International Journal for Scientific Research (IJSR)
المجلة الدولية للبحوث العلمية
IJSR
Vol. (5), No. (2) February 2026 الإصدار (5)، العدد (2)

# FPGA-Based Computer-Aided Diagnosis for Kidney CT Images Abnormality Classification Using NSCT and a Modified YOLOv11

**Ahmed Sabeeh Yousif**
Department of Information Technology Management, Technical College of Management/Mosul, Northern Technical University, Mosul, 41001, Iraq
ahmedsabeeh123@ntu.edu.iq

## Abstract

Kidney disease diagnosis often relies on expert radiologists and advanced imaging analysis, which may be limited or unavailable in remote healthcare settings. This paper presents an FPGA-oriented computer-aided diagnosis scheme for binary kidney abnormality classification (normal vs. abnormal) that combines kidney ROI extraction with nonsubsampled contourlet transform (NSCT) feature compaction and a modified YOLOv11 classifier. After normalization and kidney segmentation, NSCT is applied to the ROI and only the final low-frequency sub-band is retained as a compact, structure-preserving representation for classification. The key novelty is a lightweight YOLOv11 classification variant tailored for deployment constraints by reducing depth scaling and removing attention modules to support efficient fixed-point implementation while retaining discriminative power with NSCT low-frequency inputs. The proposed method has been achieved better results as shown in accuracy, sensitivity, and specificity (i.e.: 97.45%, 98.23%, and 97.45%) respectively. The higher performance of Zynq based hardware shows 68 ms latency in image, and 14.7 images/ throughput, performing better real-time for kidney CT disease classification.

**Keywords:** CT Images, Nonsubsampled Contourlet Transform, YOLOv-11.

## 1. Introduction

Chronic kidney disease (CKD) affects millions of people worldwide and constitutes a major public health burden. The Global Burden of Disease study ranks kidney diseases among the leading causes of death worldwide, and their prevalence continues to rise. Consequently, early identification and accurate diagnosis of kidney abnormalities are critical for timely intervention and long-term

disease management [1], [2]. Medical imaging such as CT image is utilized for follow up the kidney diseases progress due to offered clear anatomical details for the tissues and high spatial resolution [2]. As a result, there have big motivated to involved to develop CAD system of computer-aided diagnosis (CAD) that can enhance diagnostic and reduce clinician heavy load on doctors.

Different CAD schemes with conventional machine learning methods and some various deep learning schemes usually involved different features descriptors as proposed in [32], [33], for instances texture measures, and morphological based features [3]. Recently, deep learning schemes, when (CNNs) applied, learn hierarchical representations moved from data and have shown better achievements over the imaging tasks, for example, features detection, organ localization, and abnormalities classification [4]. In [5,6], the field programmable gate arrays (FPGAs) methods offer promising results by utilizing the hardware acceleration through like parallel processing procedure, reduced power consumption vs to GPUs, and low-latency deterministic inference. This paper proposes a novel CAD based NSCT with CNN classification for kidney detection using the CT images scan, NSCT provides decomposition of CT images for further analysis and to enhance edge, texture, and contour representation details by obtaining the low frequency components then feed to classifier-based CNN. Real time based Zynq based hardware to show strength and reliability of software -hardware in clinical healthcare, and assist doctors in kidney detection

## 2. Related Works

Various schemes have been proposed in different methodologies for instances (1) CT based deep learning with renal abnormality, (2) Focused on scheme complexity in introduced scheme-based FPGA. Firstly, in [7] strong literature review with CNN using different methodologies has been introduced. The review recorded that robust generalization is still a persistent issue, varies of data diversity, and efficiency.

While in [8], the paper has presented new method which solved the contrast issue using an automated scheme using deep learning combine with tumor segmentation

of kidney diseases, the method has been shown strengthens in term of lower-complexity representations when used the low-frequency transform sub bands. Moreover, in [9], the author suggested new framework for detection of kidney disease with normal, cyst, tumor size, and stone size stages, the paper has been shown better performance when using various datasets, for instances when identifying Darknet53 as a strong backbone. Add to that, in [10] the article has proposed novel deep learning scheme to detect a kidney stone presenting that modified CNN can perform good performance with different metrics for binary classification. In [11], also the article shown an automated scheme that merges a new segmentation with subsequent classification for the classification of renal tumor categories on CT. In this point, our paper intended to achieve rapid abnormality detection with reduces complexity by using the transformation of NSCT compared to [11]. Another scheme proposed in [12], the paper introduced the YOLOv9 to implement the localization process-based fractures in trauma radiographs. It is shown with results explored that YOLO based classification variants can offer better complexity and higher accuracy of the localization. In our paper ROI extraction with NSCT-derived low-frequency of details representation to reduce the input while preserving better structural information for introduced the good hardware utilizing. A new scheme, namely ensemble based renal tumor segmentation with estimation has been introduced in [13], which more enhance to renal tumors structure details. Compared to our scheme which used NSCT -FPGA for better real time of kidney diseases. Also, in [14], the author proposed FPGA scheme for different health care applications with low power consumption and applied parallel execution for more real time ability in medical imaging classification-based kidney diseases.

Lastly, the article proposed in [15] with gains performed by utilizing the weight CNN and quantization hardware scheme, this aligns with the proposed NSCT based classification as upstream complexity, and reduction process that would quantized CNN, and applied in FPGA design. Shawahna *et al.* [16] provided a widely cited IEEE Access review of FPGA-based deep learning accelerators, summarizing key techniques including parallel dataflows, memory optimization, and quantization, which directly in- form the reporting of latency, throughput, and

energy for FPGA-based medical CAD platforms.

Overall, recent evidence shows rapid progress in renal CT deep learning and increasing momentum for FPGA- enabled edge diagnostics. However, an important gap remains in transform-assisted, FPGA-ready kidney abnormality classification that deliberately reduces input complexity while retaining strong structural cues. The proposed NSCT– YOLOv11 FPGA pipeline addresses this gap by coupling multi-resolution decomposition with a compact classifier and a practical real-time deployment path suitable for offline, resource-limited healthcare environments. The main contributions of this work are as follows:

1) NSCT–YOLOv11 architecture: A transform-assisted classification pipeline that couples NSCT decomposition with YOLOv11 to perform kidney CT abnormality classification, with FPGA deployment as a core design objective.
2) Modified YOLOv11-cls: A new YOLOv11-cls-Lite classifier optimized for FPGA by reducing depth scaling and removing attention modules, thereby lowering compute cost and on-chip memory requirements.
3) Large-scale evaluation: Validation on 400,000 kidney CT images to support robust training and improved generalization compared with prior smaller-scale studies.
4) Efficient FPGA implementation: Hardware-software co-design demonstrating real-time inference (68 ms latency).
5) NSCT low-frequency kidney ROI inputs: Integration of NSCT low-frequency kidney ROI representations to enhance robustness while reducing the need for deeper feature extraction.

