Mesopotamian journal of Big Data
Vol. **(2025)**, 2025, **pp.** 136–143
DOI: https://doi.org/10.58496/MJBD/2025/009 , ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

MESOPOTAMIAN
JOURNAL OF BIG DATA

Research Article

Precise Kidney Stone Localization in Medical Imaging via a Capsule Network

Omar Hammad Jasim[1], Daniah Abdul Qahar Shakir [2], Mohammed Shihab Hamad [3], Waleed Kareem Awad [3]

[1] Department of Information Technology, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq

[2] Department of Computer Networks System, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq

[3] Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq

ARTICLEINFO

Article History

Received 2 Jun 2025
Revised 12 Jul 2025
Accepted 15Aug 2025
Published 23 Aug 2025

Keywords

Capsule Networks

Machine Learning

CNN

CapsNets

Kidney Stone

Object Detection

ABSTRACT

Traditional convolutional neural networks (CNNs) face significant limitations in medical imaging when detecting small, spatially variable objects such as kidney stones, primarily due to their inability to preserve pose information and spatial correlations through max pooling operations. While previous CNN-based studies achieved approximately 93% accuracy in kidney stone detection, they struggled with the precise localization of small or partially obscured stones, creating a critical research gap in automated urological diagnostics. This study develops and evaluates a capsule network (CapsNet) framework that leverages dynamic routing and vector-based capsules to increase kidney stone localization accuracy in computed tomography (CT) images while maintaining spatial coherence and reducing false positives. The CapsNet model incorporates convolutional layers, primary capsules, and stone capsules via dynamic routing algorithms. The approach was systematically evaluated via a publicly accessible kidney stone CT dataset from the Mendeley repository, comprising 512 anonymized abdominal CT slices preprocessed to 256×256 pixels. The dataset was partitioned into training (70%), validation (15%), and test (15%) sets. The performance was compared against that of a baseline CNN under identical conditions using 50 epochs and the Adam optimizer. The results demonstrate CapsNet's superior performance across all the metrics: 96.5% accuracy, 96% precision, 97% recall, 96% F1 score, 0.93 Dice coefficient, and 0.89 IoU, significantly outperforming the CNN baseline (92% accuracy, 0.84 Dice coefficient, 0.78 IoU). CapsNets enhance kidney stone localization and generalization by preserving spatial and pose information, improving diagnostic accuracy in medical imaging.

1. INTRODUCTION

Millions of people throughout the world suffer from kidney stone illness. Accurate identification and location of kidney stones is essential for diagnosis, treatment planning, and recurrence prevention [1]. Among the available imaging techniques, including computed tomography (CT), ultrasonography (US), and X-ray, CT is considered the gold standard because of its high sensitivity and specificity[2] [3] [4]. Nevertheless, manually interpreting medical images time-consuming, error-prone, and heavily dependent on the skill of radiologists [5]. Artificial intelligence has revolutionized medical image analysis. CNNs have shown considerable promise in tasks such as object detection, segmentation, and classification across a range of medical imaging applications [6] [7]. However, their limitations in handling pose variations and spatial hierarchies hinder the accurate localization and detection of small, irregular shaped objects like kidney stones. As a remedy for these drawbacks, CapsNets have attracted much attention as a solution to these constraints[8] [9]. It preserves spatial links and recognizes part-whole hierarchies [10]. CapsNets are especially well suited for applications that demand high accuracy because they use dynamic routing to maintain feature orientation and placement [11],[12].

Our study aims to develop a deep learning framework that employs capsule networks for precisely identifying kidney stones via medical imaging. The proposed method seeks to improve localization precision and reduce false positives by leveraging the attributes of CapsNet to maintain spatial coherence. The methodology is assessed via a publicly accessible dataset of kidney CT scans, with findings compared with those of leading CNN-based techniques to validate the efficacy of the proposed system

*Corresponding author. waleed.kareem@uoanbar.edu.iq

2. RELATED WORK

Recently, researchers have highlighted the importance of integrating AI into medical imaging. Some of these studies focused on identifying and classifying kidney stones via traditional image processing methods, but they often struggle to identify small or spatially irregular objects accurately. Despite extensive research on deep learning methods for finding and classifying kidney stones, a thorough search of major scientific databases such as PubMed, Scopus, and IEEE Xplore revealed no studies that specifically utilized capsule networks for this purpose. The literature has focused predominantly on traditional convolutional neural network (CNN) designs, such as the following:

Alper et al. [13] used a CNN in a retrospective evaluation of 2,959 unenhanced CT scans from 455 patients with kidney stones. Conventional CT achieved test accuracies of 78% (axial), 72% (coronal), and 93% (sagittal) in detecting kidney stones, with no reported F1 scores exceeding 90%.

