U-Net segmentation of the kidney on noncontrast CT for kidney stones location: a deep learning model for Mexican patient

L. A. Pérez Martínez[a,i], T. Córdova Fraga[a,ii], R. Guzman Cabrera[b,iii], A. Hérnandez Trinidad[a,iv], A. E. López Valencia[c,v], and M. A. Hernández González[c,vi]

[a]*Universidad de Guanajuato Campus León, Department of Physical Engineering, Division of Science and Engineering, Cda. Loma del Bosque 103, Lomas del Campestre, C.P. 37150, León de los Aldama, Gto., México.*
[b]*Universidad de Guanajuato Campus Irapuato-Salamanca, Irapuato-Guanajuato, 36500 Ex-Hacienda del Copal, Gto.*
[c]*Instituto Mexicano del Seguro Social, Titular 1, Blvd A. López Mateos, Obregon, 37320 León de los Aldama, Gto.*

Received 1 June 2025; accepted 2 July 2025

This work presents the implementation of convolutional neural networks for the automated detection and segmentation of kidney stones in non-contrast computed tomography (CT) medical images from patients of the Mexican Institute of Social Security (IMSS). A U-Net architecture was employed to segment the kidneys and delineate regions of interest in high-resolution CT slices, with ground truth annotations provided by medical specialists. The model was trained and validated on a curated dataset representative of the Mexican population, achieving a Dice similarity coefficient of 0.9808 ± 0.038, an IoU of 0.9623 ± 0.0073, and a validation loss of 0.0040 ± 0.008 in five-fold cross-validation. The proposed system demonstrated rapid convergence and excellent agreement with manual segmentations, confirming its utility as a reliable diagnostic aid for kidney stone detection in clinical practice.

Keywords: Kidney stones; non-contrast computed tomography; image segmentation; U-Net; convolutional neural networks; medical artificial intelligence.

DOI: https://doi.org/10.31349/RevMexFis.72.011101

1. Introduction

Renal lithiasis is one of the most prevalent urological conditions in the Mexican population, with significant clinical impact due to potential complications such as urinary tract infections, nephritic colic, hematuria, and chronic kidney failure. In regions such as Yucatán, its prevalence is particularly high, associated with factors including dietary habits, climate, high water hardness, and a high incidence of obesity [12,13]. This epidemiological context underscores the need for accurate and efficient diagnostic tools to enable timely detection of urinary stones.

Non-contrast computed tomography (CT) has become the gold standard for diagnosing kidney stones due to its high sensitivity and specificity in detecting calculi of various sizes and compositions, as well as its ability to localize them precisely within the urinary tract [15]. However, CT imaging also presents limitations related to cost, availability, and patient exposure to ionizing radiation.

Traditionally, CT image analysis is performed manually using specialized software that allows axial slice visualization in DICOM format. These tools enable contrast adjustment, precise measurement of stone size and density, and manual inspection of each image. However, this process is highly dependent on the specialist's experience and is often time-consuming, subjective, and prone to error.

Convolutional neural networks (CNNs) have demonstrated outstanding performance in medical image analysis, particularly in classification and segmentation tasks, by enabling automated identification of anatomical structures and lesions with high precision [18]. Architectures such as U-Net (Fig. 1) have been widely adopted in both magnetic resonance and computed tomography applications. Moreover, transfer learning has made it possible to use pretrained models to improve performance in data-limited clinical environments [19].

The integration of these technologies into clinical workflows is not intended to replace medical expertise, but rather to provide support tools that reduce workload, minimize

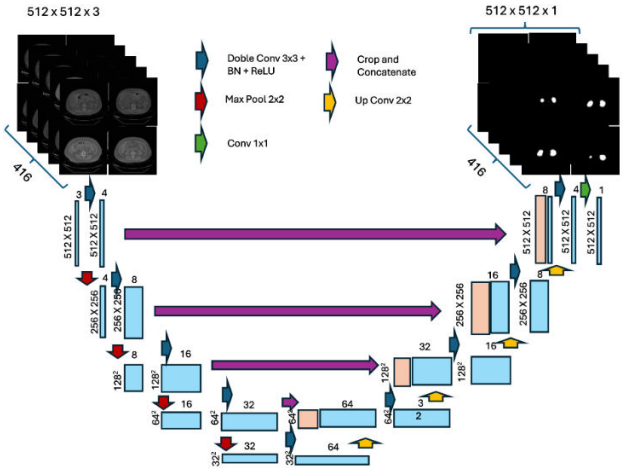

FIGURE 1. Architecture of the U-Net convolutional neural network implemented in this study.

inter-observer variability, and enhance diagnostic accuracy. In this context, the application of CNNs in renal lithiasis detection represents a meaningful step toward clinical automation and improved patient care.

