Automatic Kidney Stone Detection using Low-cost CNN with Coronal CT Images

Murillo Freitas Bouzon[1], Samuel Patrício de Oliveira[2], Oscar Eduardo Hidetoshi Fugita[3],
Paulo Sergio Silva Rodrigues[4]
[1, 2, 4]*Centro Universitário FEI*, São Bernardo do Campo, Brazil
[3]*Hospital Universitário da USP*, São Paulo, Brazil
mfbouzon@fei.edu.br[1]

***Abstract*—A fast diagnosis of kidney stones is crucial to start the correct treatment, minimizing the risks of urinary complications. Machine learning approaches are valuable for an automatic diagnosis system for kidney stones from computer tomography (CT) exams. Recently, related studies achieved high accuracy in detecting kidney stones using deep-learning neural networks. However, their approaches were highly complex and time-consuming. This paper proposes a method for automatically detecting kidney stones on CT images using low-complexity deep-learning techniques. We compared three models based on Convolutional Neural Networks (CNN): 3-layered CNN (*Conv3*), 4-layered CNN (*Conv4*), and *MobileNetV1*. They were applied to kidney stone detection using 1799 CT images divided into 80% for training, 10% for validation, and 10% for testing. The proposed model *Conv4* obtained the best performance, achieving a test accuracy of 97.2% and an F1-score of 97.6%, with a training time of 140 seconds.**

***Index Terms*—Kidney Stone Detection, Deep Learning, Convolutional Neural Network, MobileNetV1, Computed Tomography**

I. INTRODUCTION

Kidney stone is a common disease with an increased incidence rate in recent decades [1], that consists of the aggregation of minerals in the urine, forming solid substances located in the kidneys. These stones may lead to severe pain and permanent damage to the patient's kidneys [2], [3]. Fast detection and early diagnosis of these stones are essential to prevent complications and initiate an efficient treatment strategy [4].

Imaging tests are a common practice to diagnose patients suspected of kidney stones. Computed tomography (CT) is considered the gold standard for diagnosing stones because of its speed, ease of image collection, the details of the patient's internal organs, and the possibility of diagnosis of other abnormalities in the urinary system [5]. CT scans also provide detailed information on the patient's organs, allowing manipulation and 3D visualization of the patient's interior.

CT scans not only allow the detection of kidney stones but also provide characteristics of the stone, such as its location, size, density, and associated complications. Based on this information, doctors decide the best treatment strategy for the patient. However, this process can be time-consuming due to the excessive number of tests needed to be done. Additionally, errors in diagnosis can occur due to the lack of experience of the doctor in charge and the characteristics variability of the tests used.

Recently, the development of deep-learning neural network techniques applied to computer vision has benefited the medical field, achieving success in applications such as intelligent diagnostic systems [6] and medical image segmentation [7], for example. These techniques are also applied to solve problems of urology, such as the automatic diagnosis of kidney stone disease from image examinations [8], analysis of stones composition [9], prediction of treatments and surgical consequences [10].

Automatic detection of kidney stones on imaging exams is a fundamental task in developing an intelligent kidney stone diagnosis system. Usually, to perform the detection, image scans are used as input to teach a learning model to differentiate between scans of healthy patients and patients who have stones in one or both of their kidneys. Recently, other works have used machine learning techniques to detect the presence of kidney stones in CT scans.

[11] used machine learning with neural networks to detect kidney stones automatically in imaging exams. A CT image is received and preprocessed using the Discrete Wavelet Transform. [12]. Then, the method transforms the image into a feature vector extracted using the GLCM (Gray-Level Co-Occurrence Matrix) technique. An MLP (Multi-layer Perceptron) neural network [13] is used to classify each extracted feature vector, with samples with kidney stones considered abnormal and samples without kidney stones considered normal. The method obtained a classification accuracy of 98.8% on tests with 20 samples.

