Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 32, Num. 1 (2025) 61-66

RESEARCH ARTICLE

Renal Cancer Classification in Tomographic Images Using Convolutional Neural Networks to Assist Early Diagnosis

Classificação de Câncer Renal em Imagens Tomográficas Utilizando Redes Neurais Convolucionais para Auxiliar no Diagnóstico Precoce

Vinícius Meireles Pereira Santos[1]*, Ana Claudia Patrocinio[2], William Chaves de Souza Carvalho[3], Pedro Moises de Sousa[1]

Abstract: Renal cancer is one of the 13 most frequent types of cancer in Brazil, primarily affecting individuals between 50 and 70 years of age, with around 12,000 cases and 4,000 deaths recorded in 2020. Early detection is crucial to prevent the progression of the disease, which can lead to kidney transplants, dialysis, or death. However, the lack of symptoms and screening tests make diagnosis difficult. This research proposes the use of Convolutional Neural Networks (CNNs) for the detection of renal cancer in CT images, classifying them as cancerous or normal. It is expected that the application of the obtained results can assist in early diagnosis, supporting healthcare professionals in providing faster and more accurate treatments and contributing to improving the patients' quality of life.
Keywords: Convolutional Neural Network (CNN) — Kidney Cancer — Classification of Kidney Images

Resumo: O câncer renal é um dos 13 tipos mais frequentes de câncer no Brasil, afetando principalmente indivíduos com idade entre 50 e 70 anos, com aproximadamente 12.000 casos e 4.000 mortes registradas em 2020. A detecção precoce da doença é crucial para evitar sua progressão, que pode levar à necessidade de transplantes de rins, hemodiálise ou, em casos graves, à morte. No entanto, a falta de sintomas e de exames de rastreamento eficazes que permitam a identificação precoce tornam o diagnóstico mais difícil. Esta pesquisa propõe o uso de Redes Neurais Convolucionais (CNNs) para a detecção de câncer renal em imagens tomográficas, classificando-as como cancerosas ou normais (sem a presença da doença). Espera-se que a aplicação dos resultados obtidos possa auxiliar na detecção precoce da doença, ajudando os profissionais de saúde a fornecer tratamentos mais precisos e rápidos, contribuindo para a qualidade de vida dos pacientes.
Palavras-Chave: Rede Neural Convolucional (CNN) — Câncer Renal — Câncer de Rins — Classificação de Imagens Renais

[1] *Institute of Exacts and Technological Sciences – Federal University of Viçosa (UFV), Brasil*
[2] *Biomedical Lab, Faculty of Electrical Engineering - Federal University of Uberlândia (UFU), Brasil*
[3] *FACOM Computation Faculty – Federal University of Uberlândia (UFU), Brasil*
***Corresponding authors**: [1]{vinicius.m.santos, pedromoises}@ufv.br, [2,3]{ana.patrocinio, william.carvalho}@ufu.br
DOI: http://dx.doi.org/10.22456/2175-2745.143539 • **Received:** 24/10/2024 • **Accepted:** 02/12/2024

1. Introduction

The kidneys are vital organs of the human body, responsible for filtering and eliminating metabolized substances from the body, producing hormones that stimulate the production of red blood cells, and helping regulate blood pressure.
More common in people aged 50 to 70, kidney cancer is already among the 13 most frequent types of cancer in Brazil, with around 13,000 cases and over 4,000 deaths recorded in 2020 due to this type of cancer [1]. Although kidney cancer is not among the ten most frequent oncological diseases worldwide, the neoplasm is considered the most lethal type of urological cancer [13].
According to [13], the absence of clear symptoms and the lack of screening tests for kidney cancer make early diagnosis a challenge, reducing the chances of effective treatment. This highlights the need for supportive tools that can detect the disease at early stages and assist in defining therapeutic approaches.
With the advancement of machine learning technologies, tools such as Convolutional Neural Networks (CNNs) have shown great potential in analyzing medical images, enabling the detection of patterns that are imperceptible to the human eye.

This advancement may help fill the gap left by the lack of effective methods for the early detection of kidney cancer [7]. For this reason, this work proposes the identification and classification of kidney cancer in tomographic images, supported by Convolutional Neural Networks. The main objective is to assist healthcare professionals in the early identification of cancer to facilitate decision-making for each individual case and contribute to more accurate diagnoses, which may provide patients with more appropriate and targeted treatments, increasing the chances of better outcomes and improving quality of life.