Bouzon et al. [14] applied a low-complexity CNN with just four convolutional layers to coronal CT images of 1,799 kidney stones. Under an 80/10/10 /val/test split, a test accuracy of 97.2% and an F1 score of 97.6% were achieved. However, the lightweight network excels in slice-level classification, which means that it does not preserve the interslice spatial context.

Caglayan et al. [15] In this retrospective study of 455 patients (405 with stones and 50 without stones) who underwent unenhanced CT, the researchers evaluated 2,959 images across the axial, coronal, and sagittal planes via an xResNet50-based CNN. The model training accuracy exceeded 97% across all planes (98.2–99.1%), but the testing accuracy ranged from 78–70% (axial), 63–72% (coronal), and 85–93% (sagittal) depending on the stone size.

This study seeks to address this deficiency by utilizing the spatial encoding features of CapsNet to improve the localization of kidney stones in CT images. The suggested strategy enhances the strengths and mitigates the limitations highlighted in other research, providing a more dependable and interpretable instrument for clinical decision-making in urology and radiology.

3. CAPULE NEURAL NETWORKS IN MEDICAL IMAGING

CapsNets are classified as a form of deep learning system that captures part-whole relationships to enhance robust pattern recognition[16], [17]. This signifies a notable progression over traditional CNNs by mitigating one of the CNN's primary shortcomings in the degradation of spatial correlations among features [18]. CapsNets use neuron clusters (capsules) to encode the probability and pose (location, orientation, and size) of a feature [17]. This enables the model to comprehend the spatial configuration of objects with their components in a more effective manner [19],[20]. Conventional CNNs with procedures such as max pooling, which removes location information [21]. CapsNet employs dynamic routing by agreement, enabling lower-level capsules to selectively transmit outputs to higher-level capsules based on prediction coherence. [22].

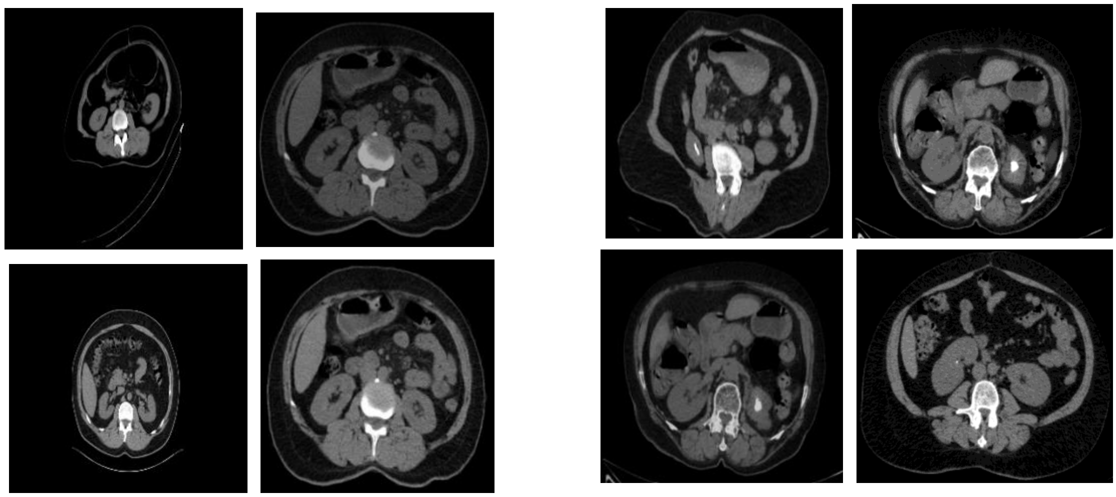
Accurate localization and structural recognition are essential for the detection of lesions, tumors, and kidney stones[23]. These structures can differ in form, size, and orientation. Unlike convolutional neural networks (CNNs)[24]. It may misclassify or overlook objects owing to spatial irregularities[25]. CapsNets preserve spatial awareness and provide enhanced generalizability on limited datasets. Consequently, utilizing CapsNets for kidney stone identification improves detection accuracy and reduces false positives, making them an ideal choice for medical imaging applications that need precision and reliability.

