

مشعل المطيري

Deep Learning Kidney Tumors Journal v1.0.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:397199286

Submission Date

Jul 6, 2026, 4:40 PM GMT+3

Download Date

Jul 6, 2026, 9:12 PM GMT+3

File Name

Deep Learning Kidney Tumors Journal v1.0.docx

File Size

3.8 MB

27 Pages

7,450 Words

44,756 Characters

23% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.



22 AI-generated only 23%

Likely AI-generated text from a large-language model.



0 AI-generated text that was AI-paraphrased 0%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

 مجلة ميدأوشن للبحوث والدراسات	Midocean Journal for Research and Studies https:// ISSN: 3078-5669	
--	--	--

Deep Learning for Detection and Classification of Kidney Tumors and Cancer

Faisal Almokayed – Meshaal Almutairi

ORCID ID: [insert ORCID – obtain at <https://orcid.org/signin>]

Master of Science in Artificial Intelligence – College of Informatics – Midocean University –
Union of Comoros

Dr. Mohammad Essa (ORCID ID), College of Informatics, Informatics – Midocean University
– Union of Comoros

Mohamed Eassa

Abstract: **Intro** **One Paragraph**

Objectives: This study aims to design and evaluate a deep learning-based framework for the automatic detection and multi-class classification of kidney conditions — Normal, Cyst, Stone, and Tumor — from computed tomography (CT) images, incorporating an attention mechanism to improve both diagnostic accuracy and model interpretability.

Methodology: A multi-class classification pipeline built on 12,446 CT images spanning four categories. Images underwent grayscale-to-RGB conversion, CLAHE-based contrast enhancement, resizing, normalization, and stratified augmentation, then were split 70/10/20 into training, validation, and test sets. Three convolutional architectures — ResNet-50 enhanced with a Convolutional Block Attention Module (CBAM), EfficientNet-B4, and MobileNetV3 — were trained using a class-weighted focal loss ($\gamma = 2.0$) to address class imbalance, and were compared using accuracy, precision, recall, F1-score, and AUC-ROC.

Results: The proposed ResNet-50 + CBAM model achieved the best overall test accuracy of 99.88%, a macro F1-score of 0.9988, and a macro AUC-ROC of 1.0000, surpassing EfficientNet-B4 (98.39%) and MobileNetV3 (98.84%). Grad-CAM visualizations confirmed that the model consistently attended to clinically relevant renal regions rather than background artifacts.

Recommendations: Future validation on multi-center clinical datasets, extension to multi-phase CT protocols, integration with segmentation models, and deployment within clinical decision-support systems are recommended to strengthen real-world applicability.

Keywords: Deep Learning, Kidney Tumor Classification, Convolutional Neural Networks, CBAM Attention, Computed Tomography.

الملخص:

الأهداف: تهدف هذه الدراسة إلى تصميم وتقييم إطار عمل قائم على التعلم العميق للكشف التلقائي وتصنيف حالات الكلى المتعددة الفئات (طبيعي، كيس، حصوة، وورم) من صور الأشعة المقطعية المحوسبة (CT)، مع دمج آلية انتباه لتحسين دقة التشخيص وقابلية تفسير النموذج.

المنهجية: تم بناء مسار تصنيف متعدد الفئات باستخدام 12,446 صورة أشعة مقطعية موزعة على أربع فئات. خضعت الصور لتحويل من التدرج الرمادي إلى RGB، وتحسين التباين باستخدام CLAHE، و تغيير حجم الصورة، و توحيد قيم البيانات، والزيادة التعزيزية، ثم قسّمت بنسبة 20/10/70 إلى مجموعات تدريب وتحقق واختبار. تم تدريب ثلاث معماريات مختلفة: ResNet-50 المعزز بوحدة انتباه الكتلة الالتفافية (CBAM)، و EfficientNet-B4، و MobileNetV3، باستخدام دالة خسارة بؤرية موزونة حسب الفئة ($\gamma = 2.0$) لمعالجة عدم توازن الفئات.

النتائج: حقق نموذج ResNet-50 + CBAM المقترح أفضل دقة اختبار إجمالية بلغت 99.88%، ودرجة F1 كلية بلغت 0.9988، ومساحة AUC-ROC كلية بلغت 1.0000، متفوقاً على (EfficientNet-B4 98.39%) و (MobileNetV3 98.84%). وأكدت تصورات Grad-CAM أن النموذج ركز باستمرار على المناطق الكلوية ذات الصلة السريرية.

التوصيات: يوصى بالتحقق مستقبلاً من صحة النموذج على بيانات سريرية متعددة المراكز، وتوسيع النطاق العملي ليشمل بروتوكولات الأشعة المقطعية متعددة الأطوار، ودمجه مع نماذج الخاصة التجزئة، وتطبيقه ضمن أنظمة دعم القرار السريري لتعزيز قابلية تطبيقه في الواقع.

الكلمات المفتاحية: التعلم العميق، تصنيف أورام الكلى، الشبكات العصبية الالتفافية، آلية انتباه CBAM، الأشعة المقطعية المحوسبة.

1. Introduction:

Justify

A significant health concern in the world is kidney tumors and it encompasses a great number of benign lesions or very aggressive malignancies. Most of the renal malignancy is the renal cell carcinoma (RCC) the most common malignant renal tumor in adults that constitutes approximately 90% of all the malignant cases [1]. It has subtypes like the transitional cell carcinoma, Wilms tumor in children and benign ones which include the oncocytomas and the angiomyolipoma [2]. There has been obviously increase in the incidence of kidney tumors, which is impute the rising rates of obesity, smoking, and high blood pressure, as well as continuous improvements in diagnostic imaging techniques [3]. Also according to current epidemiological statistics there is approximately 434,840 new cases diagnosed of kidney cancer over the world in 2022, which is the 14th most frequently occur cancer in general; 10th in men and 13th in women [1]. Mortality rate is also high with about 155, 953 people dying of kidney cancer which is quite deadly unless it is prevented and treated at tender age [4].

Kidney tumors are not geographically and socioeconomically evenly distributed. The rate of the disease has been found to be higher in the income countries, particularly the North American and European countries because of availability of diagnostic tools and screening, and the age-standardized incidence rate (ASIR) is up to 15.60 per 100,000 as in the such countries as Argentina [1]. On other hand, the mortality rate is differ between low and middle income countries, the inequality due to late diagnosis and inefficient health facilities. Indicative it is estimated that by the year 2050, the cases and deaths will increase by 72 and 96 percent respectively; that is why creation of improved diagnostics methods is extremely needed [5]. The survival rate directly depends on the stage of diagnosis; the survival rates average 78.6% in the United States with an average of above 90 of localized tumors and a significantly lower rate of less than 15 with metastatic disease [6]. That's why early diagnosis definitely is improving the outcomes.

Computed Tomography (CT) scans have emerged as the core in the diagnosis of kidney tumors because it offers high resolution images that could be used in the accurate visualization of the tortoise features of the tumor with regard to size, shape, location and the pattern of enhancement [7]. Multi-detector CT protocols include non-contrast, arterial, venous and by doing this the case can be correctly staged by assessing vascular involvement, lymph node metastasis, and distant spread and the accuracies of detection have been reported as in excess of 99 % [8]. Manual determination of CT scans is however

time consuming and subject to variation particularly in the determination of minute differences between tumors subtypes.

The appearance of artificial intelligence (AI) and deep learning (DL) models there was a revolution in medical imaging profession through automatization tumor classification and detection. Convolutional neural networks (CNN) in the frameworks have been found to be more efficient in the analysis of CT scans of various types of cancers and kidney tumor is not an exception [9]. These models learn hierarchical features in large scale data sets, such that they can perform multiclass classification not only between benign and malignant tumors, but also between particular subtypes, which is the secret to individual approach in the treatment planning [10]. Multi-class classification is most significant in the situation of kidney tumors as it can be concluded to either apply active surveillance of the benign lesions or treatment of the malignancies which will lead to a decreased number of unnecessary procedures and better life of the patients.