Two other papers also addresed an automatic system for detecting kidney stones. [14] used GLCM for feature extraction and an MLP neural network for classification. However, they pre-processed the input images with a Gaussian filter and bilateral filtering to smooth the inputs. The results of the experiments showed that the method obtained an accuracy of 96% in the classification. [15] applied a histogram equalization to the CT images used as input, followed by edge enhancement using convolutional filters. After pre-processing, the SVM (Support Vector Machine) classifier [16] was applied to differentiate kidneys with stones from those without. The authors collected 156 CT images of kidneys to test the method, 78 of which were of kidneys with stones presence, and 78 were of healthy kidneys. The method correctly detected 154 of the 156 images, achieving an accuracy of 98.71%.

Convolutional Neural Networks (CNN) became notable for solving image classification problems, including applications for medical images. [17] demonstrated the CNN application for automatic kidney stone detection. The authors trained a CNN model with 349 CT scans and tested on 88 scans, obtaining a specificity of 1.0 and an F1-score of 0.783 on their best model. [18] also used CNN for kidney stone detection, obtaining an accuracy of 90%, a sensitivity of 80%, and a F1-score of 89%. [19] explored CNN in different planes of CT images. They classified the patients into three groups based on the size of the stones, using a total of 2959 CT images to train a CNN model, obtaining a test accuracy of 63%, 72%, and 64% for each group of the coronal plane.

Other studies used approaches based on deep-learning, using highly complex neural networks and a larger dataset. [20] proposed an automatic kidney stone detector using computed tomography and deep-learning techniques. A total of 1799 2D images were acquired, of which 80% were used to train a network *XResNet-50* [21] and 20% to test the model. The model obtained an accuracy of 96.82% and 95.76% of sensitivity using 146 test samples. [22] proposed the use of *DarkNet19* [23] for the same problem. The *DarkNet19* was used to reduce the feature generation time, and INCA (*Iterative Neighborhood Component Analysis*) method [24] was applied to reduce the dimensionality of the feature vector. KNN (*K-Nearest Neighbours*) technique [25] was used in conjunction with a Bayesian optimizer to adjust its hyper-parameters for classification. Using 10-*fold* and *hold-out* cross-validation, the proposed method obtained an accuracy of 99.22% and 99.71%, respectively. [26] proposed a new CNN (Convolutional Neural Network) architecture that uses a convolution based on the product of *Kronecker* [27] to solve the problem of automatic kidney stone detection. The method obtained an accuracy of 98.56% using a 10-fold cross-validation. [28] also proposed an CNN model for kidney stone detection that achieved 99.4% of accuracy.

Despite the promising results for the automatic detection of kidney stones in CT scans, the presented deep-learning techniques are highly complex and require a high computational cost, being time-consuming on the training step. This paper investigates the use of low-cost CNN for the automatic detection of kidney stones, compared to state-of-the-art methods. The potential contributions of this paper are as follows:

- Use of a low-cost CNN architecture for automatic kidney stone detection.
- A direct comparison of different low-cost neural network architectures, including MobileNetV1.
- Evaluation of classification performance to other state-of-the-art methods.

II. Materials and Methods

The main steps of the proposed method is illustrated in Fig. 1. In the first step, the method reads the dataset of CT images and applies a preprocessing step, which consists of resizing and normalizing the intensity of the images. The preprocessed images are used as input to train deep-learning neural network models applied to kidney stone detection, classifying images into two distinct classes: normal and kidney stone.

A. Dataset

A public dataset was used (https://github.com/yildirimozal/Kidney_stone_detection) [20], and it consists of CT scans without contrast from 463 patients between 18 and 80 years old. Of the scans collected, 268 belonged to patients with kidney stones and 165 to patients considered normal. In total, 1799 coronal CT images were collected, of which 790 are samples with stones and 1009 are samples of normal patients. The images were in 8-Bits PNG format. Fig. 2 shows samples from the dataset used to train the models.