The following sections of this article are structured as follows: Section II presents a review of the current literature on the subject and analyzes works similar to what is being proposed. Section III describes the Materials and Methods used in this research, detailing the database, each of the models tested, the criteria for pre-processing the dataset, the test parameters, as well as the metrics used for model evaluation. In Section IV, we present the results obtained from the tests performed and briefly analyze them. Finally, in Section V, we conclude the article with a discussion on the results obtained and reflect on possible next steps for future work.

2. Related Work

The work [6] presents the creation of a CNN (Convolutional Neural Network) model for classifying images collected from the Kaggle platform [2] and an *ANN* (Artificial Neural Network) for classifying blood samples, with the aim of identifying kidney cancer and predicting the disease. For image preprocessing, the authors used resizing the images to 512x512x3 and applying a 3x3 filter. The results obtained by the research were - Training Accuracy: *CNN* - 96.4%, *ANN* - 99%; Validation Accuracy: *CNN* - 99.6%, *ANN* - 97%.

The work [5] used the *Inception V3* network, a Deep Learning Neural Network (*DLNN*), to predict stage one kidney cancer versus other stages. It was trained using Computed Tomography (CT) images, originally in 3D, from which 2D slices were extracted to compose the dataset. The databases used were TCGA-KIRKC from the Cancer Imaging Archive and the TCIA (maintained by the National Cancer Institute, Bethesda, Maryland, USA). In addition to extracting slices from the original 3D images, the authors resized the generated images to match the input size required by the DLNN and balanced the dataset so that the class percentages were equal. After training the network for 4,000 epochs, the best accuracy result occurred at epoch 400 (best model), with a learning rate of 0.01, achieving 91% on the validation set, 97% on training, and 90% on testing.

In [4], the authors proposed a deep learning framework that aims to segment 3D tumor and kidney stone images at a pixel level, perform multiphase alignment, and categorize kidney tumor subtypes. The dataset consisted of images from eligible patients from Seoul St. Mary's Hospital and images from the TCIA dataset for validation. The network used for training was ResNet-101 for classifying renal cancer subtypes, with the addition of a 1x1x1 convolutional layer at the start of the network and a modification to the last convolutional layer to balance the distribution among the five classes. The output showed the probability for each subtype. As a result, the model applied by the researchers achieved an average accuracy of 89.9%, surpassing the predictions of radiologists.

3. Materials and Methods

3.1 Dataset

The dataset was collected from the Kaggle platform [2] and contains 12,446 unique computed tomography (CT) images, of which 3,709 are of cysts, 5,077 are of normal kidneys, 1,377 are of kidneys with stones, and 2,283 are of kidneys with cancer. This research was conducted by training the models with images of normal kidneys and kidneys with cancer. The classification output consists of two classes: normal and tumor. Figure 1 presents sample images from the dataset.

Figure 1. Exemples of images from the dataset

3.2 Architectures

For this work, we trained four deep learning architectures: the CNNs ResNet50, AlexNet, MobileNet V2, and VGG-16.

ResNet50 (Residual Neural Network) is a Convolutional Neural Network with 50 deep layers.[1] A ResNet is an Artificial Neural Network that stacks residual blocks to form a network. This neural network architecture was first introduced by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in 2015 in their work Deep Residual Learning for Image Recognition [9].

AlexNet is a neural network that won the ImageNet Large Scale Visual Recognition Challenge 2012 by a large margin, consisting of five convolutional layers and three fully connected layers.[2] This network was the first to show the possibility of extracting features beyond those manually designed, breaking a paradigm in computer vision. It innovated by using ReLU as the activation function and GPU to process the network training [10]. [3]

MobileNet V2 is a computer vision model focused on mobile devices, meaning it operates with lower computational power.[4] This class of open-source Convolutional Neural Network uses depthwise separable convolutions, which significantly reduces the number of parameters compared to regular convolutions and the same depth in networks. Developed by Google Inc., the first version was released in 2017 [11]. The version used in this work was the second version.

