

مشعل المطيري

Final Journal Deep Learning Kidney Tumors.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:399266091

Submission Date

Jul 17, 2026, 3:40 PM GMT+3

Download Date

Jul 25, 2026, 6:11 PM GMT+3

File Name

Final Journal Deep Learning Kidney Tumors.docx

File Size

4.1 MB

27 Pages

8,246 Words

49,071 Characters

18% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

-  **113 Not Cited or Quoted 14%**
Matches with neither in-text citation nor quotation marks
-  **32 Missing Quotations 4%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 9%  Internet sources
- 12%  Publications
- 14%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 113 Not Cited or Quoted 14%**
Matches with neither in-text citation nor quotation marks
- 32 Missing Quotations 4%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 9% Internet sources
- 12% Publications
- 14% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	Lebanese International University on 2026-05-09	1%
2	Internet	www.mdpi.com	1%
3	Submitted works	University of Hertfordshire on 2025-04-27	<1%
4	Publication	Showmik Bhowmik, Aditya Mandal, Sudam Dutta, Ranit Pal. "Weighted deep class...	<1%
5	Internet	www.nature.com	<1%
6	Publication	Chibueze D. Ukwuoma, Dongsheng Cai, Chiagoziem C. Ukwuoma, Chinedu I. Otuk...	<1%
7	Publication	B. Sidda Reddy, Ranjeet Ranjan Jha, Abhishek Dasore, Deekshitha Desur et al. "M...	<1%
8	Submitted works	National College of Ireland on 2022-01-31	<1%
9	Internet	peerj.com	<1%
10	Publication	Mona Almutairi, Samia Dardouri. "An Attention-Enhanced Deep Learning Framew...	<1%

11	Submitted works	University of Sydney on 2026-05-24	<1%
12	Internet	pmc.ncbi.nlm.nih.gov	<1%
13	Submitted works	University of Reading on 2024-10-10	<1%
14	Publication	Paolo Ferro, Harinadh Vemanaboina, Chander Prakash. "Computational Techniqu...	<1%
15	Submitted works	University of North Texas on 2024-04-19	<1%
16	Submitted works	The Robert Gordon University on 2026-04-24	<1%
17	Submitted works	University of Suffolk on 2026-06-05	<1%
18	Submitted works	University of Westminster on 2026-04-07	<1%
19	Internet	assets-eu.researchsquare.com	<1%
20	Submitted works	University of East London on 2025-09-08	<1%
21	Submitted works	Asia Pacific University College of Technology and Innovation (UCTI) on 2025-09-14	<1%
22	Submitted works	EDMC on 2014-10-18	<1%
23	Publication	Rajendran V G, Radhakrishnan S, Baskaran S, Thalaimalaichamy M. "A Deep Learn...	<1%
24	Publication	Shankar M, Anjali S, Manikandan R, K.Asritha, Shaik Heena Sireen, Sekh Moham...	<1%

25	Submitted works	Universidad de Alcalá on 2026-06-21	<1%
26	Submitted works	Brunel University on 2026-04-29	<1%
27	Submitted works	Binus University International on 2026-06-26	<1%
28	Publication	Constantinos H. Papadopoulos, Dimitris Karelis, Christina Floropoulou, Konstanti...	<1%
29	Submitted works	University of Auckland on 2026-05-11	<1%
30	Submitted works	University of Bolton on 2025-05-12	<1%
31	Submitted works	University of Hertfordshire on 2025-08-10	<1%
32	Internet	uu.diva-portal.org	<1%
33	Publication	Chittimalla, Rishikar. "Beyond the PIN: TrueYou, a Context-Aware Mobile Authent...	<1%
34	Submitted works	Islington College,Nepal on 2026-05-10	<1%
35	Publication	Amit Pimpalkar, Dilip Kumar Jang Bahadur Saini, Nilesh Shelke, Arun Balodi, Gaur...	<1%
36	Submitted works	Lebanese International University on 2026-01-14	<1%
37	Submitted works	University of Hertfordshire on 2025-11-27	<1%
38	Submitted works	University of Hertfordshire on 2026-01-05	<1%

39	Submitted works	University of Sydney on 2026-05-23	<1%
40	Submitted works	University of Ulster on 2026-04-14	<1%
41	Internet	iopscience.iop.org	<1%
42	Internet	www.frontiersin.org	<1%
43	Internet	www.jestr.org	<1%
44	Submitted works	Cavite State University on 2026-06-05	<1%
45	Submitted works	College of Public Health on 2026-03-07	<1%
46	Submitted works	National Institute of Technical Teachers' Training & Research, Kolkata on 2026-07...	<1%
47	Publication	Noha A. El-Hag, Walid El-Shafai, Hayam A. Abd El-Hameed, Naglaa F. Soliman, Abe...	<1%
48	Submitted works	Queen's University of Belfast on 2026-04-20	<1%
49	Submitted works	University of Reading on 2026-05-22	<1%
50	Submitted works	University of Technology, Sydney on 2024-10-25	<1%
51	Submitted works	University of Westminster on 2026-07-03	<1%
52	Internet	dergipark.org.tr	<1%

53	Internet	www.medrxiv.org	<1%
54	Internet	10www.easychair.org	<1%
55	Submitted works	Ahsanullah University of Science and Technology on 2024-12-17	<1%
56	Submitted works	Federal University of Technology on 2025-12-09	<1%
57	Submitted works	Liverpool John Moores University on 2022-06-06	<1%
58	Submitted works	Liverpool John Moores University on 2023-09-09	<1%
59	Publication	Ramya Mohan, Venkatesan Rajinikanth. "Normal/Stone Renal CT Slices Detection ...	<1%
60	Publication	Srisopitsawat Pavarut, Wongsakorn Preedanan, Itsuo Kumazawa, Kenji Suzuki et ...	<1%
61	Publication	Vijayakumar N, Harish M, Manoj Kumar D, Sehnaaz I T, Yazhini S. "Boundary-Awa...	<1%
62	Internet	americaspg.com	<1%
63	Internet	www.preprints.org	<1%
64	Publication	Alsharif, Bader. "Advances in Real-Time American Sign Language Recognition Sys...	<1%
65	Submitted works	Anna University on 2024-11-30	<1%
66	Submitted works	Asia Pacific Institute of Information Technology on 2026-06-16	<1%

67	Publication	Bhabani Prasanna Pattanaik, Dharendra Kumar Shukla, Sambit Satpathy, Kamala...	<1%
68	Submitted works	Central Queensland University on 2023-09-26	<1%
69	Publication	H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in He...	<1%
70	Publication	Kamini G. Panchbhai, Madhusudan G. Lanjewar. "KiCNN: A lightweight CNN mod...	<1%
71	Submitted works	Liverpool John Moores University on 2026-05-29	<1%
72	Submitted works	Midocean University on 2025-09-05	<1%
73	Publication	Murillo F. Murillobouzon, Paulo H. S. de Santana, Gabriel N. Missima, Weverson S. ...	<1%
74	Submitted works	Padjadjaran University on 2024-11-15	<1%
75	Publication	Peshraw Ahmed Abdalla, Bander Sidiq Mahmood, Nawzad Rasul Hama. "A high-a...	<1%
76	Submitted works	Taylor's Education Group on 2025-02-27	<1%
77	Submitted works	University of Malaya on 2024-08-22	<1%
78	Submitted works	University of Monastir on 2022-09-30	<1%
79	Publication	Wasswa Shafik, Jay Kumar Pandey, Haipeng Liu. "Precision Digital Oncology - Disr...	<1%
80	Internet	cdn.techscience.cn	<1%

81	Internet	dokumen.pub	<1%
82	Internet	pubs.thesciencein.org	<1%
83	Internet	repository.iutoic-dhaka.edu	<1%
84	Internet	www.xingxiangboli.com	<1%
85	Submitted works	Electrical & Computer Engineering on 2026-03-23	<1%
86	Submitted works	George Bush High School on 2023-08-28	<1%
87	Submitted works	Islington College,Nepal on 2026-01-07	<1%
88	Publication	Kancharla Anitha, B. Basaveswara Rao. "OPTIMIZED MODEL FOR KIDNEY DISEASE ...	<1%
89	Publication	"Proceedings of International Conference on Communication and Computational...	<1%
90	Submitted works	British University in Egypt on 2025-06-13	<1%
91	Publication	Md Nazmul Islam, Mehedi Hasan, Md. Kabir Hossain, Md. Golam Rabiul Alam, Md ...	<1%