B. Preprocessing

Before the images were used to train the neural network models, we applied a preprocessing step on all CT images from the dataset. First, the images were resized to 224×224 to reduce the complexity of the networks to process the inputs. Then, we normalized the intensity of each pixel to range their values between the interval $[0, 1]$ to enhance the operation's precision that the neural network used during the training process.

C. Deep-Learning Neural Network Architectures

Three distinct low-cost neural network architectures, with less than 5 Million (M) trainable parameters, were used for kidney stone detection based on CT images. Two architectures are based on traditional CNN [29]. The first consists of 4 convolution layers 2D, followed by a max pooling layer of 2×2 size and using ReLu as activation function, and 2 dense layers for classification using 150 and 100 neurons, respectively. This architecture is referred to in this paper as *Conv4* and is illustrated in Fig. 3. The second architecture is similar to *Conv4* but has only 3 convolution layers, named *Conv3*. These architectures were chosen due to their lesser quantity of parameters compared to architectures used in other related papers. The last deep-learning architecture used was *MobileNetV1* [30] because it is considered a low-complexity deep neural network applied for fast and mobile applications. We trained three models of each architecture with the preprocessed CT images from the dataset. The hyperparameters used for training are shown in Table I.

III. Experiments and Results

The hardware used for training the networks included an Intel Core™i5-12400F 2.50GHz, 16GB DDR4, and an NVIDIA RTX 4070 12GB graphics card. For implementation, we used the programming language Python and the libraries TensorFlow, Keras and OpenCV.

The dataset, consisting of 1799 CT images, was separated into three groups: training (80%), validation (10%), and testing (10%). After that, we performed two validation strategies: 10-fold cross-validation, to generalize the error of the networks

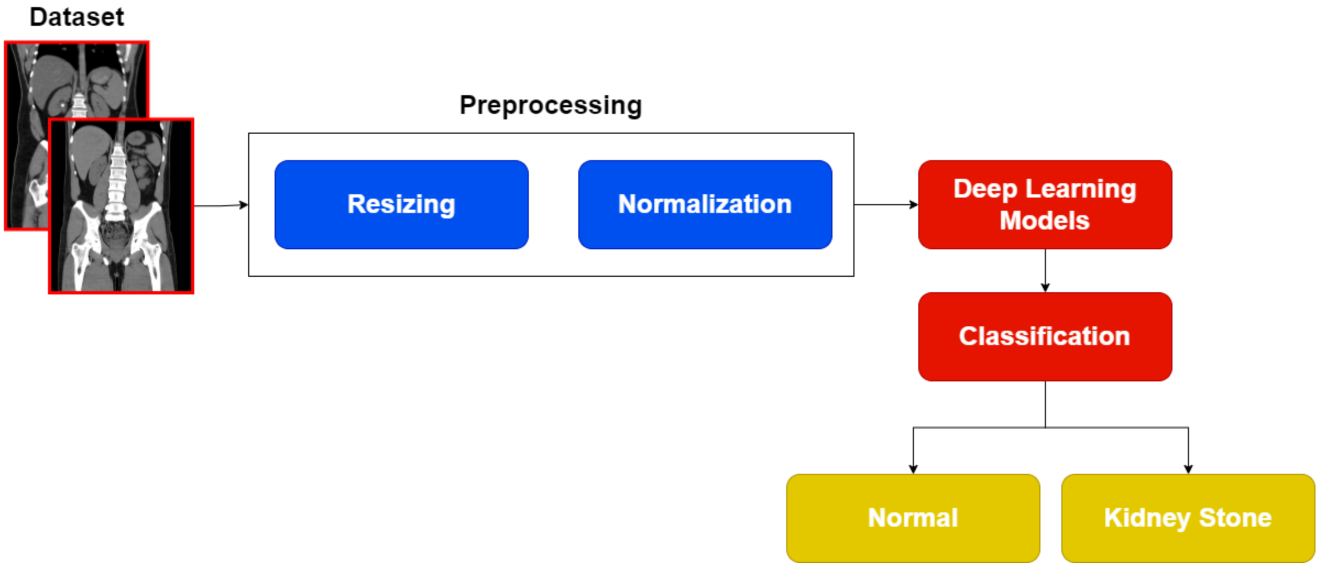

Fig. 1: The steps of the method proposed.