Recent studies have explored deep segmentation networks, including U-Net and its variants, for the detection of kidney stones in non-contrast CT scans, reporting promising Dice similarity coefficients [20,21]. However, most of these studies are based on public or non-Mexican datasets, which may not reflect the anatomical or epidemiological characteristics of Mexican patients particularly those in advanced stages, where stones are larger and more obstructive [12]. Furthermore, segmenting isodense calculi-stones with attenuation values similar to adjacent tissue-remains a persistent challenge for standard CNNs, often requiring shape-aware segmentation strategies [22].

Given these limitations, there is a growing need for context-aware models specifically trained on local clinical data. This study proposes a U-Net-based model trained on non-contrast CT images from Mexican patients, with annotations performed by urology specialists and validated through manual segmentation. The aim is to automate the detection and measurement of renal calculi, standardize diagnostic workflows, and reduce dependence on subjective clinical interpretation.

Beyond mere detection, volumetric segmentation of renal calculi provides valuable quantitative information for clinical management. Accurate estimation of stone volume has been shown to predict treatment outcomes, recurrence risk, and to inform decisions regarding the choice of surgical intervention or medical therapy, as recommended by contemporary urological guidelines [1]. Automated volumetric assessment therefore represents a key step toward personalized and evidence-based patient care in the management of urolithiasis.

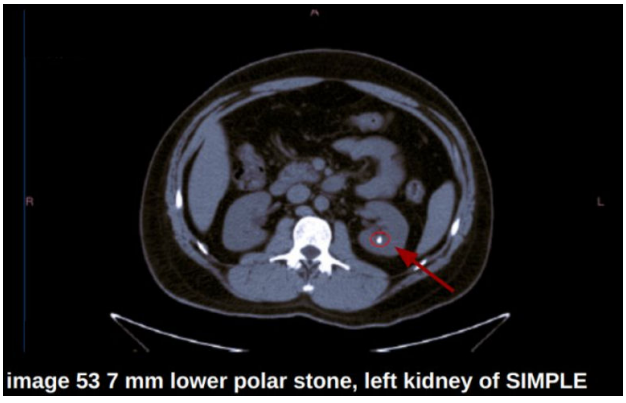

FIGURE 2. Axial CT slice labeled by medical specialists, showing the location of a renal calculus. Stone size and composition were confirmed post-extraction, allowing accurate annotation for model training.

2. Methodology

The data used in this study consisted of non-contrast computed tomography (CT) images provided by IMSS, which represent the current gold standard for the diagnosis of renal calculi due to their superior sensitivity and specificity. Each scan was acquired with a field of view (FOV) of $500\,\mathrm{mm} \times 500\,\mathrm{mm}$ and a reconstruction matrix of 512×512 pixels. Axial slices were obtained with a thickness of $1\,\mathrm{mm}$, ensuring high spatial resolution for precise kidney and stone localization. All images were stored in DICOM (Digital Imaging and Communications in Medicine) format.