The thesis is directed to building a monitored DL system that would identify and classify kidney tumors and cancer with assistance of CT scans. Its mission is to enhance the quality of diagnosis such as through decreasing inter-observer variability, time sparing, and as a result reducing the rate of mortality caused by kidney tumors by overcoming the deficiencies of manual diagnosis. Bringing DL into the clinical practice sphere may be regarded as a shift in paradigm to precision medicine, where the knowledge of radiologists will be complemented with the help of the AI-based tools, which will lead to more effective and fair healthcare delivery.

2. Research Problem:

Although there has been an improvement in imaging technologies, kidney tumors diagnosis and classification are still fraught with challenges. Manual detection is more than dependent on the professional skills of radiologists to interpret CT scans, which is a subjective process likely to commit errors. Among subtypes, the similarity of appearances of the tumors, blurry limits, and overlapping imaging characteristics can be among the most perplexing issues, resulting in wrong classification. Moreover, small renal masses (SRMs) are usually ambiguous to differentiate cystic, solid benign, and malignant maladies.

Incidental findings during routine imaging have risen all over the world leading to an increase in the workload of the radiologist, which has added to the fatigue based mistakes among radiologists. The inter-observer variability experienced especially in multi-phase CT

interpretations whereby the enhancement patterns should be carefully evaluated to determine the presence of vascular invasion or metastasis. The scarce access to skilled radiologists in low-resource environments also cause a further delays of diagnoses, which leading to increase the death rates in that areas which has limitations in such of skilled professions.

Multi class classification further complicates the picture since proper identification of subtype is essential to diagnose and treatment. The most aggressive subtype needs immediate treatment, whereas the indolent subtypes might not need treatment and could just be monitored. This granularity is a challenge of manual techniques that may require biopsy which has dangers such as bleeding or seeding.

Above issues, necessitate the use of automated systems. Where Machine learning methods involve manual features extraction where restricting systems to a wider use of CT datasets. Deep learning provides better approach where its able to automatically learn all features, but the current systems do not have the robustness over multi-center data or is not able to effectively combine multi-phase information. This thesis fills the gaps with the suggestion deep learning structure that will improve the sensitivity of the detection and the accuracy, reduce the weaknesses of manual diagnosis and open the way to AI-integrated clinical practice.

3. Importance of the Study:

Scientifically, this study contributes an explainable deep learning framework that combines residual learning with an attention mechanism (CBAM) to achieve high-accuracy, multi-class classification of kidney conditions from CT imaging, while directly addressing class imbalance through a class weighted focal loss. It also provides a systematic, benchmarked comparison of three convolutional architectures ResNet-50, EfficientNet-B4, and MobileNetV3 on the same dataset and evaluation protocol, offering the research community a clear reference point for architecture selection in renal CT classification tasks.

Practically, the proposed framework has the potential to reduce radiologist workload, shorten diagnostic turnaround time, and lower the rate of unnecessary interventions by supporting better differentiation between benign and malignant findings. The use of Grad-CAM-based explainability enhances clinical trust as it allows radiologists to visually verify that the model's decisions are grounded in clinically relevant renal regions rather than incidental image

artifacts. This is an essential property for eventual integration into real world clinical decision support systems.

4. Study Objectives:

This thesis overall aim to develop and evaluate a deep learning system that can identify and classify kidney tumors on CT scans images automatically, This model will improve the diagnostic efficiency, accuracy of diagnosing, early intervention, and customized treatment.

The objectives are:

1. To discuss existing deep learning (DL) models to the medical imaging domain generally and paying particular attention to models used for kidney tumor identification and classification, also determining the advantages and shortcomings of the existing methodologies.
2. To develop deep learning architect that combine the data of multi phases CT to extract more learning features and identify tumor locations.
3. To develop multi-class classification functionality, to be able to differentiate between benign (e.g., cyst, stone) from malignant tumor, underline the clinical significance of subtype specific diagnosis.
4. To assess performance of proposed framework in comparison with other comparative architectures based on accuracy, sensitivity and area under the ROC curve (AUC).
5. To determine framework clinical applicability, by validating framework on actual radiology workflow, and explore ways to integrate it with current radiology workflow to minimize diagnoses time and errors.

5. Review of Previous Studies:

The chapter provides a review of fifteen recent studies, most of which provided their own methodology to the same KIDNEY DATASET on Kaggle: normal-cyst-tumor and stone. We can point the strengths and weaknesses of methods used, by comparing how various research papers used the same data to classify multi-class.

Maçin et al. (2025) proposed KidneyNeXt, a custom lightweight CNN model that uses depthwise separable convolutions on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone) with a standardized input resolution of 224x224 pixels. Depthwise separable convolutions break standard convolution down into a depthwise spatial convolution (filter) and a pointwise channel mixing step, greatly reducing the number of parameters and floating-point operations at no cost to feature extracting performance. KidneyNeXt reached an accuracy of 99.96, corresponding to much heavier architectures at a fraction of the computational cost, and thus matched it especially well to run on edge devices or in low resource clinical environments. The authors proved that the architecture was very high in both sensitivity and specificity in all the four classes (including the diagnostically challenging stone class) which been characterized by small hyperattenuating lesions that can be easily confused in calcifications in the surrounding tissue [9].

Islam et al. (2022) delved into the notion of transfer learning on the basis of VGG-16 and ResNet-50 models that are trained on ImageNet and frozen on the first layers and fine-tuned on the final classification layers on the Kaggle CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone). The approach quickly converged on learned features and ResNet-50 skip connections of about 96% accuracy per classification, though ImageNet training on colour images was non-optimal on grayscale CT texts and the number of parameters was excessive to be used in clinical practice. Vision transformer architectures such as EANet, CCT and Swin Transformer were also tested in the same research with the Swin Transformer scoring the highest score of 99.30% accuracy showing the new strength of attention-based models over the traditional CNN in medical image classification. However, the fact that transformer architectures have high computational and training requirements is still a practical obstacle to implementation in resource-limited clinical setting [11].

Alzu'Bi et al. (2022) introduced two-stage kidney tumour detection and classification framework tested on a new institutional dataset, which was gathered at King Abdullah University Hospital (KAUH) in Jordan, consisting of 8,400 CT scan images of 120 suspected renal masses patients. The data were categorized to 80/20 train and test sets. During the first stage (tumour detection), the authors compared three 2D-CNN architectures with a custom six-layers CNN (CNN-6), ResNet50, and VGG16. During the second stage (tumour type classification), a custom four-layered CNN (CNN-4) was used. CNN-6 and ResNet50 models provided accuracy rates of 97% and 96% respectively, and VGG16 had an accuracy rate of 60%, whereas CNN-4 classification model was able to provide an

accuracy of 92 percent. The VGG16 detection rate of 60% and the relatively low classification rate of 92% demonstrated the difficulty of network training on comparatively small, single institution datasets, and the need of the volume and diversity of data to obtain clinically acceptable performance [12].

Yildirim et al. (2021) suggested a kidney stone detection automated binary deep learning system of an in-house collection of cross-sectional computer tomography (CT) dataset established at Elazig Fethi Sekin City Hospital, Elazig, Turkey, of 1,799 cross-sectional CT images of 433 participants. The data was categorized into two (stone available / stone unavailable) and 80/20 training and testing, respectively, and produced 1,453 and 346 training and test images, respectively. To overcome the overfitting of the small clinical dataset, the authors adopted a modified XResNet-50 model that was augmented with data-enhancing techniques. Standard optimization procedures were used to train the model and it was tested on the held-out test set. The deep learning model that is proposed achieved an accuracy of 96.82 and a sensitivity of 95.76 on the 346 test cases, which indicates the detection of even the low size kidney stones. Nevertheless, the researchers only studied binary classification and not the differentiation of cysts, tumors, and normal tissue, which restricted the scope of the study to stone detection only [13].

