



Deep Learning for Detection and Classification of Kidney Tumors and Cancer

Prepared By

Name: Faisal Almokayed

ID No.: 253001589

Name: Meshaal Almutairi

ID No.: 252000278

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics

Midocean University

Supervisor Name: Dr. Mohamed Eassa

Collage of Informatics, Midocean University, Comoros

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

Dedication and Acknowledgments

This work is dedicated to our families who have supported us through this journey by extending their love, encouragement, and patience.

Also, we would like to extend our sincere appreciation to Dr. Mohammed Issa, for his supervision and encouragement.

We are thankful to all the faculty members of the College of Informatics in Midocean University for their valuable contribution of academic knowledge.

We appreciate our colleagues and friends for their help and motivation.

Lastly, we are thankful to all those people who have supported us directly or indirectly to complete this work successfully.

Abstract

Renal cell carcinoma is one of the most serious health issues around the globe: it is assumed that 90 percent of all malignancy in the kidneys, and 434,840 new cases of the disease are estimated to occur all over the world in 2022, resulting in 155,953 deaths. Early diagnosis and proper multi-classification are very vital in the enhancement of survival rates in the occurrence of disease. Nevertheless, the process of interpreting computed tomography (CT) scans images manually is time consuming, liable to inter-observer discrepancies, and is difficult due to overlapping imaging appearances of subtypes.

The main goals are to develop a powerful deep learning framework incorporating multiphase data to extract features more effectively, perform multi-class subtype distinction, measure performance using such metrics as accuracy, sensitivity, specificity, AUC, and test the applicability of the method to clinical data. These contributions are a new explainable model based on attention mechanisms to reduce black-box problems, better management of preprocessing artifacts, and possible ability to decrease unnecessary interventions.

A multi class kidney CT classification pipeline was developed and evaluated on the publicly available Kaggle CT KIDNEY DATASET, comprising Normal, Cyst, Stone and Tumor images. The pipeline applies CLAHE based contrast enhancement, normalization, resizing and stratified data augmentation before training and comparing three convolutional architectures: ResNet50, EfficientNet-B4, and MobileNetV3, each enhanced with attention module using Convolutional Block Attention Module (CBAM). The proposed ResNet50 with CBAM model was trained with a class weighted focal loss ($\gamma = 2.0$) to address class imbalance and emphasize clinically important regions of each scan.

The ResNet50 model with CBAM attention achieved the highest overall test accuracy of 99.88%, a macro F1-score of 0.9988 and a macro AUC-ROC of 1.0000, outperforming both EfficientNet-B4 and MobileNetV3 across all evaluation metrics. Grad-CAM visualizations further confirmed that the model consistently focused on clinically relevant renal regions rather than background artifacts, supporting its interpretability and potential trustworthiness for clinical decision support. These findings demonstrate that combining residual learning with attention mechanisms yields a kidney tumor classification system that is simultaneously accurate, stable and explainable, and that is well suited to assisting radiologists and reducing unnecessary clinical interventions.

Table of Contents

Abstract	3
Chapter 1: Introduction and Overview	8
1.1 Overview	8
1.2 Problem Definition	9
1.3 Thesis Objectives	10
1.4 Thesis Contribution	11
1.5 Thesis Organization	12
Chapter 2: Background and Related Work	13
2.1 Introduction	13
2.2 Background	14
2.3 Related Work	17
2.4 Synthesis, Gap Analysis	24
Chapter 3: Methodology	26
3.1 Introduction	26
3.2 Dataset	26
3.3 Data Preprocessing Pipeline	28
3.4 Data Augmentation Strategy	28
3.5 Train, Validation, and Test Split Protocol	29
3.6 Proposed Deep Learning Architecture	29
3.6.1 Classification Models	29
3.6.2 Convolutional Block Attention Module (CBAM)	32
3.6.3 Classification Head	33
3.7 Training Protocol	34
3.8 Evaluation Metrics	35
Chapter 4: Results and Discussion	37
4.1 Preprocessing Results	37
4.2 Results of Classification Techniques	38
4.2.1 ResNet-50	38
4.2.2 EfficientNet B4	43

4.2.3	MobileNetV3	45
4.3	Comparison between Proposed Models	47
4.4	Overall Comparative Analysis.....	48
4.5	Explainability using (Grad-CAM)	50
4.6	Discussion.....	51
Chapter 5: Conclusion and Future Work.....		53
5.1	Conclusion	53
5.2	Future Work	54
References		55

List of Abbreviations

- **AdamW:** Adam Optimizer with Weight Decay.
- **AI:** Artificial Intelligence.
- **ASIR:** Age Standardized Incidence Rate.
- **AUC:** Area Under the Curve.
- **CBAM:** Convolutional Block Attention Module.
- **CCT:** Compact Convolutional Transformer.
- **CLAHE:** Contrast Limited Adaptive Histogram Equalization.
- **CNN:** Convolutional Neural Network.
- **CT:** Computed Tomography.
- **DenseNet:** Densely Network.
- **DL:** Deep Learning.
- **DNN:** Deep Neural Network.
- **EANet:** External Attention Network.
- **FC:** Fully Connected Layer.
- **GAP:** Global Average Pooling.
- **Grad-CAM:** Gradient-weighted Class Activation Mapping.
- **LR:** Learning Rate.
- **MBConv:** Mobile Inverted Bottleneck Convolution.
- **ML:** Machine Learning.
- **MobileNetV3:** Mobile Network Version 3.
- **MRI:** Magnetic Resonance Imaging.
- **RCC:** Renal cell carcinoma.
- **ReLU:** Rectified Linear Unit.
- **ResNet50:** Residual Network (50 layers).
- **ROC:** Receiver Operating Characteristic.
- **U-Net:** U shaped Network.
- **VGG:** Visual Geometry Group Network.

List of Figures

Figure 2.1: samples CT KIDNEY Images per class from DATASET	13
Figure 2.2: AI overview	14
Figure 2.3: CNN architecture	15
Figure 2.4: Architecture of DenseNet121 model www.researchgate.net	16
Figure 2.5: Architecture of the U-Net www.researchgate.net	16
Figure 3.1: Proposed Kidney CT Image Classification Framework.....	26
Figure 3.2: Class Distribution Per-Class Visualization.....	27
Figure 3.3: EfficientNet-B4 Architecture illustrating	30
Figure 3.4: ResNet50 Architecture illustrating	31
Figure 3.5: CBAM Attention Refinement Module	32
Figure 4.1: Sample Images before and after preprocessing	37
Figure 4.2: Sample Images before and after CLAHE	38
Figure 4.3: Per Class Classifications Metrics	40
Figure 4.4: Confusion Matrix for ResNet50 Model.....	41
Figure 4.5: Training History for ResNet50 Model	41
Figure 4.6: Per-class ROC curves for the ResNet-50 and CBAM model.	42
Figure 4.7: Confusion matrix for the EfficientNet-B4 and CBAM model	44
Figure 4.8: Training history for the EfficientNet-B4 and CBAM model.....	44
Figure 4.9: Per-class ROC curves for the EfficientNet-B4 and CBAM model.....	45
Figure 4.10: Confusion matrix for the MobileNetV3 and CBAM model	46
Figure 4.11: Training history for the MobileNetV4 with CBAM model.	46
Figure 4.12: Per-class ROC curves for the MobileNetV3 with CBAM model.	47
Figure 4.13: Generating Grad-CAM for all classes correctly classified	51

List of Tables

Table 2.1: Summary of reviewed studies on deep learning for kidney CT classification.....	24
Table 3.1: Class distribution of the dataset	27
Table 3.2: Data augmentation operations applied to the training set.....	28
Table 3.3: Image count in each subsets.....	29
Table 3.4: Training hyperparameters and justification for each design choice.	34
Table 4.1: Per-Class Classification Metrics for the ResNet-50 and CBAM	39
Table 4.2: Per-class classification metrics for EfficientNet-B4 and CBAM.....	43
Table 4.3: Per-class classification metrics for MobileNetV3 and CBAM.	45
Table 4.4: Comparison of Deep Learning Models for Kidney CT Classification	48
Table 4.5: Comparative Analysis with related classification work	49

Chapter 1: Introduction and Overview

1.1 Overview

A significant health of 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. It has subtypes like the transitional cell carcinoma, Wilms tumor in children and benign ones which include the oncocytomas and the angiomyolipoma [1]. 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 [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]. 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 [2]. 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.

1.2 Problem Definition

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 is 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 (e.g. clear cell RCC versus papillary or chromophobe) is essential to prognosis and treatment. The most aggressive subtype (clear cell RCC) 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 these 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.

1.3 Thesis 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. In order to develop multi-class classification functionality, to be able to differentiate between benign (e.g., oncocytoma, angiomyolipoma) and malignant subtypes (e.g., clear cell, papillary, chromophobe RCC), to underline the clinical significance of subtype-specific prognosis.
4. To assess performance of framework in comparison to manual radiologist ratings 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.

Through the accomplishment of these aims, the thesis aims to make contributions to the domain of AI-driven oncology and eventually positively affect the patient outcomes in the management of kidney tumors.

1.4 Thesis Contribution

This thesis contributes to the field of medical imaging and AI in kidney tumor diagnosis as a following:

1. Deep learning model that detect and simultaneously classify the objects into multi classes with high rates of accuracy on CT datasets.
2. Enhance subtype differentiation and reduce false positives in the benign-malignant classification.
3. Multi-center validation, which is resistant to differences in imaging regimes and patient demographics, which is a common limitation in earlier studies.
4. Possible impact on the real world by providing such of tool that could decrease the workload on radiologists, decrease unnecessary interventions, and helps the precision medicine in management of kidney cancer.

Such a contributions, facilitate integration for deep learning in the field of oncology, offering a universal solution to the global challenges encountered in healthcare.

1.5 Thesis Organization

The thesis is organized in the following way:

- Chapter 2 is a review of background literature about kidney tumors, imaging modalities, and applications of DL in medical imaging.
- Chapter 3 describes the suggested DL framework, its architecture, preprocessing of data, and training strategies.
- Chapter 4 is the results and discussions of the experiment, performance analysis, and the comparison with the benchmark.
- Chapter 5 ends with conclusion, limitations and future directions of work.

This organization facilitates a rational flow of the initial ideas to the application and analysis of the ideas.

Chapter 2: Background and Related Work

2.1 Introduction

The chapter provides the basic background of automated renal diagnosis. It introduces with clinical and pathological aspects of target abnormalities and physical principles of CT imaging. This discussion follows the history of computational diagnostics starting with the traditional machine learning that is rigid, and moving to modern deep learning structures. Lastly, the chapter reviews the fifteen latest studies, most of which are run against the same Kaggle dataset, namely, CT KIDNEY DATASET: Normal, Cyst, Tumor, and Stone (Figure 2.1). The comparative analysis of how various research groups were able to utilize the same data to perform multi-class classification allows us to easily identify the strong points of the methodology, the weaknesses, and the scientific and clinical need that this thesis fulfills.