A total of 521 images were included in the dataset, of which 416 were used for training and 105 for validation. The

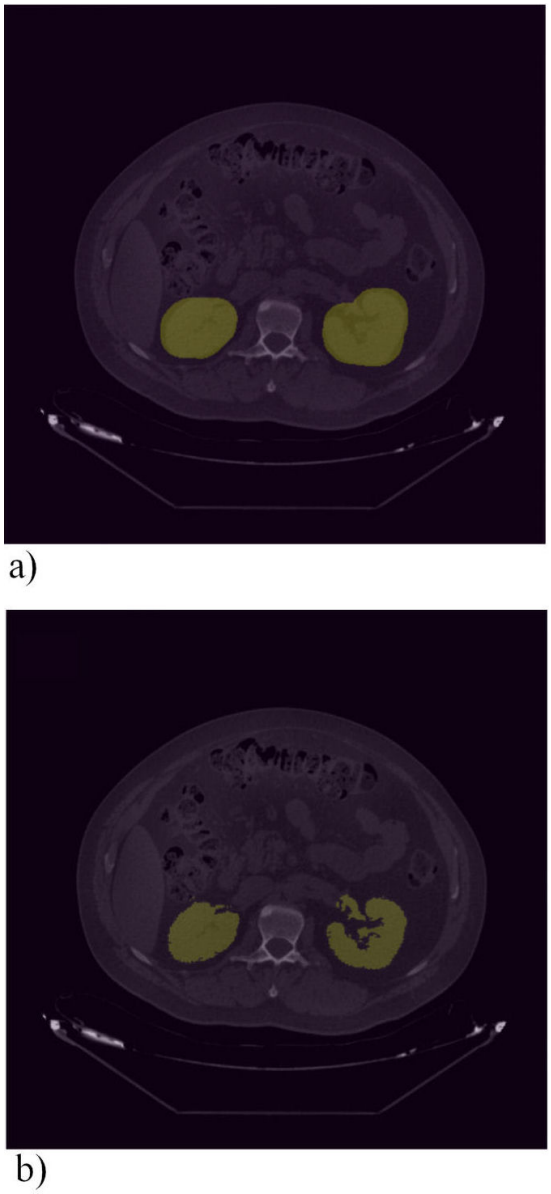

FIGURE 3. a) Original axial CT image without preprocessing. b) Image after contrast adjustment to enhance the visibility of soft tissues and high-density structures.

split was performed randomly to ensure a representative distribution of cases in both sets. No external test set was used in this work.

The images were labeled by medical specialists, who identified the maximum stone length in the axial slices of each study (Fig. 2). The algorithm was applied exclusively to the subset of CT slices corresponding to the anatomical region containing the kidneys, as identified and selected by the medical specialists. This targeted approach ensured that the analysis focused only on the relevant portion of each study, excluding slices outside the renal area.

To optimize model training, all selected images underwent contrast adjustment using an adaptive CLAHE (Contrast Limited Adaptive Histogram Equalization) filter, enhancing tissue visualization and highlighting relevant anatomical structures (Fig. 3).

2.1. Python packages

The programming and model development were performed in Python 3.9. The following main libraries were employed:

- **NumPy:** For numerical computations and array manipulations.
- **OpenCV:** For image processing and contrast adjustment.
- **Pydicom:** For reading and handling DICOM medical images.
- **PyTorch:** For building and training the U-Net convolutional neural network.
- **Matplotlib:** For plotting results and training curves.
- **ITK-SNAP:** For manual segmentation (software, not a Python library).

2.2. Thresholding and post-processing

After contrast enhancement, a global threshold of 0.65 (applied to normalized images in the range $[0, 1]$) was used to binarize each CT slice and the predicted masks. The threshold value was determined empirically to optimize the separation of kidney tissue from background.

For post-processing of the U-Net segmentation output, a contour detection algorithm was implemented to isolate the two largest connected regions, which correspond to the left and right kidneys. This step was essential to eliminate smaller spurious contours frequently generated by noise or adjacent anatomical structures. Only the two principal contours were retained for further analysis.

Finally, after kidney segmentation and contour filtering, intensity-based thresholding was used within the segmented kidney regions to highlight hyperdense areas indicative of renal stones. This workflow ensured that stone detection was

