,91	0,94

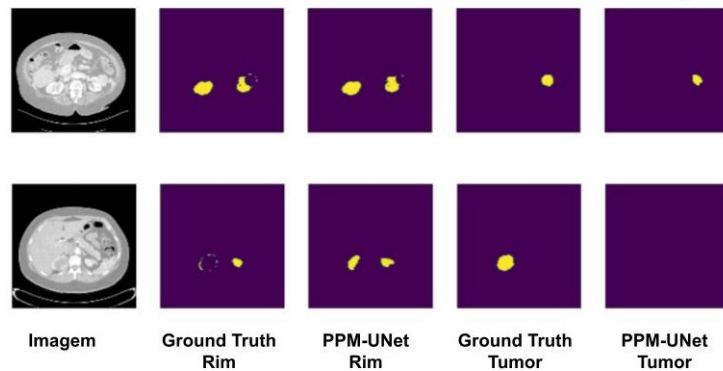


Figura 5. Resultado comparativo entre máscara e saída da rede

Levando em conta que o resultado obtido mostrado na Tabela 2 foi superior a U-Net padrão, se percebe que a U-Net pode ser melhorada sofrendo pequenas mudanças no seu campo de contração. Também é possível se observar que a rede tem um bom grau de aprendizado ao se analisar a Figura 5, onde se tem a imagem de tomografia computadorizada em primeiro, ao lado se possui a máscara do Kits19, referente ao rim, em seguida a imagem de saída de PPM-UNet referente ao rim, e por fim se tem a máscara do Kits19 referente ao tumor, e a saída PPM-UNet em relação ao tumor. Pode se observar na imagem que no primeiro caso ela se saiu bem, acertando grande parte do rim e do tumor, no entanto no segundo caso a rede não foi capaz de diferenciar o rim do tumor. Perceba que neste caso, um dos rins é praticamente ausente, o que pode ter levado a rede a entender de maneira errada a área de tumor que está dentro dele.

Comparando-se com os trabalhos publicados na literatura, apresentado na Tabela 3, observa-se na que os trabalhos de [Shahin et al. 2019] e [Hafhouf et al. 2020] que utilizaram o PPM na construção da sua rede possui resultados comparáveis a rede PPM-UNet. Verificamos que os resultados obtidos são comparáveis aos obtidos por [Isensee and Maier-Hein 2019], com ligeira vantagem para a PPM-UNet, deixando em vista que é possível aplicar melhorias na U-Net para melhorar sua assertividade.

Tabela 3. Comparação com outros trabalhos

Rede	Jaccard	Jaccard-Tumor	Jaccard-Rim	Dice
[Isensee and Maier-Hein 2019]	-	-	-	0,91
[Shahin et al. 2019]	0,83	-	-	0,90
[Hafhouf et al. 2020]	0,82	-	-	0,89
PPM-UNet	0,93	0,88	0,91	0,94

4. Conclusão

Nesse trabalho foi produzido um método automático para segmentação de rim e tumor em imagens de tomografia computadorizada, utilizando uma rede que tem como base uma U-Net padrão, sendo a modificação a adição do PPM em duas camadas da contração. A rede obteve um Iou no Tumor de 0,88 e no Rim de 0,91. Obteve resultados melhores em comparação com a UNet padrão e resultados enquadrados dentro o estado da arte.

Esse resultado mostra que a adição do PPM no lugar dos blocos de convolução aumenta a assertividade da rede, possuindo espaço pra outras melhorias.

A inserção dos blocos PPM na rede se deu de maneira empírica, onde vários experimentos foram realizados para checar a adequação do processo. O problema com esta condução consiste no tempo para se realizar as adaptações. Entendemos que a proposição de uma automatização do processo melhoraria sobremaneira a descoberta da arquitetura.

Como trabalhos futuros se pretende estudar novas modificações da rede U-Net, estudando de outras modificações positivas a rede, ou até mesmo uma junção do PPM com outras modificações e otimizar de maneira automática os locais onde devem ser inseridos tais blocos na rede.

Agradecimentos

Os autores agradecem às instituições brasileiras que ajudaram no desenvolvimento desta pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) pelo financiamento Apoio.

Referências

- da Cruz, L. B., Araújo, J. D. L., Ferreira, J. L., Diniz, J. O. B., Silva, A. C., de Almeida, J. D. S., de Paiva, A. C., and Gattass, M. (2020). Kidney segmentation from computed tomography images using deep neural network. *Computers in Biology and Medicine*, 123:103906.
- de Nefrologia, S. B. (2022). Compreendendo os rins. Disponível em: <https://www.sbn.org.br/o-que-e-nefrologia/compreendendo-os-rins/>.
- do Câncer, I. N. (2022). Ministério da saúde amplia tratamento para câncer renal. Disponível em: <https://www.inca.gov.br/noticias/ministerio-da-saude-amplia-tratamento-para-cancer-renal>.
- Hafhouf, B., Zitouni, A., Megherbi, A. C., and Sbaa, S. (2020). A modified u-net for skin lesion segmentation. In *2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*, pages 225–228.
- Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., Dean, J., Tradewell, M., Shah, A., Tejpal, R., Edgerton, Z., Peterson, M., Raza, S., Regmi, S., Papanikolopoulos, N., and Weight, C. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. Disponível em: <https://github.com/neheller/kits19>.
- Isensee, F. and Maier-Hein, K. H. (2019). An attempt at beating the 3d u-net.
- Jatobá, A., Lima, L., Amorim, L., and Oliveira, M. (2020). Cnn hyperparameter optimization for pulmonary nodule classification. In *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 25–36, Porto Alegre, RS, Brasil. SBC.
- Long, J., Shelhamer, E., and Darrell, T. (2014). Fully convolutional networks for semantic segmentation.

- Oncoguia (2022). Taxa de sobrevida para câncer de rim.
- O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.
- Shahin, A. H., Amer, K., and Elattar, M. A. (2019). Deep convolutional encoder-decoders with aggregated multi-resolution skip connections for skin lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 451–454.
- Sírio-Libanês, H. (2022). Câncer de rim: Saiba mais. Disponível em: <https://www.hospitalsiriolibanes.org.br/blog/urologia/cancer-de-rim-saiba-mais>.
- Trevisan, V., Rodrigues, D., and Rezende, E. (2021). xrayaid detecting pneumonia using artificial intelligence. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 1–12, Porto Alegre, RS, Brasil. SBC.
- Vogado, L., Veras, R., Araujo, F., Silva, R., and Aires, K. (2019). Rede neural convolucional para o diagnóstico de leucemia. In *Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 46–57, Porto Alegre, RS, Brasil. SBC.
- Yang, G., Li, G., Pan, T., Kong, Y., Wu, J., Shu, H., Luo, L., Dillenseger, J.-L., Coatrieux, J.-L., Tang, L., and Zhu, X. (2018). Automatic segmentation of kidney and renal tumor in ct images based on 3d fully convolutional neural network with pyramid pooling module. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3790–3795.
- Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2016). Pyramid scene parsing network.