VGG Networks are Convolutional Neural Networks(CNNs) designed with multiple layers. The abbreviation VGG stands for Visual Geometry Group, and each variation of a VGG

[1] ⟨https://viso.ai/deep-learning/resnet-residual-neural-network/⟩
[2] ⟨https://viso.ai/deep-learning/alexnet/⟩
[3] ⟨https://d2l.ai/chapter_convolutional-modern/alexnet.html⟩
[4] ⟨https://builtin.com/machine-learning/mobilenet⟩

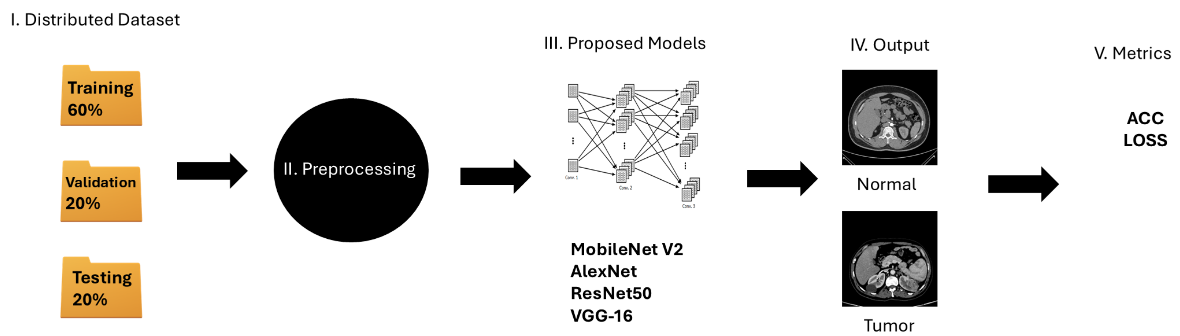


Figure 2. Overall Methodology Diagram

network differs in the number of convolutional layers the network contains. In the case of the network used in this work — VGG-16 — it has 16 convolutional layers. The model supporting these 16 layers (VGG16) was developed by A. Zisserman and K. Simonyan from the University of Oxford [12]. The VGG-16 can classify photos into up to 1000 different object categories; the model accepts images with a resolution of 224x224.7 [5].

3.3 Inclusion and Exclusion Criteria in the Dataset

For this study, as previously mentioned, we used images from the dataset that contain normal kidneys and kidneys with cancer. In the dataset, there are 5,077 images of the normal class and 2,283 images of the tumor class. In order to balance the dataset and maintain proportionality between each class, as performed by Nathan Hadjiyski in [5], 2,283 images from the normal class were randomly selected, along with all 2,283 images from the tumor class. After balancing, the dataset was divided into three parts: 60% of the images for training, 20% for validation, and 20% for testing. This division is essential for the correct classification between normal and tumor classes. Figure 3 represents the entire process of selection, balancing, and dataset division.

3.4 Preprocessing

The images in the dataset had to be preprocessed for more efficient model training. Originally, the images are 24-bit and have 3 color channels. The preprocessing consisted of converting all selected images to 8-bit grayscale. Figure 2 presents a summary of the methodology used in this work.

3.5 Training and Evaluation Parameters

We defined basic parameters in the code for executing training, validation, and testing, such as a batch size of 32, a SEED of 42, and a learning rate of 0.0001. We used Cross-Entropy Loss as the loss function, and ADAM from the PyTorch library was chosen as the optimizer. Tests were performed on all models using 100 epochs.

[5] ⟨https://encurtador.com.br/Xb51a⟩

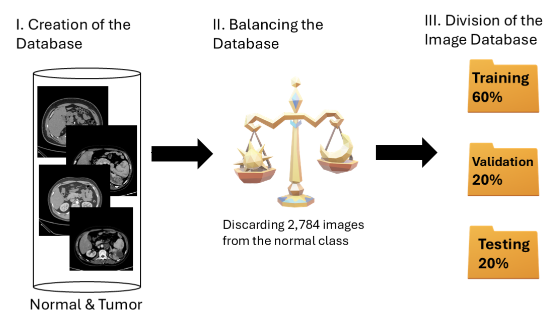


Figure 3. Dataset Balancing and Division

3.6 Model Evaluation

We evaluated the models, which were trained on the training, validation, and test sets, using accuracy and loss (Cross-Entropy Loss). Based on the data obtained, we generated accuracy and loss graphs, as well as a confusion matrix for each model. By analyzing the metrics obtained, it is possible to extrapolate the classification capability of the models for unseen samples.