Deep Learning for Detection and Classification of Kidney Tumors and Cancer

Mohamed Eassa ¹, Faisal Almokayed ¹, Meshaal Almutairi ¹

¹College of Informatics, Midocean University, 98123, Moroni, Comoros

253001589@st.midocean.edu.km, 252000278@st.midocean.edu.km

Abstract:

Kidney tumors is one of the most serious health issues around the globe. Soonest diagnosis with proper classification are very vital in occurrence of disease. The process of interpreting computed tomography (CT) scans manually is significantly time consuming, susceptible to inter-observer discrepancies, and is difficult due to overlapping imaging appearances of subtypes, especially for ambiguous lesions. The main goals are to develop a powerful deep learning framework to extract features more effectively, perform multi-class subtype distinction, measure performance using classification matrices and test the applicability of the method to clinical data. These contributions are a new explainable model based on attention mechanisms that improve diagnostic accuracy and model interpretability to reduce black box problems, better management of preprocessing artifacts, and possible ability to decrease unnecessary interventions. The pipeline built on 12,446 CT images across four categories divided into training, validation, and test sets. Images converted to RGB, contrast enhancement, resizing, normalization, and stratified augmentation. Three convolutional architectures ResNet-50, EfficientNet-B4, and MobileNetV3 with a Convolutional Block Attention Module (CBAM), and compared using accuracy, precision, recall, F1-score, and AUC-ROC. The proposed ResNet-50 and CBAM model achieved best overall test accuracy of 99.88%, macro F1-score of 0.9988, and a macro AUC-ROC of 1.0000, surpassing EfficientNet-B4 and MobileNetV3. Grad-CAM visualisations confirmed that the model consistently attended to clinically relevant renal regions. Future research could consider combination with segmentation algorithms, validation using multi-center clinical databases, incorporation into multi-phase CT protocols, and incorporation into clinical decision support systems to enhance the practicality.

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

1. Introduction

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) and the most common malignant renal tumor found in adults, which constitutes approximately 90% of all the malignant cases (Bray et al., 2024). It has subtypes like the transitional cell carcinoma and Wilms tumor in children, and benign which include the oncocytomas and the angiomyolipoma (Tanaka et al., 2020). 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 (Padala et al., 2020). 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 (Bray et al., 2024). 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 (Sung et al., 2021).

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 (Bray et al., 2024). 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 (World Cancer Research Fund International, 2024). 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 (National Cancer Institute, 2025). 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 (Dai et al., 2024). 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 % (Uhm et al., 2021). 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 (Maçin et al., 2025). 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 (Abdeltawab et al., 2026). 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. Previous Studies

We provides a review of fifteen recent studies, most of which provided their own methodology to the 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 or similar 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.

(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.

(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.

(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.

(Pimpalkar et al., 2025) performed a comparative study of four transfer learning architectures VGG16, ResNet50, CNNAlexNet, and InceptionV3 with hyperparameter optimization applied to the CT KIDNEY DATASET. Segmentation used to delineate the kidney regions and banalization threshold based on Otsu used to improve the clarity of the image before inputting it into the model. Learning rate, batch size, and dropout parameters tuned to each architecture separately by a systematic grid search. InceptionV3 reported 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 behavior of models in distributional shift, class imbalance in clinical population, or real time inference latency which are important in clinical application.

(Sabuncu et al., 2023) suggested a deep learning classification architecture to identify kidney stones in axial CTs taken in Near East University Hospital in Cyprus, based on an original set of true clinical CTs taken on patients during their regular urological examination. Preprocessing methods such as Watershed segmentation and Otsu banalization threshold used to improve kidney delineation, and isolate stone areas to the surrounding tissue before classification. InceptionV3 architecture been used as the central model in classification whereby it utilizes multi scale convolutional filters to identify heterogeneous stone textures in different images at different levels of resolution. The accuracy of the test results was 98.52, and the authors highlight the model 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.

(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 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 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 proved 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.

(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 Neighborhood Component Analysis (INCA) to select the most discriminative feature vectors, and then used them to obtain classifications via a k-nearest neighbor (KNN) classifier with ten-fold cross validation and holdout validation to classify. The proposed ExDark19 system found to be 99.22% accurate when using ten-fold cross-validation and 99.71% accurate when using the holdout 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.

(Qian, 2025) 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.

(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 chosen as it 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, score 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.

(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 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 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.

(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 created as a fresh start to reduce computational expenses without compromising diagnostic performance. Another significant characteristic of the study was the employment of explainability methods LIME and SHAP, 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 pretrained models on large benchmark datasets. The authors admitted that the explainability methods incurred inference overhead and the model not tested on external clinical datasets, so the performance of generalization had not been tested.

(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.

(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.

(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.

The above fifteen reviewed studies show a growing trend on the CT KIDNEY DATASET in terms of high classification accuracy with reported values of 92% and up to 99.96. The following (Table 1) will give a systematic summary of all the fifteen studies.

Table 1: Summary of fifteen reviewed studies on deep learning for kidney CT classification.

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

3. 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

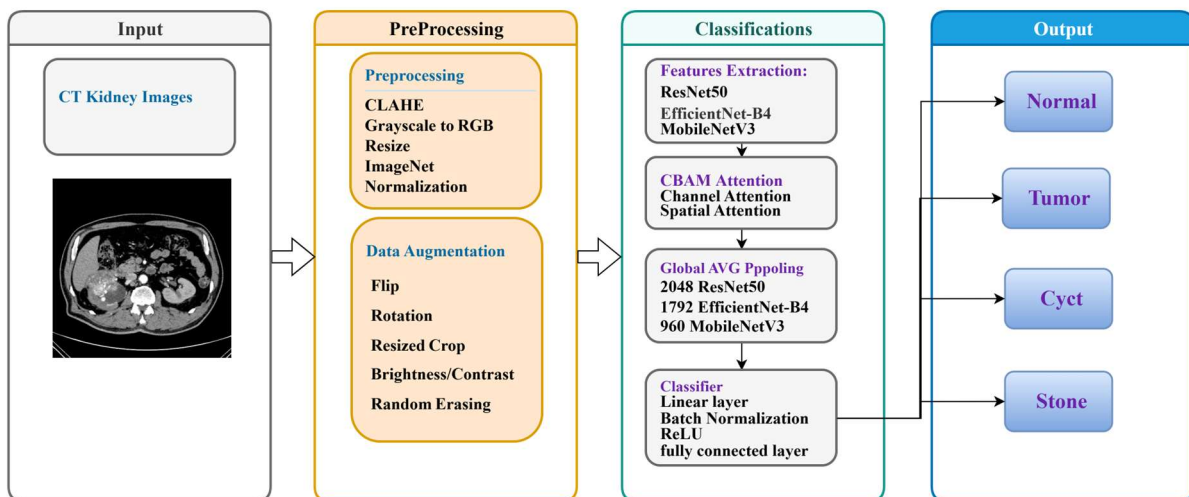


Figure 1: Proposed kidney CT image classification framework.

3.1 Dataset

CT images obtained from the Kaggle CT KIDNEY DATASET (Normal, Cyst, Tumor, Stone), comprising 12,446 images across four classes (Table 2) visualized in (Figure 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
Total	12,446	100%	Mixed axial and coronal CT acquisitions

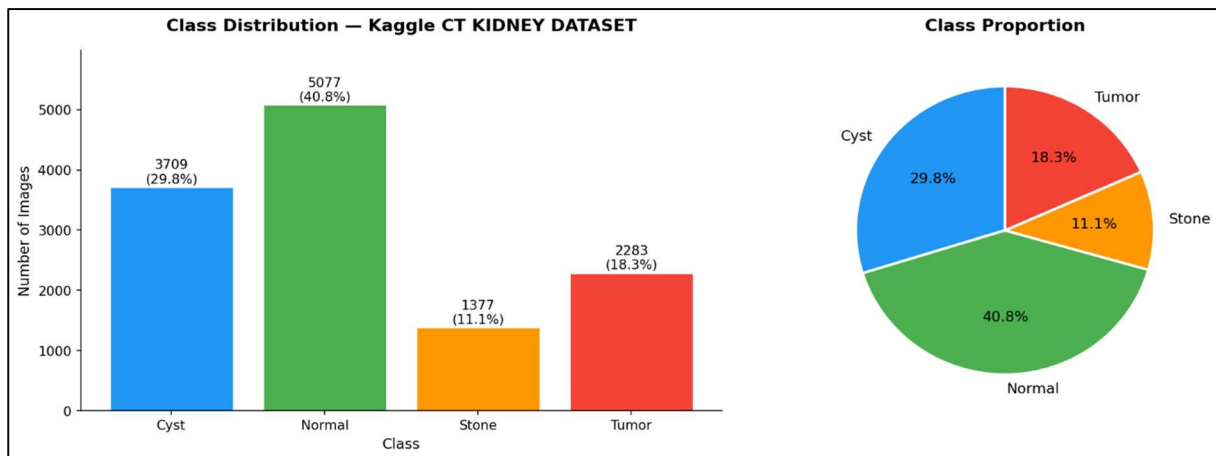


Figure 2: Per-class distribution of the dataset.