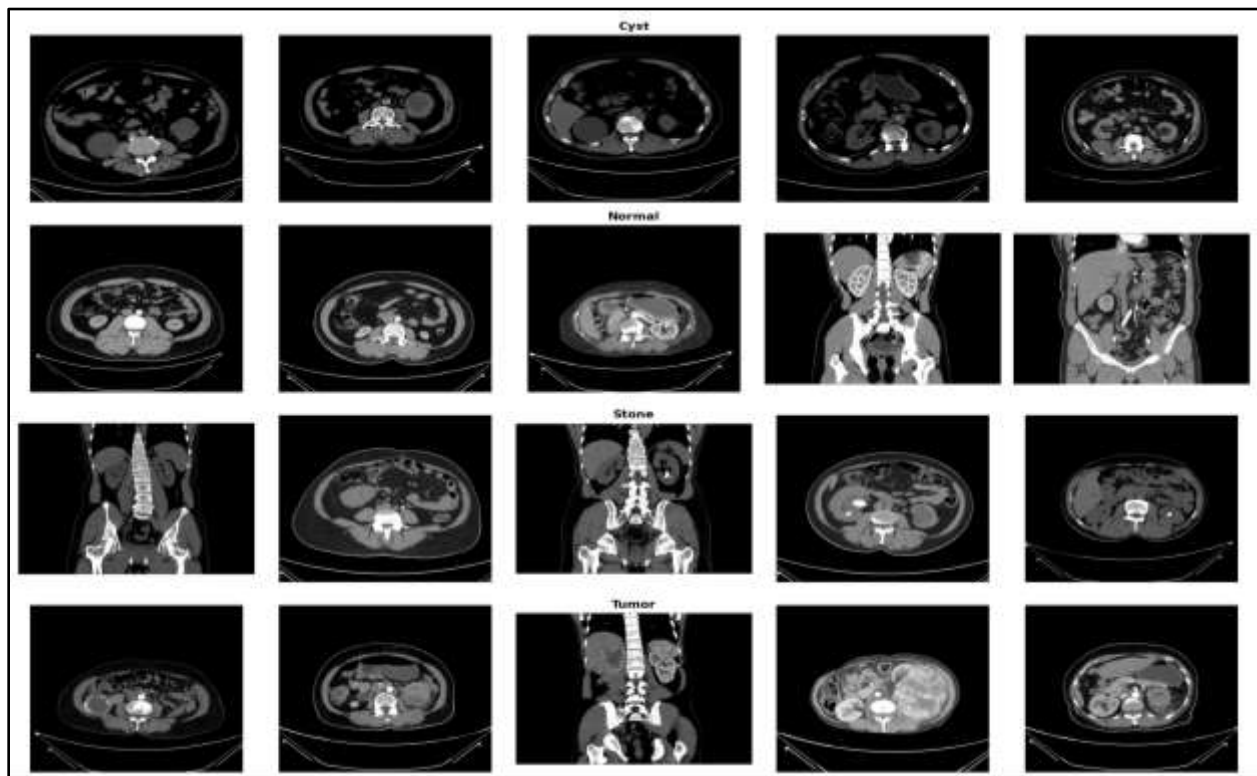


Figure 2.1: samples CT KIDNEY Images per class from DATASET

2.2 Background

Kidneys are complex retroperitoneal glands that have a role of filtering blood, eliminating metabolic waste, and homeostasis of fluids and electrolytes. In the case of structural abnormalities developing, they usually belong to a number of different categories as shown in above Figure 2.1. As the benchmark dataset and the current study suggest, four clinically relevant states are of interest, which include: Normal, Cyst, Stone, and Tumor [8].

Although ultrasound provides screening with radiation free and MRI with best contrast of soft-tissues, Computed Tomography has become the backbone of kidney tumor diagnosis because of its level of spatial resolution, fast acquisition, and characterization of masses [7].

Artificial Intelligence is an engineering which focuses on developing solutions that able to simulate human intelligence and it's falling under computer since, Where Machine Learning is a branch of artificial intelligence defined as a science of programming computers so they can learn from data and gives computers the ability to learn without being explicitly programmed [25].

Early computer-aided diagnostic systems were based on a classical machine learning where features were constructed manually. Researchers established definite regulations to identify shapes, textures (e.g., through Gray Level Co-occurrence Matrix), or certain Hounsfield thresholds. These hard techniques were fragile, and could not tolerate biological variation- tumors hardly ever develop ideal geometries and lesions have broad morphological heterogeneity [9].

Deep Learning (DL) is a branch from artificial intelligence and a subset of machine learning (Figure2.3), it's based on multi-layered neural networks to learn patterns by automatically analyzing medical image data, including CT scans. Deep learning can enhance a kidney cancer diagnosis with higher accuracy by revealing disguised tumor features not necessarily visible to a human being. It facilitates the detection, classification and risk evaluation of renal tumors at early stage which helps to improve clinical outcomes. The rising number of cancer cases of the kidneys in the world has found deep learning critical towards helping to increase efficiency in the diagnosis process and

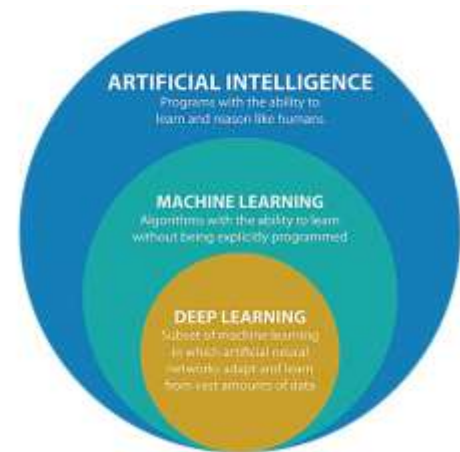


Figure 2.2: AI overview

subsequently lower the number of deaths [2]. The breakthrough was gained with the introduction of deep learning, especially Convolutional Neural Networks (CNNs).

Convolutional Neural Networks (CNNs) are a popular method of kidney tumor detection because this type of neural networks can extract spatial features of medical images. They include convolutional, pooling and fully connected layers (Figure 2.3) which detect features like shape and texture of tumors. CNN based models have been proven to be very effective in identifying renal abnormalities, such as tumors, cysts, and stones [9]. They are a core method of automated kidney disease diagnosis systems because their effectiveness in medical imaging. Within a CNN, the initial convolutional layers process gradients and edges and other basic components with the help of filters. These are gradually absorbed into higher-level representations, e.g. clustered tumor textures or the edges of kidney stones. Pooling reduces the number of spatial dimensions to lessen the computational load but maintain the information of significance. Last fully connected layers generate probability distributions of classes [21].

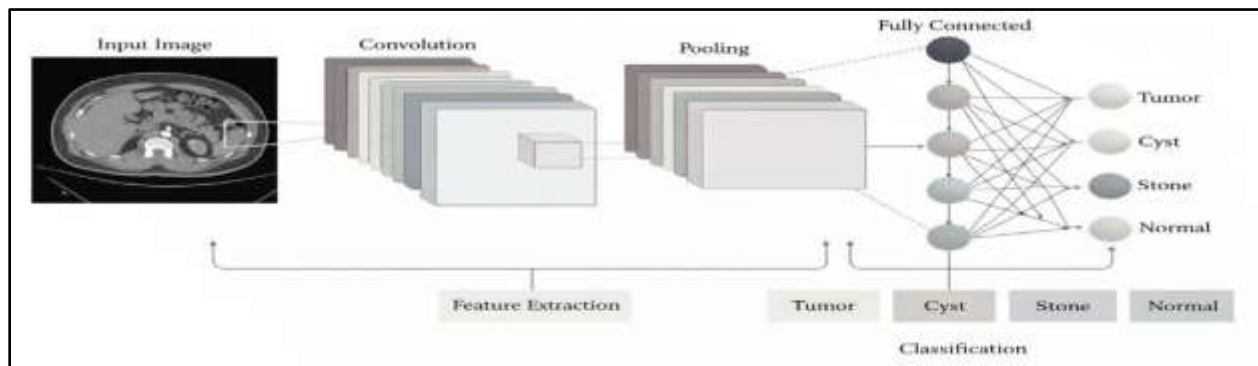


Figure 2.3: CNN architecture

DenseNet 121 is a neural network structure that enhances propagation of features, each layer linked with every other layer as shown in Figure 2.4. The structure decrease redundancy and increase efficiency in learning process. DenseNet-121 can be useful in kidney tumor classification because it plays an important role in generating detailed features of the CT images, which result a high degree of accuracy. It is often used in addition to other models for better performance in multi-class tumor classification tasks [18].

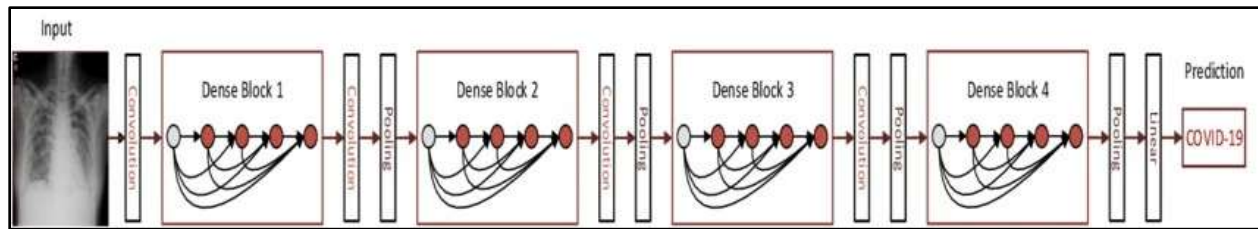


Figure 2.4: Architecture of DenseNet121 model www.researchgate.net

U-Net is a type of deep learning architecture, which is designed for a particular purpose to solve the issue of clinical images segmentation. It uses an encoder and decoder design (Figure 2.5) to precisely locate and marks kidneys tumor areas in CT images. This strict localization is indispensable to diagnosis and planning of treatment. Alternated U-Net architectures demonstrated better results in the segmentation of kidney structures and abnormalities in small datasets. The fact that they can produce high-quality segmentation maps also makes them very useful in the analysis of medical images [24].

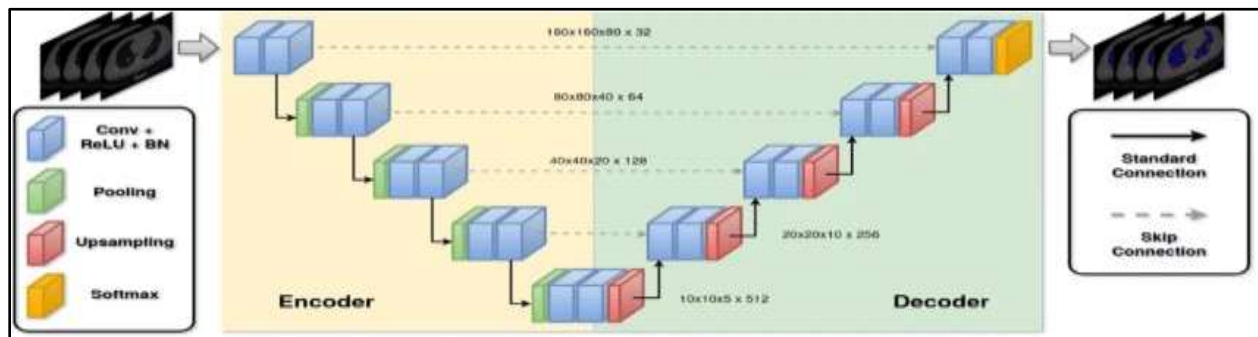


Figure 2.5: Architecture of the U-Net www.researchgate.net

Deep network training is associated with an enormous challenge, with severe class imbalance prevalent in medical data being the most notable. Naive training leads to models preferring majority classes (e.g., Normal) to reduce the total loss. Mitigation can be used with other deep learning techniques which can disproportionately punish errors of minority classes (stones and tumors) and encourage equal predictive accuracy in all categories.