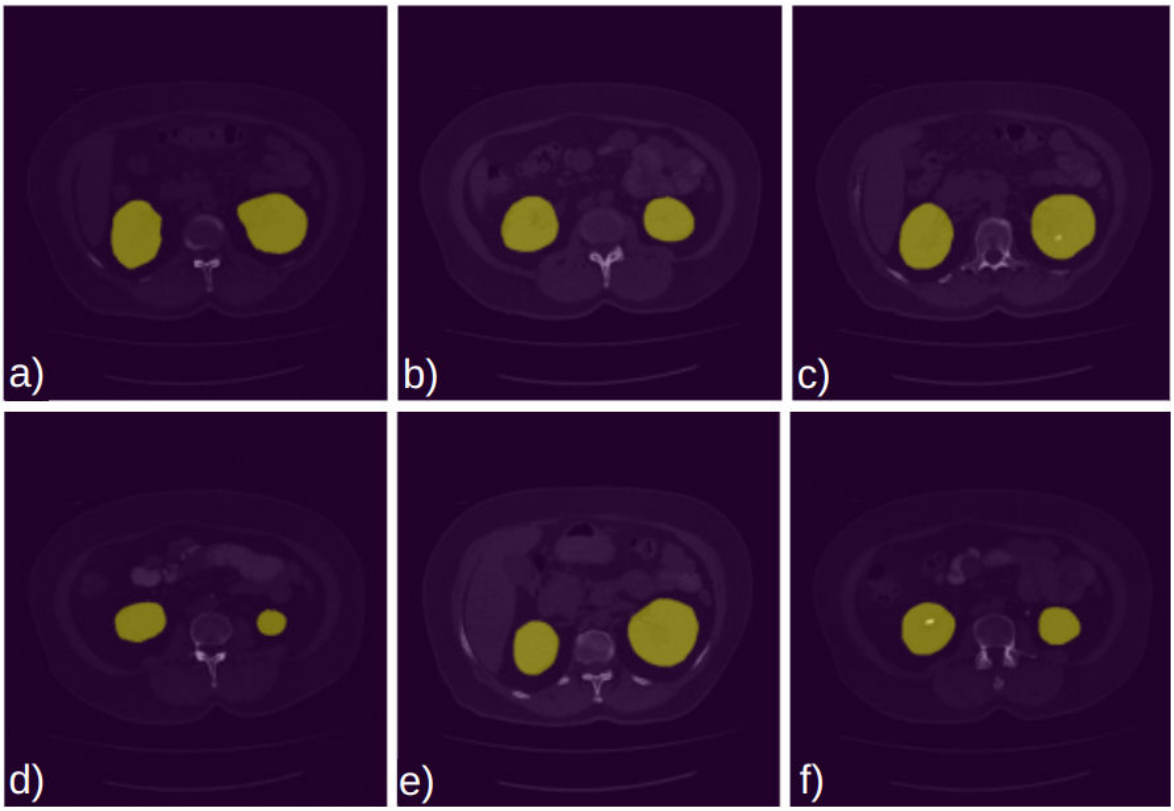

FIGURE 4. Training images and their corresponding manual masks.

restricted to the relevant anatomical context, minimizing false positives from extrarenal tissues.

It should be noted that segmentation was performed directly on the raw DICOM images, without applying fixed window or level settings prior to contrast enhancement. This design choice allows the model to remain robust to clinical variations in image display parameters, but also means that soft tissue contrast may vary slightly between studies. As such, quantitative comparisons between datasets or scanners should be interpreted with caution. Future work could explore the effect of standardized window/level preprocessing to further harmonize tissue appearance and potentially improve segmentation performance.

Kidney detection was carried out using a U-Net convolutional neural network specifically designed for segmentation tasks in medical imaging. This architecture can accurately identify contours and generate masks that automatically delimit regions of interest.

For manual segmentation —a fundamental step in training— the open-source software ITK-SNAP was used. This Python-based tool enables precise segmentation of DICOM images.

The contours generated allowed the creation of final masks identifying multiple regions in the presence of stones. If no calculi were present, the mask remained dark, covering only the segmented area (Fig. 4).

3. Results

The model was trained on 521 medical images provided by IMSS, each accompanied by its corresponding manual mask. Figure 5 illustrates the comparison between manual and automatic segmentations, as well as the results of the post-processing steps.

The automatic masks enabled the application of intensity-based filters to highlight areas of high calcium concentration, thereby reducing false positives from anatomical structures such as ribs. Subsequent contour analysis allowed precise estimation of the maximum stone length (Fig. 6).