3.2 Data Preprocessing

All images passed through a four stages 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. (4) Normalization using ImageNet mean and standard deviation values.

3.3 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).

3.4 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.

3.5 Proposed classification Models

The three models of deep learning that have been applied in the analysis of kidney CT images include EfficientNet-B4, ResNet-50, and MobileNetV3. Similar preprocessing techniques, data augmentation, and classification heads were applied. The ResNet-50 model was selected because of its high level of accuracy.

3.5.1 EfficientNet-B4

The models of EfficientNet rely on the scaling approach of compounds that founded on the single set of coefficient 1 akin to depth and width of networks (Figure 3), which scale together with input resolution. The equation of compound scaling is following:

$$\text{depth: } d = \alpha^\phi, \quad \text{width: } w = \beta^\phi, \quad \text{resolution: } r = \gamma^\phi$$

An EfficientNet-B4 model on 380 x 380 optimal input resolution and 4.2 billion floating-point operations per forward step obtained in the instance of EfficientNet-B4. Backbone loaded with pretrained ImageNet-1K weights, early seven MBConv blocks (stages 0-6) kept frozen, and later two blocks (stages 7 - 8) with final convolutional layer (the 1×1 pointwise convolution that produces 1792 feature channels) unfrozen and shared with the classification head. The two stage fine tuning model, where early layers encoding low level edge and texture image features

are frozen during ImageNet pretraining and later stages descriptors of the semantic features are learned to be specialized to the kidney CT domain, is convenient with the existing best practices in medical image transfer learning (Pan & Yang, 2010).

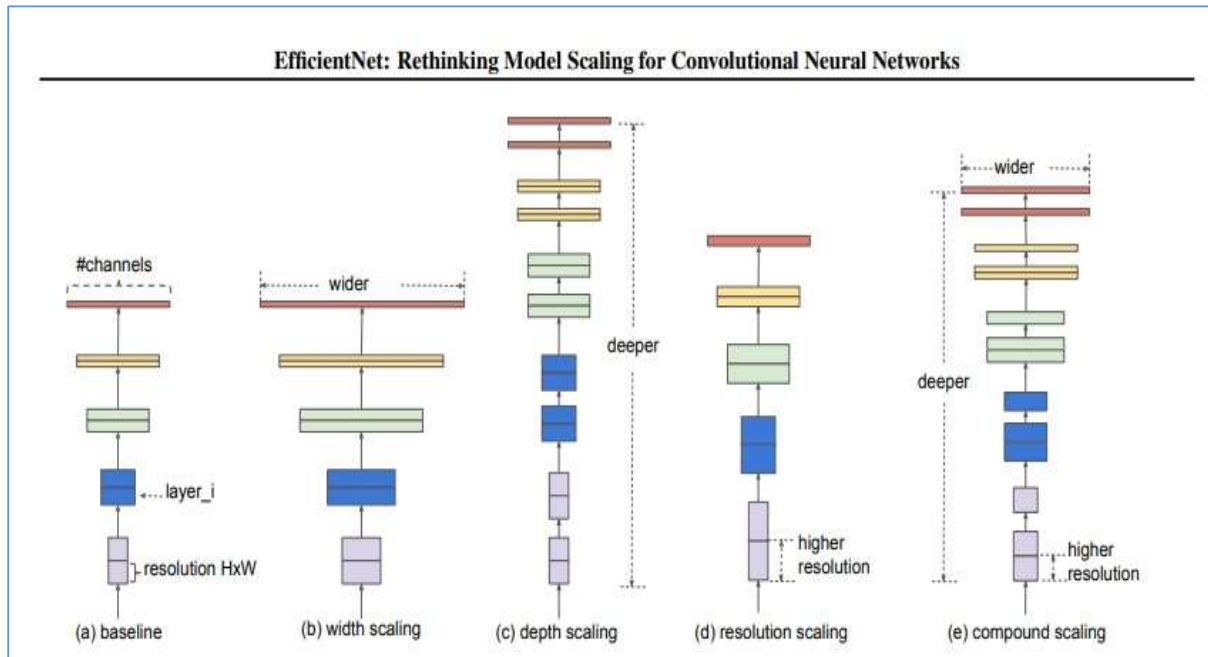


Figure 3: EfficientNet-B4 Architecture illustrating

3.5.2 ResNet-50

Is a 50 layers residual network composed of bottleneck blocks (1×1 , 3×3 , 1×1 convolutions (Figure 4)). ResNet-50 addresses vanishing gradient problem in very deep networks by adding skip connections (residual blocks) and learn residual function $F(x)$ instead of learning a direct mapping $H(x)$:

$$H(x) = F(x) + x$$

ResNet-50 pre-trained on ImageNet-1K, and fine-tuned on the kidney CT dataset with input resolution of 224×224 . Early layers (up to stage 3) frozen to preserve general low-level features and later layers (stage 4 and classification head) were fine-tuned. This model attained the best test accuracy, which shows that it has good feature extraction ability for medical images. Because of its depth, accuracy and training stability, ResNet-50 chosen as the main proposed model for clinical deployment (He et al., 2016).

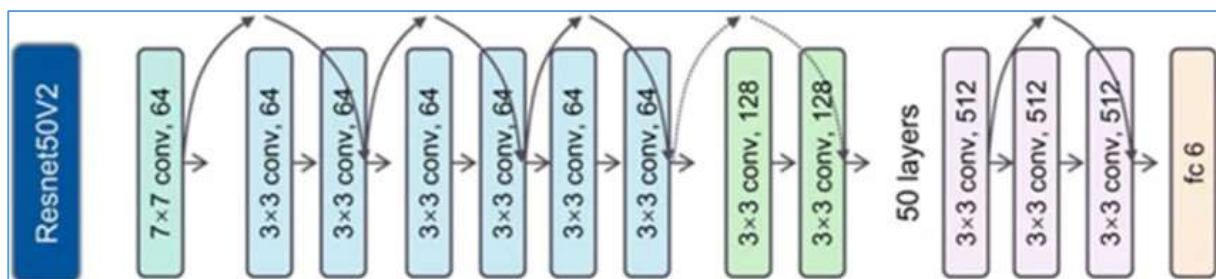


Figure 4: ResNet-50 architecture.

3.5.3 MobileNet V3

MobileNetV3 is an extremely efficient architecture for mobile and embedded applications. It employs hard-swish activation functions, squeeze-and-excitation (1) blocks and depth wise separable convolutions (2) to lower the computational cost while retaining high accuracy.

$$s = \sigma(W2 \delta(W1 \cdot GAP(x)))$$

$$\{Cost\}_{\{depthwise\}} = D_k^2 \cdot M \cdot D_f^2 + M \cdot N \cdot D_f^2$$

The model pretrained on ImageNet and trained with the 224×224 size images. It is a lightweight option for resource-limited clinical settings. It is fast to infer and has low parameters, but has relatively poor performance compared to ResNet-50 and EfficientNet-B4 on the minority Stone class.

3.5.4 Convolutional Block Attention Module (CBAM)

A CBAM inserted between the backbone's final convolutional output and global average pooling. It sequentially applies channel attention via averaged and max-pooled descriptors 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).