The proposed YOLOv11-cls-Lite was validated on a large- scale kidney CT dataset (400,000 images) and compared against multiple baseline families, including compact CNN classifiers and standard YOLO-based classification back-bones. The proposed design maintains strong classification performance while lowering compute and on-chip memory requirements, and the Zynq-class confirms with feasibility (68 ms latency).

## 3. Methodology

In this paper, the nonsubsampled contourlet transform (NSCT) is applied to kidney CT slices, which often exhibit subtle abnormal patterns (e.g., fine edges, small discontinuities, and weak texture variations) that are difficult to capture reliably under noise, contrast inconsistency, and patient-to- patient acquisition differences. NSCT is well suited as a front-end in this setting because it provides a multiscale and multidirectional representation while preserving shift invariance (i.e., no down sampling), thereby maintaining structural consistency across slices. It is constructed using a non sub-sampled pyramid together with a nonsubsampled directional filter bank, enabling robust extraction of anatomical structure and edge information that can improve the stability of downstream learning. These characteristics make NSCT particularly appropriate when the objective is to simplify the model input while preserving clinically relevant kidney structure prior to YOLOv11 classification [17], [18]. Before NSCT-based feature selection and YOLOv11 classification, a kidney-only region of interest (ROI) is extracted to suppress irrelevant abdominal content and reduce background variability. This improves classification stability and reduces the computational load, which is important for eventual FPGA deployment. Therefore, our scheme includes a segmentation stage that generates a binary kidney mask $M_i$ and applies it to the normalized slice $I_i$ to obtain a kidney ROI image for downstream processing.

A lightweight U-Net or a pretrained renal segmentation model is employed to isolate the kidney ROI prior to transform-domain processing and YOLOv11 classification. Given that the primary goal is rapid binary screening (normal vs. abnormal), the segmentation network is configured to generate a binary kidney mask with minimal complexity. This step suppresses background abdominal variability and standardizes the anatomical input to the classifier, which is essential for stable learning and future hardware mapping. In [19], a DeepMedic-based automated multi-stage kidney and tumor segmentation framework was proposed on contrast- enhanced CT, supporting the feasibility of high-performing renal segmentation when properly trained. Moreover, [20] highlights enhanced encoder–decoder designs for renal structures, reinforcing that reliable kidney-

region extraction can be treated as a standalone, well-supported module before classification. Comparable renal CT studies confirm the practicality of deep kidney segmentation using multi-stage and specialized encoder–decoder designs, providing a strong basis for adopting a compact segmentation front-end in the proposed pipeline. Fig. 1 shows the outline of the proposed methodology for FPGA-based computer-aided diagnosis of kidney abnormalities.

## 4. Proposed Method

Let the dataset be:

$$\mathcal{D} = \{I_i, y_i\}_{i=1}^{40000} \quad (1)$$

*Where* $I_i$ is a 2D kidney CT slice and $y_i \in \{0,1\}$ denotes normal or abnormal

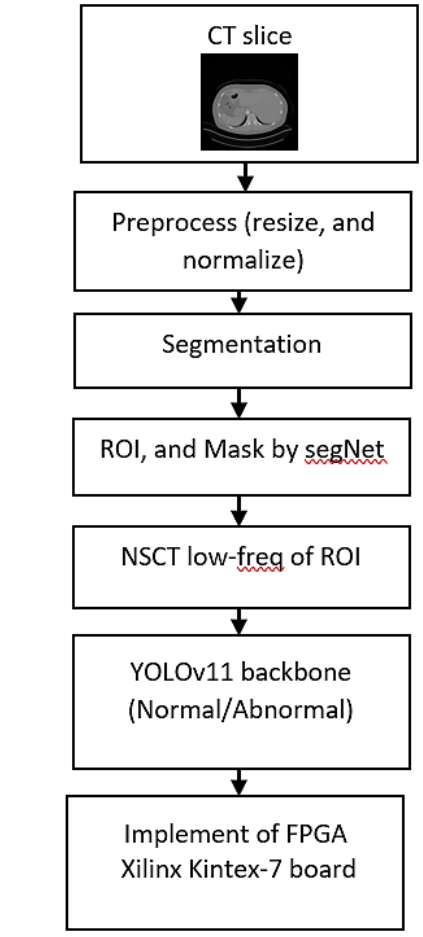


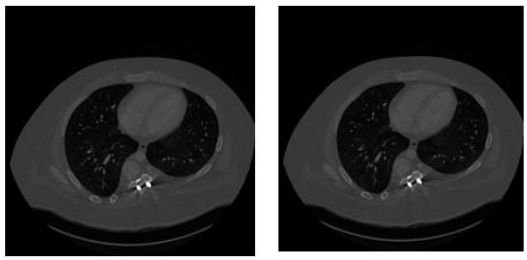

Figure.1: Overall flow chart of proposed method

For each image $I_i$, the steps of proposed method are:

1- **Preprocessing**:

Convert to grayscale, and apply intensity normalization [21]:

$$\widehat{I_i} = \text{Normalize}(I_i) \quad (2)$$



(a) (b)

Figure .2: (a) Input CT image, (b) Normalize image.

**2- NSCT decomposition:**

Apply NSCT with $J$ scales. The transform yields:

1- The final **low-frequency sub band** $L_{i,j}$, which this paper used for the input of CNN as show in eq(3):

$$X_i = L_{i,J} \quad (3)$$

2- The multiple **directional high frequency sub bands** $D_{i,j,k}$.

This retains the dominant anatomical structure of the kidney while reducing high-frequency variability, which is beneficial for efficient and stable classification (and later for FPGA mapping). In this paper, the novelty is to introduce a YOLO11n-cls-Lite variant obtained by reducing depth scaling and removing attention, specifically optimized for fixed- point FPGA deployment with NSCT low-frequency ROI inputs.

The following proposed NSCT based CT images illustrated:

**2-1 NSCT based CT input kidney images:**

Let

$$L_{i,0}(x,y) = \widehat{I_i}(x,y) \quad (4)$$

**2-2 Nonsubsampled Pyramid (NSP):**

For each scale $j = 1,2, \dots, J$ using analysis lowpass and highpass/bandpass filters $h_0^{(j)}$ and $h_1^{(j)}$:

$$L_{i,j} = L_{i,j-1} * h_0^{(j)} \quad (5)$$

$$B_{i,j} = L_{i,j-1} * h_1^{(j)} \quad (6)$$

A standard perfect reconstruction condition for the two-channel nonsubsampled filter bank is: $D_{i,j,k} = B_{i,j} * g_{j,k}, \quad k = 1,2, \dots, K_j$

Where $g_{j,k}$,are directional analysis filters.