2.3 Related Work

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 [1 to 15]. We can easily be able to pinpoint the strengths and weaknesses of methods used by comparing how various research papers used the same data to classify multi-class. The scientific and clinical gap to be filled in the present thesis is directly expressed through this analysis.

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].

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 dataset. The approach quickly converged on learned features and ResNet-50 skip connections of about 96% accuracy per classification, though ImageNet training on color 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 tumor 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 (tumor detection), the authors compared three 2D CNN architectures with a custom six layers CNN (CNN-6), ResNet50, and VGG16. During the second stage (tumor 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].

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].

Bhandari et al. (2023) came up with a lightweight, custom CNN model that classifies kidney CTs (four classes: Normal, Cyst, Stone and Tumor) on 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 same Kaggle datasets, 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].

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 binarisation thresholding used to improve kidney delineation and isolate stone areas to 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].

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 Kaggle datasets. 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 optimization, attention mechanisms, etc.) is needed [18].

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 Kaggle benchmark. 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].

Pimpalkar et al. (2025) performed a comparative study of four transfer learning architectures VGG16, ResNet50, CNNAlexNet, and InceptionV3 with hyperparameter optimization being applied to the Kaggle benchmark dataset (12,446 images: Normal, Cyst, Stone and Tumor). 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].

Obaid et al. (2025) on the Kaggle benchmark dataset (12,446 images: Normal, Tumor, Cyst 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].

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 Kaggle benchmark dataset. Preprocessing stage used reflection removal and contrast normalization 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].

Maçin et al. (2025) proposed KidneyNeXt, a custom lightweight CNN model that uses depthwise separable convolutions on the same dataset 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 is characterized by small hyperattenuating lesions that can be easily confused in calcifications in the surrounding tissue [9].

Khan et al. (2025) suggested a ConvNeXt and EfficientNetV2B0 dual-architecture transfer learning strategy on the Kaggle benchmark dataset 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 dataset of CT KIDNEY 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].

2.4 Synthesis, Gap Analysis

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 will give a systematic summary of all the fifteen studies reviewed in Section 2.3: Table 2.1 below gives a summary of the dataset, technique, the accuracy obtained, the venue of publication as well as whether the work will be used to derive a comparison with the proposed framework in this thesis.

Table 2.1: Summary of reviewed studies on deep learning for kidney CT classification.

Ref	Author	Year	Dataset	Technique Used	Accuracy
[9]	Maçın et al.	2025	Kaggle CT KIDNEY dataset	KidneyNeXt (depthwise separable CNN, 224×224)	99.96%
[11]	Islam et al.	2022	Kaggle CT KIDNEY dataset	VGG-16, ResNet-50, Swin Transformer (transfer learning)	96% (ResNet-50); 99.30% (Swin)
[12]	Alzu'Bi et al.	2022	Institutional (KAUH Jordan, 8,400 CT)	CNN-6, ResNet50, VGG16 detection; CNN-4 classification	92% classification
[13]	Yildirim et al.	2021	Private NCCT (1,799 CT, Turkey)	XResNet-50 + data augmentation	96.82%
[14]	Pimpalkar et al.	2025	Kaggle CT KIDNEY dataset	VGG16, ResNet50, InceptionV3 with hyperparameter optimisation	99.96% (InceptionV3)
[15]	Sabuncu et al.	2023	Clinical axial CT (Near East Univ. Hospital, Cyprus)	InceptionV3	98.52%
[16]	Obaid et al.	2025	Kaggle CT KIDNEY dataset	DarkNet53 + multi-model ensemble fusion	99.69%
[17]	Baygin et al.	2022	Private NCCT (1,799 CT, Turkey)	DarkNet19 + INCA feature selection + KNN (ExDark19)	99.22%-99.71%
[18]	Qian.	2025	Kaggle CT KIDNEY dataset	DenseNet121 (standard, no augmentation)	93.51%

Ref	Author	Year	Dataset	Technique Used	Accuracy
[19]	El-Hag et al.	2025	Kaggle CT KIDNEY dataset	EfficientNet-B2 + specular-free preprocessing	98.27%
[20]	Almuayqil et al.	2024	Kaggle CT KIDNEY dataset	KidneyNet (8-layer custom CNN)	99.88% / 99.92% precision
[21]	Bhandari et al.	2023	Kaggle CT KIDNEY dataset	Lightweight custom CNN + LIME + SHAP explainability	99.52%
[22]	Bingol et al.	2023	Kaggle CT KIDNEY dataset	29-layer CNN + Relief feature selection	99.37%
[23]	Khan et al.	2025	Kaggle CT KIDNEY dataset	ConvNeXt + EfficientNetV2B0 + selective data trimming	99.73% (6.2M params)
[24]	Murillobouzon et al.	2025	KSSD2025 (838 annotated abdomen CT)	U-Net pixel-level segmentation	97.06% (Dice score)

Chapter 3: Methodology

3.1 Introduction

This chapter outlines the methodology developed to meet the identified gaps from Chapter

2. The proposed architecture is composed of three main components:

- Model backbone is used for extracting the features.
- Convolutional Block Attention Module (CBAM).
- Custom classification head.

The proposed system is based on the ResNet-50 model as described in Figure 3.1 and compared to two other models: EfficientNet-B4 and MobileNetV3. The aim was to create an accurate, efficient, and interpretable deep learning model to classify kidney CT images into four classes: Normal, Cyst, Stone, and Tumor.

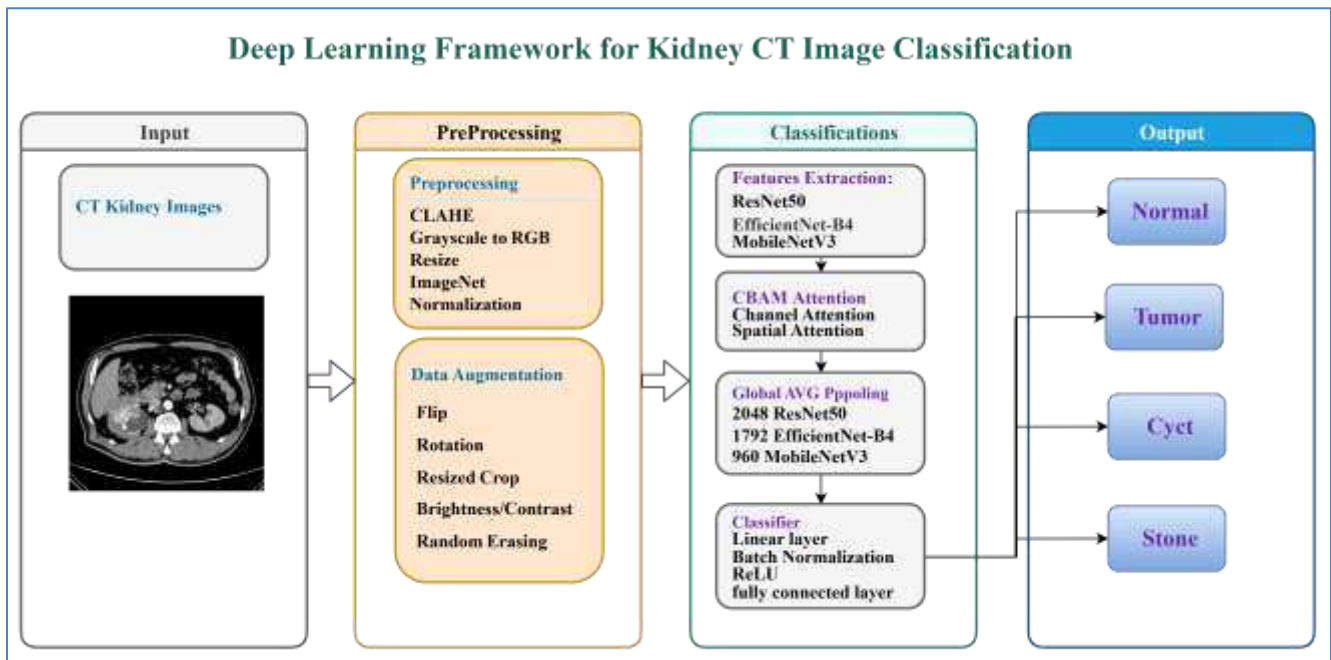


Figure 3.1: Proposed Kidney CT Image Classification Framework

3.2 Dataset

In this research, CT images are taken from Kaggle CT KIDNEY DATASET (Normal, Cyst, Tumor, and Stone) which contains 12,446 CT images in four classes as described in Table 3.1 and Figure 3.2.

Table 3.1: Class distribution of the dataset

Class	Number of Images	Percentage of Dataset	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

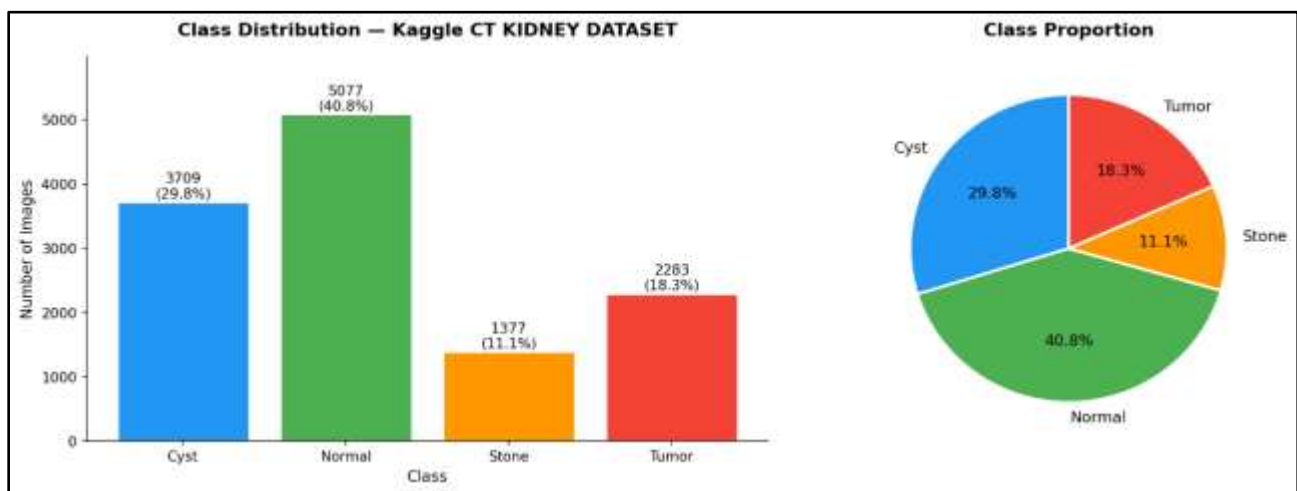


Figure 3.2: Class Distribution Per-Class Visualization

There is a huge imbalance of the classes in the dataset: the Normal is 40.8 percent of all images (5,077 samples), whereas the Stone is 11.1 percent (1,377 samples) only, which is about 3.7:1. This imbalance is typical in clinical practice of radiology. The dataset consists of both axial and coronal CT reconstruction images and has both contrast-enhanced and non-contrast studies, therefore. All the pictures are of JPEG format (Figure 3.3), and difference in original residential spatial space of the picture before standardized in preprocessing.