The accuracy and loss are defined in Equations 1 and 2, respectively:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right] \tag{2}$$

Where TP is the True Positive, TN is the True Negative, FP is the False Positive, FN is the False Negative, N is the number of samples in the batch, y_i is the label of the i-th sample, and p_i is the predicted probability for the positive class in the i-th sample.

3.7 Computational Resources

The tests were carried out in the Google Colab environment, using a NVIDIA Tesla T4 GPU with 15 GB of dedicated memory and 12.7 GB of system memory; tests were also conducted on a computer with an Intel Core i5-13450HX processor and 16 GB of DDR5 memory, with an NVIDIA GeForce RTX 3050 Ti GPU with 6 GB of dedicated memory. The development environments were all based on the Python programming language, using PyTorch with CUDA (in the case of the second environment, versions 2.4.1 and 12.1, respectively). The TorchVision, NumPy, Scikit-Learn, and MatPlotLib libraries were also used.

4. Results and Discussion

Table 1 contains the results obtained with each of the dataset sets — training, validation, and testing, respectively — for each of the models used: ResNet50, MobileNet V2, AlexNet, and VGG-16. The table shows the accuracy (ACC) and loss (LOSS) recorded after running 100 epochs for each of the models, with a batch size of 32 for the training and validation sets, and the evaluation of the best model and best weights using a batch size of 1 with the test set. In Figure 4, we present the training and validation accuracy graph generated after running 100 epochs for each model; Figure 5 shows the training and validation loss graph for each of the models; and in Figure 6, we display the Confusion Matrices generated after evaluating the models using the test set.

Table 1. Performance of the Architectures with ACC and LOSS for Different Sets

Architecture	Set	ACC (%)	LOSS (%)
ResNet50	Training	91.10%	26.57%
	Validation	92.90%	24.49%
	Test	95.45%	20.50%
MobileNet V2	Training	**96.75%**	10.20%
	Validation	**98.78%**	**06.10%**
	Test	**99.89%**	01.94%
AlexNet	Training	95.72%	**10.78%**
	Validation	97.34%	08.51%
	Test	99.67%	01.48%
VGG-16	Training	95.20%	13.08%
	Validation	96.56%	08.27%
	Test	99.56%	**01.32%**

As a result of the tests, we can observe that in all cases, using any of the models or data set sets, the accuracies are always above 90%. In all the networks used, the behavior observed was an improvement in results after training, with consistently increasing values when evaluating the model with the validation and test sets.

The best result was obtained by testing the MobileNet V2 model, achieving an accuracy of 99.89% when evaluating the model with the test set and a loss of 1.94%. A slightly lower result was observed with AlexNet, which had an accuracy of 99.67%, and with VGG-16, whose best recorded accuracy was 99.56% with a loss of 1.32%. In contrast, with ResNet-50, we obtained the worst results from our tests, with the best accuracy of the model reaching 95.45%. This model also recorded the highest loss values across all sets, with the best record being 20.50%.

The study achieved better results when compared to other works, as presented in Table 2. In the model where we obtained the best results, MobileNet V2, the training accuracy was 96.75%. In comparison, the CNN tested in [6] achieved a training accuracy of 96.40%, and all the models compared in [7] reached a maximum training accuracy of 97%. When comparing the test and validation accuracies, *MobileNet V2* reached a test accuracy of 99.89%, surpassing the validation accuracy of 99.60% achieved in [6] with the used model. Furthermore, the best test accuracy reported in [7] was 97.70% with the CNN-4 model.

5. Limitations on the Study

The present work has some limitations that should be acknowledged. First, the study utilized only one public database for training, testing, and validation purposes, which may limit the generalizability of the results. Additionally, as this represents an initial phase of research, the current implementation serves as a prototype, with future phases planned to incorporate additional databases and potential heat map implementations to assist radiologists in cancer identification.