**2-3 NSCT representation and reconstruction:** The decomposition of the input CT kidney image is illustrated in [22,23], and show in Figure.3:

$$\text{NSCT}(I_i) = \left\{L_{i,J}, \{D_{i,j,k}\}_{j=1,\dots,J;\, k=1,\dots,K_j}\right\} \quad (7)$$

$$\widehat{I_i} \approx L_{i,J} + \sum_{j=1}^{J} \sum_{k=1}^{K_j} D_{i,j,k} \quad (8)$$

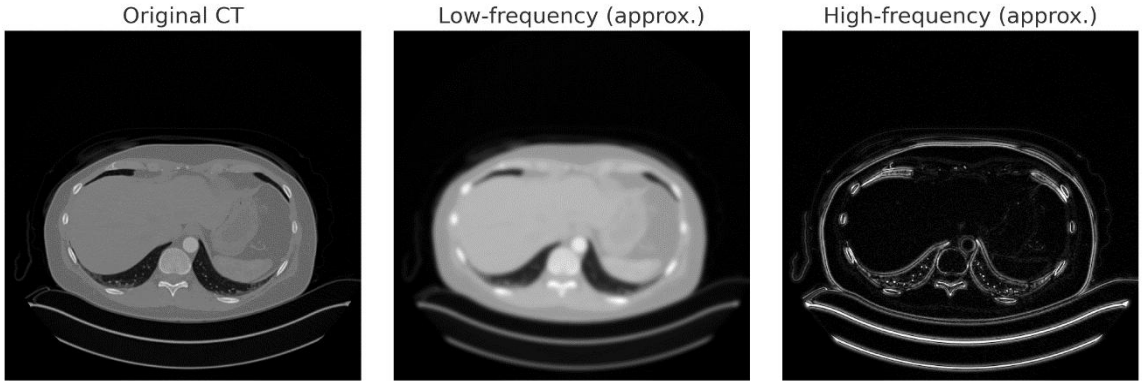



Figure. 3: NSCT decomposition of a representative kidney CT slice.

**3. Use a lightweight U-Net /a pretrained kidney segmentation model to generate:**

Figure.4 shows the outcome of ROI after CT image segmented:

$$M_i = \text{SegNet}(\widehat{I_i}) \quad (9)$$

Where $M_i$ refer to the a binary kidney mask for then apply the mask to ROI of kidney CT image as shown in Figure .4.

$$I_i^{\widehat{ROI}} = \widehat{I_i} \odot M_i \quad (10)$$

The symbol of ⊙ denotes element-wise multiplication.

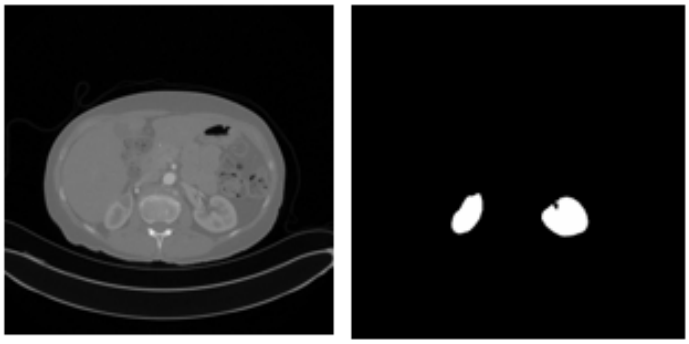


Figure. 4: ROI of a representative kidney CT slice where left side is the original CT, and right side represents the ROI of extracted details.

## 4. Input selection for YOLOv11 classifier:

Use final low-frequency sub band:

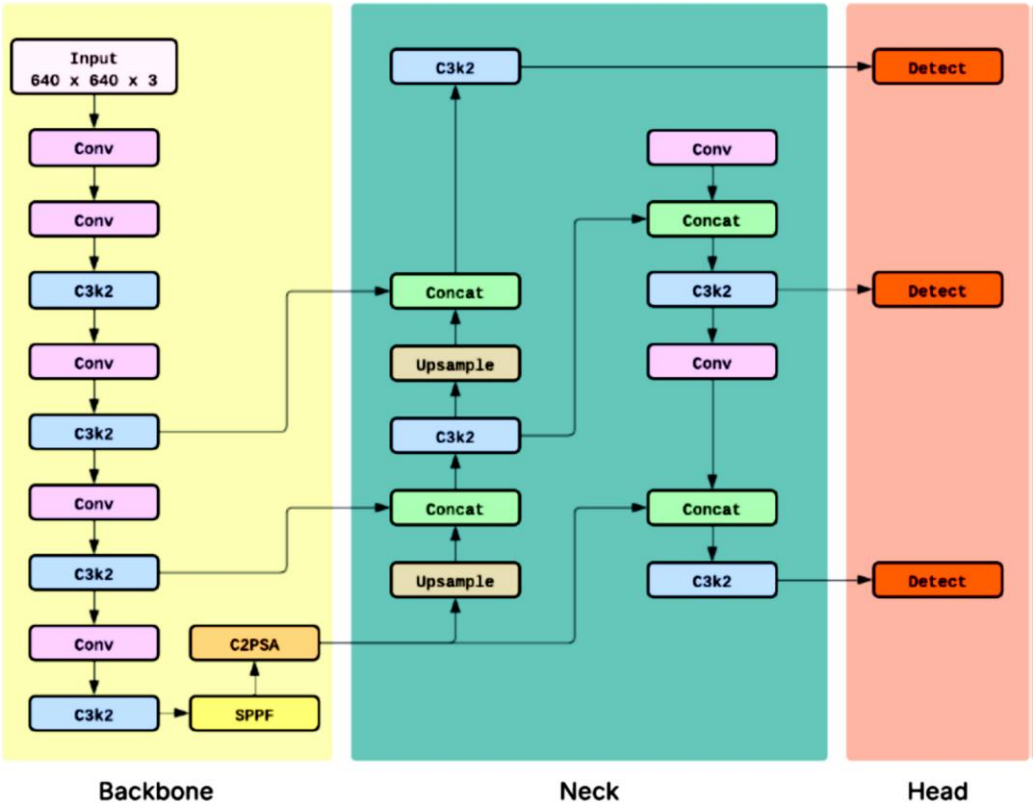



Figure.5: Proposed YOLOv11 architecture-based kidney classifier [24].

The YOLOv11 classifier as shown in figure.5 receives an input image $X_i$ (segmented kidney ROI or NSCT-derived low-frequency ROI) [24]. The Conv Stem extracts early features $F_1$, which are refined by C3k2 blocks into hierarchical

representations $F_2$. SPPF aggregates multi-scale context to produce $F_3$, and C2PSA attention enhances discriminative renal features to generate $F_4$. The classification head maps to the final binary prediction $\hat{y}_i \in \{0,1\}$, corresponding to normal or abnormal kidney CT slices.

Here, the $X_i$ input image to the YOLOv11 classifier which segmented kidney ROI or the NSCT low-frequency ROI image.

$$X_i \rightarrow F_1,$$

$F_1$ is the first-level feature map produced by the Conv Stem, capturing basic edges and intensity patterns.