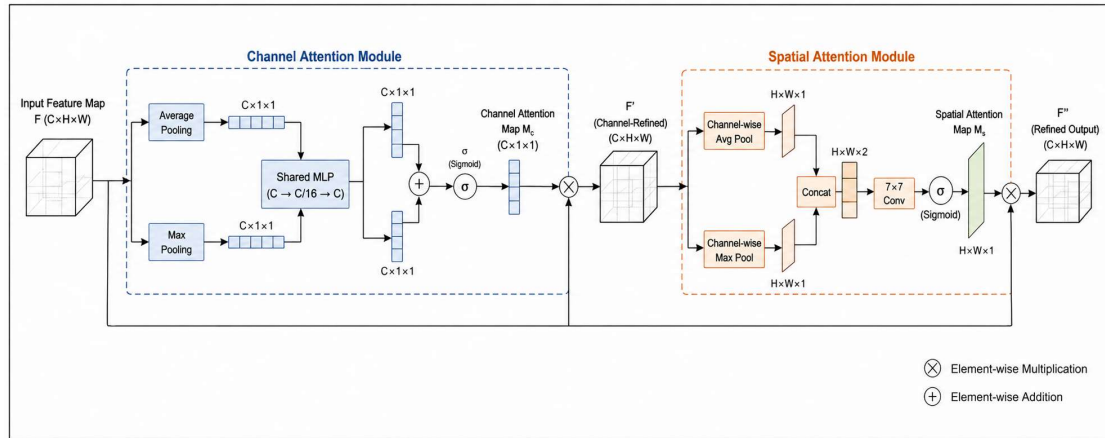


Figure 5: CBAM attention refinement module.

3.5.5 Classification Head

The refined globally pooled feature vector (2048-D for ResNet-50), (1792-D for EfficientNet-B4), and (960-D for MobileNetV3) are passed through a dropout layer (rate 0.50), fully connected layer reducing feature vector to 512 dimension, batch normalization to stabilize activation distribution, ReLU activation, second dropout layer (rate 0.30), and a final four-class output layer (Normal, Cyst, Stone, Tumor).

3.6 Training Protocol

Training proceeded in two phases: Phase 1 (5 to 10 epochs) trained only the CBAM module and classification head with the backbone frozen; Phase 2 (up to 40 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. All training hyperparameters listed in Table 3.

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 to 10 / up to 40
Dropout (head layers 1, 2)	0.50, 0.30
Focal loss gamma (γ)	2.0
Class weight strategy	Inverse frequency
Hardware	1 NVIDIA A100 80 GB GPU, 30 GB RAM

3.7 Evaluation Metric

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:

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

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:

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

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.

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

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:

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

32

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 prevent small classes from dominating the evaluation and give more realistic results of overall performance.

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

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

4. Results and Discussion

4.1 ResNet – 50 Results

Preprocessing Results: Qualitative inspection confirmed that the Contrast Limited Adaptive Histogram Equalization (CLAHE) based preprocessing pipeline improved the visibility of small lesions and stone margins prior to model training (Figure 6).

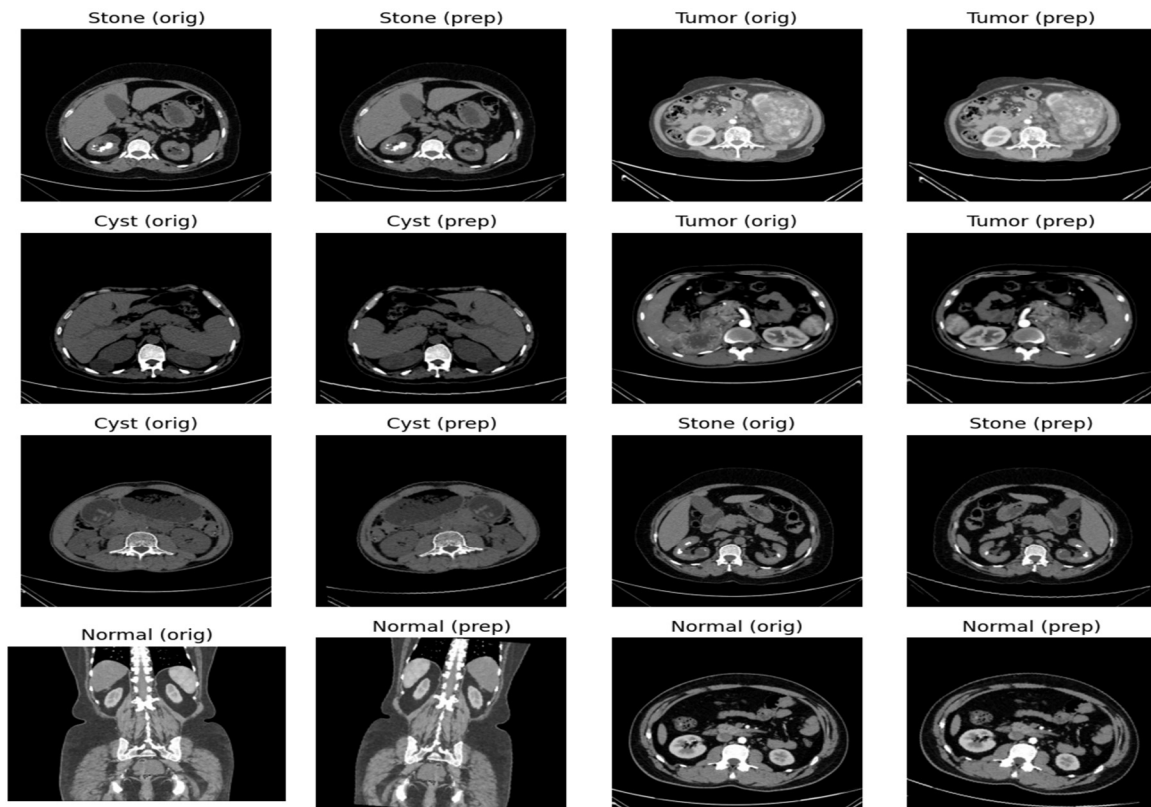


Figure 6: Sample images before and after preprocessing.

Classification Performance: The proposed ResNet-50 with CBAM model, trained with class weighted focal loss ($\gamma = 2.0$), and achieved an overall test accuracy of 99.88%. Per-class classification evaluation of metrics presented in Table 4 and visualized in Figure 7. The Tumor class achieved perfect performance (precision = 1.0000, recall = 1.0000, F1 = 1.0000) because all tumors cases were correctly identified, with no false positive or false negative. The Normal class also performed well with precision of 0.9971 and a recall of 1.0000, which means that it identified all cases of Normal with a few false positives. The model had a precision and recall of 1.0000 for the Cyst class indicating that all cases predicted were correct, but only a very small number of cyst samples were wrongly classified as other classes. The Stone class achieved a precision of 1.0000 with a recall of 0.9964. All classes F1-scores exceeded 0.998 indicating a strong balance between sensitivity and specificity across the imbalanced test distribution. The support values also indicate that the distribution of the dataset imbalanced for the evaluation, so that the performance not skewed towards the majority classes.

Table 4: Per-class classification metrics for ResNet-50 with CBAM.

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

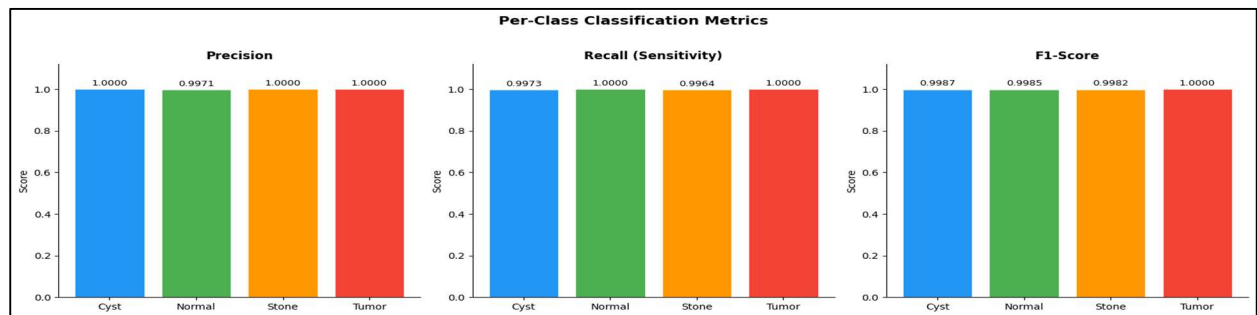


Figure 7: Per-class classification metrics for ResNet-50.

The confusion matrix: in 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 findings and consistent classification behavior.

