



***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 Collage of Informatics Midocean University, UAE.

Supervisor Name: Dr. Mohamed Eassa

Collage of Informatics, Midocean University, Comoros

Table of Contents

<i>Deep Learning for Detection and Classification of Kidney Tumors and Cancer</i>	1
Abstract	3
1. Chapter 1: Introduction and Overview	4
1.1 Overview	4
1.2 Problem Definition.....	5
1.3 Thesis Objectives	6
1.4 Thesis Contribution.....	7
1.5 Thesis Organization	8
2. Chapter 2: Background and Related Work	9
2.1 Introduction.....	9
2.2 Background	10
2.3 Related Work	14
2.4 Synthesis, Gap Analysis.....	21
References	23

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 metastatic disease. Nevertheless, the process of interpreting computed tomography (CT) scans manually is time-consuming, susceptible 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.

1. 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 [1]. It has subtypes like the transitional cell carcinoma, Wilms tumor in children and benign ones which include the oncocytomas and the angiomyolipoma [2]. There has been obviously increase in the incidence of kidney tumors, which is impute the rising rates of obesity, smoking, and high blood pressure, as well as continuous improvements in diagnostic imaging techniques [3]. Also according to current epidemiological statistics there is approximately 434,840 new cases diagnosed of kidney cancer over the world in 2022, which is the 14th most frequently occur cancer in general; 10th in men and 13th in women [1]. Mortality rate is also high with about 155, 953 people dying of kidney cancer which is quite deadly unless it is prevented and treated at tender age [4].

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

Computed Tomography (CT) scans have emerged as the core in the diagnosis of kidney tumors because it offers high resolution images that could be used in the accurate visualization of the tortoise features of the tumor with regard to size, shape, location and the pattern of enhancement [7]. Multi-detector CT protocols include non-contrast, arterial, venous and by doing

this the case can be correctly staged by assessing vascular involvement, lymph node metastasis, and distant spread and the accuracies of detection have been reported as in excess of 99 % [8]. Manual determination of CT scans is however time consuming and subject to variation particularly in the determination of minute differences between tumors subtypes.

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

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

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 the development in the field of 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, pre-processing of data, and training strategies.
- Chapter 4 is the results of the experiment, performance analysis, and the comparison with the benchmark.
- Chapter 5 ends with discussions, limitations and future directions of work.

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

2. 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 data, namely, CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone. 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.