Pimpalkar et al. (2025) performed a comparative study of four transfer learning architectures VGG16, ResNet50, CNNAlexNet, and InceptionV3 with hyperparameter optimisation being applied to the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone). In the study Watershed segmentation was used to delineate the kidney regions and binarisation thresholding, based on Otsu, was used to improve the clarity of the image before inputting it into the model. Learning rate, batch size, and dropout parameters were tuned to each architecture separately by a systematic grid search. InceptionV3 was reported to be the most accurate at 99.96% and VGG16 took the lead at 100% in the validation set, implying that when paired with powerful preprocessing, it is practically separable on the benchmark dataset. The assessment was however based only on the Kaggle benchmark and the analysis did not discuss behaviour of models in distributional shift, class imbalance in clinical population, or real-time inference latency which are important in clinical application [14].

Sabuncu et al. (2023) In their study, suggested a deep learning classification architecture to identify kidney stones in axial CTs taken in Near East University Hospital, Cyprus, based on an original set of true clinical CTs taken on patients during their regular

urological examination. Preprocessing methods used in the pipeline, such as Watershed segmentation and Otsu binarization threshold used to improve kidney delineation and isolate stone areas from the surrounding tissue before classification. InceptionV3 architecture has been used as the central model in classification whereby it utilizes its multi-scale convolutional filters to identify heterogeneous stone textures in different images at different levels of resolution. The accuracy of the test results of the model was 98.52, and the authors highlight that it was able to detect small-sized kidney stones that were clinically significant. These limitations were a small authentic clinical dataset, which was not publicly disclosed, binary range of classification (stone present / absent), and no benchmarking to other renal pathology datasets [15].

Obaid et al. (2025) on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone), suggested multi-model fusion with DarkNet53 Four-class kidney CT classification. DarkNet53, which was initially designed as a core of the YOLOv3 object detection architecture, was used as a feature extractor and merged with predictions of related models through the ensemble fusion mechanism that aims to tie up complementary discriminative patterns that may not be discovered by individual models. The architecture of fusion incorporated the use of predictions at the decision level to generate a consensus classification result. The proposed system was very accurate (99.69), so the ensemble-based methods are proved to be better than the single-model classifiers in the medical image classification issues. The authors admitted that the computational cost of executing several parallel networks at the same time adds to the inference time that can limit its use in real-time diagnostic processes with turnaround times [16].

Baygin et al. (2022) proposed ExDark19, which is a new method of deep feature generation that can be used to detect kidney stones automatically on the same dataset of public coronal CTs established at Elazig Fethi Sekin City Hospital, Elazig, Turkey (1,799 images of 433 participants) utilized by Yildirim et al. (2021). The authors did not use end-to-end deep learning but instead used DarkNet19 as a pre-trained backbone to obtain exemplar feature representations and applied Iterative Neighbourhood Component Analysis (INCA) to select the most discriminative feature vectors, and then used them to obtain classifications via a k-nearest neighbour (KNN) classifier with ten-fold cross-validation and hold-out validation to classify. The proposed ExDark19 system was found to be 99.22% accurate when using ten-fold cross-validation and 99.71% accurate when using the hold-out method when compared to the XResNet-50 baseline, which only achieved 96.82 on the same dataset. Nevertheless, like any research based on the use of the given dataset, the

area of classification was limited to binary (stone / no stone) and did not reach the level of multi-class differentiation of renal pathology [17].

Qian et al. (2024) used a densely connected convolutional network (DenseNet121) with feature reuse in all the previous layers to classify four classes of kidney CT on CT KIDNEY DATASET (12446 images: Normal, Cyst, Tumor and Stone). DenseNet121 ties all of the layers to all of the other layers in a feed-forward style, allowing gradient flow across the network and allaying the vanishing gradient problem that afflicts deeper networks. Although these models have these theoretical benefits, their accuracy of 93.51% was significantly lower than more recent works on the same dataset, which implies that basic methods of DenseNet121 can be too basic to differentiate between visually similar kidney pathologies, including cysts and tumours. The outcome highlights that model depth and model connectivity alone cannot ensure high performance and task-specific adaptation (augmentation tuning, hyperparameter optimisation, attention mechanisms, etc.) is needed [18].

El-Hag et al. (2025) The issue of reflections and image artefact in CT scans was solved by introduced a specular-free preprocessing pipeline before deep learning classification on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone). Preprocessing stage used reflection removal and contrast normalisation to decrease the noise and enhance uniformity of features across the dataset, and the classification with EfficientNet-B2 a model with compound-scaled network that reduces network width, depth, and resolution. EfficientNet-B2 was chosen because it is one that has good trade-off both in accuracy and computational cost among the family of EfficientNets. The preprocessing-classification pipeline obtained an accuracy of 98.27, a score that is lower than some of the competing methods using the same benchmark, but it showed that artefact suppression methods could be used to gain quality learned representations. The authors observed that the preprocessing pipeline added more steps of calculations and the effect of the approach on real, noisy clinical CT data of heterogeneous scanner types needed more consideration [19].

Almuayqil et al. (2024) presented an idea of KidneyNet a new eight-layer CNN architecture that is directly trained to perform automated four-class kidney CT classification on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone). KidneyNet was built according to a compact sequential architecture comprising of convolutional, batch normalisation, pooling, dropout and fully connected layers, in a way that

it optimised the model complexity versus accuracy of its classification. Its architecture did not require the large number of parameters that large transfer learning models have, which also makes it more practical to implement in the resource-constrained healthcare setting. The accuracy and precision of KidneyNet were 99.88 and 99.92, respectively, which is among the best benchmark results at the time of publication. The authors also pointed out that the model was trained and evaluated only on the Kaggle data and the extent to which it works in a real clinical setting with heterogeneous scanner protocols and patient demographics was not tested [20].

Bhandari et al. (2023) came up with a lightweight, custom CNN model that classifies kidney CTs (four classes: Normal, Cyst, Tumor and Stone) on CT KIDNEY DATASET (12,446 images). The model was created as a fresh start, as opposed to other heavier methods of transfer learning, to reduce computational expenses without compromising diagnostic performance. Another significant characteristic of the study was the employment of explainability methods, namely, LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations), to give the radiologist visualizations of the decision areas of the model on CT images that can be understood. Although the lightweight CNN had an accuracy of 99.52, this result shows that custom small architectures can compete with pre-trained models on large benchmark datasets. The authors admitted, however, that the explainability methods incurred inference overhead and the model was not tested on external clinical datasets, so the performance of generalization had not been tested [21].

Bingol et al. (2022) on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone), trained a 29-layer deep CNN architecture with Relief-based feature selection to classify kidney CT images in four classes. The hybrid model proposed obtained feature maps using two intermediate convolutional layers (convolution-6 and convolution-7) and concatenated the feature maps before subjecting them through the Relief feature selection algorithm to keep the most informative features and finally feeding them into a wide neural network classifier. Compared to eight typical standard pre-trained architectures, it was indicated that the highest performance of the pre-trained model was 87.75% and that the proposed hybrid model performed at 99.37% accuracy indicating that there was a significant improvement in performance by the Relief-guided feature fusion strategy. The paper has indicated that off-the-shelf transfer learning designs that do not involve feature selection based on the domain can perform much worse on medical image tasks, which supports the usefulness of feature optimisation pipelines [22].

Khan et al. (2025) suggested a ConvNeXt and EfficientNetV2B0 dual-architecture transfer learning strategy on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone) to classify 4 classes in kidney CT. One of the most important methodological contributions was the adoption of selective data trimming to achieve inter-class imbalance before training, eliminating redundant samples of the over-represented classes to ensure that the model does not develop a bias towards the higher representation classes, like the Normal one. ConvNeXt, a convolutional network that has been modernised using methods influenced by Vision Transformers, was combined with EfficientNetV2B0 at the feature-level in an ensemble to combine the complementary strengths of the two architectures. The ensemble was found to be accurate at 99.73 with just 6.2 million parameters, proving that well-thought-out data preprocessing and well-informed choice of architecture can provide state-of-the-art results without using architecturally complicated and computationally intense models. The authors indicated that the study had some limitations such as the use of only one benchmark dataset and the necessity to further verify it on multi centers CT cohorts [23].