4. METHODOLOGY

The approach for developing and testing a capsule network system for precisely identifying kidney stones is described in this section. This methodology includes four main stages: data preparation, image preprocessing, model engineering, and evaluation metrics.

4.1 Data Preparation

The dataset used in this investigation was sourced from the kidney stone CT dataset accessible in the Mendeley data repository [26]. The dataset comprises anonymized abdominal CT slices obtained from individuals diagnosed with nephrolithiasis. The photos are grayscale and exhibit varying resolutions, although the data are organized appropriately for binary classification and localization tasks. To ensure consistency, all pictures were reduced to 256 × 256 pixels, and labels were transformed into binary masks as necessary to provide pixel-level assessment measures, including the Dice coefficient and Intersection over union (IoU). The dataset was partitioned as follows: training set: 70% of the data (358 photos), validation set: 15% (77 images), and test set: 15% (77 images). The division was detailed to preserve the ratio of positive and negative instances throughout all the subgroups. Some samples of images are shown in Figure 1.



(a) (b)

Fig. 1: Samples of the Kindey Stone Dataset: (a) nonstone, (b) stone

4.2 Image Preprocessing

Before feeding images into the network, a sequence of preprocessing procedures is implemented to improve image quality and diminish noise: normalization and contrast enhancement, acoustic attenuation, and data augmentation[25]. The following figure shows the preprocessing of a sample image.

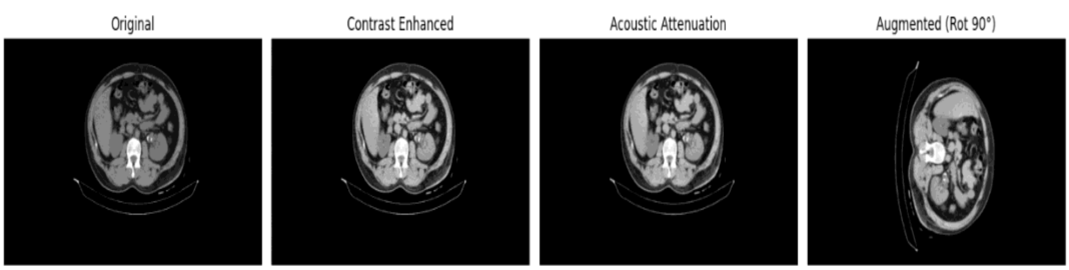


Fig. 2: Preprocessing procedures

4.3 Capsule Network Architecture

The foundation of the proposed system is a tailored capsule network influenced by the original design presented by Sabour [17]. The model consists of the following layers:

- Convolutional Layer: Extracts fundamental characteristics from the input CT picture via a 9x9 kernel to capture high-level edges and texture information.
- Primary Capsules: Clusters of convolutional units are transformed into vector capsules, which encapsulate an 8-dimensional feature vector. These capsules encode low level patterns such as shape fragments, edge groupings and potential object contours.
- Digit Capsules (Stone Capsules): Advanced capsules indicate the existence and positioning of kidney stones. Each capsule in this layer outputs a 16-dimensional vector that encodes both the probability and pose the information of a potential stone.
- Dynamic Routing: The routing-by-agreement technique facilitates communication between lower and higher level capsules based on consistent feature concordance, enabling the model to preserve pose and spatial linkages.

The margin of the loss function promotes accurate categorization while mitigating erroneous predictions. Localization maps are generated from the activation patterns of the capsules, indicating the stone's position within the picture.