Fig. 2: Coronal CT scans of patients with kidney stones - (a) and (b) - and without kidney stones - (c) and (d) -

TABLE I: Hyperparameters used for training the deep-learning architectures.

Hyperparameters	Details
Epochs	150
Batch size	16
Learning rate	0.01
Optimizer	Adam
Activation function	ReLU
Loss function	Cross-entropy
Dropout	0.25

on unseen data during the training step, and test validation, to evaluate the performance of the best models of each of the three architectures. Thus, we trained the models *Conv3*, *Conv4*, and *MobileNetV1* with 1590 images (training and validation) and tested them with 209 images.

For each cross-validation fold, we evaluated the mean and the standard deviation of the following classification metrics: accuracy, precision, sensitivity, specificity, and F1-Score. The results of the 10-fold cross-validation are presented in Table II. Overall, we observed that the proposed architecture *Conv4* achieved higher values on almost every metric evaluated, except for the sensitivity, where the *MobileNetV1* achieved the highest mean value of 99.3%, with a difference of 0.4% from the sensitivity of *Conv4*. Also, the *Conv4* model obtained a value above 96% on all metrics evaluated, indicating a notable performance on kidney stone detection based on CT images.

On the test samples, we evaluated the same metrics of the 10-fold cross-validation, with the results of this experiment presented in Table III. The model *Conv4* remained the best of the three models tested, achieving values above 96% on all metrics again, and obtained a test accuracy of 97.2% and F1-Score of 97.6%. Observing the training time, the *Conv3*

