UNITECH SELECTED PAPERS

International Scientific Conference
UNITECH`2024

APPLICATION OF THE YOLO ALGORITHM IN THE AUTOMATED DETECTION OF KIDNEY STONES FROM MEDICAL IMAGING DATA

Stefan Ćirković*, Katarina Karić, Nikola Stanić

Faculty of Technical Sciences Čačak, University of Kragujevac, Serbia;
** Corresponding author: stefan.cirkovic@ftn.kg.ac.rs*

Abstract

Kidney stones are one of the most common urological issues worldwide, and timely diagnosis is crucial for effective treatment and reducing complications. Traditional diagnostic methods, such as ultrasound and computed tomography (CT), require manual interpretation of images by specialists, which can be time-consuming. With the development of artificial intelligence, neural networks are increasingly being applied in medicine. This paper presents the application of the YOLO (You Only Look Once) algorithm on a dataset of medical images for the detection and classification of kidney stones. This approach can serve as support for specialists in making faster and more informed decisions, with the potential to reduce the burden and costs in the healthcare system, but it does not replace professional medical diagnosis.

Keywords: Stone, YOLO, Kidney, Annotated.

INTRODUCTION

Kidney stones represent one of the most common urological conditions worldwide, and timely diagnosis is crucial for effective treatment and the reduction of potential complications. Traditional diagnostic methods, such as ultrasound and computed tomography (CT), rely on manual interpretation of images by specialists, which can be time-consuming and subject to subjective assessments. With the advent of artificial intelligence and its application in medicine, new conditions have been created for the automation of medical data analysis, enabling the introduction of faster, more objective, and more accurate diagnostic tools [2].

Algorithms based on convolutional neural networks (CNN), such as YOLO (You Only Look Once), are increasingly being applied in the field of detection and classification of objects in medical images. The YOLO algorithm, originally developed for real-time object detection in general computer vision applications, shows great potential in medicine due to its speed, efficiency, and ability to detect objects in a single pass through an image [6].

This study applies the latest version of the algorithm, YOLOv11, which brings improvements in detection accuracy and processing efficiency compared to previous versions. YOLOv11 implements innovative architectural elements such as the C3k2 block and SPPF (Spatial Pyramid Pooling - Fast), allowing for better recognition of objects of varying sizes and positions within medical images [6].

The goal of this research is to evaluate the effectiveness of the YOLOv11 algorithm on a set of medical data for the automatic detection of kidney stones, as well as its potential application as support for specialists in making quicker and more accurate diagnostic decisions. The proposed method has the potential to reduce subjective errors, speed up the processing time, and lower costs within the healthcare system, but it is not a substitute for professional medical diagnosis. This study contributes to a comprehensive examination of the application of advanced deep learning models in the analysis of medical images, paving the way for further research and optimization of tools in modern medical practice.

doi: www.doi.org/10.70456/SHCE4087

RELATED WORK AND COMPARISON

In recent years, research on kidney stone detection using machine learning techniques, particularly object detection algorithms like YOLO, has become increasingly prevalent in the medical community. Studies such as Bayram et al. [1] and Siener [2] demonstrate the significant use of deep learning in identifying cysts and stones in the kidney, with results including a mean average precision (mAP) of 0.85 and an F1 score of 0.854 for the YOLOv7 model. The research by Nguyen et al. [3] proposes super-resolution techniques combined with YOLOv7, achieving a sensitivity of 87.6% and a precision of 92.2%, while Rabby et al. [4] showed the superiority of YOLOv5 and YOLOv7 compared to traditional models like CNN and SVM.

Given the latest advancements in this field, this research utilizes the YOLOv11 model, which provides a more advanced architecture compared to previous versions. The model was tested and achieved the following results: an overall precision of 0.922 and a recall of 0.82, suggesting high performance in detecting all instances. When analyzing specific classes, the results for kidney detection demonstrate an exceptional precision of 0.987 and a recall of 0.993, while the results for kidney stones were achieved with a precision of 0.857 and a recall of 0.647.

These results are significantly better compared to most related studies that rely on earlier versions of the YOLO architecture, such as YOLOv5 or YOLOv7. Although previous works have reported good performance, this research demonstrates further advancements in accuracy and precision due to the use of YOLOv11. For example, while some studies reached an mAP of 0.85 or lower, my research achieved an mAP of 0.871, indicating improved capabilities of the model in recognizing and classifying kidney stones and other structures.