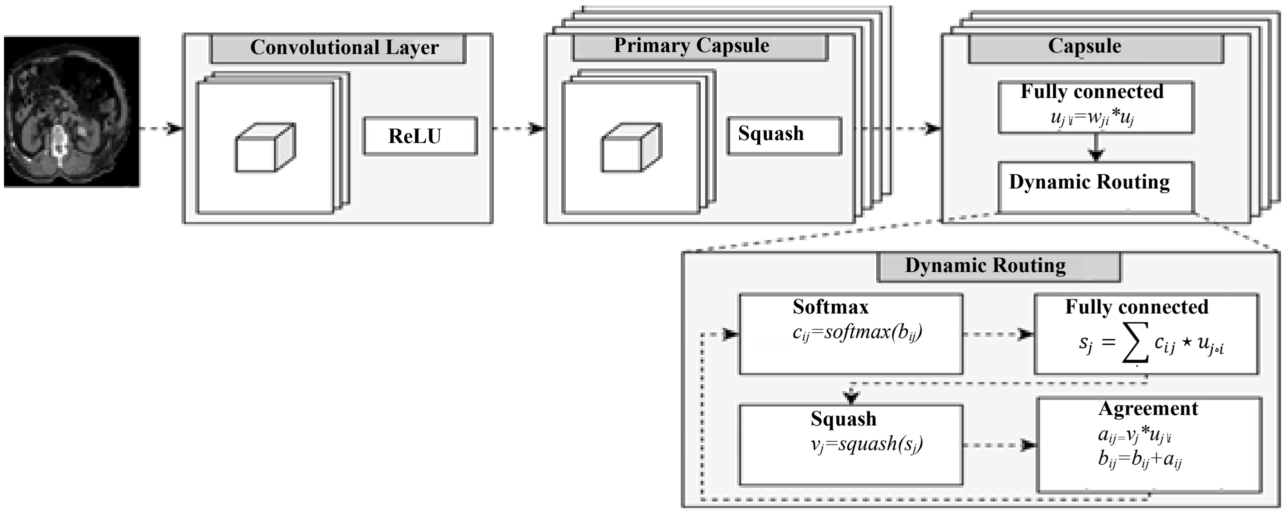


Fig. 3: Capsule network architecture

where

$u_{j|i}$ *Prediction vector from capsule i to capsule j*
b_{ij} *Logit (initial log probability) between capsule i and capsule j , updated during routing*
c_{ij} *Coupling coefficient (output of softmax) indicating how much capsule i contributes to capsule j*
s_j *Total input vector to capsule j (weighted sum of predictions)*
v_j *Output vector of capsule j after squash function (encodes presence and pose)*
a_{ij} *Agreement score (dot product) between v_j and $u_{j|i}$*
W_{ij} *Weight matrix used to transform u_j to prediction*

5. TRAINING CONFIGURATION

A baseline CNN model is constructed with comparable parameters to measure the performance enhancement provided by CapsNet. This facilitates an equitable assessment of localization accuracy, detection precision, and noise and picture distortion robustness. Both models are trained on an identical dataset with the same training methodologies and learning rates. The capsule network and the CNN baseline were trained under the same conditions to guarantee a fair comparison. The subsequent parameters employed for both models shown in Table 1.

TABLE I: TRAINING PARAMETERS

Epochs: 50	Loss Function For CNN: Cross-Entropy Loss
Batch Size: 16	Cross-Validation: 5-Fold Cross Validation applied to improve statistical significance
Learning Rate: 0.001	Loss Function for CapsNet: Margin Loss
Optimizer: Adam[27]	Validation Split: 20% of training data used for validation

6. RESULTS AND DISCUSSION

The experimental outcomes of the Capsule Network (CapsNet) model are compared with those of a baseline Convolutional Neural Network (CNN), emphasizing the precision of kidney stone localization. The findings are analysed through quantitative metrics, qualitative visualization, and performance robustness.

6.1 Quantitative Results

The performance of both models was assessed via an identical dataset and assessment procedure. Table 2 encapsulates the principal metrics derived from the test set.