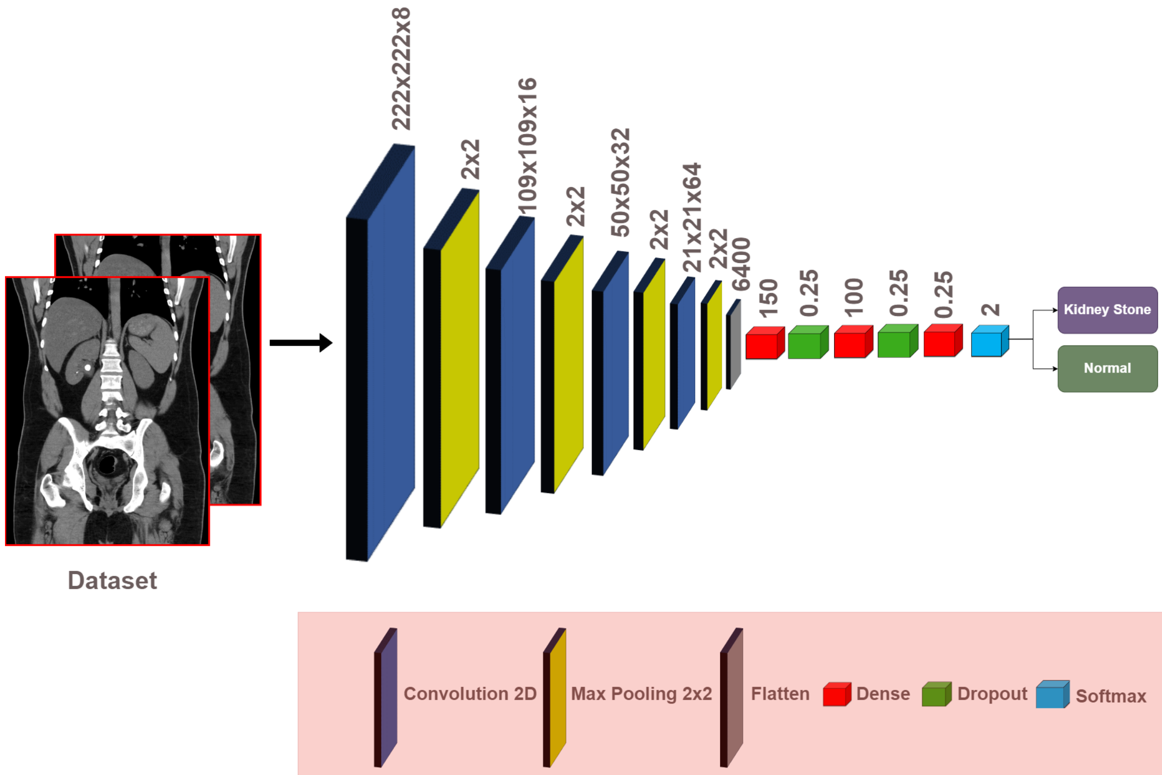

Fig. 3: Architecture of the network *Conv4*.

model was only 2 seconds faster than *Conv4* but achieved lower results, while the *Conv4* took 340 seconds less than the *MobileNetV1* to complete the training process, achieving the highest results.

IV. Discussion

We compared three deep-learning neural network architectures for automatic kidney stone detection from coronal CT images. The trained models are based on CNN and are considered low-cost neural network architectures since they have less than 5M trainable parameters. The best model was the proposed *Conv4*, which achieved a 10-fold cross-validation mean accuracy of 97.9%, test accuracy of 97.2%, and 140 seconds of training time.

Other studies also proposed deep-learning-based methods for kidney stone detection on the same dataset used in this work. We compared the classification metrics of our best-performing method with similar papers in the literature [20], [26], as shown in Table IV. Our method performed better than the method proposed by [20], the paper that introduced the dataset used, and achieved lower values compared to [26].

In addition to classification metrics, we analyzed the trainable parameters and training time in seconds of each method to compare their complexity. Table V shows this analysis. Despite our method achieving lower classification values than [26], it resulted in a training time of 140 seconds, being 2982 seconds faster. The processing time may be relative to the hardware used, however, the amount of parameters in our method is still less than a half compared to the Deep Kronecker Neural Network. Furthermore, according to [26], Kronecker convolution is more computationally expensive than traditional convolution. Also, compared to XResNet50, our method trained faster and had 23x less parameters. Other papers used the same dataset, but we did not consider them in our analysis because they did not specify the complexity of their methods.

The proposed neural network model demonstrated high classification performance, closer to other methods in the related literature, with a faster training time. According to the American Urological Association, the estimated sensitivity and specificity of clinician diagnosis of kidney stones on CT exams are $\sim 95\%$ and $\sim 98\%$, respectively [31]. Compared to the results obtained by our method, the sensitivity obtained is 1.2% higher and the specificity, 0.7%. This tiny difference indicates that the architecture proposed may aid as a clinical diagnosis tool and help to reduce the time elapsed for radiologists to detect kidney stones in CT exams.

Our paper has some limitations. The CNN architecture is relatively simple, with only 4 convolution layers and 2 dense layers, and was tested with only 180 samples from the dataset, besides the 10-fold cross-validation. These factors compromise the method's robustness, as no tests were made on other datasets. Furthermore, although the proposed architecture has low complexity, for applications that require

TABLE II: Results of the 10-fold cross-validation.

	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Conv3	$94.4 \pm 1.9\%$	$93.5 \pm 2.4\%$	$96.6 \pm 1.4\%$	$91.7 \pm 2.8\%$	$95.0 \pm 1.8\%$
Conv4	$97.9 \pm 1.0\%$	$97.4 \pm 1.3\%$	$98.9 \pm 1.2\%$	$96.6 \pm 1.7\%$	$98.1 \pm 0.9\%$
MobileNetV1	$96.7 \pm 2.3\%$	$95.1 \pm 3.2\%$	$99.3 \pm 1.1\%$	$93.4 \pm 4.5\%$	$97.1 \pm 2.0\%$

TABLE III: Results of the test validation.