Furthermore, similar studies do not include the application of the latest version of the YOLO model in the context of kidney stone detection, thus contributing new insights to this area. Based on the results and innovative application of YOLOv11, this research can be considered to set new standards in precise diagnostics and has the potential for significant impact on clinical practices in identifying renal pathologies.

YOLO (You Only Look Once)

YOLO (You Only Look Once) represents one of the most significant innovations in object detection within computer vision, introduced in 2015 by Redmon and colleagues. A key feature of the YOLO approach is treating object detection as a regression problem, allowing the model to simultaneously predict bounding box coordinates and object classes in a single pass through the image. This "single-stage" detection approach, as opposed to two-stage methods requiring multiple passes, makes the YOLO algorithm highly efficient and suitable for real-time applications [5].

The YOLO architecture, based on convolutional neural networks (CNN), consists of three main components: the backbone extracts key features, the neck processes these features at multiple scales to detect objects of various sizes, and the head produces bounding box coordinates and class labels for accurate detection [5].

Fig 1. illustrates the basic architecture of the YOLO model, encompassing these components: backbone, neck, and head.

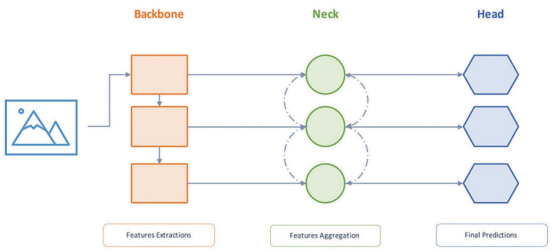

Fig 1. The architecture of YOLO consists of a backbone, neck, and head [5]

YOLOv11: Innovations and Advantages

YOLOv11, the latest in the YOLO series, brings notable enhancements in accuracy, speed, and multi-tasking, enabling object detection, segmentation, pose estimation, and oriented object detection (OBB) [6]. Key innovations include:

- **C3k2 Block:** an advanced variant of the Cross Stage Partial (CSP) with smaller convolutional kernels, providing faster processing and a reduced parameter count.
- **SPPF (Spatial Pyramid Pooling – Fast):** enables efficient multi-scale feature extraction for detailed image analysis.
- **C2PSA Block:** a parallel spatial attention module allowing the model to focus on crucial image regions, improving the detection of objects of various sizes and positions.

Studies indicate that YOLOv11 surpasses previous versions in mean Average Precision (mAP), while reducing parameters by approximately 22% compared to YOLOv8m, making it highly efficient for applications requiring speed without sacrificing accuracy [6].

METHODOLOGY

The application of the YOLO algorithm is investigated for the automated detection of kidney stones based on medical images. A selected dataset obtained from the Kaggle [7] platform was collected from PACS (Picture Archiving and Communication System) across various hospitals in Dhaka, Bangladesh. This dataset contains images of patients with previous kidney diagnoses, including stones, cysts, normal findings, and tumors.

The dataset comprises 12,446 unique images, including 3,709 cysts, 1,377 stones, 2,283 tumors, and 5,077 normal findings. For model training, 500 annotated images were used, classified into two main classes: *stone* and *kidney*. This methodology has the potential to accelerate the diagnostic process and facilitate specialists in making quicker and more informed decisions.