TABLE II: PERFORMANCE COMPARISON BETWEEN CAPSNET AND CNN

Metric	Capsule Network	CNN Baseline
Accuracy (%)	96.5	92
Precision (%)	96	90
Recall (%)	97	89
F1-Score (%)	96	90
Dice Coefficient	0.93	0.84
IoU (Localization)	0.89	0.78

The table compares the efficacy of kidney stone detection between the capsule network and a CNN baseline. The capsule network surpasses capsule network outperforms the CNN across all measures, with an accuracy of 96.5%, along with enhanced precision, recall, and F1score, and markedly improved localization metrics (Dice coefficient of 0.93 and IoU of 0.89), demonstrating superior proficiency in classification and spatial localization.

6.2 Qualitative Analysis

The results validate that CapsNet augments detection precision and elevates localization accuracy. Furthermore, its capacity to generalize effectively on a minimal dataset renders it appropriate for medical fields where labelled data are deficient.

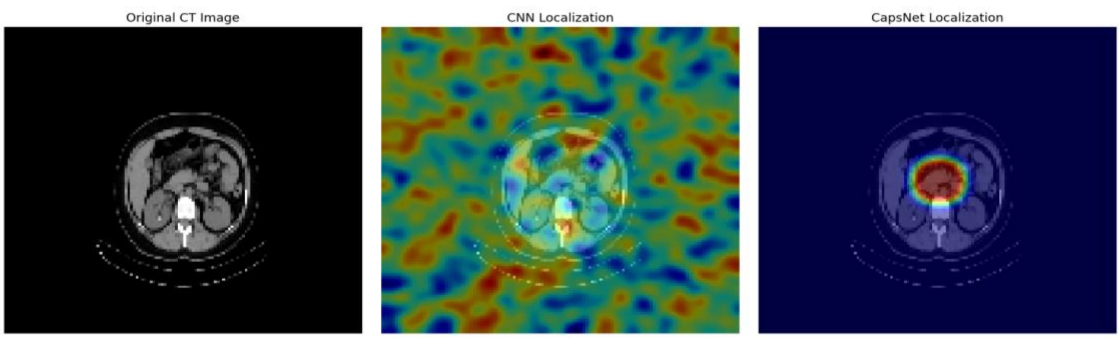


Fig. 4: Localization heatmaps generated by both algorithms.

The confusion matrices for the capsule network and the CNN baseline are presented. This graphic presents a comprehensive comparison of the performance of each model in categorizing photos with and without kidney stones. The CapsNet model attained greater count of true positives (69) and true negatives (70) while reducing false positives and false negatives to only two and three instances, respectively. The results validate that CapsNet augments detection precision and increase localization accuracy. Furthermore, its capacity to generalize effectively on a minimal dataset renders it appropriate for medical fields where data are deficient. In contrast, the CNN baseline exhibited more classification errors, including six false positives and seven false negatives. The discrepancies are significant in medical diagnostics, as even a minor incidence of false negatives can lead to missed diagnoses and delayed treatment. as shown in the following figure.