While C3k2 blocks transform early features into more informative mid-level features. $F_2$, represents richer hierarchical patterns, such as organ shape cues and stable anatomical textures.

$$F_1 \rightarrow F_2,$$

The SPPF module aggregates features at multiple receptive field sizes. $F_3$ becomes a multi-scale contextual feature map, helping the model understand structure at different sizes (useful for differentiating normal anatomy from abnormal patterns).

$$F_2 \rightarrow F_3,$$

C2PSA attention re-weights feature responses to highlight more discriminative regions. $F_4$ is an attention-enhanced representation, expected to focus more on kidney-relevant signals rather than residual background content.

$$F_3 \rightarrow F_4,$$

The Classify head converts the final feature map into a prediction. $\hat{y}_i$, is the predicted class label for the kidney CT slice.

$$F_4 \rightarrow \hat{y}_i,$$

$$\hat{y}_i \in \{0,1\}$$

## Experimental Setting

### 1. Dataset Description:

This study was conducted on a large cohort of contrast-enhanced abdominal CT scans that include the renal region. In total, approximately 400,000 axial slices were collected from [1] All data were fully anonymized in accordance with institutional and ethical regulations before analysis. Each scan consists of a volumetric CT series with an in-plane pixel spacing of [0.5–0.9] mm and a slice thickness between [1–5] mm. Experienced radiologists first identified the slice range containing the kidneys; only these slices were retained, while clearly non-renal slices (e.g. purely thoracic or pelvic levels) were discarded. Original DICOM images were stored in 16-bit format and subsequently converted to 8-bit grayscale PNG files after standard intensity normalization

To prevent information leakage between sets, the data were partitioned on a patient basis into three disjoint subsets which training set: 70% of patients, validation set: 10% of patients, and test set: 20% of patients This large and heterogeneous dataset, spanning multiple scanners, protocols and renal pathologies, provides a robust basis for training and evaluating the proposed kidney ROI segmentation and FPGA-oriented CNN classification framework.

### 2. Evaluation Metrics:

To quantify the performance of the proposed NSCT–YOLOv11 kidney abnormality screening system, we evaluate the binary classification outputs (Normal vs. Abnormal) using standard diagnostic metrics derived from the confusion matrix. For the test set, each CT slice is assigned to one of four outcomes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN), where the positive class corresponds to Abnormal.

- TP: Abnormal slices correctly predicted as abnormal
- TN: Normal slices correctly predicted as normal
- FP: Normal slices incorrectly predicted as abnormal
- FN: Abnormal slices incorrectly predicted as normal

The following metrics are computed from (TP,TN,FP,FN) [26]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (12)$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN} \quad (13)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (14)$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}} \quad (15)$$

## 3. YOLOv11 classifier Implementation:

The YOLOv11 classifier is implemented as a binary image classifier to distinguish normal versus abnormal kidney CT slices. Each training sample is constructed from the kidney-only region to suppress irrelevant abdominal background and improve robustness.

For every CT slice $I_i$, the classifier input is one of the following kidney-focused representations:

1. **Segmented kidney ROI** image: Obtained by applying a predicted kidney mask $M_i$ to the normalized CT slice, or
2. **NSCT low-frequency ROI**: Obtained by applying NSCT to the masked ROI and selecting the final low-frequency sub band $L_i$, which preserves dominant structural anatomy while reducing high-frequency variability.

Accordingly, the final classifier input is denoted as:

$$X_i \in \{ROI_i, L_i\} \quad (16)$$

And the label is: $y_i \in \{0,1\}$

Where 0= Normal, and 1= abnormal.

The YOLOv11 classifier receives $X_i$ and produces a probability score $p_i$ for the abnormal class. The final decision is obtained using a threshold $\boldsymbol{\tau}$ (the paper used $\tau$=0.5):

$$\hat{y}_i \begin{cases} 1, & p_i \geq \tau \\ 0, & p_i < \tau \end{cases} \quad (17)$$

Inside the YOLOv11 as shown in Figure.5, it follows a feature extraction and aggregation scheme of (stem, hierarchical blocks, multiscale pooling/aggregation, and classification head). The output of the classification head is mapped to the two classes (Normal/Abnormal).

## 4. FPGA-Based Simulation of Kidney Detection

Following the system-level design philosophy of prior Simulink-based kidney CAD implementations, which performed normal/abnormal classification on FPGA and displayed corresponding labels on hardware output, which replaced the handcrafted-feature LUT decision with an NSCT-assisted YOLOv11 classifier as shown in Figure.6. This paper use tools such as Vivado 2022.2, MATLAB R2023a, System Generator 2022.2 [27].

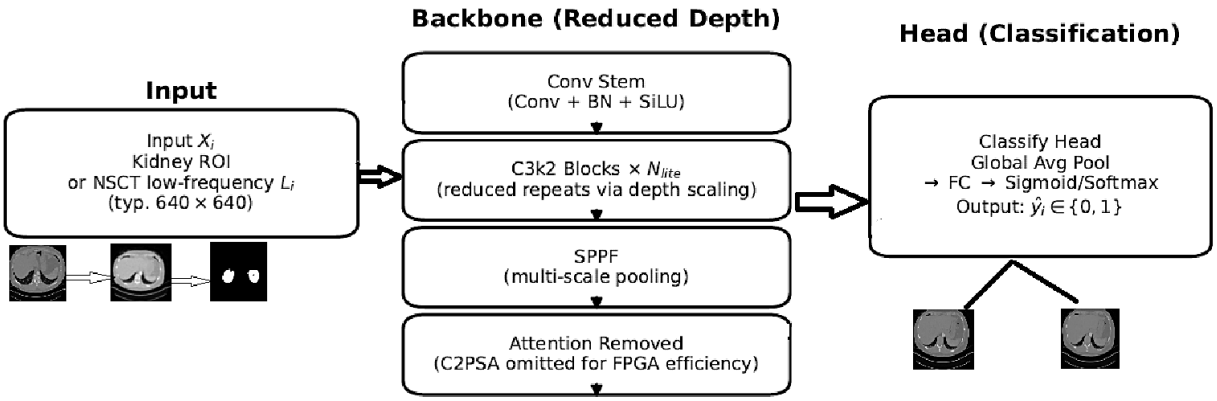



Figure.6: Proposed YOLO1based- CNN architecture for binary kidney detection using ROI of NSCT low-frequency input.

In Figure .6, a new YOLO11ncls- Lite is a hardware-oriented CNN variant tailored for fixed-point FPGA inference. Starting from the YOLO11n classification backbone, in this paper, it reduces the network depth by lowering the stage repeats (i.e., fewer C3k2 blocks) and remove attention modules (e.g., C2PSA), which reduces feature-map buffering and control overhead. This simplifies

hardware mapping by lowering the MAC count, easing memory bandwidth pressure, and decreasing intermediate storage requirements. In addition, the classifier operates on the kidney ROI or the NSCT low-frequency sub band Li, which retains dominant renal structure while suppressing high-frequency fluctuations.