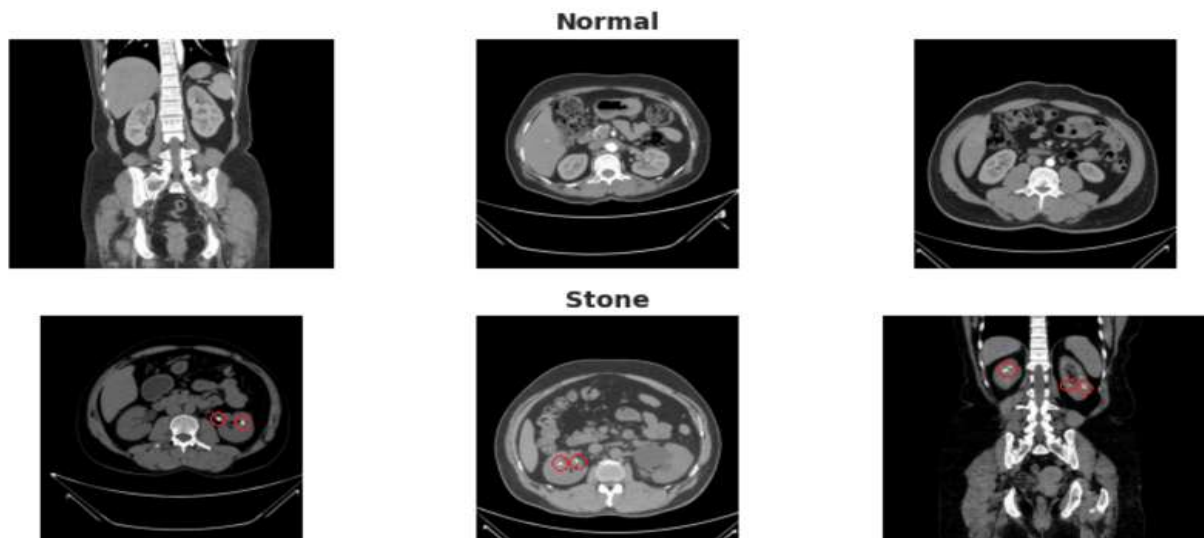


Figure 2.1 CT-KIDNEY from DATASET

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 the benchmark dataset and the current study suggest, four clinically relevant states are of interest, which include: Normal, Cyst, Stone, and Tumor [8].

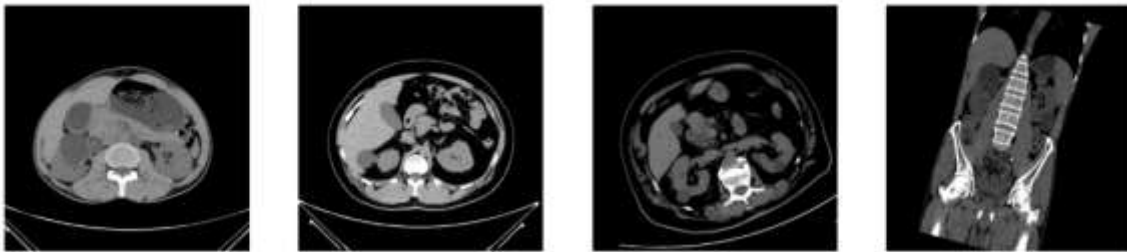


Figure 2.2 CT-KIDNEY from DATASET

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 subsequently lower the number of deaths [1]. The breakthrough was gained with the introduction of deep learning, especially Convolutional Neural Networks (CNNs).

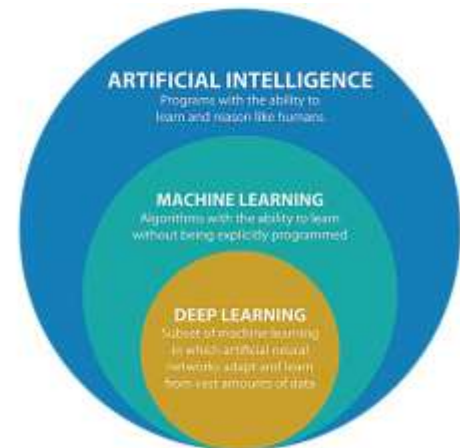


Figure2.3

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 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].

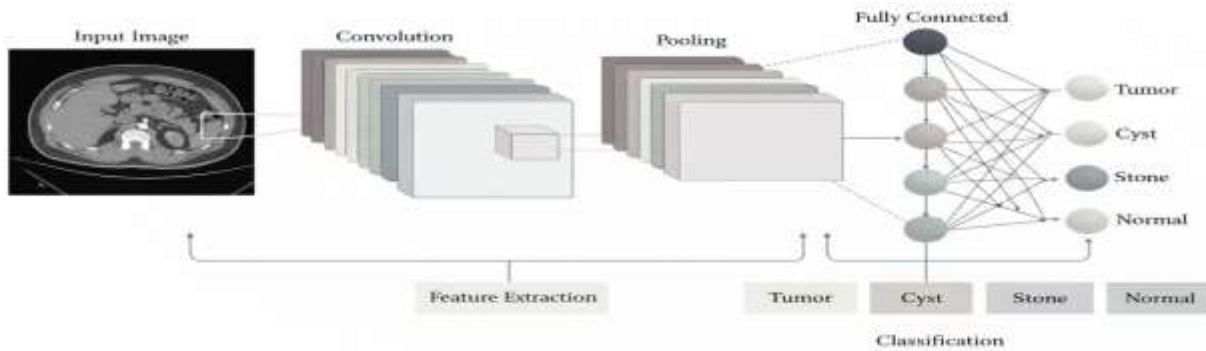


Figure 2.3 CNN Architect

DenseNet-121 is a neural network structure that enhances propagation of features, each layer linked with every other layer. 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].