	Accuracy	Precision	Sensitivity	Specificity	F1-Score	Training time (s)
Conv3	91.7%	90.9%	95.2%	86.7%	93.0%	138
Conv4	**97.2%**	**99.0%**	**96.2%**	**98.7%**	**97.6%**	**140**
MobileNetV1	96.1%	97.1%	**96.2%**	96.0%	96.0%	480

TABLE IV: Classification metrics comparison with similar works.

Paper	Method	Accuracy	Precision	Sensitivity	Specificity	F1-Score
[20]	XresNet50	96.8%	97.5%	95.7%	97.8%	96.4%
[26]	Deep Kronecker Neural Network	**98.5%**	**99.1%**	**98.1%**	**99.0%**	**98.6%**
This paper	CNN (4 Convolution Layers)	97.2%	99.0%	96.2%	98.7%	97.6%

TABLE V: Complexity comparison with similar works.

Paper	Parameters	Time (s)
[20]	∼23M	1920
[26]	∼2.7M	3122
This paper	**∼1M**	**140**

greater accuracy, high-cost techniques that result in a more accurate classification are preferable. We intend to expand the dataset used, adding samples with varying characteristics from different CT scanners. Besides the detection of kidney stones in CT exams, we plan to work in the future with other applications related to urology, extending our study to applications in organs of the urinary system other than the kidney and developing applications related to urinary stone segmentation, stone location and size estimation, and path planning of urological invasive surgeries for urinary stone treatment.

V. Conclusion

We investigated and compared three deep-learning neural network architectures for kidney stone detection on coronal CT images. The best model was the proposed *Conv4*, obtaining a test accuracy of 97.2%, a sensitivity of 96.2%, and an F1-score of 97.6%. Although our method was not the best-performing model compared with the related literature, the proposed architecture is lower in complexity, with approximately 1M trainable parameters and a training time of 140 seconds. Despite the results obtained so far, we plan to improve the model developed, exploring different planes of CT images on a larger dataset to train the deep-learning model developed. Also, we intend to extend our studies to other conditions related to the urinary system, exploring kidney stone segmentation, stone volume and composition analysis, and path planning of kidney stone removal surgeries.

Acknowledgment

The authors would like to thank the Centro Universitário FEI, HU-USP, FAPESP, CAPES and CNPq for the support they have given to this work.

References

[1] R. Bartoletti, T. Cai, N. Mondaini, F. Melone, F. Travaglini, M. Carini, and M. Rizzo, "Epidemiology and risk factors in urolithiasis," *Urologia internationalis*, vol. 79, no. Suppl. 1, pp. 3–7, 2007.

[2] A. A. Shokeir, "Renal colic: new concepts related to pathophysiology, diagnosis and treatment," *Current opinion in urology*, vol. 12, no. 4, pp. 263–269, 2002.

[3] J. P. Ingimarsson, A. E. Krambeck, and V. M. Pais, "Diagnosis and management of nephrolithiasis," *Surgical Clinics*, vol. 96, no. 3, pp. 517–532, 2016.

[4] N. Rasulova, A. Aminova, and F. Ismailova, "Improvement of early diagnosis and prevention measures of kidney stone diseases among the population," *Science and innovation*, vol. 2, no. D3, pp. 61–66, 2023.

[5] Y. Andrabi, M. Patino, C. J. Das, B. Eisner, D. V. Sahani, and A. Kambadakone, "Advances in ct imaging for urolithiasis," *Indian journal of urology: IJU: journal of the Urological Society of India*, vol. 31, no. 3, p. 185, 2015.

[6] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Medical physics*, vol. 47, no. 5, pp. e218–e227, 2020.

[7] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of digital imaging*, vol. 32, pp. 582–596, 2019.