3.3 Data Preprocessing Pipeline

All images go through a standard four stages preprocessing pipeline before being fed into the model:

- **Grayscale to RGB Conversion:** CT images are single-channel. They are converted to three channels by duplicating the grayscale values so that pre-trained models like EfficientNet-B4 and ResNet-50 can be used.
- **CLAHE (Contrast Limited Adaptive Histogram Equalization):** Applied with clip limit = 2.0 and tile grid size = 8×8 to improve visibility of small lesions and stones Figure.
- **Resizing:** Images are resized to 380×380 pixels (matching EfficientNet-B4 input size) and 224×224 pixels (matching ResNet-50 input size) using bilinear interpolation.
- **Normalization:** Pixel intensities are normalized using ImageNet mean and standard deviation values.

3.4 Data Augmentation Strategy

The concept of data augmentation is implemented only to the training set in order to artificially increase diversity potential and decrease overfitting. Validation or test sets are not augmented in any way but only undergo preprocessing transformations as outlined in above section 3.3. All the augmentations used are in Table 3.2.

Table 3.2: Data augmentation operations applied to the training set

Augmentation Technique	Parameters
Random horizontal flip	p = 0.5
Random rotation	±15°
Random brightness / contrast jitter	factor 0.2
Random resized crop	scale 0.85–1.0
Random erasing (Cutout)	p = 0.1, scale 0.02–0.1

3.5 Train, Validation, and Test Split Protocol

The 12446 images are further divided into training, validation and test subsets with stratified split to preserve the class distribution across three subsets. The division of the ratio will be between 70/10/20 in which the total number of images they arrive at 8,712 training images, 1,244 validation images and 2,490 test images Table 3.3.

Table 3.3: Image count in each subsets

Subset	Images counts
Training set	8,712 images
Validation set	1,244 images
Test set	2,490 images
Train Split class distribution	Cyst: 2596, Normal: 3554, 'Stone: 964, Tumor: 1598
Validation Split class distribution	Cyst: 371, Normal: 507, Stone: 138, Tumor: 228
Test Split class distribution	Cyst: 742, Normal: 1016, Stone: 275, Tumor: 457

3.6 Proposed Deep Learning Architecture

3.6.1 Classification Models

The three deep learning models evaluated for kidney CT images classification are: EfficientNet-B4, ResNet-50 and MobileNetV3. The same preprocessing, data augmentation and classification head were used to ensure fair comparisons across all models.

ResNet-50 was chosen as the primary proposed model achieved an excellent overall accuracy on the test set.

A. EfficientNet-B4

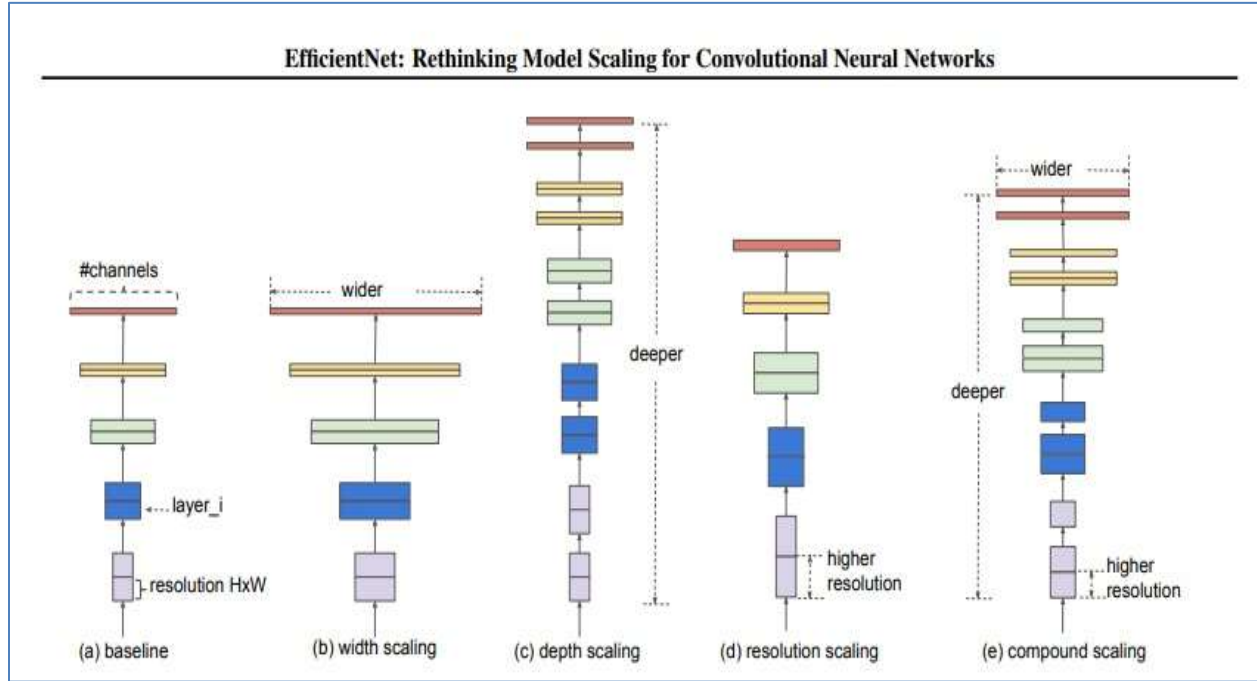


Figure 3.3: EfficientNet-B4 Architecture illustrating

The models of EfficientNet rely on the scaling approach of compounds that is founded on the single set of coefficient 1 akin to depth and width of networks, which scale together with input resolution (Figure 3.3). The equation of compound scaling (1) is following:

$$\text{depth: } d = \alpha^\varphi, \text{ resolution: } r = \gamma^\varphi, \text{ width: } w = \beta^\varphi \quad (1)$$

An EfficientNet-B4 model on 380 x 380 optimal input resolution and 4.2 billion floating-point operations per forward step is obtained in the instance of EfficientNet-B4, which allows it to compete with significantly lower computational cost than MobileNetV3, ResNet50.

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 [26].

B. ResNet-50 (Strong Comparative Model)

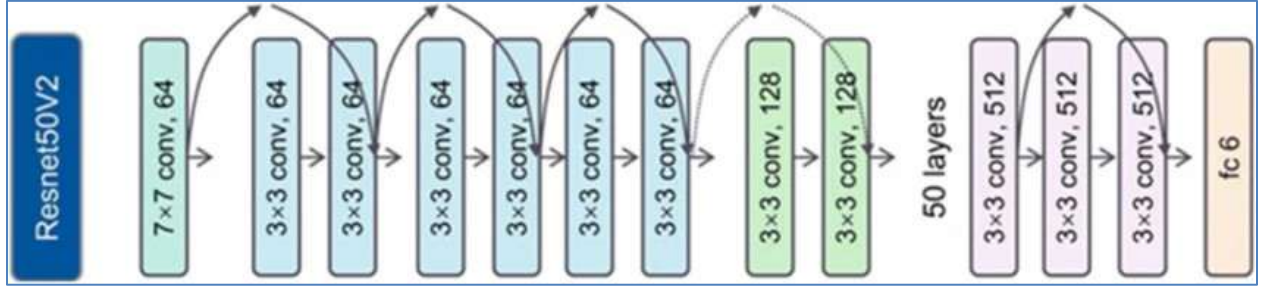


Figure 3.4: ResNet50 Architecture illustrating

ResNet-50 is a very deep residual network, it has 50 layers of bottlenecks with 1×1 , 3×3 , and 1×1 convolutions (Figure 3.4). 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 $H(x)$ as in following equation (2) :

$$H(x) = F(x) + x \quad (2)$$

ResNet-50 was pre-trained on ImageNet-1K and fine-tuned on the kidney CT data set with input resolution of 224×224 . Early layers (up to stage 3) were 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 was chosen as the main proposed model for clinical deployment [28].

C. MobileNet V3

MobileNetV3 is an extremely efficient architecture for mobile and embedded applications. It employs hard-swish activation functions, squeeze-and-excitation (3) blocks and depth wise separable convolutions (4) to lower the computational cost while retaining high accuracy.

$$s = \sigma(W2 \delta(W1 \cdot GAP(x))) \quad (3)$$

$$\{Cost\}_{\{depthwise\}} = D_k^2 \cdot M \cdot D_f^2 + M \cdot N \cdot D_f^2 \quad (4)$$

The model was pre-trained 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.

3.6.2 Convolutional Block Attention Module (CBAM)

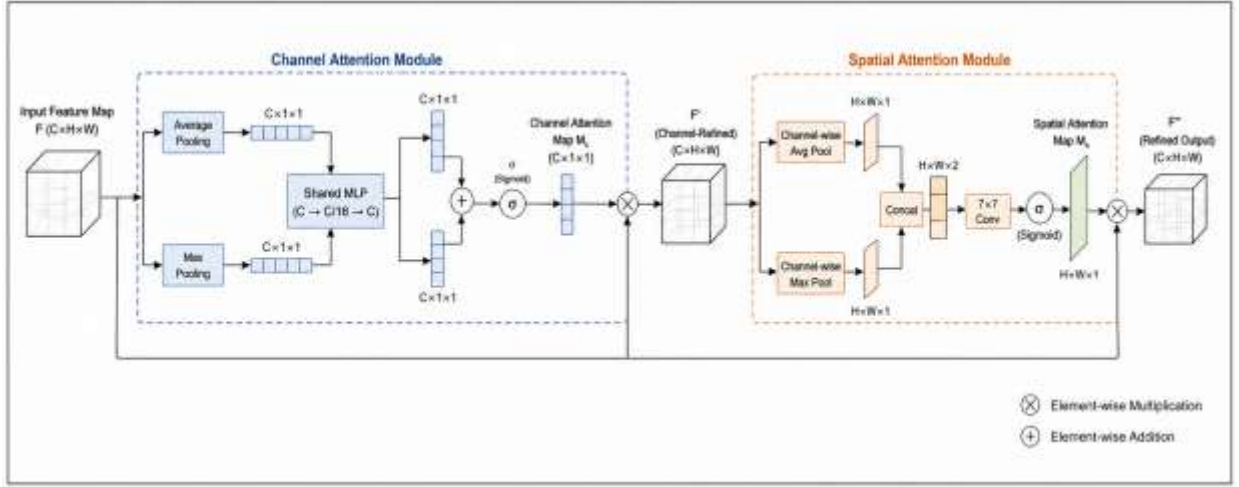


Figure 3.5: CBAM Attention Refinement Module

A CBAM is incorporated between the convolutional output and global averaging pooling layer [27]. CBAM applies two consecutive attention tasks of channel attention and subsequently spatial attention of the now generated feature map (Figure 3.5), which has undergone EfficientNet-B4 performing backbone on 380 x 380 inputs, while ResNet50 and MobileNetV3 on 224 x 224 inputs.