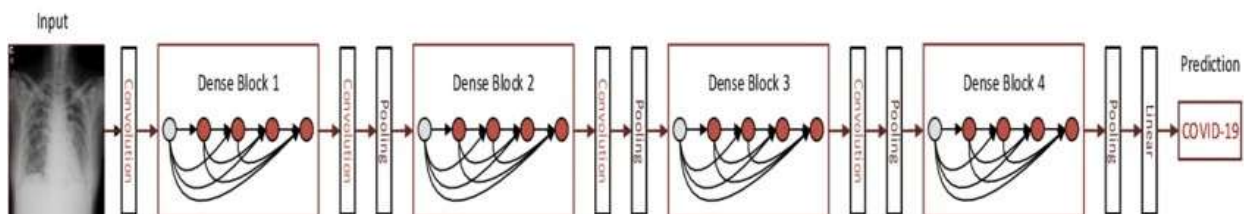


Figure 2.4 Architecture of the DenseNet121 model www.researchgate.net

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 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].

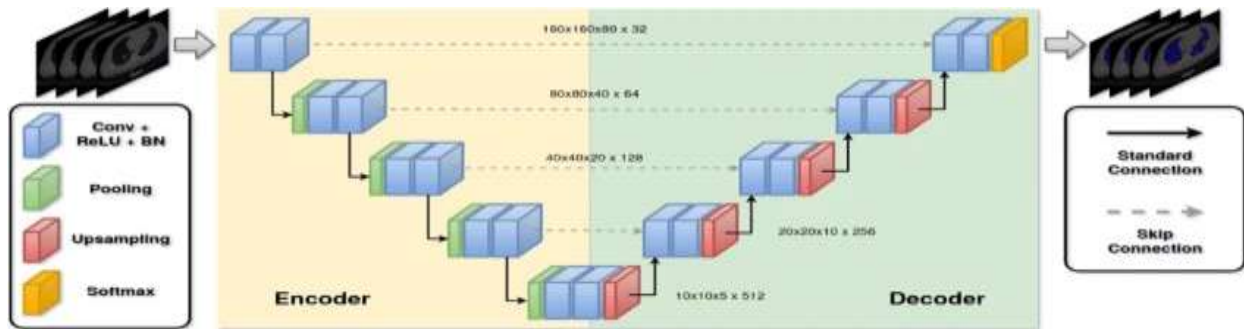


Figure 2.6 Architecture of the U-Net www.researchgate.net

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 [1–15]. 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 CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone). The approach quickly converged on learned features and ResNet-50 skip connections of about 96% accuracy per classification, though ImageNet training on colour images was non-optimal on grayscale CT texts and the number of parameters was excessive to be used in clinical practice. Vision transformer architectures such as EANet, CCT and Swin Transformer were also tested in the same research with the Swin Transformer scoring the highest score of 99.30% accuracy showing the new strength of attention-based models over the traditional CNN in medical

image classification. However, the fact that transformer architectures have high computational and training requirements is still a practical obstacle to implementation in resource-limited clinical setting [11].

Alzu'Bi et al. (2022) introduced two-stage kidney tumour detection and classification framework tested on a new institutional dataset, which was gathered at King Abdullah University Hospital (KAUH) in Jordan, consisting of 8,400 CT scan images of 120 suspected renal masses patients. The data were categorized to 80/20 train and test sets. During the first stage (tumour detection), the authors compared three 2D-CNN architectures with a custom six-layers CNN (CNN-6), ResNet50, and VGG16. During the second stage (tumour type classification), a custom four-layered CNN (CNN-4) was used. CNN-6 and ResNet50 models provided accuracy rates of 97% and 96% respectively, and VGG16 had an accuracy rate of 60%, whereas CNN-4 classification model was able to provide an accuracy of 92 percent. The VGG16 detection rate of 60% and the relatively low classification rate of 92% demonstrated the difficulty of network training on comparatively small, single institution datasets, and the need of the volume and diversity of data to obtain clinically acceptable performance [12].