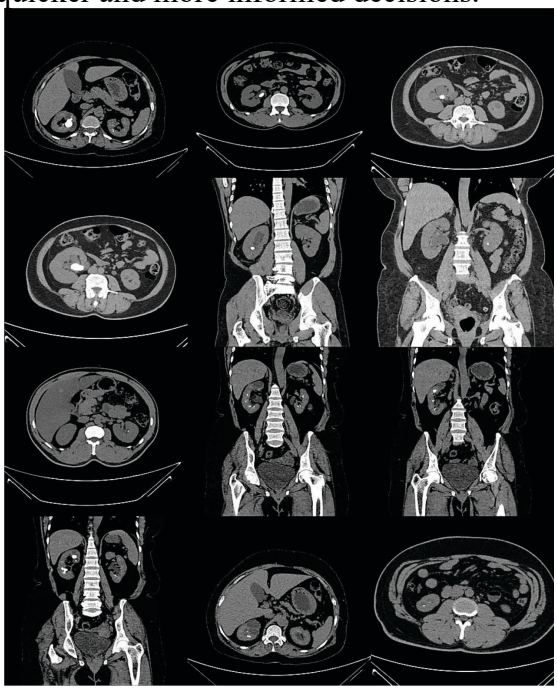

Fig 2. Dataset

The images were pre-processed to achieve optimal size and quality for further analysis. During this process, each image was resized to dimensions of 224x224 pixels. To enable the application of the YOLO algorithm for classification, appropriate annotations were required. Each image was manually annotated using the Labelme tool, available as a Python package. After annotation, a corresponding JSON file with the same name was created for each image. These JSON files contain coordinates that accurately mark the relevant parts of each image, allowing for more precise object recognition. Fig 3. shows the annotation of an image in the Labelme tool.

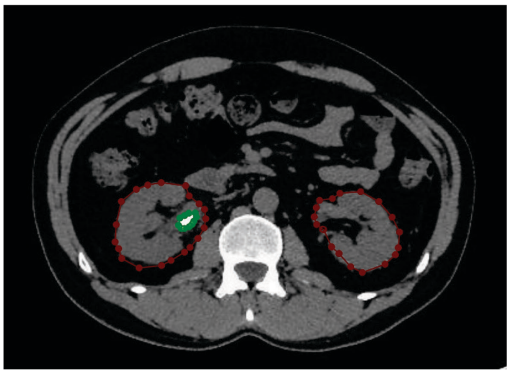

Fig 3. Annotated image in the Labelme tool

Images unsuitable for annotation were removed from the dataset to ensure higher model training accuracy. Since the YOLO algorithm does not use JSON format for training but rather TXT format, annotations were converted from JSON to TXT using the labelme2yolo tool. The prepared dataset was divided into three parts — training, testing, and validation — in the standard 80:10:10 ratio, providing an appropriate data distribution for each training step and performance assessment.

YOLO model training was conducted on an NVIDIA GeForce 1080 graphics card, using key optimization parameters: 100 epochs and a batch size of 16 samples. The training process lasted a total of 1 hours and 39 minutes, during which optimal classification results were achieved through iterative model refinement.

RESULT AND INTERPRETATION

This chapter presents the results and interpretation obtained after training the YOLO model for automated kidney stone detection on medical images. Based on the confusion matrix shown in Fig 4., the model's performance metrics were analyzed as it classified images into three categories: kidney, stone, and background. The "background" category represents parts of the image that were not annotated. The model correctly classified 145 cases as kidney but made one error, misclassifying a stone case as kidney. In stone classification, the model accurately identified 151 cases but made 13 errors, labeling them as background. Since the background was not annotated in the training data, the model lacked a basis to learn how to distinguish it from other categories, resulting in misclassification of 90 stone cases and one kidney case as background.

It is important to emphasize that the aim of this research was not background detection but rather precise identification and classification of parts of the image representing kidneys and kidney stones. Although the model made certain errors in background classification, this did not significantly impact the primary objective of the research. For more precise differentiation of the background in future analyses, it is recommended that it be annotated and included in the training process; however, this is not essential for the current task. These results indicate that the model provides good and acceptable preliminary results for kidney stone detection, where positive findings can be further analyzed by a radiologist.

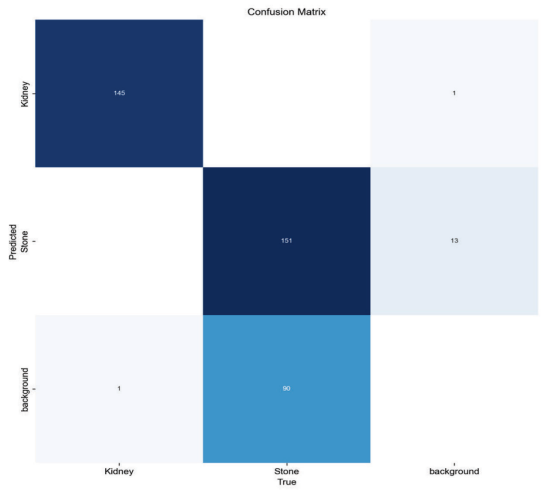

Fig 4. Confusion Matrix

Fig 5. displays the model's precision-confidence curve for all classes. Precision is defined as the ratio of true positive predictions (True Positives, TP) to the total number of positive predictions, including false positives (False Positives, FP), as shown in Eq (1):