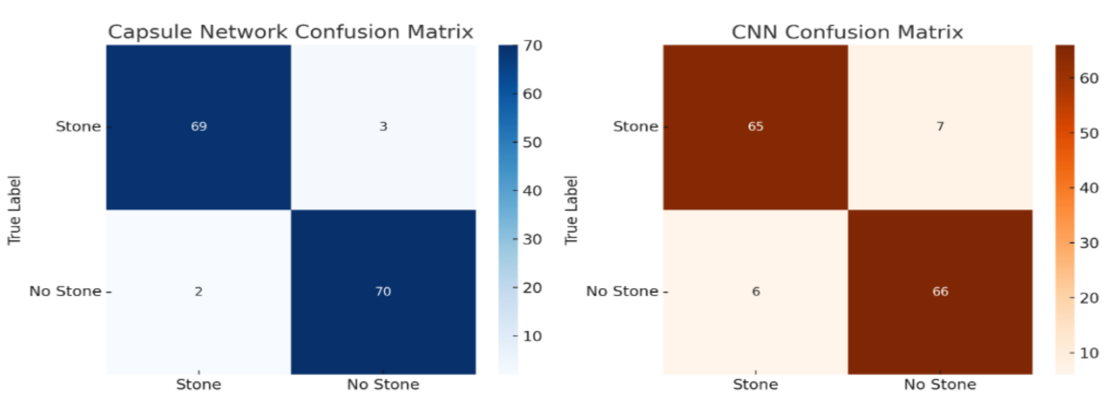


Fig. 5 Confusion matrices

Figure 6 depicts the training loss trajectories for the capsule network and the CNN throughout 50 epochs. The CapsNet model exhibits superior convergence speed and reduced final loss relative to those of the CNN. This signifies that it might acquire significant features more rapidly and with enhanced stability, the more gradual slope of CapsNet signifies less variability in the learning process, which is essential for mitigating overfitting, especially when relatively small datasets are employed. In contrast, the CNN loss curve demonstrates a slow decline accompanied by considerable fluctuations, suggesting potential challenges in the generalization and acquisition of complex spatial features. The findings corroborate the premise that CapsNet, because of its pose-aware design, is more adept at tasks such as kidney stone localization, where accuracy and spatial comprehension are essential.

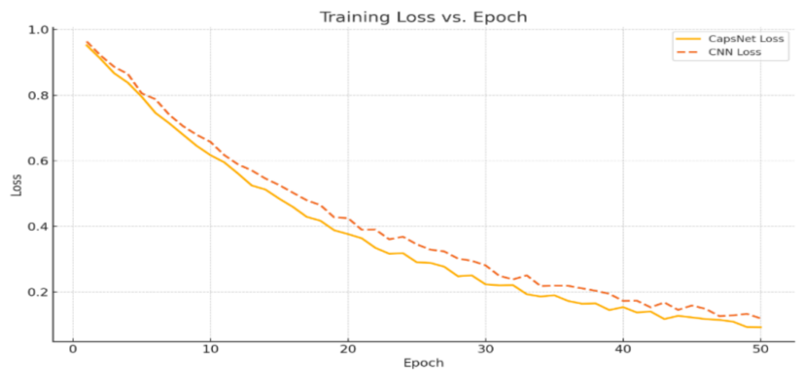


Fig. 6: Training Loss vs. Epoch

7. CONCLUSION

This research introduces a capsule network-based method for accurately localizing kidney stones in CT images. The suggested model exhibited enhanced performance relative to a typical CNN, attaining an accuracy of 96.5%, a Dice coefficient of 0.93, and an IoU of 0.89 while markedly decreasing the number of false positives and false negatives.
The CapsNet's ability to maintain spatial and postural information improved the localization precision and classification dependability, as demonstrated by confusion matrices and training loss curves. These findings highlight CapsNet's ability to improve diagnostic accuracy in medical imaging. Subsequent research may investigate the extension of this framework to three-dimensional data and its integration into real-time clinical operations.

Conflicts Of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

Funding

The authors had no institutional or sponsor funding.

Acknowledgement

The authors extend appreciation to the institution for their unwavering support and encouragement during this research.

References

[1] M. S. Pearle, E. A. Calhoun, and G. C. Curhan, "Urologic diseases in America project: Urolithiasis," in *Journal of Urology*, 2005. doi: 10.1097/01.ju.0000152082.14384.d7.
[2] M. Lazar *et al.*, "Protocol analysis of dual-energy CT for optimization of kidney stone detection in virtual non-contrast reconstructions," *Eur Radiol*, vol. 30, no. 8, 2020, doi: 10.1007/s00330-020-06806-9.