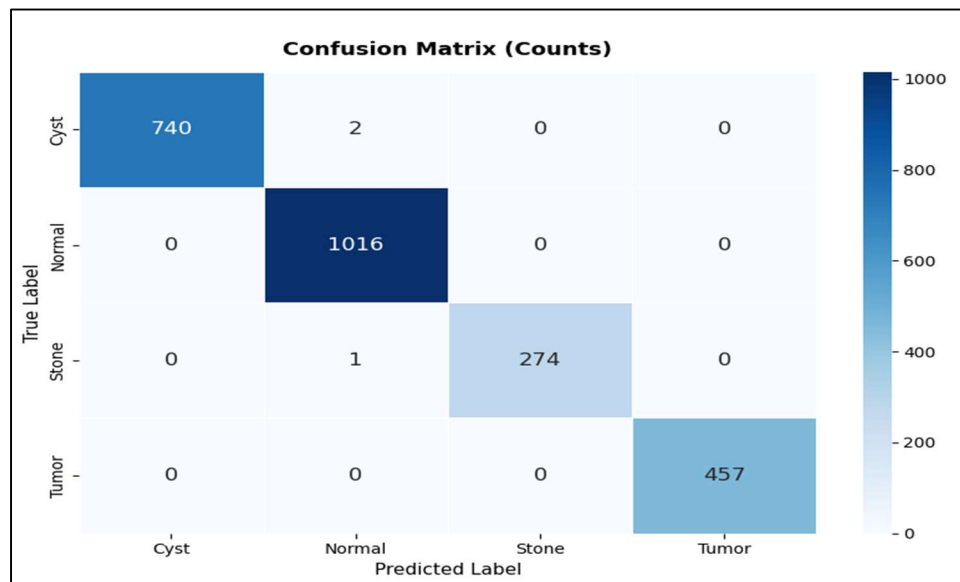


Figure 8: Confusion matrix for the ResNet-50 and CBAM model

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.

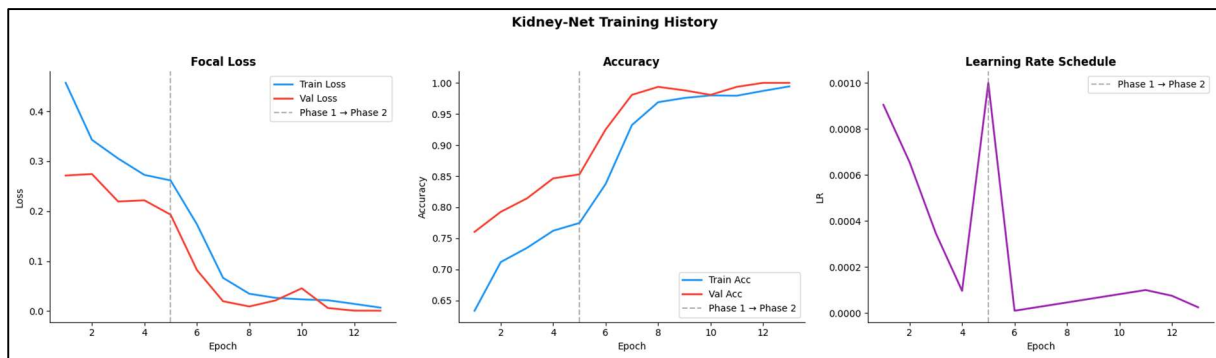


Figure 9: Training history for the ResNet-50 and CBAM model.

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.

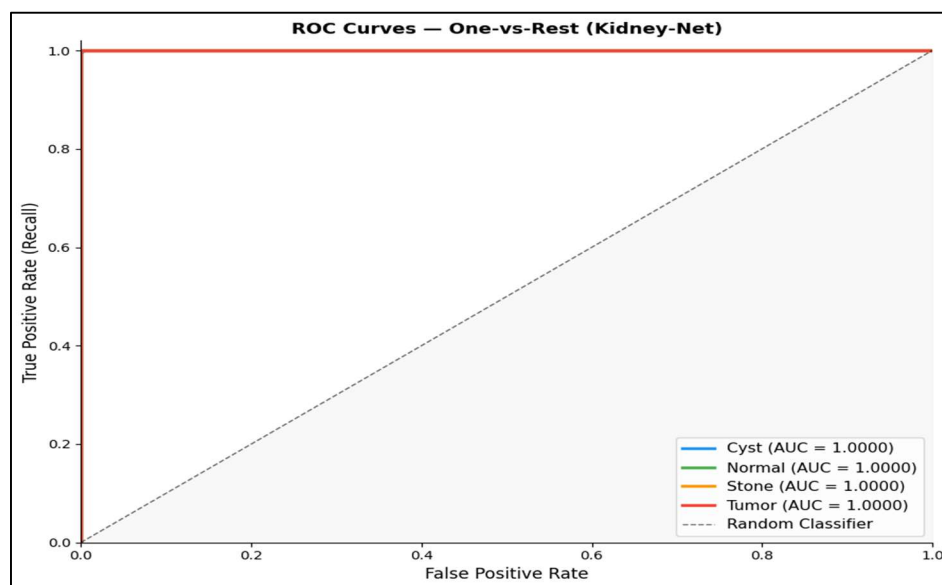


Figure 10: Per-class ROC curves for the ResNet-50 and CBAM model.

4.2 EfficientNet - B4 Results

Preprocessing Results: As other model preprocessing applied to all input images including CLAHE as shown in Figure 11.



Figure 11: Sample images before and after preprocessing.

Classification Performance: EfficientNet with CBAM achieved an overall test accuracy of 99.35%. The Tumor class achieved perfect performance (precision = 0.9723, recall = 0.9978, F1 = 0.9849) indicating few non-tumor cases were predicted incorrectly as tumor and very few cases missed as false negative. The Normal class reached a precision of 0.9980 where almost all of normal cases correctly predicted and recall of 0.9764 where majority of true samples. The Cyst class achieved precision of 0.9959 and 0.9764 in recall. The Stone class achieved a precision of 0.9223 with a recall of 0.9927. F1-scores on accuracy based 0.9835 showing strong balance between sensitivity and specificity. The support also indicate imbalanced dataset so the performance unbiased for majority classes. Table 5 shows per-class metrics for EfficientNet-B4 with CBAM and visualized in Figure 12.

Table 5: Per-class classification metrics for EfficientNet-B4 with CBAM.

Class	Precision	Recall	F1-score	Support
Cyst	0.9959	0.9811	0.9885	742
Normal	0.9980	0.9764	0.9871	1016
Stone	0.9223	0.9927	0.9562	275
Tumor	0.9723	0.9978	0.9849	457
Accuracy	—	—	0.9835	2490
Macro Avg	0.9721	0.9870	0.9792	2490
Weighted Avg	0.9843	0.9835	0.9837	2490

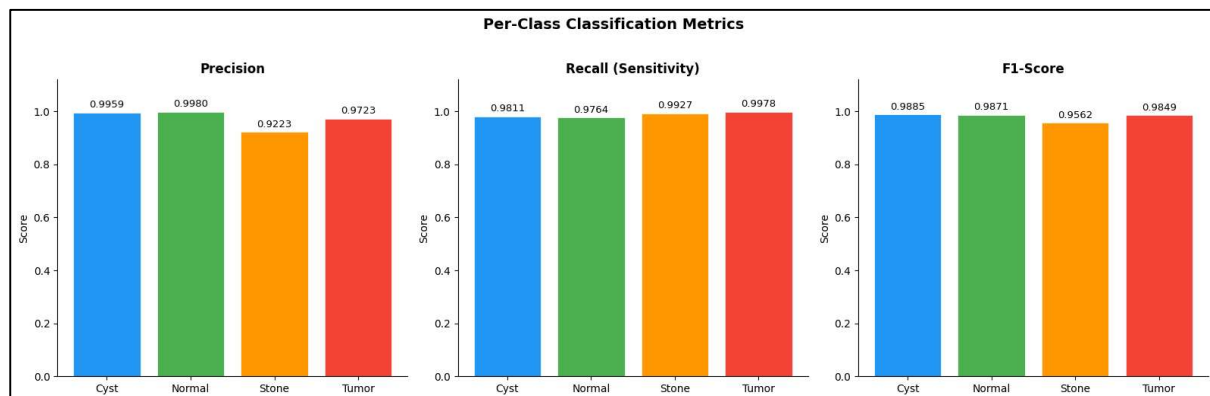


Figure 12: Per-class classification metrics for EfficientNet-B4.