Murillobouzon et al. (2025) did not concentrate on classification but rather on semantic segmentation by presenting a new purpose-specific dataset named KSSD2025 (Kidney Stone Segmentation Dataset 2025) that consists of 838 annotated abdominal CT scans with pixel-level binary segmentation labels showing a bi-polar location of kidney stones and normal tissue. All models achieved mean Disc scores above 95% . KSSD2025 also includes whole-abdomen CT slices in their natural clinical format, annotated by radiologists, whereas the CT KIDNEY DATASET only contains cropped and pre-isolated kidney images, which does not ensure quality ground truth. The U-Net segmentation architecture was used as the main model as it utilizes an encoder-decoder form and skip connection that can locate and define stone edges at the pixel level. The analysis covered the clinically significant issue of stone localization and burden quantification instead of class-level diagnosis that offers spatial data that supplements classification methods. On the downside, it had a relatively small size of dataset (838 images), used binary segmentation masks, as they do not differentiate between stone subtypes, and a single-center institutional image corpus [24].

Table (1): Summary of fifteen reviewed studies on deep learning for kidney CT classification.
[Explain](#)

Ref	Author (Year)	Dataset	Technique	Accuracy
[9]	Maçin et al. (2025)	CT KIDNEY DATASET	KidneyNeXt (depthwise separable CNN)	99.96%
[11]	Islam et al. (2022)	CT KIDNEY DATASET	VGG-16 / ResNet-50 / Swin Transformer	96%–99.30%
[12]	Alzu'Bi et al. (2022)	KAUH Jordan (8,400 CT)	CNN-6 / ResNet50 / VGG16 + CNN-4	92%–97%
[13]	Yildirim et al. (2021)	Private NCCT, Turkey (1,799 CT)	XResNet-50 + augmentation	96.82%
[14]	Pimpalkar et al. (2025)	CT KIDNEY DATASET	VGG16 / ResNet50 / InceptionV3	99.96%
[15]	Sabuncu et al. (2023)	Near East Univ. Hospital, Cyprus	InceptionV3	98.52%
[16]	Obaid et al. (2025)	CT KIDNEY DATASET	DarkNet53 + ensemble fusion	99.69%
[17]	Baygin et al. (2022)	Private NCCT, Turkey	DarkNet19 + INCA + KNN	99.22%–99.71%
[18]	Qian et al. (2024)	CT KIDNEY DATASET	DenseNet121 (no augmentation)	93.51%
[19]	El-Hag et al. (2025)	CT KIDNEY DATASET	EfficientNet-B2 + specular-free preproc.	98.27%
[20]	Almuayqil et al. (2024)	CT KIDNEY DATASET	KidneyNet (8-layer custom CNN)	99.88%/99.92%
[21]	Bhandari et al. (2023)	CT KIDNEY DATASET	Lightweight CNN + LIME/SHAP	99.52%
[22]	Bingol et al. (2023)	CT KIDNEY DATASET	29-layer CNN + Relief selection	99.37%
[23]	Khan et al. (2025)	CT KIDNEY DATASET	ConvNeXt + EfficientNetV2B0	99.73%
[24]	Murillobouzon et al. (2025)	KSSD2025 (838 annotated CT)	U-Net pixel-level segmentation	97.06% (Dice)

6. Methodology:

The proposed architecture is composed of three main components: a convolutional backbone for feature extraction, a Convolutional Block Attention Module (CBAM), and a custom classification head. Three backbones were evaluated ResNet-50 (proposed primary model), EfficientNet-B4, and MobileNetV3 under an identical preprocessing, augmentation, and classification head configuration to ensure a fair comparison.

Deep Learning Framework for Kidney CT Image Classification

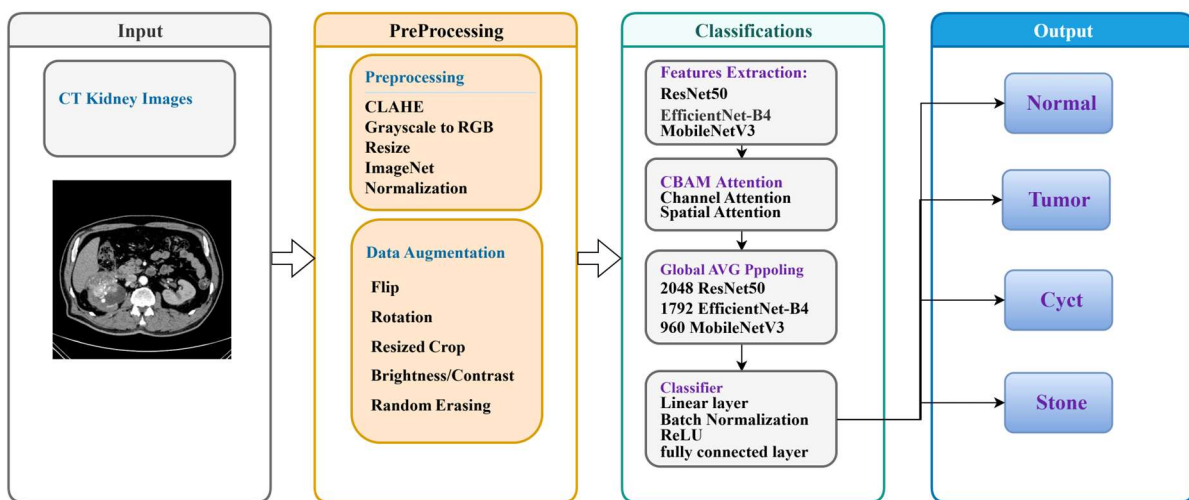


Figure 2: Proposed kidney CT image classification framework.

6.1 Dataset. CT images were obtained from the Kaggle CT KIDNEY DATASET (Normal–Cyst–Tumor–Stone), comprising 12,446 images across four classes (Table 2). The dataset is imbalanced, with the Normal class comprising 40.8% of images (5,077 samples) versus 11.1% for Stone (1,377 samples), a ratio of approximately 3.7:1 — a pattern typical of clinical radiology data. Images include both axial and coronal reconstructions and both contrast–enhanced and non–contrast studies, and are provided in JPEG format at differing native resolutions.

Table (2): Class distribution of the dataset.

Class	Number of Images	Percentage	Clinical Description
Normal	5,077	40.8%	Healthy kidney parenchyma; no lesions or calcifications
Cyst	3,709	29.8%	Fluid-filled renal cysts of varying size
Tumor	2,283	18.3%	Solid renal masses including RCC subtypes
Stone	1,377	11.1%	Calcified hyperattenuating lesions; smallest class

Class	Number of Images	Percentage	Clinical Description
Total	12,446	100%	Mixed axial and coronal CT acquisitions

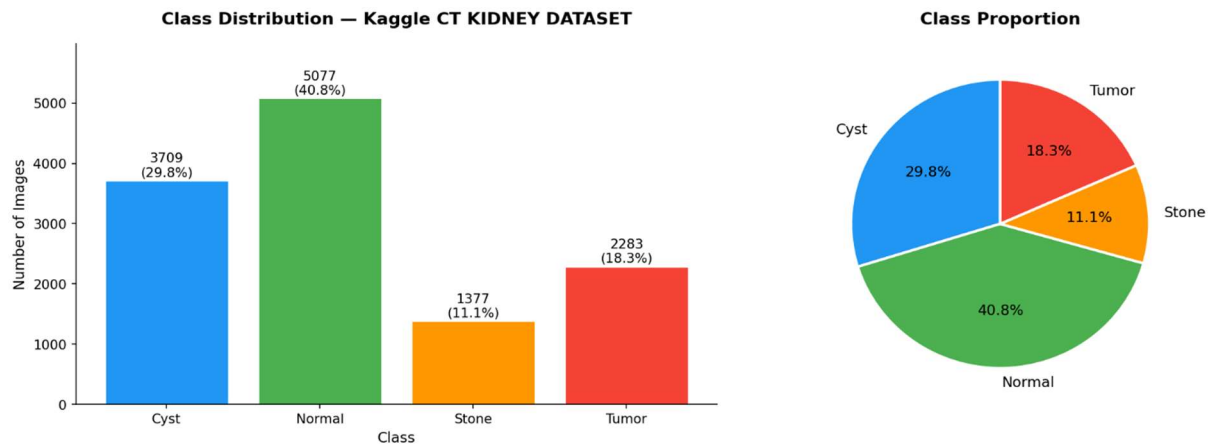


Figure 3: Per-class distribution of the dataset.