[3] P. F. Fulgham, D. G. Assimos, M. S. Pearle, and G. M. Preminger, "Clinical effectiveness protocols for imaging in the management of ureteral calculous disease: AUA technology assessment," 2013. doi: 10.1016/j.juro.2012.10.031.
[4] S. Y. Mohammed, N. S. Jumaah, E. Abbas, H. M. Ali, and M. Aljanabi, "Healthcare Intelligence and Decision Making: Big Data's Role in Predictive Analytics for Clinical Decision-Making," *Mesopotamian Journal of Big Data*, vol. 2024, pp. 223–229, Dec. 2024, doi: 10.58496/MJBD/2024/016.
[5] C. Yan and N. Razmjooy, "Kidney stone detection using an optimized Deep Believe network by fractional coronavirus herd immunity optimizer," *Biomed Signal Process Control*, vol. 86, 2023, doi: 10.1016/j.bspc.2023.104951.
[6] E. T. Mahdi, W. K. Awad, M. M. Rasheed, and A. T. Mahdi, "Proposed Security System for Cities Based on Animal Recognition Using IOT and Clouds," in *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 834–839. doi: 10.1109/DeSE60595.2023.10469597.
[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, 2017, doi: 10.1145/3065386.
[8] "bibtex (1)".
[9] S. J. Pawan and J. Rajan, "Capsule networks for image classification: A review," 2022. doi: 10.1016/j.neucom.2022.08.073.
[10] G. H. Alashour and N. K. El Abbadi, "Advances and Insights in Image Texture Analysis: A Review," *Mesopotamian Journal of Big Data*, vol. 2025, pp. 108–135, Aug. 2025, doi: 10.58496/MJBD/2025/008.
[11] M. E. Paoletti *et al.*, "Capsule Networks for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, 2019, doi: 10.1109/TGRS.2018.2871782.
[12] W. K. Awad, E. T. Mahdi, and M. N. Rashid, "IJTPE Journal FEATURES EXTRACTION OF FINGERPRINTS BASED BAT ALGORITHMS," *International Journal on "Technical and Physical Problems of Engineering" (IJTPE) Issue*, vol. 53, pp. 274–279, 2022, [Online]. Available: www.iotpe.com
[13] 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, no. 5, 2022, doi: 10.1590/S1677-5538.IBJU.2022.0132.
[14] M. F. Bouzon, S. P. de Oliveira, O. E. H. Fugita, and P. S. S. Rodrigues, "Automatic Kidney Stone Detection using Low-cost CNN with Coronal CT Images," 2024. doi: 10.5753/wvc.2023.27527.
[15] 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, no. 5, pp. 830–839, Sep. 2022, doi: 10.1590/S1677-5538.IBJU.2022.0132.
[16] R. LaLonde, Z. Xu, I. Irmakci, S. Jain, and U. Bagci, "Capsules for biomedical image segmentation," *Med Image Anal*, vol. 68, 2021, doi: 10.1016/j.media.2020.101889.
[17] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in Neural Information Processing Systems*, 2017.
[18] S. John, "Advancing Capsule Networks: Addressing CNN Limitations for Hierarchical Feature Learning," *International Journal of Emerging Trends in Computer Science and Information Technology*, doi: 10.63282/30509246/IJETCSIT-V1I3P101.
[19] D. Wang and Q. Liu, "An optimization view on dynamic routing between capsules," in *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings*, 2018.
[20] W. K. Awad and E. T. Mahdi, "Tasks Scheduling Techniques in Cloud Computing," in *3rd Information Technology to Enhance e-Learning and Other Application, IT-ELA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 94–98. doi: 10.1109/IT-ELA57378.2022.10107956.
[21] "scholar".
[22] E. T. Mahdi, W. J. Al-Kubaisy, and M. Mahmood, "Capsule Networks for Rice Leaf Disease Classification," *Journal of Intelligent Systems and Internet of Things*, vol. 14, no. 2, pp. 1–7, 2025, doi: 10.54216/JISIoT.140201.
[23] M. M. Abu-Faraj and M. Zubi, "Analysis and implementation of kidney stones detection by applying segmentation techniques on computerized tomography scans," *Italian Journal of Pure and Applied Mathematics*, vol. 43, 2020.
[24] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. N. Prasad, "Diabetic retinopathy detection and classification using capsule networks," *Complex and Intelligent Systems*, vol. 9, no. 3, 2023, doi: 10.1007/s40747-021-00318-9.

[25] W. K. Awad, E. T. Mahdi, and A. A. Nafea, "Accurate Rice Disease Detection Using Hybrid Convolutional Neural Networks and Transformer Models," *Passer Journal of Basic and Applied Sciences*, vol. 7, no. 1, pp. 336–346, 2025, doi: 10.24271/psr.2025.490265.1825.

[26] P. A. Abdalla, M. Y. Shakor, A. K. Ameen, B. S. Mahmood, and N. R. Hama, "Kidney stone detection via axial CT imaging: A dataset for AI and deep learning applications," *Data Brief*, vol. 59, p. 111446, 2025, doi: https://doi.org/10.1016/j.dib.2025.111446.

[27] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.