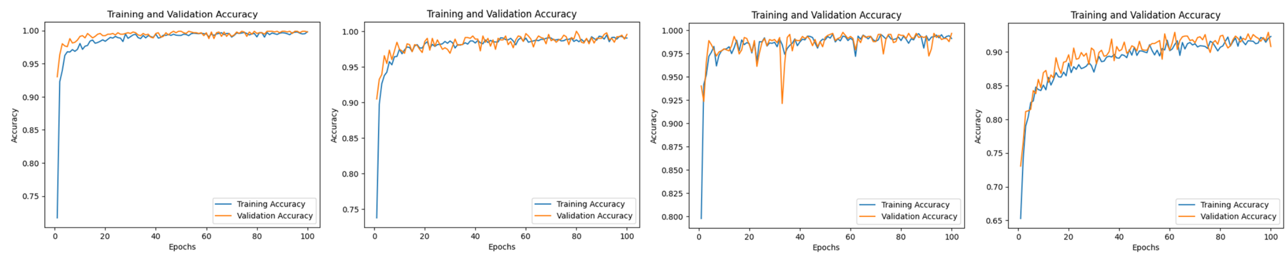
6. Conclusion

In this work, we conducted various analyses using a medical image dataset to identify the presence or absence of cancer in the kidneys. To achieve this, we utilized four pre-trained Convolutional Neural Networks (CNNs): ResNet50, AlexNet, VGG-16, and MobileNet V2. The preprocessing strategy consisted solely of normalizing the images in the dataset, followed by balancing the dataset and dividing it into training, validation, and test sets. Using this strategy and the aforementioned parameters, including training for 100 epochs, we obtained results averaging 96.58% accuracy and a mean loss of 11.10%. However, the ResNet50 model performed below average in all execution scenarios.

The results demonstrate that MobileNet V2, AlexNet, and VGG-16 models are promising tools to assist healthcare professionals in classifying tomographic images for kidney cancer identification, potentially providing more accurate prognoses and aiding in clinical decision-making. The implications of this study are significant as it aims to reduce false negatives in kidney cancer detection, potentially increasing patients' survival chances. The developed system is designed to support radiologists in their decision-making process, not

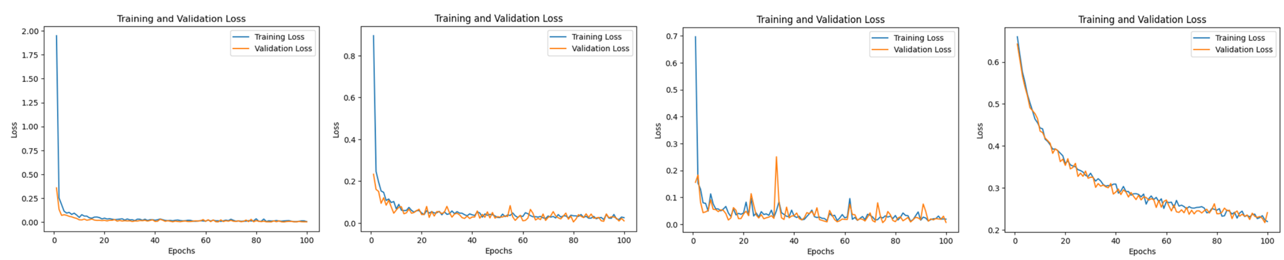
Table 2. Comparison of Accuracy Achieved by the Proposed Methods with State-of-the-Art Methodology

Metrics	**This Work**				**[6]. 2023**	**[7]. 2022**			
	ResNet50	***MobileNet V2***	***VGG-16***	***AlexNet***	***CNN***	***VGG-16***	***ResNet50***	***CNN-6***	***CNN-4***
Training Accuracy	91.10%	96.75%	95.20%	95.72%	96.40%	60.00%	96.00%	97.00%	92.00%
Validation Accuracy	92.90%	98.78%	96.56%	97.34%	99.60%	–	–	–	–
Test Accuracy	95.45%	99.89%	99.67%	99.67%	–	59.30%	97.40%	95.30%	97.70%



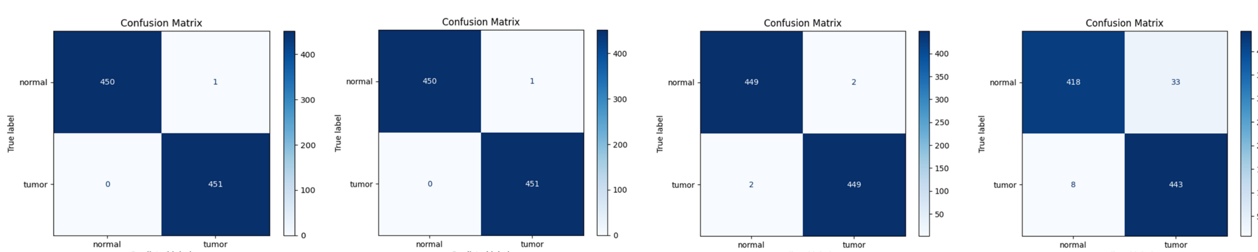
MobileNet V2 AlexNet VGG-16 ResNet50

Figure 4. Training and Validation Accuracies of All Models After Running 100 Epochs