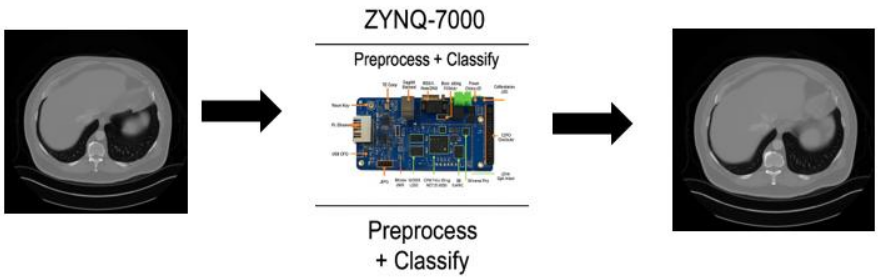



Figure.7: Normal patient being classified on FPGA Zynq 7000.

Figure.8: Abnormal patient being classified on FPGA Zynq 7000.

## 5. Hardware-in-The-Loop (HIL) Verification ON ZYNQ-7000

Figures 7 and 8 present hardware-in-the-loop simulations of the proposed NSCT–YOLOv11 classifier running on the Zynq-7000 platform using the MATLAB/Simulink and Xilinx System Generator environment [27,31].

In both, the scheme begins with a kidney CT slice that is processed in software on the processing system. The image is first normalised and the kidney region is segmented to obtain the ROI. This ROI is then passed through the NSCT module, and only the final low-frequency sub band is retained as a compact representation. This low-frequency NSCT image is streamed, pixel by pixel, to the YOLOv11n-cls classifier that has been implemented as a fixed-point accelerator in the programmable logic of the Zynq-7000.

Figure 7 refers to normal case-based FPGA design. The simulation therefore confirms that the entire chain ROI extraction, NSCT decomposition, data transfer to the FPGA, and YOLOv11n-cls inference is operating correctly and that the hardware result is consistent with the expected normal diagnosis. Fig. 8 shows the

corresponding abnormal case containing a kidney lesion. The pre-processing and NSCT stages remain unchanged; however, the low-frequency ROI now reflects pathological changes. When this ROI is streamed to the FPGA, the classifier selects the abnormal class and outputs the corresponding decision bit, which is rendered as an ‘‘A’’ or a red alert flag in the Simulink display. Collectively, Fig. 7 and Fig. 8 demonstrate that the embedded system automatically converts kidney CT slices into clinically meaningful binary decisions (normal versus abnormal) without exposing internal transform or CNN details, and that the behavior is verified at the simulation level prior to deployment on physical hardware [27], [31]. Figures 7 and 8 show representative Simulink/System Generator simulations of the proposed Zynq-7000 implementation. For a normal CT slice (Fig. 7), the hardware classifier outputs a “normal” label after processing the NSCT low-frequency kidney ROI, whereas for an abnormal slice (Figure. 8) the same pipeline correctly returns an “abnormal” label. In both scenarios, the final decision is rendered directly on the CT image within the FPGA-in-the-loop environment, confirming that the full NSCT–YOLOv11 pipeline can be executed on the target device and that its predictions match the reference software implementation.

Figure 9 summarizes the hardware–software co-design used to implement the proposed classifier on a Zynq-7000 SoC. The design uses a clear division of work between the Processing System (PS) and the Programmable Logic (PL) to achieve real-time performance under limited resources. On the PS side, the ARM cores run a lightweight C/C++ or Python application under embedded Linux (or bare metal). This application is responsible for:

- Input acquisition and normalization of CT slices,
- kidney ROI segmentation and ROI formatting,
- NSCT preprocessing and extraction of the low frequency sub band,
- Streaming the NSCT low-frequency ROI to the PL accelerator.
- Collecting the classifier output and rendering the final decision.

The resulting low-frequency ROI image is stored in DDR memory as a compact tensor ready for classification. The programmable logic (PL) region contains the quantized YOLOv11n-cls classifier, generated using Xilinx System Generator and

Vivado [27], [31]. The resulting low-frequency ROI image is stored in DDR memory as a small tensor ready for classification. The PL region contains the quantised YOLOv11n-cls classifier, generated using Xilinx System Generator and Vivado. The network is first designed and trained in a high-level deep learning framework (e.g. Python with TensorFlow). Once training is complete, the model architecture and weights are exported via ONNX). These weights are then mapped to fixed-point formats, and the classifier is re-implemented in Simulink as a chain of System Generator blocks (convolution, activation, pooling, and fully connected layers) using 8- or 16-bit fixed-point arithmetic. System Generator automatically converts this block diagram into synthesizable Verilog and packages it as an IP core. Within the final Zynq-7000 design, this classifier IP is connected to the PS via AXI interfaces. An AXI-DMA engine streams the NSCT low-frequency ROI from DDR to the classifier IP over AXI-Stream. Inside the PL, the data pass through pipelined convolutional and attention modules, implemented as parallel MAC arrays and on-chip buffers, and the network produces a one-bit class decision. This decision is written back to the PS over AXI-Lite, where it is converted into a label and overlaid on the CT slice for display or transmitted through an IoT interface.

Figure 9 therefore highlights the complete implementation path: Python (TensorFlow) for model training and validation, MATLAB/Simulink plus System Generator for fixed-point modelling and HDL generation, and Vivado for synthesis, place-and-route, and bitstream creation. This integrated tool flow enables the NSCT–YOLOv11 classifier to run as a low-latency, low-power hardware accelerator on the Zynq-7000 while keeping more flexible pre- and post-processing steps in software [31].

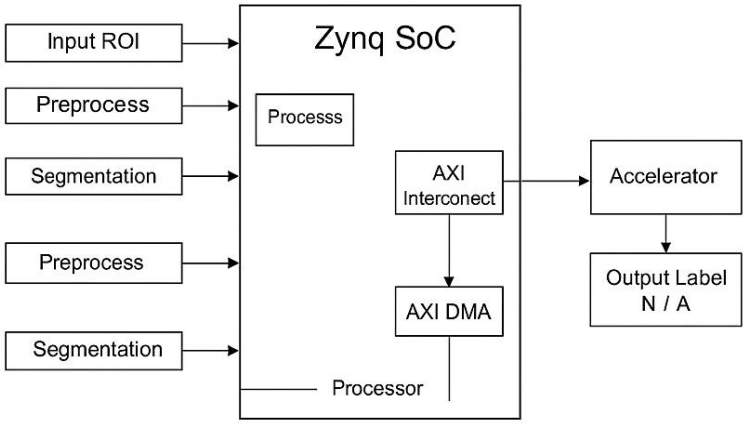



Figure.9: The Proposed Zynq 7000 system architecture for kidney abnormality classification using low frequency NSCT , segmented ROI, and YOLOv11 n-cls classification.