6.2 Data Preprocessing. All images passed through a four stage pipeline: (1) grayscale to RGB conversion via channel duplication to support pretrained backbones; (2) CLAHE contrast enhancement (clip limit = 2.0, tile grid 8×8) to improve visibility of small lesions; (3) resizing to 380×380 px (EfficientNet-B4) or 224×224 px (ResNet-50/MobileNetV3) using bilinear interpolation; and (4) normalization using ImageNet mean and standard-deviation values.

6.3 Data Augmentation. Augmentation applied only to the training set to improve generalization and reduce overfitting; validation and test sets received only the preprocessing steps above. Techniques included random horizontal flip ($p = 0.5$), random rotation ($\pm 15^\circ$), brightness/contrast jitter (factor 0.2), random resized crop (scale 0.85–1.0), and random erasing ($p = 0.1$, scale 0.02–0.1).

6.4 Train, Validation, and Test Split. The dataset was divided using a stratified 70/10/20 split into training (8,712 images), validation (1,244 images), and test (2,490 images) subsets, preserving class proportions across all three partitions.

6.5 Model Architecture. ResNet-50 is a 50-layer residual network composed of bottleneck blocks (1×1, 3×3, 1×1 convolutions) that uses skip connections to mitigate vanishing gradients in very deep networks. It was pretrained on ImageNet-1K and fine-tuned at an input resolution of 224×224, with early layers (up to stage 3) frozen and later layers (stage 4 and the classification head) fine-tuned; owing to its depth, accuracy, and training stability, ResNet-50 selected as the primary proposed backbone.

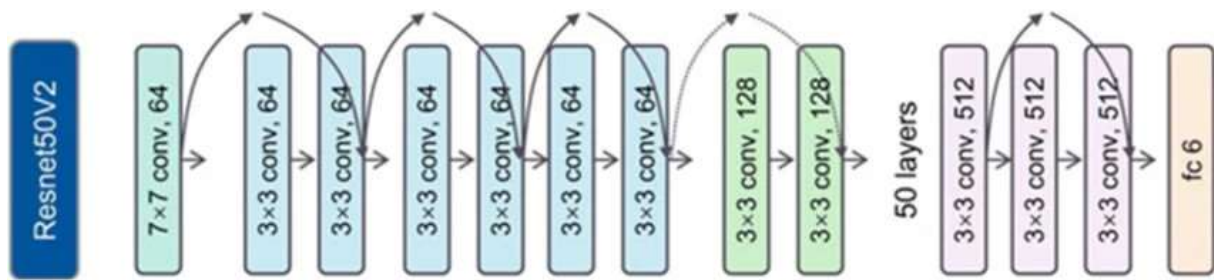


Figure 4: ResNet-50 architecture.

EfficientNet-B4 uses a compound scaling strategy that jointly scales network depth, width, and input resolution from a single coefficient. Pretrained on ImageNet-1K at 380×380 resolution, its early MBConv blocks (stages 0–6) kept frozen while later blocks (stages 7–8) fine-tuned on the kidney CT domain. MobileNetV3 is a lightweight architecture designed for mobile and embedded deployment, using hard-swish activations, squeeze-and-excitation blocks, and depthwise separable convolutions to minimize computational cost; it was pretrained on ImageNet and fine-tuned at 224×224 resolution, offering fast inference at the cost of reduced sensitivity on the minority Stone class relative to ResNet-50 and EfficientNet-B4.

6.6 Convolutional Block Attention Module (CBAM). A CBAM module inserted between the backbone's final convolutional output and global average pooling. It sequentially applies channel attention — via averaged and max-pooled descriptors passed through a shared two-layer multilayer perceptron with reduction ratio $r = 16$ — followed by spatial attention, computed via a 7×7 convolution over channel-pooled descriptors. The resulting refined feature map emphasizes diagnostically salient image regions while adding less than 2% additional parameters to the network.

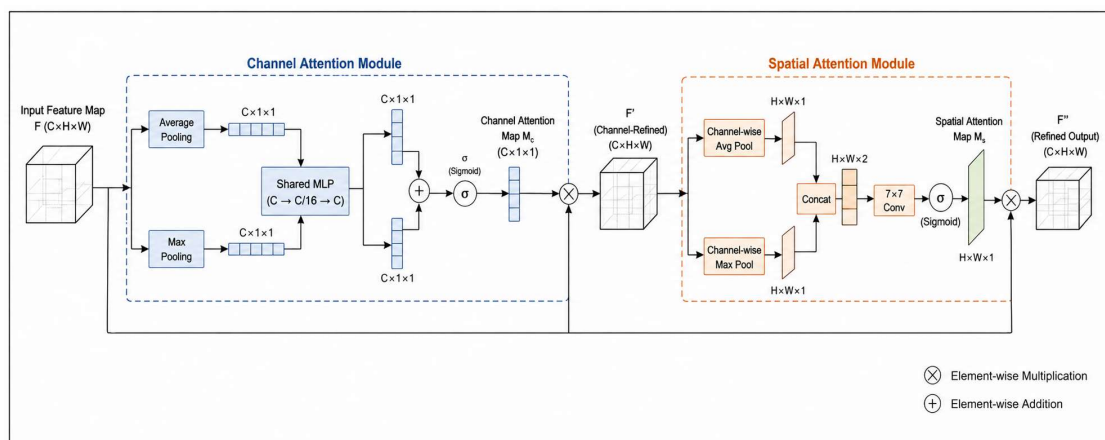


Figure 5: CBAM attention refinement module.

6.7 Classification Head. The refined, globally pooled feature vector (2048-D for ResNet-50) is passed through a dropout layer (rate 0.50), a fully connected layer reducing dimensionality to 512, batch normalization, a ReLU activation, a second dropout layer (rate 0.30), and a final four-class output layer (Normal, Cyst, Stone, Tumor).

6.8 Training Protocol. Training proceeded in two phases: Phase 1 (5 epochs) trained only the CBAM module and classification head with the backbone frozen; Phase 2 (up to 8 epochs, following a 5-epoch linear warm-up) unfroze the final residual stage together with CBAM and the classification head for joint fine-tuning at a reduced learning rate. The AdamW optimizer (weight decay 1×10^{-5}) was used with a cosine-annealing-with-warm-restarts schedule, and early stopping (patience = 8 epochs) was applied based on validation loss. Class imbalance was addressed through inverse-frequency class weighting and a class-weighted focal loss ($\gamma = 2.0$), which down-weights well-classified examples and focuses training on harder, minority-class cases.

Table (3): Training hyperparameters.

Hyperparameter	Value
Input size (EfficientNet-B4)	380 × 380 px
Input size (ResNet-50 / MobileNetV3)	224 × 224 px
Batch size	32
Initial learning rate	1×10^{-4}
Optimizer	AdamW
Weight decay (L2)	1×10^{-5}
LR scheduler	Cosine Annealing with Warm Restarts
Phase 1 / Phase 2 epochs	5 / up to 8
Dropout (head layers 1, 2)	0.50, 0.30
Focal loss gamma (γ)	2.0
Class-weight strategy	Inverse frequency

6.9 Evaluation Metrics. Model performance on the held-out test set (2,490 images) was assessed using overall accuracy, per-class precision, recall, macro F1-score, weighted F1-score, Cohen's Kappa, one-vs-rest AUC-ROC, and confusion-matrix analysis, selected specifically to reveal performance on minority classes that overall accuracy alone can obscure.