[8] B. Z. Hameed, A. V. S. Dhavileswarapu, S. Z. Raza, H. Karimi, H. S. Khanuja, D. K. Shetty, S. Ibrahim, M. J. Shah, N. Naik, R. Paul *et al.*, "Artificial intelligence and its impact on urological diseases and management: A comprehensive review of the literature," *Journal of Clinical Medicine*, vol. 10, no. 9, p. 1864, 2021.

[9] U. S. Kim, H. S. Kwon, W. Yang, W. Lee, C. Choi, J. K. Kim, S. H. Lee, D. Rim, and J. H. Han, "Prediction of the composition of urinary stones using deep learning," *Investigative and Clinical Urology*, vol. 63, no. 4, p. 441, 2022.

[10] R. Suarez-Ibarrola, S. Hein, G. Reis, C. Gratzke, and A. Miernik, "Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer," *World journal of urology*, vol. 38, pp. 2329–2347, 2020.

[11] M. Akshaya, R. Nithushaa, N. S. M. Raja, and S. Padmapriya, "Kidney stone detection using neural networks," in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*. IEEE, 2020, pp. 1–4.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[14] J. R. S. Antony, U. S. RonyJoseph, P. Vijayaragavan, and A. Jerrinsimla, "Refine observation of kidney stones using neural network," 2020.

[15] A. Soni and A. Rai, "Kidney stone recognition and extraction using directional emboss & svm from computed tomography images," in *2020 Third International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*. IEEE, 2020, pp. 57–62.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[17] M. Längkvist, J. Jendeberg, P. Thunberg, A. Loutfi, and M. Lidén, "Computer aided detection of ureteral stones in thin slice computed tomography volumes using convolutional neural networks," *Computers in biology and medicine*, vol. 97, pp. 153–160, 2018.

[18] O. Joseph and W. O. Apena, "Development of segmentation and classification algorithms for computed tomography images of human kidney stone," *Journal of Electronic Research and Application*, vol. 5, no. 5, pp. 1–10, 2021.

[19] A. Caglayan, M. O. Horsanali, K. Kocadurdu, E. Ismailoglu, and S. Guneyli, "Deep learning model-assisted detection of kidney stones on computed tomography," *International braz j urol*, vol. 48, pp. 830–839, 2022.

[20] K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Deep learning model for automated kidney stone detection using coronal ct images," *Computers in biology and medicine*, vol. 135, p. 104569, 2021.

[21] B. Jou and S.-F. Chang, "Deep cross residual learning for multitask visual recognition," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 998–1007.

[22] M. Baygin, O. Yaman, P. D. Barua, S. Dogan, T. Tuncer, and U. R. Acharya, "Exemplar darknet19 feature generation technique for automated kidney stone detection with coronal ct images," *Artificial Intelligence in Medicine*, vol. 127, p. 102274, 2022.

[23] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[24] T. Tuncer, S. Dogan, F. Özyurt, S. B. Belhaouari, and H. Bensmail, "Novel multi center and threshold ternary pattern based method for disease detection method using voice," *IEEE Access*, vol. 8, pp. 84 532–84 540, 2020.

[25] E. Fix, *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF school of Aviation Medicine, 1985, vol. 1.

[26] K. K. Patro, J. P. Allam, B. C. Neelapu, R. Tadeusiewicz, U. R. Acharya, M. Hammad, O. Yildirim, and P. Pławiak, "Application of kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal ct images," *Information Sciences*, vol. 640, p. 119005, 2023.

[27] T. Wu, S. Tang, R. Zhang, J. Cao, and J. Li, "Tree-structured kronecker convolutional network for semantic segmentation," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2019, pp. 940–945.

[28] D. B. Adarkar, A. Lokapur, J. Porwal, and P. Mali, "Chronic kidney disease prediction," *International Journal for Research in Applied Science and Engineering Technology*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258448985

[29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[30] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[31] W. Brisbane, M. R. Bailey, and M. D. Sorensen, "An overview of kidney stone imaging techniques," *Nature Reviews Urology*, vol. 13, no. 11, pp. 654–662, 2016.