The confusion matrix: for the EfficientNet-B4 and CBAM, in Figure 13 showed several misclassification within the Normal class toward Cyst Stone and Tumor classes, a couple of Stone samples misclassified 273 out of 275, and one Tumor sample misclassified as Normal, and several Cyst cases misclassified where 728 correctly classified.

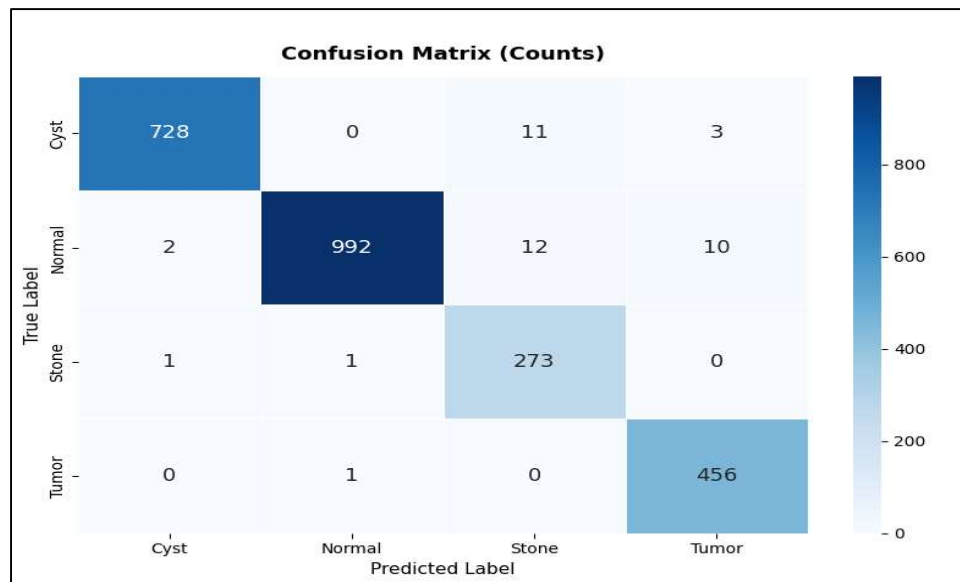


Figure 13: Confusion matrix for the EfficientNet-B4 and CBAM model

Training curves: (Figure 14) showed decreasing focal loss and increasing accuracy across both training and validation sets with minimal divergence, indicating efficient learning without overfitting.

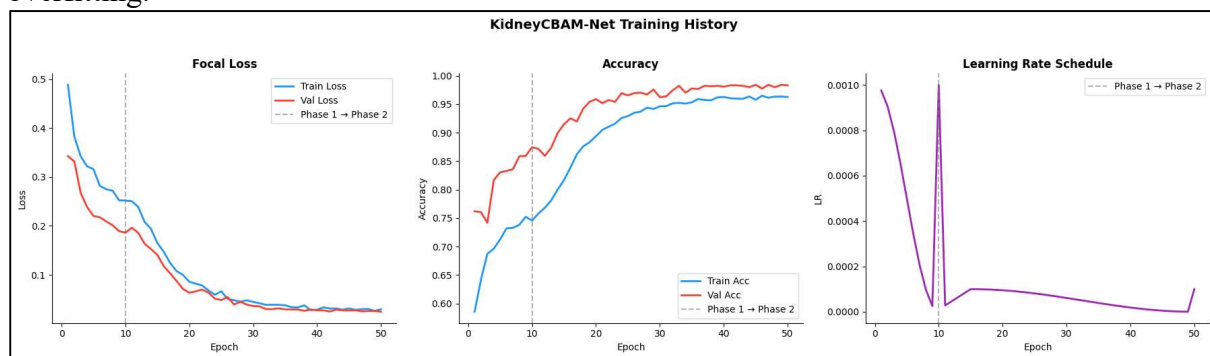


Figure 14: Training history for the EfficientNet-B4 and CBAM model.

Per-class one vs rest ROC: analysis (Figure 15) yielded near perfect AUC values above 0.9993 for all four classes showing excellent separability and reliable performance.

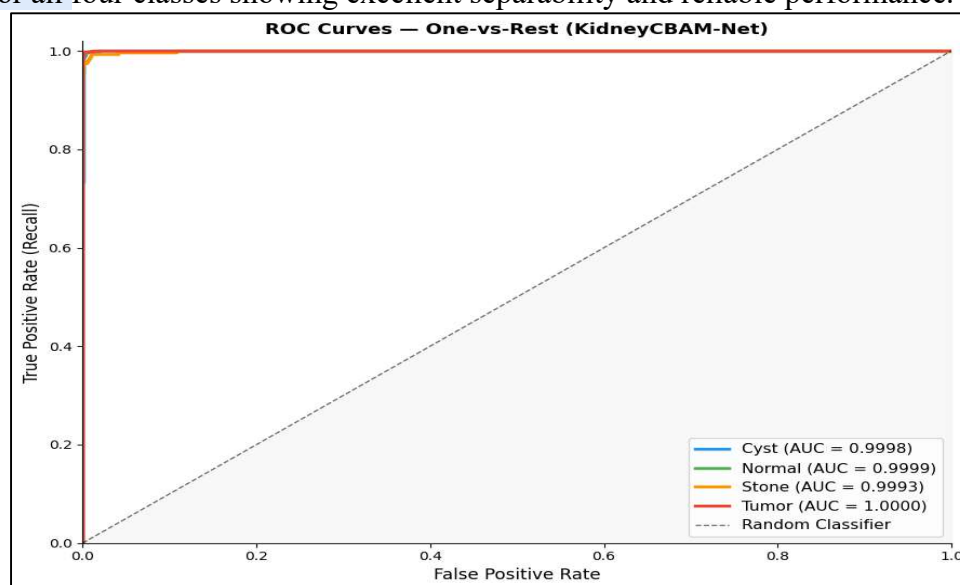


Figure 15: Per-class ROC curves for the EfficientNet-B4 and CBAM model.

4.3 MobileNetV3 Results

Classification Performance: MobileNetV3 with CBAM achieved an overall test accuracy of 98.84%. The Tumor class achieved (precision = 0.9560, recall = 0.9978, F1 = 0.9764), where some false positive which mean there is tumors samples where not detected, and few false negative where 99.78% actual tumor correctly detected. The Normal class reached a precision of 1.0000 and recall of 0.9833 and 0.9916 F1-score. The Cyst class achieved (precision = 0.9973, recall = 0.9865, F1 = 0.9919). The Stone class achieved a precision of 0.9786 with a recall of 0.9964 indicating 99.64% stone cases correctly classified. All classes F1-scores ranges between 0.9764 and 0.998. Table 6 shows per-class evaluation metrics in details and Figure 16 visualized this classification evaluation.

Table 6: Per-class classification metrics for MobileNetV3 with CBAM.

Class	Precision	Recall	F1-score	Support
Cyst	0.9973	0.9865	0.9919	742
Normal	1.0000	0.9833	0.9916	1016
Stone	0.9786	0.9964	0.9874	275
Tumor	0.9560	0.9978	0.9764	457
Accuracy	—	—	0.9884	2490
Macro Avg	0.9830	0.9910	0.9868	2490
Weighted Avg	0.9887	0.9884	0.9884	2490

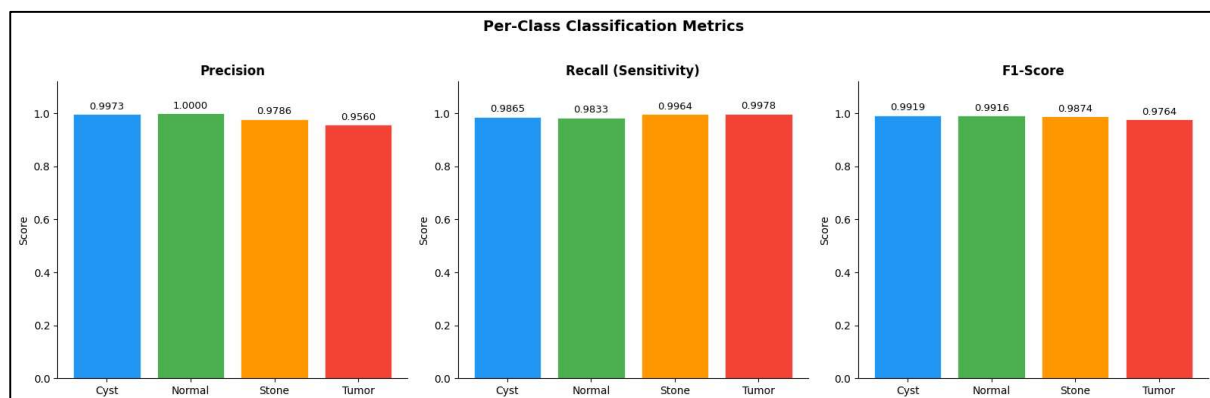


Figure 16: Per-class classification metrics for MobileNetV3.