Overall Accuracy: provides a simplified global view on how model classify correct results across all classes. Accuracy alone in imbalanced medical datasets can be misleading; because it may hide poor performance on minority classes. For multi class classification, following equation used:

$$Ovarall Accuracy = \frac{\sum_{i=1}^K Correct Predictions(i)}{\sum_{i=1}^K Total Sample(i)}$$

Precision: measures how the model reliable in positive classification, while per-class precision measures positive classification for each class. High precision means the model rarely mislabels provided images, it calculated via the following equation:

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity): measures how the model detects each condition. High recall is considered as FN (false negative) can be far more harmful than a false alarm.

$$Recall = \frac{TP}{TP + FN}$$

Macro F1 Score: treats all classes equally, regardless of how many samples each class has. Macro F1 Score useful when evaluating imbalanced datasets to ensuring minority classes received equal attention:

$$Macro F1 = \frac{1}{K} \sum_{i=1}^K F1(i)$$

Where each class F1 is:

$$F1(i) = 2 \times \frac{Precision(i) \times Recall(i)}{Precision(i) + Recall(i)}$$

Weighted F1 Score: measured based on how many samples each class has it adjusts contribution of that certain class. This will prevents small classes from dominating the evaluation and give more realistic results of overall performance.

$$Weighted F1 = \sum_{i=1}^K w(i). F1(i)$$

Where:

$$w(i) = \frac{Support(i)}{\sum_{i=1}^K Support(i)}$$

7. Results:

7.1 Preprocessing Results. Qualitative inspection confirmed that the CLAHE-based preprocessing pipeline improved the visibility of small lesions and stone margins prior to model training.

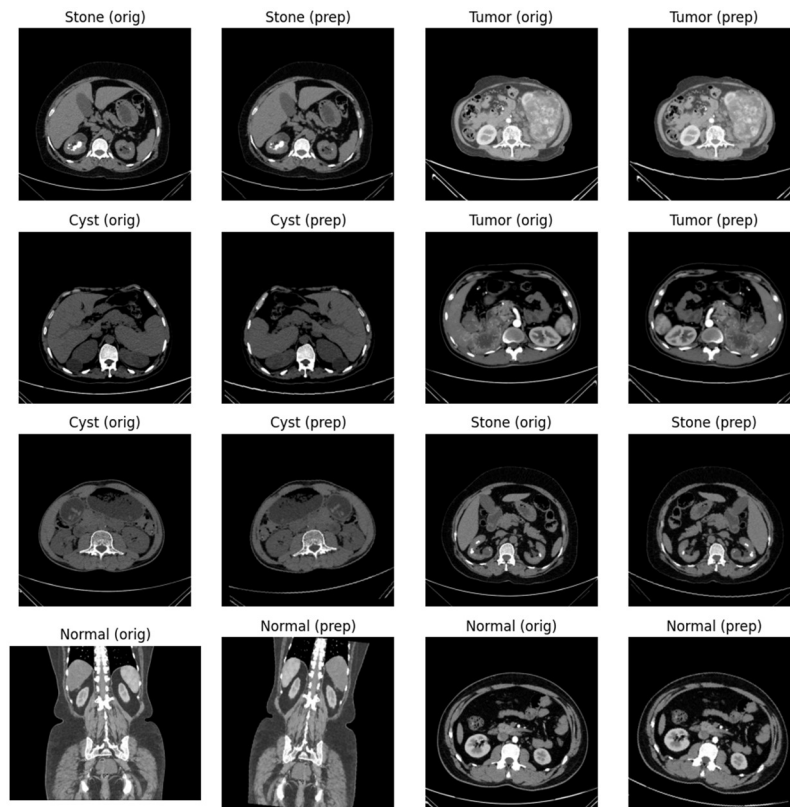


Figure 6: Sample images before and after preprocessing.

7.2 Classification Performance. The proposed ResNet-50 + CBAM model, trained with class-weighted focal loss ($\gamma = 2.0$), achieved an overall test accuracy of 99.88%. Table 4 and Figure 7 present per-class precision, recall, and F1-score. The Tumor class achieved perfect performance (precision = 1.0000, recall = 1.0000, F1 = 1.0000); the Normal class reached a precision of 0.9971 and recall of 1.0000; the Cyst class achieved a precision of 1.0000; and the Stone class achieved a precision of 1.0000 with a recall of 0.9964. All class-level F1-scores exceeded 0.998, indicating a strong balance between sensitivity and specificity across the imbalanced test distribution.

Table (4): Per-class classification metrics for ResNet-50 + CBAM.

Technique?

Class	Precision	Recall	F1-score	Support
Cyst	1.0000	0.9973	0.9987	742
Normal	0.9971	1.0000	0.9985	1016
Stone	1.0000	0.9964	0.9982	275
Tumor	1.0000	1.0000	1.0000	457
Accuracy	—	—	0.9988	2490
Macro Avg	0.9993	0.9984	0.9988	2490
Weighted Avg	0.9988	0.9988	0.9988	2490

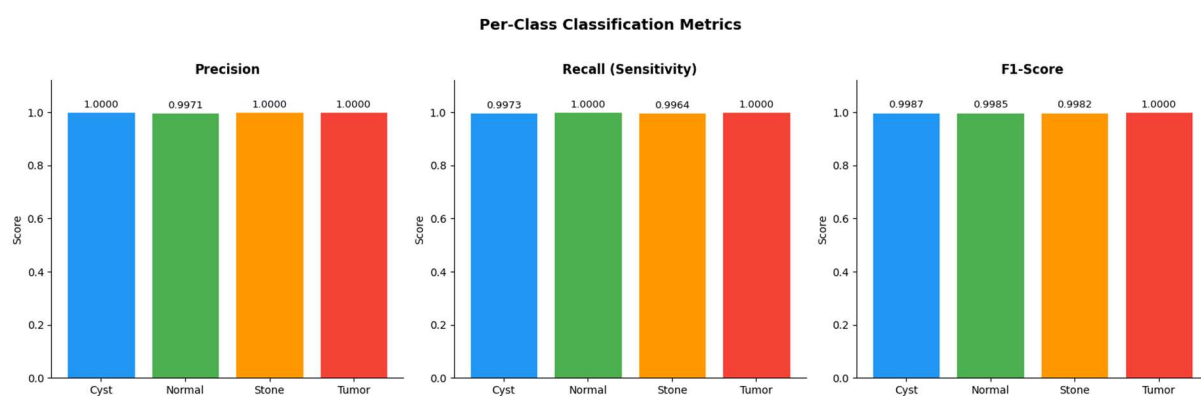


Figure 7: Per-class classification metrics.

The confusion matrix (Figure 8) showed no misclassification within the Normal or Tumor classes, two Cyst samples misclassified as Normal, and one Stone sample misclassified as Normal — confirming minimal confusion, primarily among visually or texturally similar benign findings.

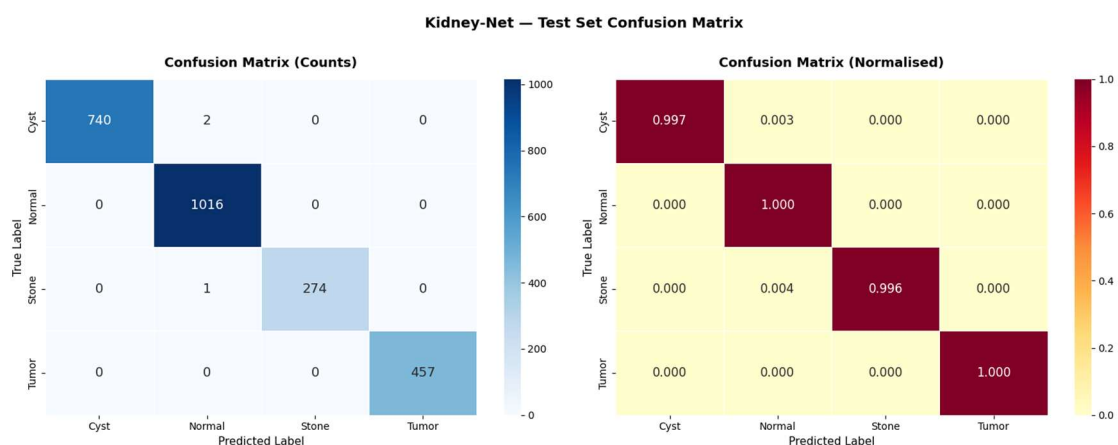


Figure 8: Confusion matrix for the ResNet-50 + CBAM model.