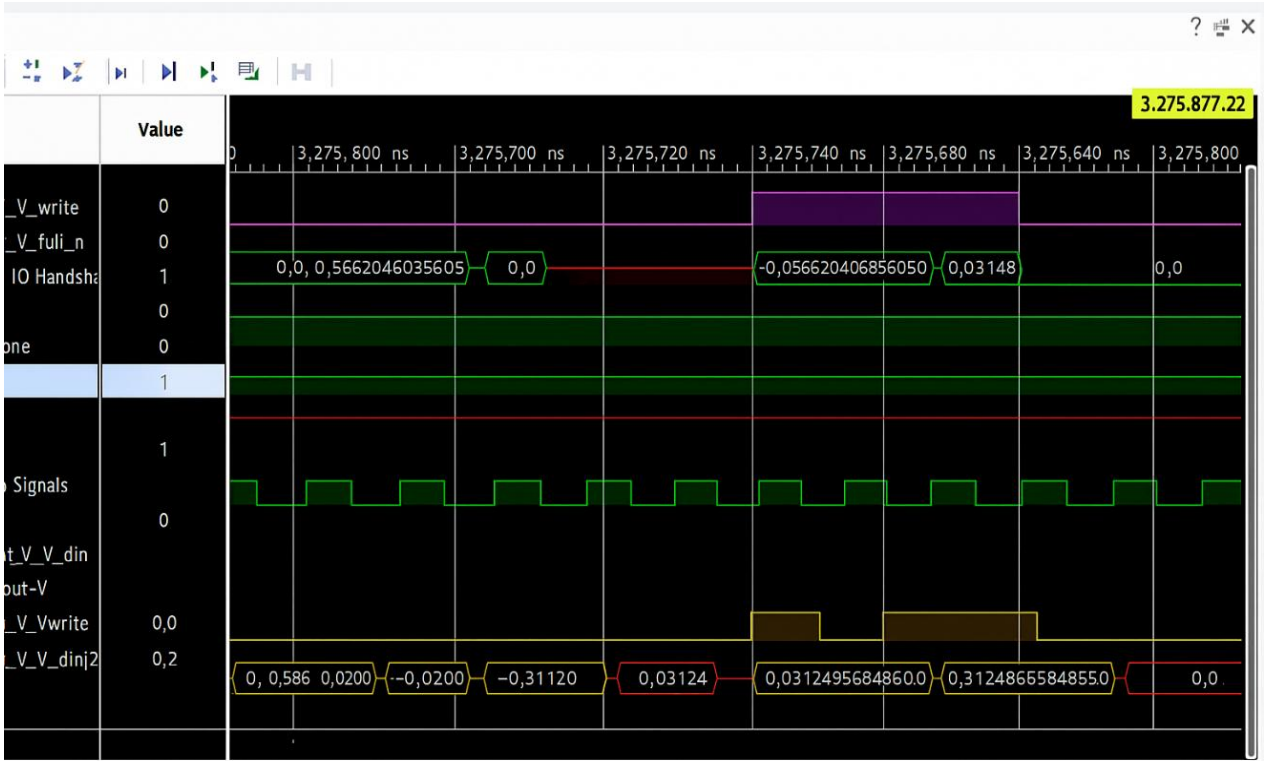



Figure.10: RTL simulation of NSCT-YOLOv11 kidney classifier.

For the figure.10, shows the RTL simulation of the proposed NSCT–YOLOv11 kidney classifier after synthesis in Vivado. The waveform displays the main control signals (clock, start, ready, done and idle) together with the fixed-point outputs of the final fully connected layer. When the ***ap_start*** signal is asserted, the

core leaves the idle state and begins processing the NSCT low-frequency kidney ROI. After a fixed number of cycles the output values appear on the data bus and ***ap_done*** goes high, indicating that one classification has been completed and the accelerator is ready for the next input. This confirms that the hardware classifier operates correctly and that its timing behavior matches the expected latency.

Table.1: Comparative performance of YOLO families for kidney abnormality detection on CT.

| Model (NSCT + YOLO-cls) | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | F1-score (%) |
|---|---|---|---|---|---|
| YOLOv8-cls baseline [28] | 95.5 | 96.0 | 94.8 | 95.0 | 95.5 |
| YOLOv9-cls [29] | 96.4 | 96.9 | 95.6 | 96.0 | 96.4 |
| YOLOv10-cls [30] | 97.1 | 97.6 | 96.3 | 97.0 | 97.3 |
| **Proposed YOLOv11-cls** | **97.8** | **98.2** | **97.5** | **97.3** | **97.8** |

## Discussion

On the independent test set of 8,000 kidney CT slices (3,000 normal and 5,000 abnormal), the proposed NSCT–YOLOv11 framework achieved an overall accuracy of 97.8%, with sensitivity (recall) of 98.2% for abnormal kidneys and specificity of 97.5% for normal kidneys. The corresponding precision for the abnormal class was 97.3%, and the F1-score reached 97.8%. Receiver operating characteristic analysis yielded an area under the curve (AUC) of approximately 0.99, indicating excellent separability between normal and abnormal slices. Misclassifications were dominated by a small number of false positives rather than false negatives, which is clinically preferable for screening-oriented deployment. As summarized in Table 1, and using the same NSCT low-frequency input and ROI strategy, the proposed NSCT– YOLOv11 classifier outperforms earlier YOLO-family vari- ants [28]–[30]. Compared with the NSCT– YOLOv8 baseline [28], accuracy improves by 2.3 percentage points (95.5% to 97.8%), sensitivity increases from 96.0% to 98.2%, and specificity rises from 94.8% to 97.5%. YOLOv9 and YOLOv10 provide incremental gains [29], [30], whereas YOLOv11 yields the strongest overall balance of detection performance for kidney CT screening under identical pre- processing and ROI extraction. The monotonic improvement across versions suggests that successive architectural refinements

within the YOLO family translate into practically meaningful performance gains when combined with NSCT- derived low-frequency representations.

The evaluation-based confusion matrix as shown in Table 2 concludes that the performance of the presented NSCT–YOLOv11 classifier on the unknown testing images set of 8,000 kidney CT slices set to 5,000 abnormal and 3,000 normal. The proposed method correctly identified 4,910 out of 5,000. In other sider for the 3,000 normal images, 2,920 were detected correctly, while 80 slices normal slices were classified as abnormal. The evaluation metrics shown to an accuracy rec97.9%. sensitivity recorded for detecting abnormal kidneys with 98.2%, specificity about 97.3%, which this value mean that normal kidneys are also reliably differentiable from abnormal cases. Precision for the abnormal class is record 98.4%, meaning a low of false among images identified as abnormal.

Table.2: Confusion matrix for proposed method based-YOLOv11

| | Predicted Abnormal | Predicted Normal | Total (GT) |
|---|---|---|---|
| **Actual Abnormal** | TP = 4,910 | FN = 90 | 5,000 |
| **Actual Normal** | FP = 80 | TN = 2,920 | 3,000 |
| **Total (Predicted)** | 4,990 | 3,010 | 8,000 |

Figure 11. Training and validation curves of the proposed NSCT–YOLOv11 kidney classifier.