Channel Attention: Considering the feature map $F \in \mathbb{R}^{C \times H \times W}$, the channel attention will produce a 1-dimensional feature map $M_c \in \mathbb{R}^{C \times 1 \times 1}$, which represents the significance of every channel of features. The two spatial dimensions are fed independently on the global average pooling and the global max pooling, resulting in two C -dimensional vectors which are independently fed into a single two layer multilayer perceptron (MLP) with a reduction ratio of $r = 16$, which then has a resultant representation size of $C/16 = 112$ then projected back to: ($C = 1792$, EfficientNet, $C = 2048$ for ResNet-50, and $C = 960$ for MobileNetV3). The channel attention map is obtained with the help of the two MLPs and the output of one of them by a sigmoid activation in following equation (5):

$$M_{c(F)} = \sigma \left(MLP(AvgPool(F)) + MLP(MaxPool(F)) \right) \quad (5)$$

Spatial Attention: The refined channel based feature map F' is once more conveyed to the spatial attention sub-module. The channel dimension is averaged and maximized to come up with

two $H \times W$ Descriptor maps, which are combined together and convoluted with a 7×7 convolutional filter to come up with the spatial attention map $M_s \in \mathbb{R}^{(H \times W)}$ as following (6):

$$M_s(F') = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}_c(F'); \text{MaxPool}_c(F')])) \quad (6)$$

The final CBAM output is (7):

$$F'' = M_{s(F')} \otimes F' \quad (7)$$

A serial channel-then-space attention refinement process is employed to delineate the serial channels of the irrelevant background tissue, and emphasis the attention on the space components of the CT image that elicit the most discriminative knowledge on the classification of kidney pathology the region out of interest (ROI) shared by a CT scan as examined by a radiologist. Despite that CBAM is adding small parameters (less than 1% to 2%) it remains lightweight and end to end trainable just like the rest of the network [27].

3.6.3 Classification Head

After global average pooling is applied to the refined feature map, there is 2048 dimensional feature vector in the ResNet50 model that is passed through a custom classification head consists of the following sequential components:

- Dropout layer with a rate of 0.50.
- Fully connected layer reducing the feature vector from 2048 to 512 dimensions.
- Normalization to stabilize the activation distribution.
- ReLU.
- Second dropout layer with a rate of 0.30.
- Output: 4 classes logits (Normal, Cyst, Stone, Tumor).

3.7 Training Protocol

All the training hyperparameters and their reasoning are illustrated in Table 3.4.

Table 3.4: Training hyperparameters and justification for each design choice.

Hyperparameter	Value
Input image size for EfficientNet-B4	380×380 px
Input image size for ResNet50 and 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 epochs	5
Phase 2 epochs	Upto 8
Dropout (head, layer 1)	0.50
Dropout (head, layer 2)	0.30
Focal loss gamma (γ)	2.0
FC hidden dimension	512
Class weight strategy	Inverse frequency

The training process done in two phases. Phase 1 (5 to 10 epochs), only the custom classifier head and CBAM block were trained, while the whole ResNet-50 architecture was frozen. This phase provided adaptation of the new layers without changing any pretrained convolutional filters in ImageNet. In Phase 2 (up to 40 epochs), the last two residual layers of ResNet-50 (layer3 and

layer4) were unfrozen along with CBAM block and the classifier head, which allows joint fine-tuning at a smaller learning rate. Phase 2 started from a 5 epoch linear warm-up stage when the learning rate was smoothly increased up to its maximum value. AdamW optimizer with a weight-decay coefficient equal to of 1×10^{-5} was used. The early stopping approach was applied with respect to the validation loss and a patience of 8 epochs, in case of absence of improvement during 8 consecutive epochs, training was stopped, and the network was evaluated on the checkpoint with the lowest validation loss.

Also, during the training loss function and class imbalance was handled as following:

Class weight. The weight assigned to each class based on the frequency of occurrence of class in the training set as (8):

$$w_c = \frac{N_{total}}{(N_{classes} \times N_c)} \quad (8)$$

Focal loss. based on the class-weighted cross-entropy, the formula down weights of easy examples that are correctly identified, and focusing the training on hard examples that are incorrectly identified in following (9) equation:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (9)$$

3.8 Evaluation Metrics

The held-out test set (2,490 images) is used to evaluate the final model as a part of the evaluation approach, which includes accuracy, precision, recall, F1 score, Cohen's Kappa score, AUC-ROC (one-vs-rest) score and a confusion matrix to visualize class-wise performance in below:

Overall Accuracy: 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 (10) equation will be used:

$$Overall Accuracy = \frac{\sum_{i=1}^K Correct Predictions(i)}{\sum_{i=1}^K Total Sample(i)} \quad (10)$$

Precision: 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 (11) equation:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

Recall (Sensitivity): Recall or 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, it calculated via the following (12) equation.

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

Macro F1-Score: 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, it calculated via the following (13) equation:

$$Macro\ F1 = \frac{1}{K} \sum_{i=1}^K F1(i) \quad (13)$$

Where each class F1 is:

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

Weighted F1-Score: 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 (14).

$$Weighted\ F1 = \sum_{i=1}^K w(i).F1(i) \quad (14)$$

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

Chapter 4: Results and Discussion

4.1 Preprocessing Results

Preprocessing is essential step to ensure that input data are corrected and efficient to be fed into deep learning system, it prepares the raw CT images to be clean, consistent, and reliable before fed into the model. As in below Figure 4.1 and Figure 4.2 that visualize sample images before and after applying preprocessing mechanize used in proposed framework:

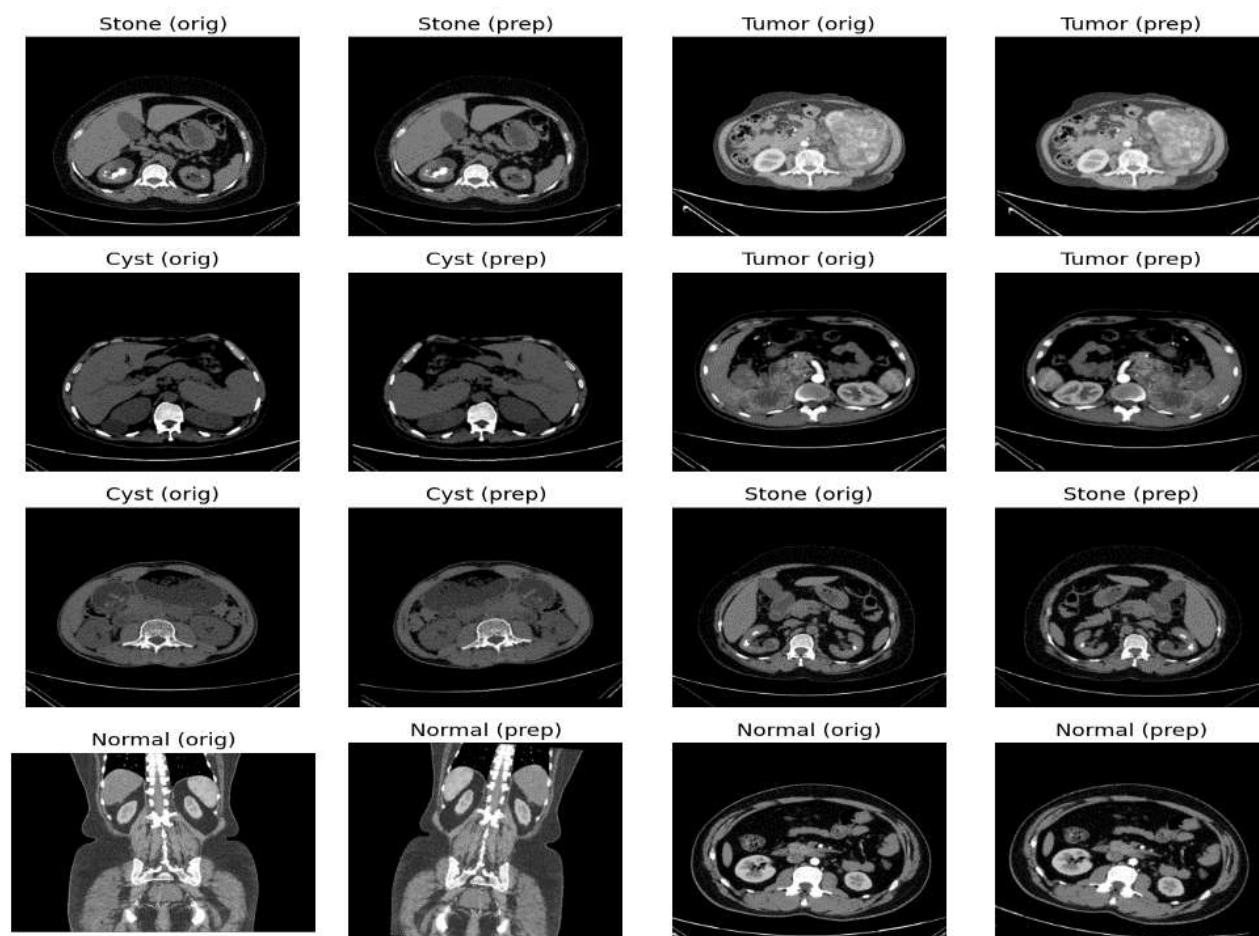


Figure 4.1: Sample Images before and after preprocessing



Figure 4.2: Sample Images before and after CLAHE

4.2 Results of Classification Techniques

In the designed deep learning architecture, the performance of EfficientNet-B4, MobileNetV3 and ResNet-50 along with CBAM attention block is assessed for classifying the kidney CT images. The reason behind selecting these models is that they are good at extracting features and have been proved efficient for the classification of medical images. For EfficientNet-B4, compound scaling was employed, which helps in balancing the depth, width and resolution of the network and hence makes it effective in terms of computation. MobileNetV3-Large has been selected for its light weight nature along with its architecture having depthwise separable convolution and squeeze-and-excitation blocks. Finally, ResNet-50 along with CBAM has been selected because of the stability offered by residual learning and also because CBAM is useful in focusing on clinically important areas of CT scan.

4.2.1 ResNet-50

The proposed ResNet-50 architecture with CBAM and trained with class weighted focal loss ($\gamma = 2.0$) resulted in excellent performance on the test set. As shown in Table 4.1. Model's accuracy was 99.88% when tested, showing it to be very well generalizable across all 4 kidney classes.