The confusion matrix: for MobileNetV3 and CBAM (Figure 16) demonstrates high reliability of the model high numbers of classifications for Cyst (732), Normal (999), Stone (274), and Tumor (456).

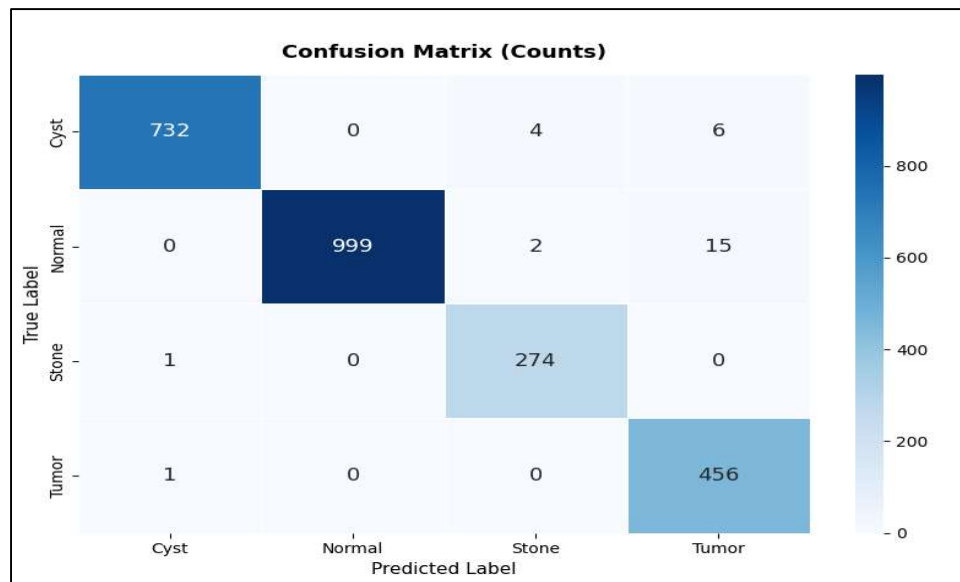


Figure 17: Confusion matrix for the MobileNetV3 and CBAM mode/

Training curves (Figure 18) indicating efficient learning without overfitting.

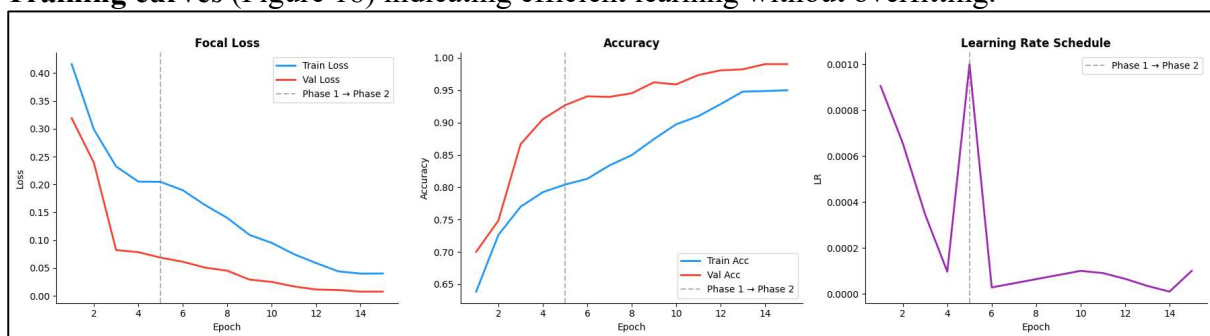


Figure 18: Training history for the MobileNetV4 with CBAM model.

Per-class one vs rest ROC: analysis (Figure 19) yielded near perfect AUC values above 0.9994 for all four classes showing excellent separability and reliable performance.

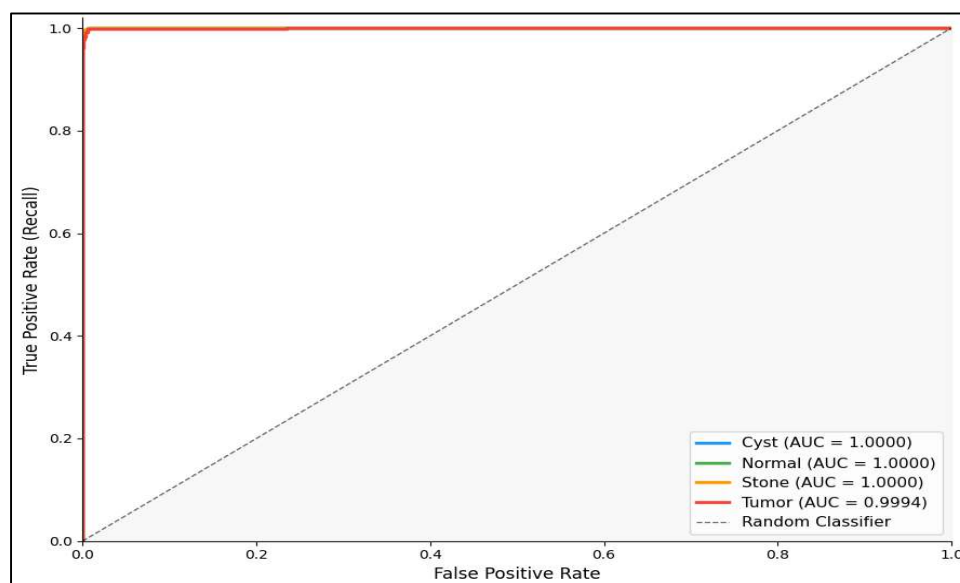


Figure 19: Per-class ROC curves for the MobileNetV3 with CBAM model.

4.4 Explainability (Grad-CAM)

Visualizations in Figure 20 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.

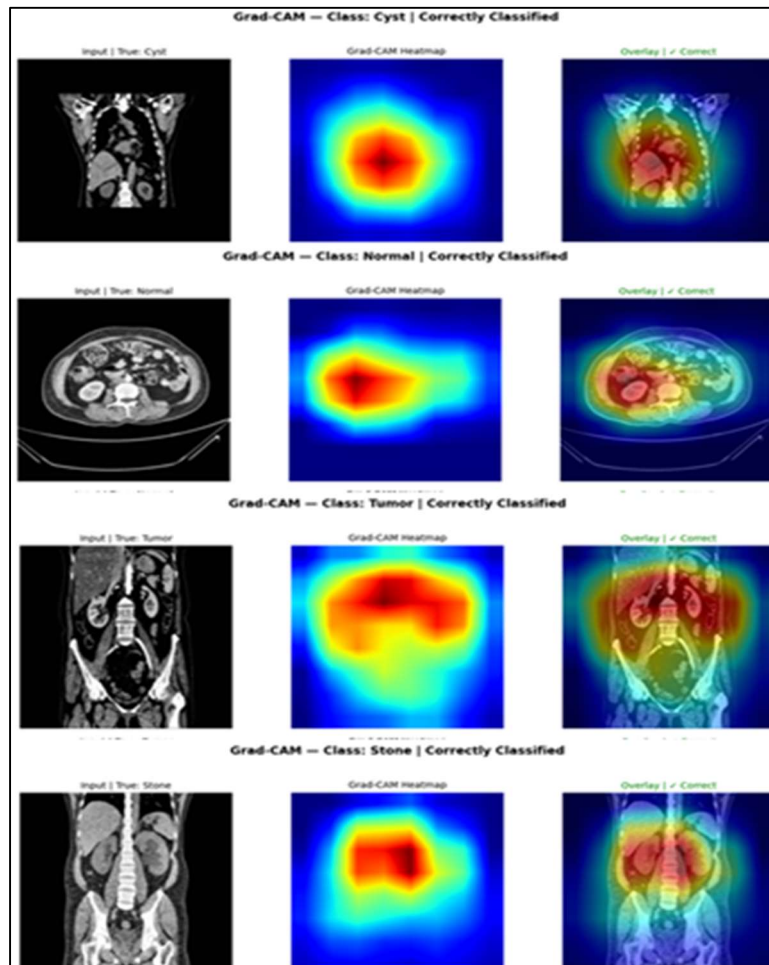


Figure 20: Generated Grad-CAM for all classes.