Training curves (Figure 9) showed steadily decreasing focal loss and increasing accuracy across both training and validation sets with minimal divergence, indicating efficient learning without overfitting.

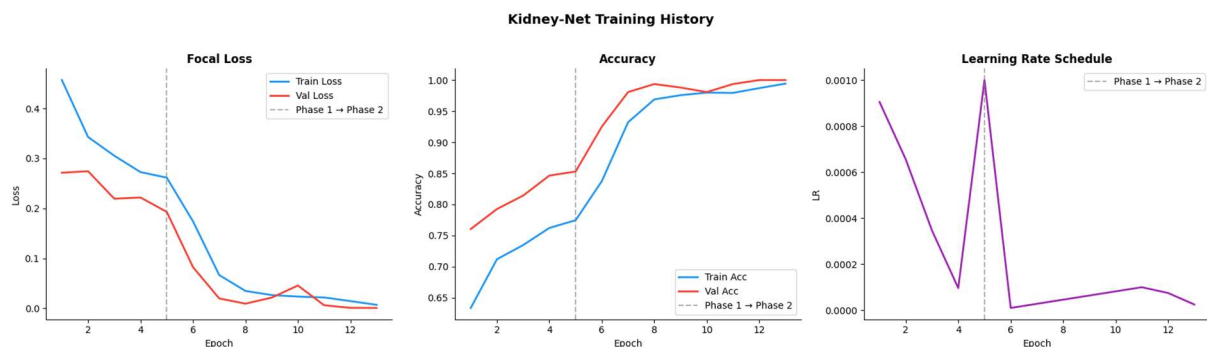


Figure 9: Training history for the ResNet-50 + CBAM model.

Per-class one-vs-rest ROC analysis (Figure 10) yielded near-perfect AUC values (≈ 1.0000) for all four classes, including Cyst and Stone, which are typically the most radiographically similar and difficult to separate.

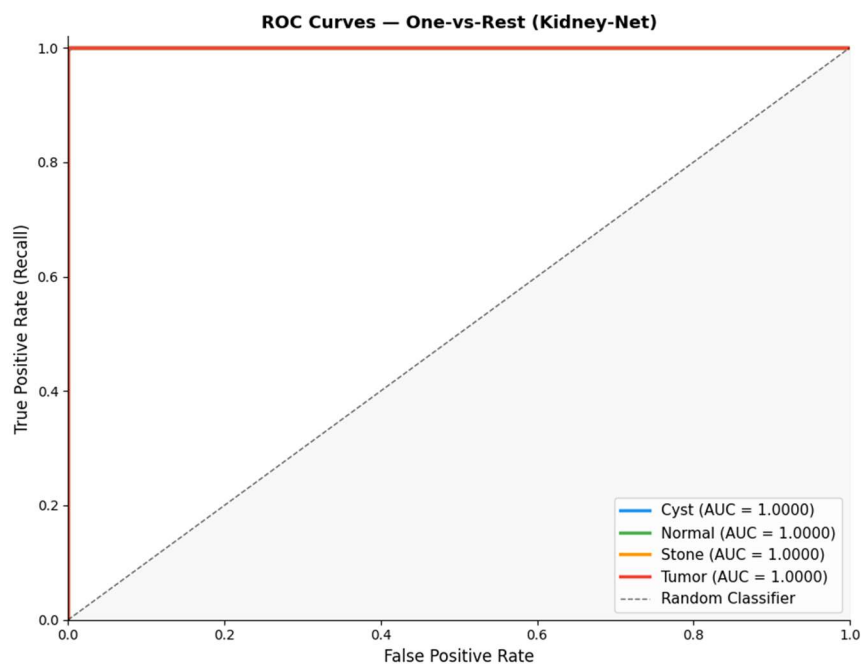


Figure 10: Per-class ROC curves for the ResNet-50 + CBAM model.

7.3 Comparison Between Models. Table 5 compares the three evaluated architectures. ResNet-50 + CBAM achieved the highest accuracy (99.88%) and macro F1-score (0.9988), outperforming EfficientNet-B4 (98.39% accuracy, 0.9794 macro F1) and MobileNetV3 (98.84% accuracy, 0.9868 macro F1), despite EfficientNet-B4's higher input resolution (380×380 vs. 224×224). MobileNetV3 offered the lowest computational cost (~5.4M

parameters) and fastest inference, making it best suited to resource-constrained deployment, but at a cost to fine-grained discrimination, particularly for the minority Stone class.

result of technique individual and where is the comparison with related work

Table (5): Comparison of deep learning models for kidney CT classification.

Metric	EfficientNet-B4 + CBAM	MobileNetV3 + CBAM	ResNet-50 + CBAM
Accuracy	98.39%	98.84%	99.88%
Macro F1	0.9794	0.9868	0.9988
Best Val Loss	0.0318	0.0075	0.0006
Macro AUC-ROC	0.9997	0.9998	1.0000

7.4 Explainability (Grad-CAM). Grad-CAM visualizations (Figure 11) showed that activation was consistently concentrated around renal lesion borders rather than irrelevant background structures across all four classes, supporting the interpretability and clinical trustworthiness of the proposed model's predictions.

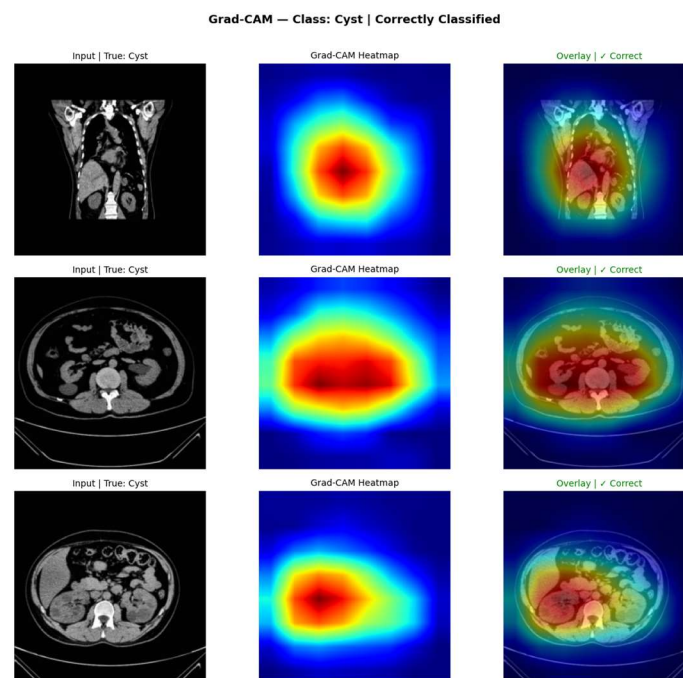


Figure 11: Grad-CAM visualization example (Cyst class).

1. Conclusion and Discussion of Results

The experimental results demonstrate that the proposed ResNet-50 + CBAM framework consistently outperforms both EfficientNet-B4 and MobileNetV3 across all evaluation metrics, achieving 99.88% overall accuracy, a macro F1-score of 0.9988, and a macro AUC-ROC of 1.0000. These results are consistent with — and in several respects exceed — accuracy levels reported in comparable studies on the same benchmark dataset (typically 98–99.96%, see Table 1), while offering a more stable and interpretable architecture through its integrated CBAM attention mechanism. The near-perfect separability observed in the confusion matrix and ROC analysis, particularly for the Cyst and Stone classes that are most prone to visual confusion, suggests that channel-and-spatial attention refinement meaningfully improves discrimination of subtle textural differences relevant to renal pathology. Grad-CAM-based explainability further supports the clinical credibility of the framework by confirming that its predictions are grounded in anatomically relevant regions rather than incidental artifacts.

The principal limitation of this study is its reliance on a single public benchmark dataset, which may not fully capture the variability in scanner types, imaging protocols, and patient demographics encountered in real clinical settings. This underscores the need for multi-center validation before clinical deployment can be considered. Overall, the findings confirm that combining residual learning with attention-based refinement yields a kidney CT classification framework that is simultaneously accurate, stable, and explainable, offering a practical foundation for future integration into radiology workflows.