Overall, the results demonstrate a high level of agreement between manual segmentations and those produced by the trained model, supporting its potential as a diagnostic aid for kidney stone detection in medical imaging.

To rigorously evaluate the model's generalizability, a five-fold cross-validation was performed on the dataset. Table I summarizes the primary performance metrics across all folds.

The model exhibited rapid and stable convergence throughout the 25 training epochs across all 10 folds. Dice coefficients exceeded 0.95 within the first 25 epochs and continued improving, reaching average values above 0.98 by the end of training. These results highlight the robustness and consistency of the segmentation approach across different data partitions.

It is important to note that the high average Dice and IoU coefficients-both exceeding 0.96-along with low standard deviations, indicate stable model performance and strong agreement with manual segmentation, regardless of the specific data split. The training curves showed that performance plateaued after approximately 20 epochs, confirming the efficiency and reliability of the proposed methodology.

TABLE I. Cross-validation performance metrics of the U-Net model (mean $\pm$ standard deviation across 10 folds).

Metric	Mean	Std. Dev.
Dice Coefficient	0.9808	0.0038
IoU	0.9623	0.0073
Validation Loss	0.0040	0.0008

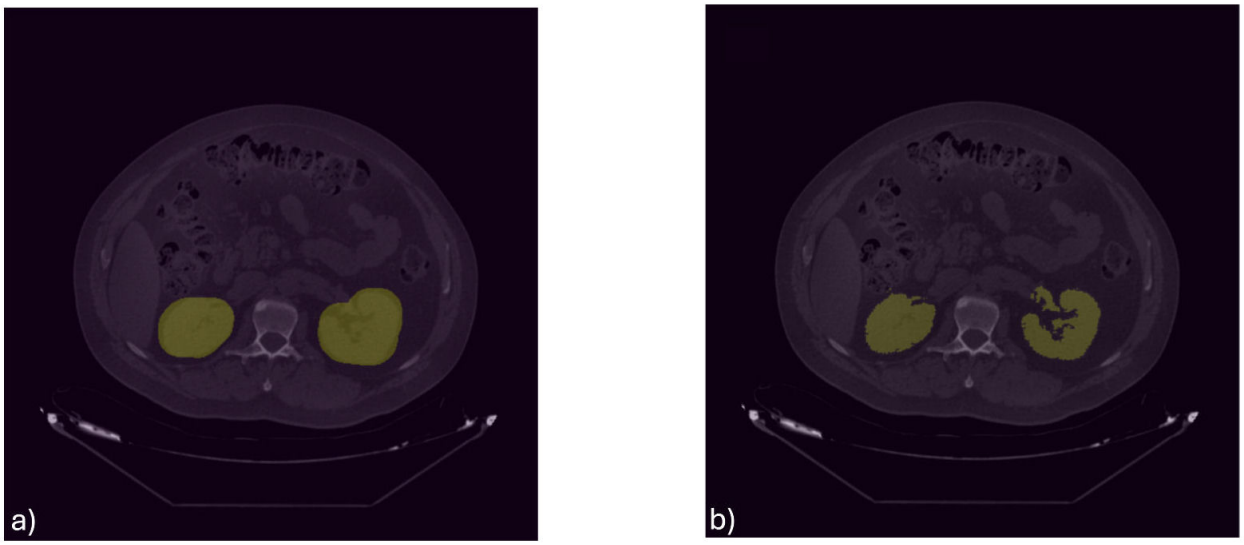

FIGURE 5. Comparison between manual and automatic masks generated by the trained model. a) Image with manually generated mask. b) Image with automatic mask and post-processing, including thresholding to highlight tissue and extraction of major contours.