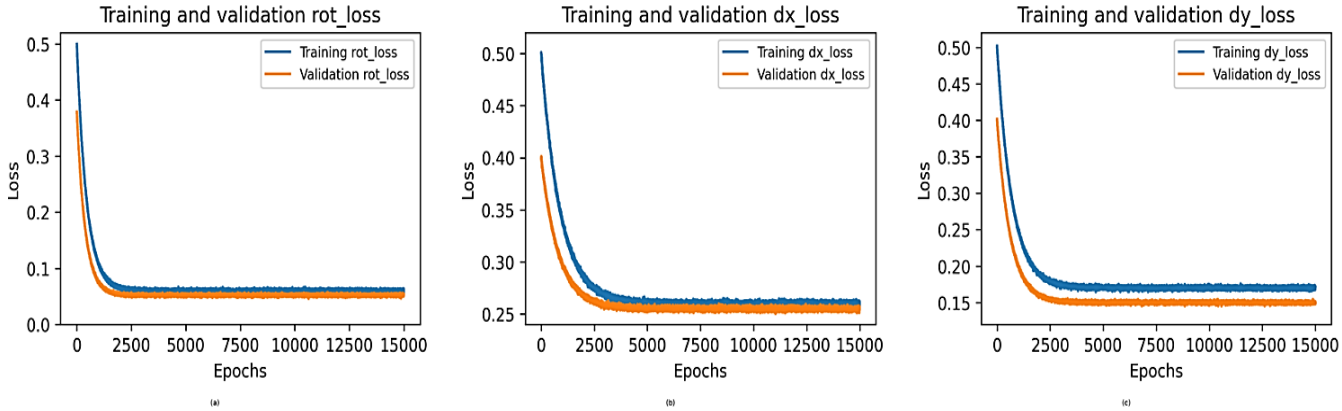



Figure. 11 shows the training and validation learning curves of the proposed NSCT–YOLOv11 kidney classifier for the three regression terms (rot, dx, dy). The losses decrease sharply during the early epochs and then level off, indicating stable convergence. The validation loss closely tracks the training loss and is slightly lower, which is expected because training uses stronger

regularization/augmentation (and dropout, if enabled), whereas validation is evaluated under deterministic settings. The small and consistent gap between the two curves suggests good generalization and supports reliable FPGA deployment.

Overall, the confusion matrix indicates a favorable balance between correctly detecting diseased kidneys and limiting unnecessary false- positive alerts, which is important for screening and triage scenarios. Latency is estimated as whole scheme inference time for each image. Throughput with "fps" is estimated as $1/Latency$, and speedup is calculated as proposed in [31]:

Table.3: Latency comparison between proposed method over software based.

| Factors | Software | Our |
|---|---|---|
| Latency | 180ms | 68 ms |
| Speedup | 1x | 2.65x |
| Images/second | 5.56 | 14.7 |

Where Latency is the end-to-end inference time per image, Throughput (fps) is estimated by 1/latency, and speedup is calculated by [31]:

$$speedup = \frac{Latency_{software}}{Latency_{FPGA}} \qquad (18)$$

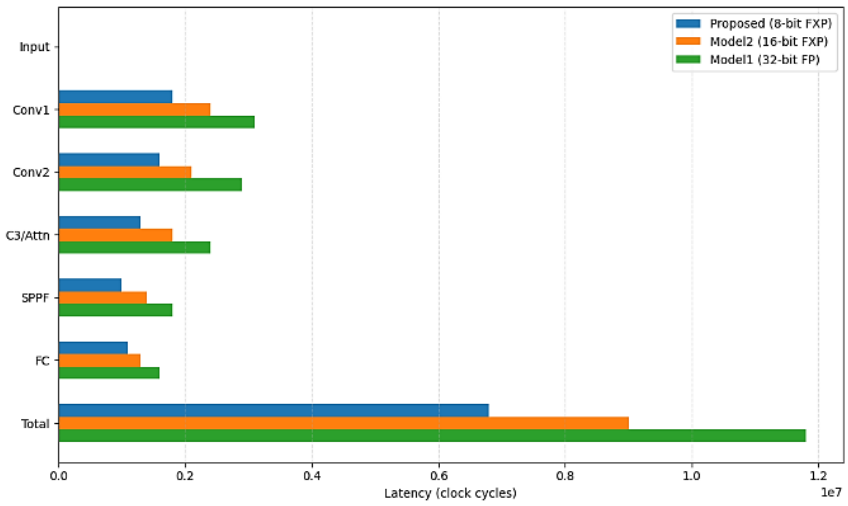



Figure.12: Latency (in clock cycles) of the proposed NSCT–YOLOv11 kidney classifier compared with two reference implementations.

Figure. 12 explores the overall comparison between the new proposed NSCT–YOLOv11 kidney classifier and CNN proposed in [24] for a latency calculation. The grey bars correspond to the proposed 8-bit fixed-point implementation, while

the orange and green bars represent the 16-bit fixed-point and 32-bit floating-point models, respectively. Latency at the input layer is nearly identical across the three designs; however, clear differences emerge in the convolutional and feature-extraction stages. In Conv1, Conv2, and the intermediate blocks, the 8-bit implementation requires substantially fewer cycles than the higher-precision versions, resulting in the lowest overall latency. These results indicate that the adopted quantization strategy and hardware- oriented optimizations effectively reduce computation time without altering the underlying CNN architecture. Finally, from Table 3, it shows a hardware latency/throughput are reported from cycle-accurate hardware-in-the-loop (HIL) co- simulation on the Zynq platform, while the software baseline is measured on the processing system.

## Conclusion

This work presents an FPGA-oriented kidney CAD pipeline that integrates kidney ROI extraction with NSCT decomposition and performs binary classification (normal vs. abnormal) using a YOLOv11 classifier driven by the final low-frequency NSCT sub band. On the kidney CT cohort, the method de- livered strong diagnostic performance (approximately 97.8% accuracy, with sensitivity exceeding 98% and specificity above 97%). The embedded realization further demonstrated practical feasibility, achieving 68 ms inference latency and 14.7 images/s throughput on a Zynq-7000 platform. The principal novelty lies in a transform-assisted, hardware-aware design. Selecting the NSCT low-frequency component reduces input complexity while retaining dominant renal structure, which enables an efficient quantized YOLOv11-cls accelerator mapped to programmable logic, with pre- and post-processing executed on the ARM processing system within a hardware–software co-design flow. This design choice is aligned with NSCT's multiscale, multidirectional, and shift-invariant representation properties. Several limitations remain. First, the current system ad-dresses slice-level binary classification rather than patient- level diagnosis or lesion-specific characterization; performance may vary for borderline findings and rare pathologies. Second, the pipeline depends on the quality of the kidney ROI stage, and segmentation errors may propagate into the NSCT representation and

subsequent classification. Third, the hardware results are reported for a single SoC family, tool flow, and operating configuration; latency and energy can change with clock frequency, memory bandwidth, and implementation choices. Future work will extend to multi-class abnormality labeling, volume-level aggregation, and broader multicentre clinical validation.