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

At the displayed confidence threshold of 0.994, the model achieved a precision of 100%, meaning there were no false positive predictions in this instance. These results indicate a high degree of model reliability and accuracy when higher confidence thresholds are used. This level of precision is particularly important for applications in medical diagnostics, where minimizing false positives is critical to reducing the need for additional diagnostic procedures

and optimizing accuracy in detecting actual pathological changes.

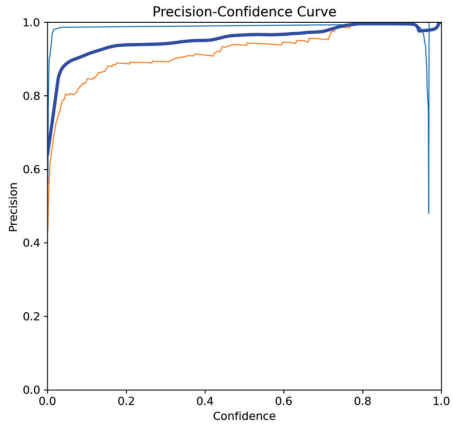

Fig 5. Precision Confidence Curve

Fig 6. shows the model's recall-confidence curve, which measures the model's capacity to identify all positive instances. Recall becomes particularly important in situations where minimizing false negatives is critical, as it quantifies the proportion of actual positives that were correctly identified by the model. In this example, the model achieves a recall of 0.86 (86%) when the confidence threshold is 0.000, indicating a high probability of correctly identifying true positive instances while practically overlooking very few.

The recall metric is defined by the following Eq. (2):

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

This high recall rate reflects the model's sensitivity in capturing positive cases, which is crucial in medical applications to ensure that significant cases are not missed.

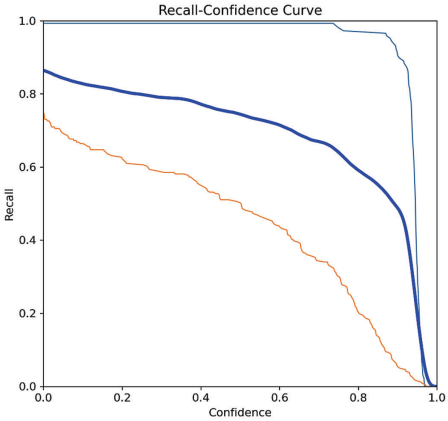

Fig 6. Recall-Confidence Curve

In Fig 7, the F1-confidence curve is shown, indicating that the F1-score for all classes reached a value of 0.87 at a confidence threshold of 0.152. The F1-score represents the harmonic mean between precision and recall and is used as an overall performance measure of the model, especially in cases where there is an imbalance between classes. It is calculated using the following Eq. (3):

$$F1 = 2\ x\ \left(\frac{Precision\ x\ Recall}{Precision + Recall}\right) \qquad (3)$$

The achieved F1-score value indicates that the model has high overall performance in classifying instances when the confidence threshold is set at a low value. This result is particularly significant for applications where both precision and recall are essential, as it provides a balance between minimizing false positives and false negatives, which is crucial in the context of medical recognition of kidney stones and other anomalies.

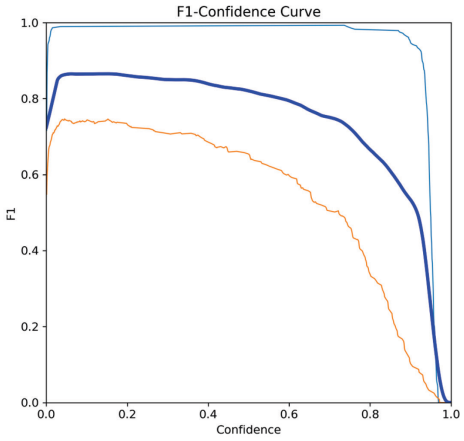

Fig 7. F1 Confidence curve

The YOLO model was tested on a validation set consisting of 75 images, as shown in Table 1. The model achieved an overall precision of 0.922 and a recall of 0.82 across all instances, indicating strong performance in detecting kidney stones.