MobileNet V2 AlexNet VGG-16 ResNet50

Figure 5. Training and Validation Losses of All Models After Running 100 Epochs



MobileNet V2 AlexNet VGG-16 ResNet50

Figure 6. Confusion Matrices of All Models Generated After Validation on the Test Set

to replace their expertise. The ultimate expected impact of this research is to save more lives through improved cancer detection accuracy.

Future work includes expanding the evaluation by incorporating additional metrics such as Recall, Precision, F1-Score, and others. We also plan to investigate improvements through image preprocessing techniques like Gaussian filtering and other noise reduction methods. An important enhancement will be the implementation of Class Activation Maps to provide visual explanations of the model's decisions, making the system more interpretable for medical professionals. Additionally, we aim to validate our approach on multiple kidney cancer databases to enhance the robustness and generalization capability of the system. These improvements will help develop a more comprehensive and reliable tool for clinical applications.

Acknowledgements

All authors express their gratitude to the Federal University of Viçosa (UFV), Rio Paranaíba Campus, and the Federal University of Uberlândia (UFU) for the institutional support provided throughout this research. Their contribution, through infrastructure and academic resources, was fundamental to the successful completion of this work.

Additionally, Pedro Moises de Sousa, professor and supervisor of this research, and Vinícius Meireles Pereira Santos would like to thank their colleagues from the scientific initiation group at the UFV Rio Paranaíba campus for their valuable discussions, feedback, and support throughout the development of this research. Their collaboration and shared insights greatly contributed to the work.

Author contributions

Vinícius M. P. Santos: Conceptualization, Methodology, Code Development, Experiments, Analysis of Results, Writing (Original Draft and Review & Editing).
Ana Claudia Patrocinio: Assisted in Conceptualization, Reviewed and Approved the Final Manuscript, Supervision.
William C. S. Carvalho: Assisted in Conceptualization, Reviewed and Approved the Final Manuscript, Supervision.
Pedro Moises de Sousa: Conceptualization, Methodology, Code Development, Analysis of Results, Writing (Review & Editing), Supervision.

References

1 ”Câncer de rim”. Pfizer Brasil. [Online]. Avaliable: Câncer de Rim [Acessed on Augusr 8, 2024].

2 ”CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone”. Kaggle. [Online]. Avaliable: CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone [Acessed on September 4, 2024].

3 ABDELRAHMAN, Abubaker; VIRIRI, Serestina. Efficientnet family u-net models for deep learning semantic segmentation of kidney tumors on ct images. Frontiers in Computer Science, v. 5, p. 1235622, 2023.

4 UHM, Kwang-Hyun et al. Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography. NPJ precision oncology, v. 5, n. 1, p. 54, 2021.

5 HADJIYSKI, Nathan. Kidney cancer staging: Deep learning neural network based approach. In: 2020 International Conference on e-Health and Bioengineering (EHB). IEEE, 2020. p. 1-4.

6 RAJKUMAR, K. et al. Kidney Cancer Detection using Deep Learning Models. In: 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2023. p. 1197-1203.

7 ALZU'BI, Dalia et al. Kidney tumor detection and classification based on deep learning approaches: a new dataset in CT scans. Journal of Healthcare Engineering, v. 2022, n. 1, p. 3861161, 2022.

8 ABDELTAWAB, Hisham et al. A pyramidal deep learning pipeline for kidney whole-slide histology images classification. Scientific reports, v. 11, n. 1, p. 20189, 2021.

9 HE, Kaiming et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

10 Russakovsky, O., Deng, J., Huang, Z., Berg, A. C., Fei-Fei, L. (2013). Detecting avocados to zucchinis: what have we done, and where are we going? International Conference on Computer Vision (ICCV).

11 HOWARD, Andrew G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

12 SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

13 ”Câncer renal: doença é silenciosa, mas pode ser prevenida”. Portal Drauzio Varella. [Online]. Avaliable: Câncer Renal [Acessed on September 6, 2024].