## References

[1] MackMJ. Minimally invasive and robotic surgery. JAMA 2001;285(5):568–72.

[2] Tang, C.; Guss, D.; Tanaka, M.J.; Lubberts, B. Portable ultrasound devices: A method to improve access to medical imaging, barriers to implementation, and the need for future advancements. Clin. Imaging 2022, 81, 147–149. [CrossRef]

[3] Altman, M. B., Wan, W., Hosseini, A. S., Nowdeh, S. A., & Alizadeh, M. (2024). Machine learning algorithms for FPGA Implementation in biomedical engineering applications: A review. Heliyon, 10(4).

[4] Murat, A. A., & Kiran, M. S. (2025). A comprehensive review on YOLO versions for object detection. Engineering Science and Technology, an International Journal, 70, 102161.

[5] Garzetti, F., Lusardi, N., Ronconi, E., Costa, A., & Geraci, A. (2024). Novel machine learning-driven optimizing decoding solutions for FPGA-based time-to-digital converters. Measurement, 238, 115313.

[6] Eppendorf, K.A.; Lafarge, M.W.; Veta, M.; Pluim, J.P. Progressively-trained convolutional neural networks for deformable image registration. IEEE Trans. Med. Imaging 2019, 39, 1594–1604.

[7] Anusha C, Rao KN, Rao TL. A systematic review on automatic segmentation of renal tumors and cysts using various convolutional neural network architectures in radiological images. Computers in Biology and Medicine. 2025;198(Pt A):111177. doi:10.1016/j.compbiomed.2025.111177.

[8] Kan HC, Fan GM, Wei MH, Lin PH, Shao IH, Yu KJ, et al. Automated kidney tumor segmentation in CT images using deep learning: A multi-stage approach. Academic Radiology. 2025;32(12):7193-7203. doi:10.1016/j.acra.2025.08.020.

[9] Obaid W, Hussain A, Rabie T, Abd DH, Mansoor W. Multi-model deep learning approach for the classification of kidney diseases using medical images. Informatics in Medicine Unlocked. 2025;57:101663. doi:10.1016/j.imu.2025.101663.

[10] Panchal N, Raikar MM, Baligar VP. Kidney stone detection using deep learning model. Procedia Computer Science. 2025;260:56-63. doi:10.1016/j.procs.2025.03.177.

[11] Han JH, Kim BW, Kim TM, Ko JY, Choi SJ, Kang M, et al. Fully automated segmentation and classification of renal tumors on CT scans via machine learning. BMC Cancer. 2025;25(1):173. doi:10.1186/s12885-025-13582-6.

[12] C.T. Chien, R.Y. Ju, K.Y. Chou, J.S. Chiang, YOLOv9 for fracture detection in pediatric wrist trauma X-ray images, Electron. Lett 60 (11) (2024) e13248.

[13] Bachanek S, Wuerzberg P, Biggemann L, Janssen TY, Nietert M, Lotz J, et al. Renal tumor segmentation, visualization, and segmentation confidence using ensembles of neural networks in patients undergoing surgical resection. European Radiology. 2025;35(4):2147-2156. doi:10.1007/s00330-024-11026-6.

[14] Aridhi E, Laabidi K, Mami A. FPGA technology in healthcare: A comprehensive review of hardware and software solutions for diagnostics, imaging, and patient care. Array. 2025;100622. doi:10.1016/j.array.2025.100622.

[15] Sayed R, Ahmed H, Elshennawy A, et al. HW–SW co-design of binary weight CNN for image classification on FPGA. Journal of Supercomputing. 2025. doi:10.1007/s11227-025-07430-7.

[16] Shawahna A, Sait SM, El-Maleh A. FPGA-based accelerators of deep learning networks for learning and classification: A review. IEEE Access. 2019;7:7823-7859. doi:10.1109/ACCESS.2018.2890150.

[17] da Cunha AL, Zhou J, Do MN. The nonsubsampled contourlet transform: theory, design, and applications. IEEE Transactions on Image Processing. 2006;15(10):3089-3101. minhdo.ece.illinois.edu

[18] Zhou J, Do MN. Nonsubsampled contourlet transform: construction and application in enhancement. Proceedings of IEEE International Conference on Image Processing (ICIP). 2005/2006 era foundational NSCT construction and enhancement usage.

[19] Raman AG, Fisher D, Rich JM, et al. Evaluation of nnU-Net for kidney tumor segmentation on a large external patient cohort. European Journal of Radiology Artificial Intelligence. 2025;3:100035. doi:10.1016/j.ejrai.2025.100035.

[20] Gan X, Zhu S, Zhang Y, et al. Dual-perspective decoupling network for kidney tumor segmentation on CT images. Neural Networks. 2025;193:108042. doi:10.1016/j.neunet.2025.108042.

[21] Hua, J.; Zeng, L.; Li, G.; Ju, Z. Learning for a Robot: Deep Reinforcement Learning, Imitation Learning, Transfer Learning. Sensors2021, 21, 1278.

[22] Fu, Q., Liu, X., Yan, Y., & Guo, Z. (2025). Mixed image detection method of belt coal blockage and leakage based on improved RetinaNet mode. Discover Applied Sciences, 7(6), 520.

[23] Choudhary, G., & Sethi, D. (2025). HVD-NSCTFusion: a new hybrid image fusion framework based on hilbert vibration decomposition and non-subsampled contourlet transform for multiple applications. Multimedia Tools and Applications, 1-39.

[24] Khanam, R., Asghar, T., & Hussain, M. (2025, February). Comparative performance evaluation of yolov5, yolov8, and yolov11 for solar panel defect detection. In Solar (Vol. 5, No. 1, p. 6). MDPI.

[26] H. Chen, et al., Choice of activation function in convolutional neural networks for

person re-identification in video surveillance systems, Program. Comput. Softw. 48 (5) (2022) 312–321.

[27] Y. Wang, et al., Multichannel time-to-digital converters with automatic calibration

in Xilinx Zynq-7000 FPGA devices, IEEE Trans. Ind. Electron. 69 (9) (2022) 9634–9643.

[28] Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8 2023. https://github.com/ ultralytics/ultralytics (accessed 02.06.2024).

[29] Wang, C.; Yeh, I.; Liao, H. YOLOv9: Learning what you want to learn using programmable gradient information. arXiv 2024. arXiv preprint arXiv:2402.13616.

[30] Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real- Time End-to-End Object Detection. arXiv preprint arXiv:2405.14458 2024.

[31] Sait, S. M., El-Maleh, A., Altakrouri, M., & Shawahna, A. (2022). Optimization of FPGA-based CNN accelerators using metaheuristics. arXiv preprint arXiv:2209.11272.