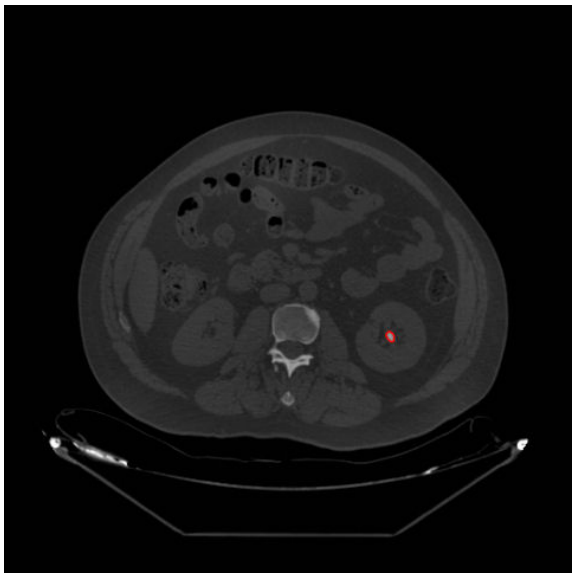

FIGURE 6. Identification of contours through automated segmentation and intensity filtering.

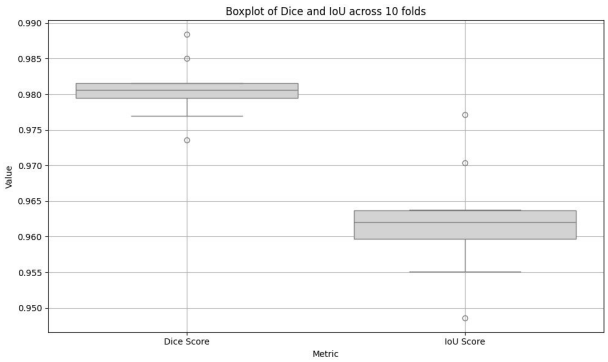

FIGURE 7. Distribution of performance metrics (Dice, IoU, and validation loss) across the 10 cross-validation folds.

While the current approach performs segmentation on individual axial slices, the pipeline can be extended to true 3D volumetric segmentation, which is particularly valuable for preoperative planning and post-treatment monitoring [1].

4. Conclusion

This study demonstrates the feasibility of automating kidney stone detection in axial CT scans using convolutional neural networks. The proposed U-Net model achieved highly accurate kidney segmentation and reliably identified hyperdense regions corresponding to renal calculi, providing a valuable diagnostic tool for the assessment of renal lithiasis.

The results confirm that deep learning-based segmentation offers fast and consistent preliminary visualization, with clear potential for continuous improvement as additional labeled data become available. Unlike fixed-rule systems, neural networks are capable of generalizing and adapting, increasing their effectiveness and scalability for real clinical applications.

To ensure the robustness of the methodology, a five-fold cross-validation protocol was employed, yielding consistently high Dice and IoU coefficients across all data partitions. This robust assessment confirms the generalizability of the U-Net architecture for automated kidney segmentation in clinical CT studies from the Mexican population.

Nevertheless, further improvements can be achieved by expanding the dataset, exploring advanced architectures, and integrating additional comparative methodologies. Ongoing collaboration between developers and healthcare professionals will be essential for refining the system and facilitating its adoption in real-world clinical settings.

Acknowledgments

The authors extend their appreciation to the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti-Scholarshop CVU 1344507), the Instituto Mexicano del Seguro Social and the Universidad de Guanajuato for funding this research.

i. https://orcid.org/0009-0002-4059-3480
ii. https://orcid.org/0000-0002-6486-7530
iii. https://orcid.org/0000-0002-9320-7021
iv. https://orcid.org/0000-0002-5619-6810
v. https://orcid.org/0009-0001-2184-9957
vi. https://orcid.org/0000-0002-6903-2233

1. C. D. Scales Jr. *et al.*, Prevalence of kidney stones in the United States, *Eur. Urol.* **62** (2012) 160. https://doi.org/10.1016/j.eururo.2012.03.052

2. K. Stamatelou, D. S. Goldfarb, Epidemiology of kidney stones, *Healthcare* **11** (2023) 424. https://doi.org/10.3390/healthcare11030424

3. B. J. Young *et al.*, Is the economic impact and utilization of imaging studies for pediatric urolithiasis across the United States increasing?, *Urology* **94** (2016) 208, https://doi.org/10.1016/j.urology.2016.05.019

4. J. M. Soucie, M. J. Thun, R. J. Coates, W. Mc-Clellan, H. Austin, Demographic and geographic variability of kidney stones in the United States, *Kid- ney Int.* **46** (1994) 893, https://doi.org/10.1038/ki.1994.347

5. A. Chewcharat, G. Curhan, Trends in the prevalence of kidney stones in the United States from 2007 to 2016, *Urolithiasis* **49** (2021) 27, https://doi.org/10.1007/s00240-020-01210-w.