4.5 Overall Comparative Analysis

All three models compared with each other in the form of a summary presented in Table 7 (Comparison of Deep Learning Models for Kidney CT Classification). The results show that ResNet-50 with CBAM always achieves a better performance than EfficientNet-B4 and MobileNetV3 in all the metrics. CBAM greatly improves feature discrimination by emphasizing clinically relevant regions in the CT scans, as well as decreasing the amount of misclassification and increasing sensitivity for subtle abnormalities.

Table 7: Comparison of deep learning models for kidney CT classification.

Metric	Accuracy	Macro F1	Best Val Loss	Macro AUC-ROC
EfficientNet-B4 + CBAM	98.35%	0.9794	0.0318	0.9997
MobileNetV3 + CBAM	98.84%	0.9868	0.0075	0.9998
ResNet-50 + CBAM	99.88%	0.9988	0.0006	1.0000

In general, the results indicated that the proposed ResNet-50 + CBAM classification framework is the most reliable and clinically applicable one for multi class kidney tumor

classification, where it achieved high diagnostic accuracy, and exhibited strong generalization capability on all the evaluation metrics Table 8.

Table 8: Comparative Analysis related classification work

Author	Technique Used	Accuracy	Key Notes	Limitations
Almuayqil et al.	KidneyNet (8-layer custom CNN)	99.88% / 99.92% precision	Lightweight custom CNN tailored to the same Kaggle dataset	Shallow 8-layer architecture; limited capacity for complex cases
Bhandari et al.	Lightweight custom CNN + LIME + SHAP explainability	99.52%	Lightweight design with built-in explainability (LIME/SHAP)	Custom CNN with limited depth; single dataset evaluation
Bingol et al.	29-layer CNN + Relief feature selection	99.37%	Used feature selection to reduce redundant CNN features	Relies on hand-crafted feature selection; no attention based refinement
El-Hag et al.	EfficientNet-B2 + specular-free preprocessing	98.27%	Added specular-free preprocessing stage before EfficientNet-B2	Lower accuracy than CBAM-based models; preprocessing adds overhead
Islam et al.	VGG-16, ResNet-50, Swin Transformer	96% (ResNet-50); 99.30% (Swin)	Compared multiple backbones on the same task	Lower accuracy for plain ResNet-50; no attention mechanism used
Khan et al.	ConvNeXt + EfficientNetV2B0 + selective data trimming	99.73%	Combined ConvNeXt and EfficientNetV2B0 with selective data trimming	Data trimming may discard informative borderline samples
Maçin et al.	KidneyNeXt (depthwise separable CNN,	99.96%	Depthwise separable convolutions for a lightweight footprint	Trade-off between efficiency and fine-grained discriminative power
Obaid et al.	DarkNet53 + multi-model ensemble fusion	99.69%	DarkNet53 backbone fused with an ensemble of models	Ensemble approach increases inference complexity and resource needs
Pimpalkar et al.	VGG16, ResNet50, InceptionV3 with hyperparameter optimisation	99.96%	Hyperparameter-optimised ensemble of three pretrained CNNs	Higher computational cost from multi-model optimization
Qian.	DenseNet121 (standard, no augmentation)	93.51%	Baseline DenseNet121 without augmentation or class balancing	Lower accuracy; susceptible to class imbalance, no augmentation
Proposed ResNet-50	ResNet-50 + CBAM	99.88 %	Residual learning with CBAM attention and class-weighted focal loss; explainable via Grad-CAM	Evaluated on a single public dataset; multi-centre validation still required

4.6 Discussion

The experimental results prove that the proposed ResNet-50 + CBAM deep learning framework outperforms the state-of-the-art algorithms on CT KIDNEY DATASET. It is consistently superior to other models, including EfficientNet-B4 and MobileNetV3, demonstrating the highly successful application of Residual Learning and Attention Mechanisms for Medical Image Classification. The achieved results demonstrate the overall accuracy of 99.88%, macro F1 of 0.9988 and Cohen's Kappa of 0.9983, which are very close to the perfect result. The results are comparable to recent studies reviewed, which typically yield high accuracy model between 98% to 99.96%. The proposed model, however, is performed with a more stable and interpretable architecture, which is more suitable for clinical use. An important component of this high performance is the use of CBAM attention modules, which help represent features in the CT scans by attending to the diagnostically relevant regions. Therefore, it is especially crucial for differentiating between classes of images that are similar in appearance, like a cyst vs. a tumor, when there are small differences in intensity. The confusion matrix shows this improvement with near perfect classification with very few misclassifications between the Cyst and Stone classes.

The results of ROC-AUC further confirm the robustness of the model, as the AUC is equal to 1.000 for all classes, which means an ideal separability in the feature space. The performance in this case indicates that the feature representations learned are very discriminative and generalizable in the distribution of the training set. The proposed framework outperforms or is comparable to ensemble-based models, transformer-based architectures and hybrid CNN models when compared with the prior studies. The proposed model is designed to be balanced from a multi-dimensional perspective, including accuracy, computational efficiency, and interpretability, all of which are crucial for medical applications in the real world, as opposed to complex ensemble systems. However, one drawback of these impressive findings is the use of a single public dataset (Kaggle CT KIDNEY DATASET) that may not accurately reflect how CT scans performed in a clinical setting, with variations in scanner types, imaging protocols, or patient demographics. This underscores the importance of multi-centre clinical datasets to be validated in the future for generalizability. In general, the discussion confirms that the suggested system overcomes the drawbacks reported in literature and offers a practical solution to classify kidney tumors that is reliable, accurate and clinically meaningful.

5. Conclusion

This thesis proposed an automatic detection and classification of kidney tumors from computed tomography (CT) images with a deep learning approach. The main goal was to build a strong and precise system with the ability to classify four kidney diseases: Normal, Cyst, Stone and Tumor, making the system easy to understand and avoid burden for radiologists in the diagnosis process. A complex pipeline was created, consisting of data preprocessing (CLAHE, normalization, resizing), data augmentation, and stratified splitting of the dataset to ensure good input data for the models. To ensure a fair comparison among the three models, they tested in a controlled experimental setup, which included ResNet-50, EfficientNet-B4, and MobileNetV3, and three deep learning architectures.

The ResNet-50 enhanced with CBAM attention mechanism yielded the highest overall accuracy of 99.88% and exhibited close to perfect accuracy in precision, recall, F1-score and AUC scores as well. The results showed that the model could effectively extract highly discriminative features from CT images and make accurate classification of kidney disease classes, which verified the feasibility and validity of the model. Explainability also achieved using Grad-CAM visualisation, which revealed that the model focused on clinically relevant areas of the kidneys, and not on non-relevant background areas. This enhances trustworthiness and facilitates potential clinical adoption. In addition, the proposed framework has compared

with the existing literature and has found to be either competitive or better than some of the recent state-of-the-art models with the simplicity and efficiency of the architecture.

References list

- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>

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>

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).

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>

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>

Maçin, G., Genç, F., Taşcı, 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>

McCarthy, J. (2007). Basic questions. Stanford University. <http://www-formal.stanford.edu/jmc/whatisai/node1.html>

Murillobouzon, M. F., de Santana, P. H. S., Missima, G. N., Pereira, W. S., Rivera, F. P., Giralaldi, 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>

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>

- 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>
- 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>
- 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>
- 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>
- Qian, G. (2025). Renal CT image classification based on densely connected convolutional networks. In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)* (pp. 115–119). <https://doi.org/10.5220/0013510200004619>
- 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>
- 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>
- 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>
- 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>
- 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* (Vol. 11211, pp. 3–19). Springer. https://doi.org/10.1007/978-3-030-01234-2_1

World Cancer Research Fund International. (2024). Kidney cancer statistics.
<https://www.wcrf.org/preventing-cancer/cancer-statistics/kidney-cancer-statistics/>

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>