A. Per Class Classifications Metrics: The performance of the proposed Kidney-Net model for the four categories of kidney disease (Cyst, Normal, Stone and Tumor) is detailed in the per-class evaluation of metrics presented in Table 4.1 and visualized in Figure 4.3. These include measures such as precision, recall (sensitivity), F1-score and support. In the overall classification results, the precision, recall and F1 scores were essentially high for all categories with almost perfect classification results, which shows the strong balance between sensitivity and specificity. The Tumor class had a

perfect performance, with a precision of 1.0000, a recall of 1.0000 and an F1 score of 1.0000 because all the tumors cases were correctly identified, with no false positive or false negative. The Normal class also performed well, with a 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. Again, the Stone class did well with a precision of 1.0000 and a recall of 0.9964. The F1-scores for all classes are still very high (0.9982 – 1.0000), indicating a good balance of precision and recall. The support values also indicate that the distribution of the dataset was used in a realistic and imbalanced way for the evaluation, so that the performance is not skewed towards the majority classes. Generally, these results confirm the high, stable and clinically accurate classification results obtained by the proposed ResNet-50 + CBAM framework, while causing the least misclassification for all the kidney disease categories.

Table 4.1: Per-Class Classification Metrics for the ResNet-50 and 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

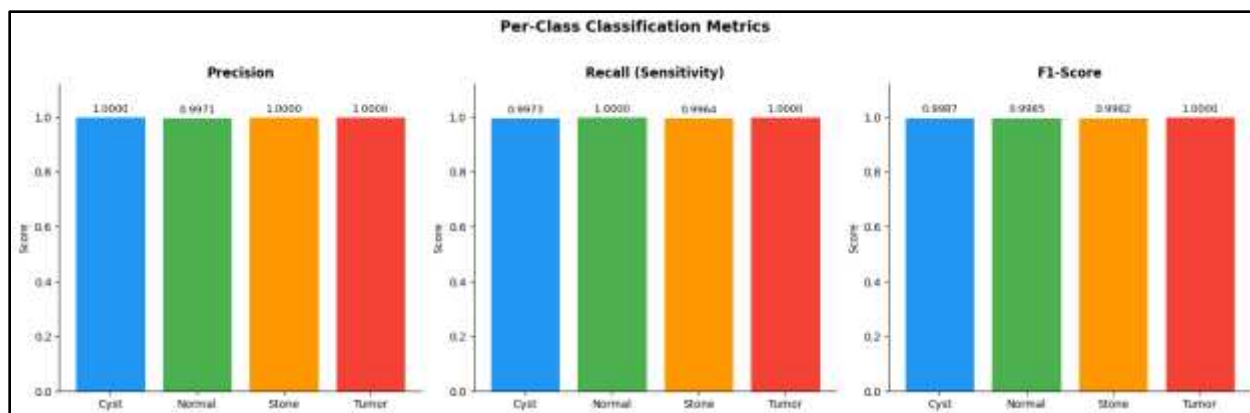


Figure 4.3: Per Class Classifications Metrics

B. Confusion Matrix: The confusion matrix in Figure 4.4 represents the classification accuracy of (ResNet-50 + CBAM) model for the held out test set expressed in absolute terms as well as normalized. The outcomes show that the predictive performance is very accurate and well balanced for all four kidney classes (Cyst, Normal, Stone, and Tumor).

The model was able to correctly classify most of the test samples in the count-based confusion matrix. There was no misclassification in Normal class with 1,016 predictions made correctly. Likewise, the Tumor class classification had also a perfect result with 100% of the data correctly classified, showing good discriminative capability of the malignant cases, as well. Within the Cyst class, 740 samples were correctly classified and 2 samples were misclassified as Normal (very little confusion between benign renal tissues). The Stone class had 274 correct answers, and just 1 answer was misclassified as Normal, meaning that the accuracy of the detection of the Stone class was high, but with some interference from similarity with the other classes, notably low intensity regions in CT scans. These results are further supported by the normalized confusion matrix which demonstrates almost perfect diagonal dominance, with almost 1 recall value for Cyst, Normal, Stone and Tumor respectively. The results show that the model is very sensitive in all categories and the loss in classification between classes is negligible. Overall, the confusion matrix shows a high level of robustness of the proposed architecture, excellent class separability, and little misclassification, especially for clinically relevant tumor detection (no false negative in the test set).

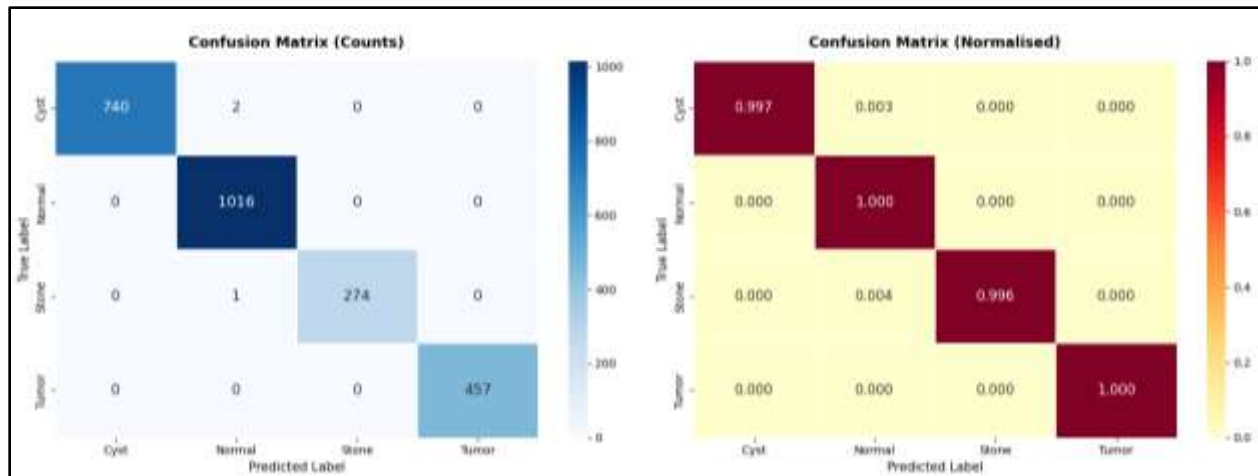


Figure 4.4: Confusion Matrix for ResNet50 Model

C. Training Curves As shown in below from training history plots Figure 4.5, the Kidney-Net network shows well defined convergence characteristics over both training and validation curves. As seen in the graph depicting focal loss, there is a clear and constant reduction in loss during training and validation phase, with little divergence between the two, suggesting that the model learns efficiently without overfitting. In addition, the plot of accuracy over the epochs clearly shows an increasing trend over time with little difference in training and validation accuracy, thus depicting good generalization of the model on unseen data. The plot of learning rate schedule reinforces this phenomenon, demonstrating a gradual warming up and decay phase of learning. Through below graphs, it can be seen that the Kidney-Net network achieves balanced learning characteristics in both phases of training.

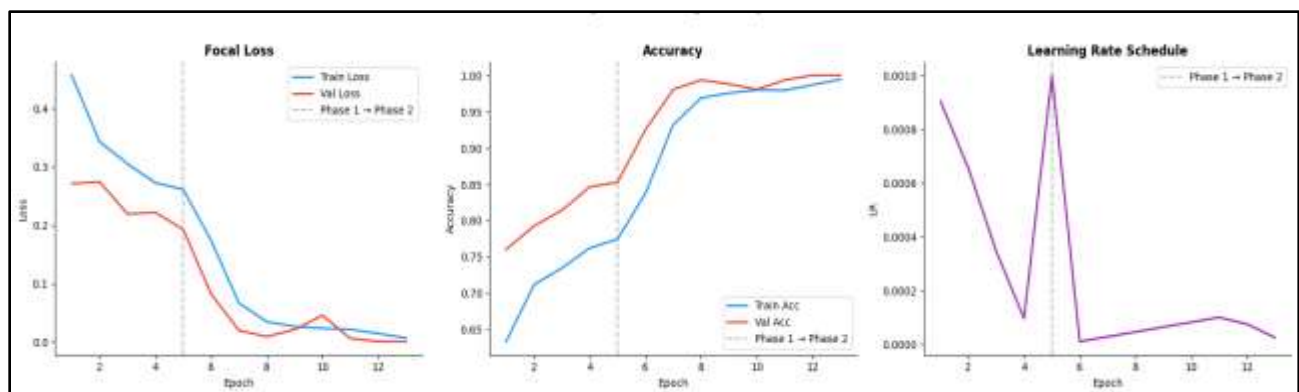


Figure 4.5: Training History for ResNet50 Model

D. ROC Curves Per Class: A Receiver Operating Characteristic (ROC) curve that is depicted. To demonstrate the diagnostic performance of the proposed model (ResNet-50 + CBAM), the classification approach is adopted as one-vs-rest for each kidney disease classification: Cyst, Normal, Stone, and Tumor. It is clearly seen from the results that there is a perfect separation of all the classes as the value of AUC is almost 1.0000, meaning it has a near-ideal classification performance. All four ROC curves are close to the top left corner of the graph, indicating a very high sensitivity (true positive rate) and a very low false positive rate for all decision thresholds. The Tumor class ROC curve is of special interest because it has no overlap with any other class and is a perfect discriminator, meaning that the model is very useful to distinguish tumors from other classes. The same is true of the Normal class, which is also separated perfectly, thus showing that the healthy kidney is always separated from the pathological cases. The Cyst and Stone classes are usually overlapping and more difficult to differentiate radiographically, and also yield perfect AUCs, demonstrating that the CBAM-based feature refinement has a significant impact on facilitating the model's ability to pick out subtle variations in texture and intensity within CT images. Overall, it was found that the proposed model can classify the different classes of kidney tumors with excellent class separability, high sensitivity, and very small misclassification rates, making it a very suitable model for the clinical diagnostic support of kidney tumor classification.

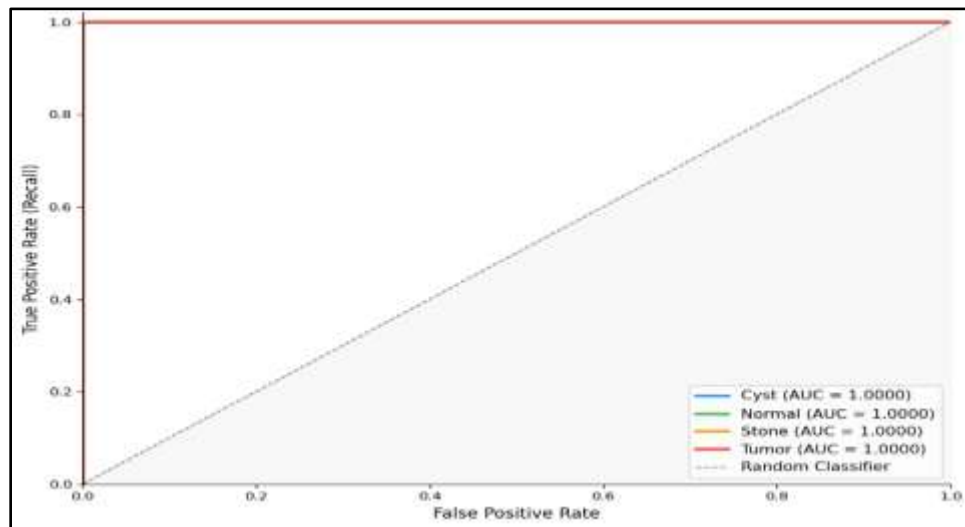


Figure 4.6: Per-class ROC curves for the ResNet-50 and CBAM model.