6. L. Moftakhar et al., Prevalence and risk factors of kidney stone disease in population aged 40-70 years old in Kharameh cohort study: a cross-sectional population-based study in southern Iran, *BMC Urol*. **22** (2022) 205, https://doi.org/10.1186/s12894-022-01161-x.
7. J. Shoag, G. E. Tasian, D. S. Goldfarb, H. Eisner, Brian, The new epidemiology of nephrolithiasis, *Adv. Chronic Kidney Dis.* **22** (2015) 273, https://doi.org/10.1053/j.ackd.2015.04.004.
8. J. B. Ziemba, B. R. Matlaga, Epidemiology and economics of nephrolithiasis, *Investig. Clin. Urol*. **58** (2017) 299, https://doi.org/10.4111/icu.2017.58.5.299
9. V. Gómez Dos Santos and F. J. Burgos, Litiasis en el origen de insuficiencia renal crónica, *Nefrología* **25** (2005) 82.
10. V. G. González, Litiasis renal: estudio y manejo endocrinológico, *Rev. Méd. Clín. Las Condes* **24** (2013) 798, https://doi.org/10.1016/S0716-8640(13)70226-8.
11. A. A. Herrera Muñoz *et al*., Nefrolitiasis: Una revisión actualizada, *Rev. Clín. Esc. Med. Univ. Costa Rica* **10** (2020) 11.
12. M. Medina-Escobedo, M. Zaidi, E. Real-de León, S. Orozco-Rivadeneyra, Prevalencia y factores de riesgo en Yucatán, México, para litiasis urinaria, *Salud Públ. Méx.* **44** (2002) 541.
13. M. Medina-Escobedo, R. Alcocer-Dzul, J. López- López, J. S. Villanueva, Obesidad como factor de riesgo para alteraciones metabólicas en adultos con litiasis urinaria, *Rev. Méd. Inst. Mex. Seguro Soc.* **53** (2015) 692.
14. Y. L. R. Salazar, E. Gómez Montañez, Litiasis renal: una entidad cada vez más común, *Expresiones Médicas* **9** (2021) 30.
15. W. Brisbane, M. R. Bailey, M. D. Sorensen, An overview of kidney stone imaging techniques, *Nat. Rev. Urol.* **13** (2016) 654, https://doi.org/10.1038/nrurol.2016.154
16. M. Malaki, The comparison of ultrasound and non-contrast helical computerized tomography for children nephrolithiasis detection, *Urol. Ann.* **6** (2014) 309, https://doi.org/10.4103/0974-7796.140991.
17. A. J. Portis and C. P. Sundaram, Diagnosis and initial management of kidney stones, *Am. Fam. Physician* **63** (2001) 1329.
18. B. Kayalibay, G. Jensen, P. van der Smagt, CNN-based segmentation of medical imaging data, arXiv 1701.03056 (2017). https://doi.org/10.48550/arXiv.1701.03056.
19. A.W. Salehi, *et al.*, A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope, *Sustainability* **15** (2023) 5930, https://doi.org/10.3390/su15075930
20. D. Li, *et al.*, Deep segmentation networks for segmenting kidneys and detecting kidney stones in unenhanced abdominal CT images, *Diagnostics* **12** (2022) 1788, https://doi.org/10.3390/diagnostics12081788
21. D.-H. Jang *et al*., Kidney, ureter, and urinary bladder segmentation based on non-contrast enhanced computed tomography images using modified U-Net, *Sci. Rep.* **14** (2024) 15325, https://doi.org/10.1038/s41598-024-66045-6.
22. T. Hatsutani, A. Ichinose, K. Nakamura, Y. Kitamura, Segmentation of Kidney Tumors on Non- Contrast CT Images Using Protuberance Detection Network, *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.* **(MICCAI)** (2023) 13, https://doi.org/10.1007/978-3-031-43990-22.