Table 1. Results at the validation set

Class	Instances	Precision	Recall
All	387	0.922	0.82
Kidney	146	0.987	0.993
Stone	241	0.857	0.647

In the following figure, Fig 8., the detection of kidney stones during the validation process is depicted.

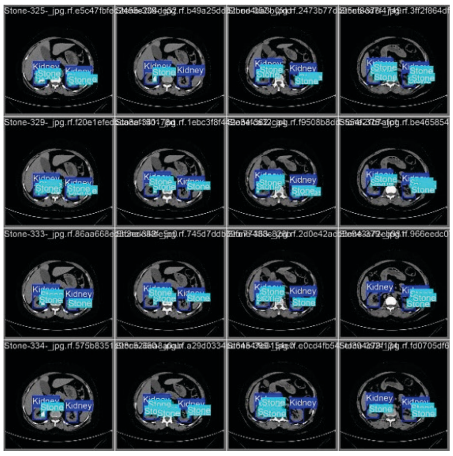

Fig 8. Kidney stone detection during validation

CONCLUSON

In this paper, the application of the YOLO (You Only Look Once) algorithm for the automatic detection and classification of kidney stones from medical images was investigated. The results showed high precision (92.2%) and a good recall (82%) in detection. These results indicate the potential of the YOLOv11 algorithm as an efficient tool that can support specialists in making faster decisions, thereby reducing the burden and costs within the healthcare system.

Given the increasing application of artificial intelligence in medicine, further research could focus on enriching the model with additional kidney anomalies, such as cysts and cancer, in collaboration with specialist physicians. Training the model on larger and higher-quality datasets of CT images, previously verified by experts, could further enhance detection accuracy and enable its integration into medical systems. This integration could significantly facilitate the work of urologists and improve the quality of healthcare.

We expect that further research in this direction will contribute to the development of advanced solutions in the automatic analysis of medical images, providing a foundation for improving diagnostic processes and the effectiveness of urological treatment.

ACKNOWLEDGEMENT

This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-66/ 2024-03/200132 with University of Kragujevac – Faculty of Technical Sciences Čačak.

REFERENCE

[1] A. F. Bayram, C. Gurkan, A. Budak, and H. Karatas, "A Detection and Prediction Model Based on Deep Learning Assisted by Explainable Artificial Intelligence for Kidney Diseases," European Journal of Science and Technology, no. 40, pp. 67-74, 2022. doi: 10.31590/ejosat.1171777.

[2] R. Siener, "Nutrition and Kidney Stone Disease," *Nutrients*, vol. 13, no. 6, Article 1917, 3 Jun. 2021. doi: 10.3390/nu13061917.

[3] M. T. P. Nguyen, V. T. Le, H. T. Duong, and V. T. Hoang, "Detection of Kidney Stone Based on Super Resolution Techniques and YOLOv7 Under Limited Training Samples," in *Lecture Notes on Data Engineering and Communications Technologies*, 2023, doi: 10.1007/978-3-031-46749-3_3.

[4] S. Rabby, F. Hossain, S. Das, I. Rahman, S. Das, J. Soeb, and M. F. Jubaye, "An automated approach for the kidney segmentation and detection of kidney stones on computed tomography using YOLO algorithms," *Journal of Ideas in Health*, vol. 6, no. 4, pp. 963-970, Nov. 2023, doi: 10.47108/jidhealth.vol6.iss4.313.

[5] M. G. Ragab, S. J. Abdulkadir, A. Muneer, A. Alqushaibi, E. H. Sumiea, R. Qureshi, S. M. Al-Selwi, and H. Alhussian, *A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023),* IEEE Access, vol. 12, pp. 1–15, 2024. doi: 10.1109/access.2024.3386826.

[6] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," arXiv, vol. 2410.17725v1 [cs.CV], Oct. 23, 2024.

[7] M. N. Islam, M. Hasan, M. Hossain, M. Alam, R. G. Rabiul, M. Z. Uddin, and A. Soylu, "CT Kidney Dataset: Normal-Cyst-Tumor and Stone," Kaggle, 2022. [Online]. Available: https://www.kaggle.com/datasets/nazmul0087/ct-kidney. [Accessed: Oct. 27, 2024].