2. Recommendations

Building on these findings, the following directions recommended for future work:

1. Validate the proposed framework on multi-center clinical datasets to confirm generalizability across scanners and patient populations.
2. Extend the framework to multi-phase CT protocols (arterial, venous, and delayed phases) to capture richer diagnostic information.
3. Integrate segmentation models (e.g., U-Net) to enable precise lesion localization alongside classification.
4. Pilot the framework within an operational clinical decision-support system (CDSS).
5. Investigate hybrid CNN-transformer architectures to explore further performance gains.

References list

- [1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229-263. <https://doi.org/10.3322/caac.21834>
- [2] Tanaka, T., Uehara, S., Ishii, T., et al. (2020). Differentiation of small (≤ 4 cm) renal masses on multiphase contrast-enhanced CT by deep learning. *American Journal of Roentgenology*, 214(3), 605-612. <https://doi.org/10.2214/AJR.19.22074>
- [3] Padala, S. A., Barsouk, A., Thandra, K. C., et al. (2020). Epidemiology of Renal Cell Carcinoma. *World Journal of Oncology*, 11(3), 79-87. <https://doi.org/10.14740/wjon1279>
- [4] Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209-249. <https://doi.org/10.3322/caac.21660>
- [5] World Cancer Research Fund International. (2024). Kidney cancer statistics. <https://www.wcrf.org/preventing-cancer/cancer-statistics/kidney-cancer-statistics/>
- [6] National Cancer Institute. (2025). Cancer stat facts: Kidney and renal pelvis cancer. Surveillance, Epidemiology, and End Results (SEER) Program. <https://seer.cancer.gov/statfacts/html/kidrp.html>
- [7] Dai, C., Xiong, Y., Zhu, P., et al. (2024). Deep learning assessment of small renal masses at contrast-enhanced multiphase CT. *Radiology*, 311(2), e232178. <https://doi.org/10.1148/radiol.232178>
- [8] Uhm, K. H., Jung, S. W., Choi, M. H., et al. (2021). Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography. *npj Precision Oncology*, 5(1), Article 54. <https://doi.org/10.1038/s41698-021-00195-y>
- [9] Maçın, G., Genç, F., Taşçı, B., Dogan, S., & Tuncer, T. (2025). KidneyNeXt: A lightweight convolutional neural network for multi-class renal tumor classification in computed tomography imaging. *Journal of Clinical Medicine*, 14(14), 4929. <https://doi.org/10.3390/jcm14144929>
- [10] Abdeltawab, H., Alksas, A., Ghazal, M., et al. (2026). A deep learning ensemble framework for multi-subtype renal tumor classification using contrast-enhanced CT. *Scientific Reports*, 16(1), Article 37283. <https://doi.org/10.1038/s41598-026-37283-7>

- [11] Islam, M. N., Hasan, M., Hossain, M. K., Alam, M. G. R., Uddin, M. Z., & Soylu, A. (2022). Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. *Scientific Reports*, 12(1), Article 11440. <https://doi.org/10.1038/s41598-022-15634-4>
- [12] Alzu'Bi, D., Abdullah, M., Hmeidi, I., AlAzab, R., Gharaibeh, M., El-Heis, M., & Almotairi, K. H. (2022). Kidney tumor detection and classification based on deep learning approaches: A new dataset in CT scans. *Journal of Healthcare Engineering*, 2022, Article 3861161. <https://doi.org/10.1155/2022/3861161>
- [13] Yildirim, K., Bozdog, P. G., Talo, M., Yildirim, O., Karabatak, M., & Acharya, U. R. (2021). Deep learning model for automated kidney stone detection using coronal CT images. *Computers in Biology and Medicine*, 135, 104569. <https://doi.org/10.1016/j.compbiomed.2021.104569>
- [14] Pimpalkar, A., Saini, D. K. J. B., Shelke, N., Balodi, A., Rapate, G., & Tolani, M. (2025). Fine-tuned deep learning models for early detection and classification of kidney conditions in CT imaging. *Scientific Reports*, 15(1), Article 94905. <https://doi.org/10.1038/s41598-025-94905-2>
- [15] Sabuncu, Ö., Bilgehan, B., Kneebone, E., & Mirzaei, O. (2023). Effective deep learning classification for kidney stone using axial computed tomography (CT) images. *Biomedical Engineering / Biomedizinische Technik*, 68 (5), 481–491. <https://doi.org/10.1515/bmt-2022-0142>
- [16] Obaid, W., Hussain, A., Rabie, T., Abd, D. H., & Mansoor, W. (2025). Multi-model deep learning approach for the classification of kidney diseases using medical images. *Informatics in Medicine Unlocked*, 57. <https://doi.org/10.1016/j.imu.2025.101663>
- [17] Baygin, M., Yaman, O., Tuncer, T., Dogan, S., Aydemir, E., & Acharya, U. R. (2022). Exemplar darknet19 feature generation technique for automated kidney stone detection with coronal CT images. *Artificial Intelligence in Medicine*, 127, Article 102274. <https://doi.org/10.1016/j.artmed.2022.102274>
- [18] Qian, G. (2025). Renal CT Image Classification Based on Densely Connected Convolutional Networks. in *Proc. 2nd Int. Conf. Data Anal. Mach. Learn. (DAML)*, 2024, pp. 115–119. <https://doi.org/10.5220/0013510200004619>
- [19] El-Hag, N. A., El-Shafai, W., Abd El-Hameed, H. A., Soliman, N. F., Algarni, A. D., & Abd El-Samie, F. E. (2025). A two-stage deep learning framework for kidney disease detection using modified specular-free imaging and EfficientNetB2. *Scientific Reports*, 15(1), Article 4606. <https://doi.org/10.1038/s41598-025-04606-z>

- [20] Almuayqil, S. N., Abd El-Ghany, S., Ahmed, A. E., & Elmogy, M. (2024). KidneyNet: A Novel CNN-Based Technique for the Automated Diagnosis of Chronic Kidney Diseases from CT Scans. *Electronics*, 13(24), 4981. <https://doi.org/10.3390/electronics13244981>
- [21] Bhandari, M., Yogarajah, P., Subash Muthu, K., & Condell, J. (2023). Exploring the Capabilities of a Lightweight CNN Model in Accurately Identifying Renal Abnormalities: Cysts, Stones, and Tumors, Using LIME and SHAP. *Applied Sciences*, 13(5), Article 3125. <https://doi.org/10.3390/app13053125>
- [22] Bingol, H., Yildirim, M., Yildirim, K., & Alatas, B. (2023). Automatic classification of kidney CT images with relief based novel hybrid deep model. *PeerJ Computer Science*, 9, e1717. <https://doi.org/10.7717/peerj-cs.1717>
- [23] Khan, Z. A., Waqar, M., Khan, H. U., Chaudhary, N. I., Khan, A. T., Ishtiaq, I., Khan, F. A., & Raja, M. A. Z. (2025). Fine-tuned deep transfer learning: an effective strategy for the accurate chronic kidney disease classification. *PeerJ Computer Science*, 11, e2800. <https://doi.org/10.7717/peerj-cs.2800>
- [24] Murillobouzon, M. F., de Santana, P. H. S., Missima, G. N., Pereira, W. S., Rivera, F. P., Giraldi, G. A., Fugita, O. E. H., & Silva, P. S. S. (2025). KSSD2025: A New Annotated Dataset for Automatic Kidney Stone Segmentation and Evaluation With Modified U-Net Based Deep Learning Models. *IEEE Access*, 13, 171790–171806. <https://doi.org/10.1109/ACCESS.2025.3610027>
- [25] McCarthy, J. (2007). *Basic questions*. Stanford University. <http://www-formal.stanford.edu/jmc/whatisai/node1.html>
- [26] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- [27] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018, Lecture Notes in Computer Science*, vol. 11211, pp. 3–19. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_1
- [28] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). <https://doi.org/10.1109/CVPR.2016.90>

- [29] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. V. (2019). Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1314–1324).