4.2.2 EfficientNet B4

A. 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 4.2 shows per class metrics for EfficientNet-B4 with CBAM.

Table 4.2: Per-class classification metrics for EfficientNet-B4 and 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

B. The confusion matrix: for the EfficientNet-B4 and CBAM, in Figure 4.7 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.

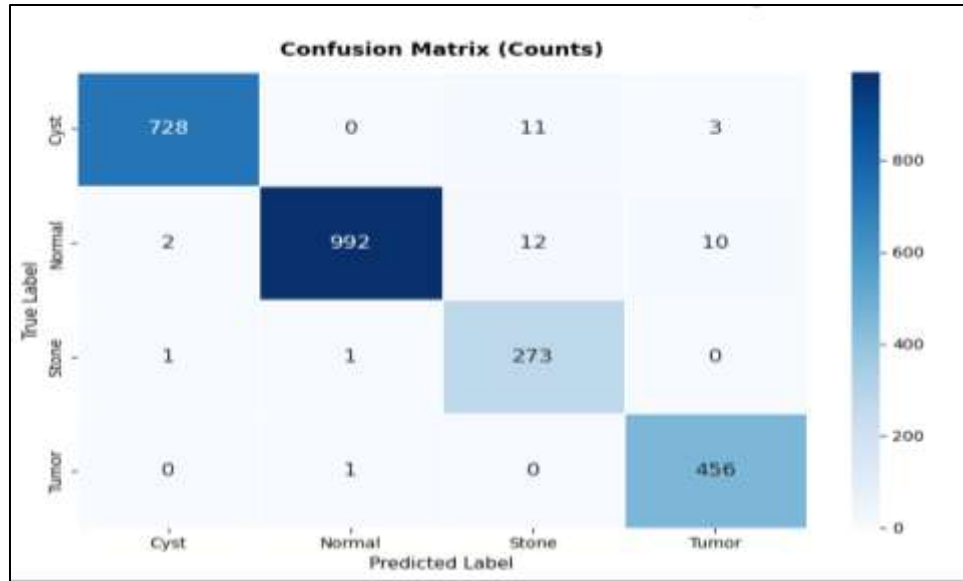


Figure 4.7: Confusion matrix for the EfficientNet-B4 and CBAM model

C. Training curves: Figure 13 showed decreasing focal loss and increasing accuracy across both training and validation sets with minimal divergence, indicating efficient learning without overfitting.

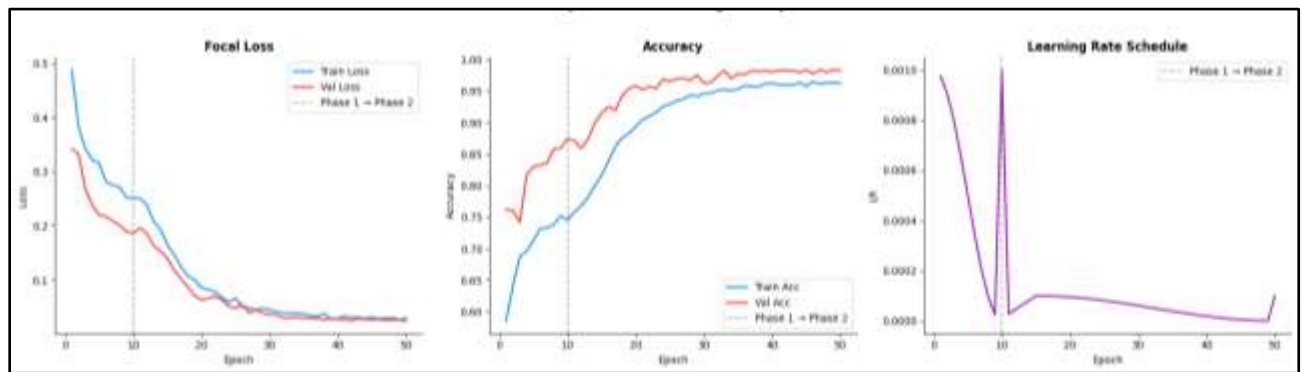


Figure 4.8: Training history for the EfficientNet-B4 and CBAM model.

D. Per-class one vs rest ROC: analysis Figure 4.9 yielded near perfect AUC values above 0.9993 for all four classes showing excellent separability and reliable performance.

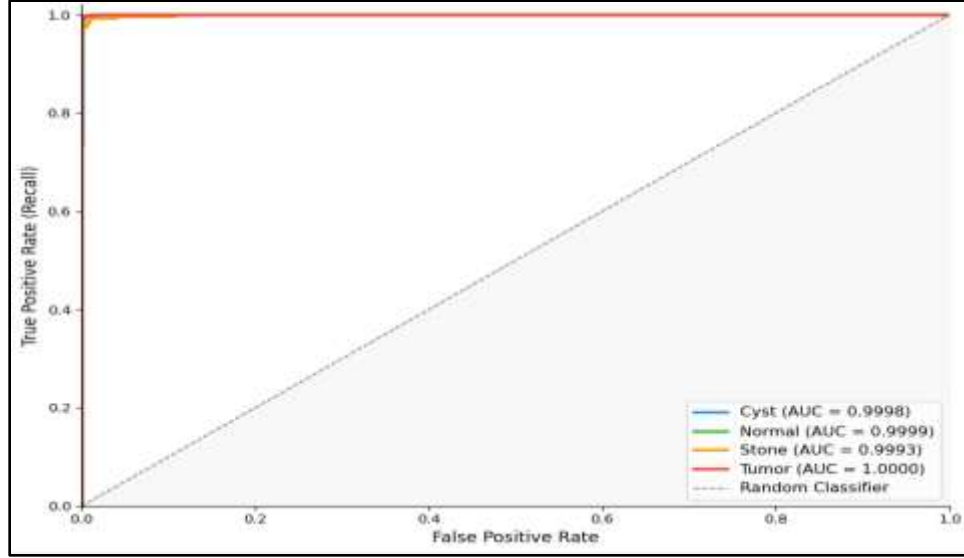


Figure 4.9: Per-class ROC curves for the EfficientNet-B4 and CBAM model.

4.2.3 MobileNetV3

A. 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 4.3 shows per-class evaluation metrics in details.

Table 4.3: Per-class classification metrics for MobileNetV3 and 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

B. The confusion matrix: for MobileNetV3 and CBAM, Figure 4.10 demonstrates high reliability of the model high numbers of classifications for Cyst (732), Normal (999), Stone (274), and Tumor (456).

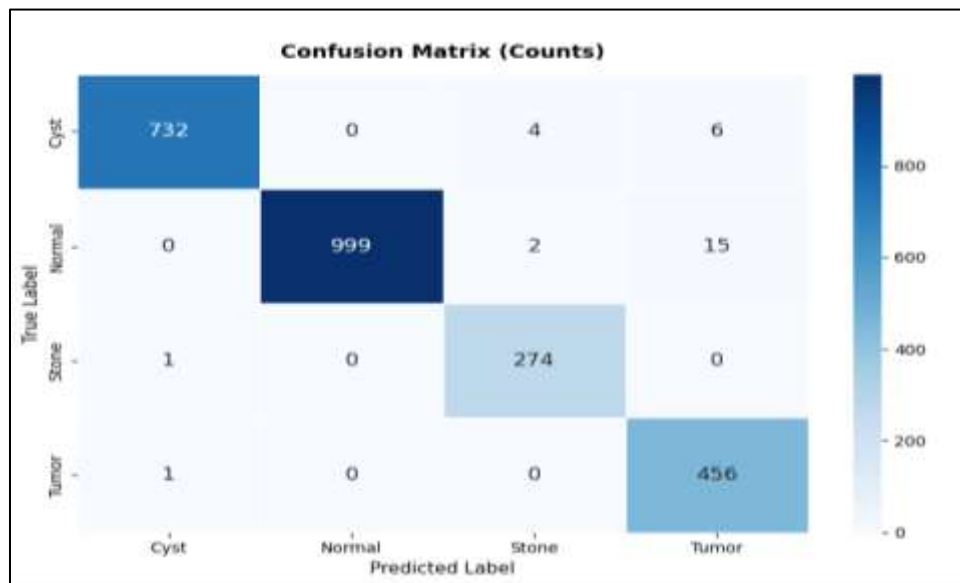


Figure 4.10: Confusion matrix for the MobileNetV3 and CBAM model

C. Training curves in Figure 4.11 indicating efficient learning without overfitting.

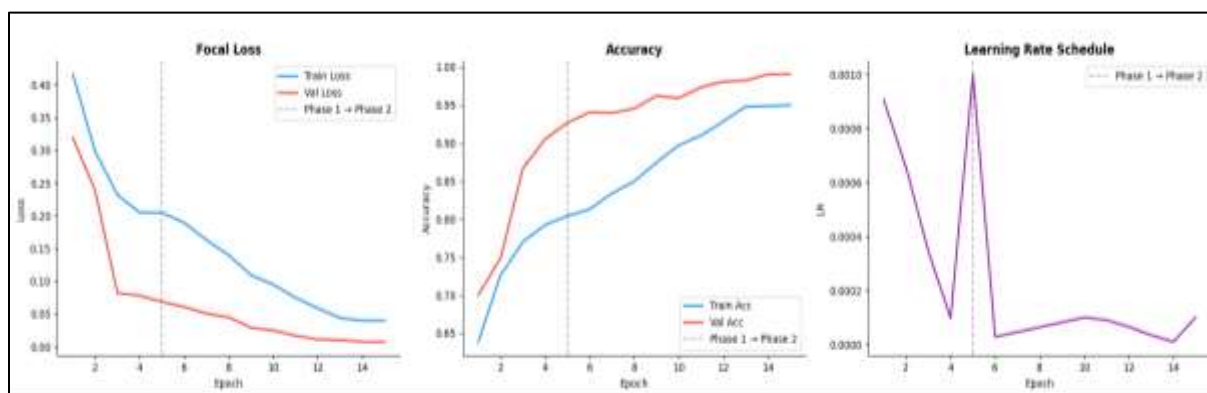


Figure 4.11: Training history for the MobileNetV4 with CBAM model.

D. Per-class one vs rest ROC: analysis for Figure 4.12 yielded near perfect AUC values above 0.9994 for all four classes showing excellent separability and reliable performance.

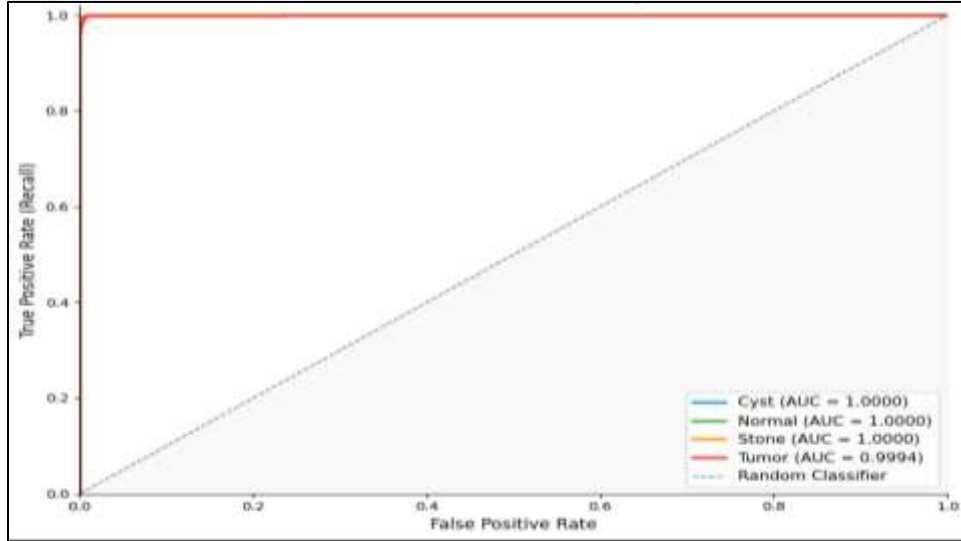


Figure 4.12: Per-class ROC curves for the MobileNetV3 with CBAM model.

4.3 Comparison between Proposed Models

The comparative study for the three deep learning models (ResNet-50, EfficientNet-B4, and MobileNetV3) is presented in Table 4.4 based on various evaluation metrics such as accuracy, computational efficiency, feature extraction, and clinical deployment. The proposed ResNet-50 model results in the highest test accuracy of 99.88%, which was better than EfficientNet-B4 (98.35%) and MobileNetV3 (98.84%). The results show that the proposed approach of residual learning and CBAM attention has a great effect in the discriminative feature learning of kidney CT image classification. In terms of input resolution, EfficientNet-B4 has a higher resolution of 380×380 , which can capture more detailed spatial information, whereas ResNet-50 and MobileNetV3 are based on 224×224 resolution. Although it is not as high a resolution, ResNet-50 is still superior in accuracy, which leads to its high representational power. In terms of complexity, ResNet-50 has around 25.6 million parameters, while EfficientNet-B4 has around 19 million parameters and MobileNetV3 has around 5.4 million parameters. The lightest and most efficient variant of MobileNet is MobileNetV3, but in return it has lower classification performance, especially in more complicated scenarios like fine discrimination of cyst vs tumor.

On the high-resolution feature extraction scenario, efficientNet-B4 is very efficient thanks to compound scaling, whereas on the clinical classification scenario, it is slightly less robust than ResNet-50. In terms of deployment, MobileNetV3 is found to be the most appropriate solution for edge and resource-constrained environments as it has high speed of inference and low computational requirements. Its shallow feature extraction ability, however, makes it unsuitable for high precision medical diagnosis. In the CBAM compatibility analysis, both ResNet-50 and EfficientNet-B4 are good with attention mechanisms, whereas MobileNetV3 is moderate because of its light structure. Overall, the comparison shows that ResNet-50 (with CBAM) yields the best accuracy-stability-clinical reliability trade-off and is the most appropriate main model for kidney CT classification, while EfficientNet-B4 and MobileNetV3 are good comparative and lightweight options.

Table 4.4: 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

4.4 Overall Comparative Analysis

All three models were compared with each other in the form of a summary presented in Table 4.2(Comparison of Deep Learning Models for Kidney CT Classification).The results in 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.

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 4.5.

Table 4.5: Comparative Analysis with related classification work

Author	Technique Used	Accuracy	Key Notes
Almuayqil et al. [20]	KidneyNet (8-layer custom CNN)	99.88% / 99.92% precision	Lightweight custom CNN tailored to the same Kaggle dataset
Bhandari et al. [21]	Lightweight custom CNN + LIME + SHAP explainability	99.52%	Lightweight design with built-in explainability (LIME/SHAP)
Bingol et al. [22]	29-layer CNN + Relief feature selection	99.37%	Used feature selection to reduce redundant CNN features
El-Hag et al. [19]	EfficientNet-B2 + specular-free preprocessing	98.27%	Added specular-free preprocessing stage before EfficientNet-B2
Islam et al. [11]	VGG-16, ResNet-50, Swin Transformer	96% (ResNet-50); 99.30% (Swin)	Compared multiple backbones on the same task
Khan et al. [23]	ConvNeXt + EfficientNetV2B0 + selective data trimming	99.73%	Combined ConvNeXt and EfficientNetV2B0 with selective data trimming
Maçin et al. [9]	KidneyNeXt (depthwise separable CNN,	99.96%	Depthwise separable convolutions for a lightweight footprint
Obaid et al. [16]	DarkNet53 + multi-model ensemble fusion	99.69%	DarkNet53 backbone fused with an ensemble of models
Pimpalkar et al. [14]	VGG16, ResNet50, InceptionV3 with hyperparameter optimisation	99.96%	Hyperparameter-optimised ensemble of three pretrained CNNs
Qian. [18]	DenseNet121 (standard, no augmentation)	93.51%	Baseline DenseNet121 without augmentation or class balancing
Proposed Methodology	ResNet-50 + CBAM	99.88 %	Residual learning with CBAM attention and class-weighted focal loss; explainable via Grad-CAM

4.5 Explainability using (Grad-CAM)

Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize the areas of the CT images that strongly influenced the predictions made by the proposed model to make the model more interpretable. The visual results show that the model is able to precisely target clinically relevant regions of the kidney structure, especially those which are prone to cyst formation. The regions of high activation (shown in red and yellow) correspond to the regions of abnormality in the kidney, whereas the regions of low activation (shown in blue) represent the irrelevant background tissues.

In every sample displayed, the model makes correct classification, which is confirmed by the overlay drawings “Correct” as in Figure 4.13. The highlighted activation regions are always grouped around the borders of the renal lesions, showing that the model is not learning from irrelevant image artifacts, but rather important pathological characteristics. The results of this explainability analysis agree that the use of CBAM attention mechanisms alongside the feature extraction using ResNet-50 improves the model's focus on the spatial aspects of the images, helping to identify the relevant regions for the disease. The results of the alignment of activations from Grad-CAM with radiological structures further confirm the reliability of the proposed system in a clinical context. The results of Grad-CAM indicate that the model is transparent, interpretable, and clinically trustworthy, critical to its use in medical decision-support systems in real-world applications.

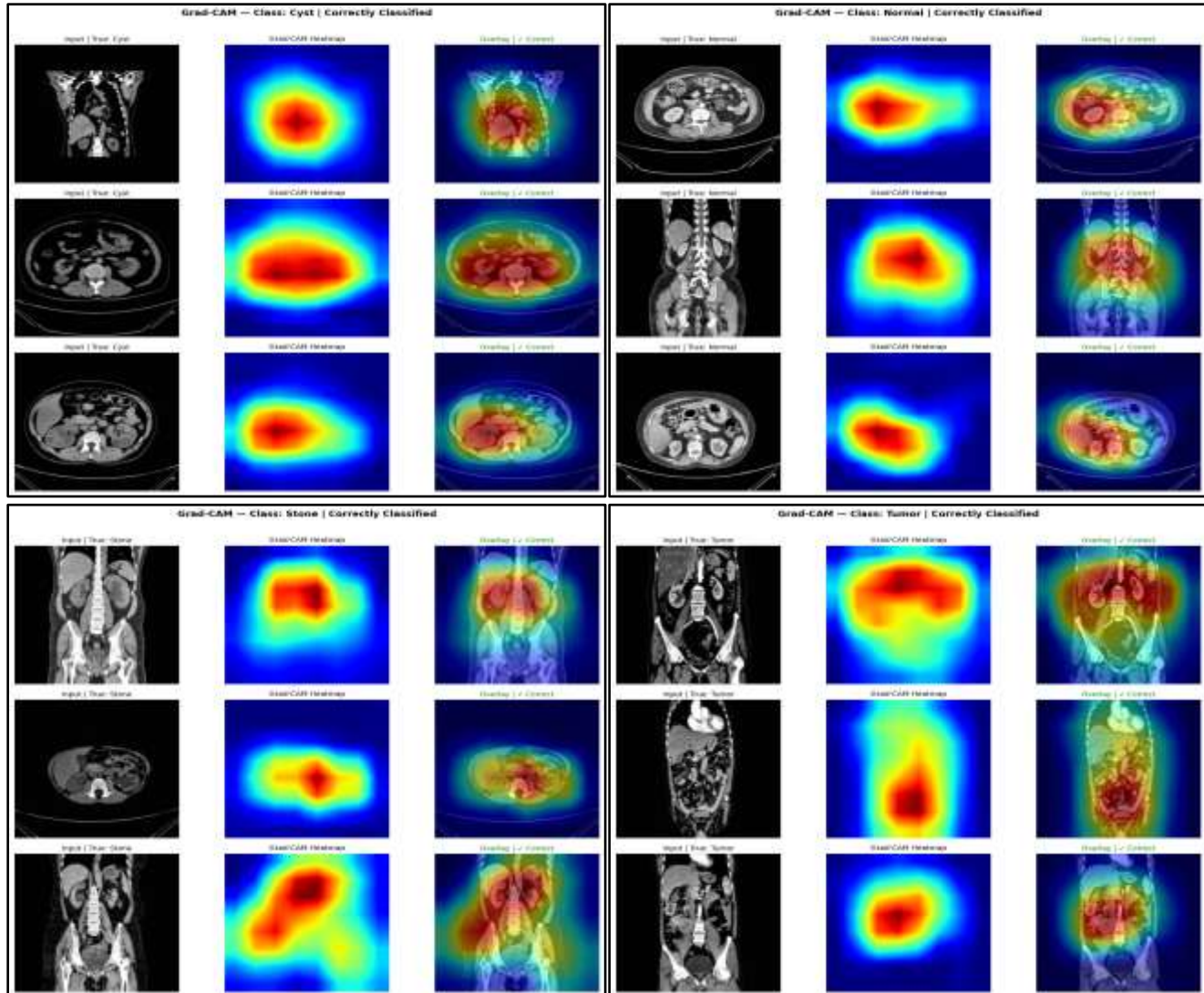


Figure 4.13: Generating Grad-CAM for all classes correctly classified

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 in the Chapter 2, 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 are 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.

Chapter 5: Conclusion and Future Work

5.1 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 were 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 was also achieved using Grad-CAM visualization, which revealed that the model is 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 been compared with the existing literature and has been 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.

5.2 Future Work

The results are very promising, but the following can be done to improve the results:

- The model has been validated in multi centre clinical data to enhance generalization.
- Expansion of the system to multi-phase CT scans (arterial, venous, delayed phases).
- Incorporating segmentation models (e.g., U-Net) into localization of lesions
- Implementation of the model in an actual clinical decision support system (CDSS)
- Investigating hybrid architectures using transformers for additional performance

References

- [1] 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>
- [2] 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>
- [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., 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>
- [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., & Soylyu, 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).