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

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

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

Almuayqil et al. (2024) presented an idea of KidneyNet a new eight-layer CNN architecture that is directly trained to perform automated four-class kidney CT classification on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone). KidneyNet was built according to a compact sequential architecture comprising of convolutional, batch normalisation, pooling, dropout and fully connected layers, in a way that it optimised the model complexity versus accuracy of its classification. Its architecture did not require the large number of parameters that large transfer learning models have, which also makes it more practical to implement in the resource-constrained healthcare setting. The accuracy and precision of KidneyNet were 99.88 and 99.92, respectively, which is among the best benchmark results at the time of publication. The authors also pointed out that the model was trained and evaluated only on

the Kaggle data and the extent to which it works in a real clinical setting with heterogeneous scanner protocols and patient demographics was not tested [20].

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

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

El-Hag et al. (2025) The issue of reflections and image artefact in CT scans was solved by introduced a specular-free preprocessing pipeline before deep learning classification on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone). Preprocessing stage used

reflection removal and contrast normalisation to decrease the noise and enhance uniformity of features across the dataset, and the classification with EfficientNet-B2 a model with compound-scaled network that reduces network width, depth, and resolution. EfficientNet-B2 was chosen because it is one that has good trade-off both in accuracy and computational cost among the family of EfficientNets. The preprocessing-classification pipeline obtained an accuracy of 98.27, a score that is lower than some of the competing methods using the same benchmark, but it showed that artefact suppression methods could be used to gain quality learned representations. The authors observed that the preprocessing pipeline added more steps of calculations and the effect of the approach on real, noisy clinical CT data of heterogeneous scanner types needed more consideration [19].

Maçin et al. (2025) proposed KidneyNeXt, a custom lightweight CNN model that uses depthwise separable convolutions on the CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone) with a standardized input resolution of 224x224 pixels. Depthwise separable convolutions break standard convolution down into a depthwise spatial convolution (filter) and a pointwise channel mixing step, greatly reducing the number of parameters and floating-point operations at no cost to feature extracting performance. KidneyNeXt reached an accuracy of 99.96, corresponding to much heavier architectures at a fraction of the computational cost, and thus matched it especially well to run on edge devices or in low resource clinical environments. The authors proved that the architecture was very high in both sensitivity and specificity in all the four classes (including the diagnostically challenging stone class) which 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 CT KIDNEY DATASET (12,446 images: Normal, Cyst, Tumor and Stone) to classify 4 classes in kidney CT. One of the most important methodological contributions was the adoption of selective data trimming to achieve inter-class imbalance before training, eliminating redundant samples of the over-represented classes to ensure that the model does not develop a bias towards the higher representation classes, like the Normal one. ConvNeXt, a convolutional network that has been modernised using methods influenced by Vision Transformers, was combined with EfficientNetV2B0 at the feature-level in an ensemble to

combine the complementary strengths of the two architectures. The ensemble was found to be accurate at 99.73 with just 6.2 million parameters, proving that well-thought-out data preprocessing and well-informed choice of architecture can provide state-of-the-art results without using architecturally complicated and computationally intense models. The authors indicated that the study had some limitations such as the use of only one benchmark dataset and the necessity to further verify it on multi centers CT cohorts [23].

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

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 fifteen reviewed studies on deep learning for kidney CT classification.

Ref	Author	Year	Dataset	Technique Used	Accuracy
[13]	Yildirim et al.	2021	Private NCCT (1,799 CT, Turkey)	XResNet-50 + data augmentation	96.82%
[11]	Islam et al.	2022	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
[17]	Baygin et al.	2022	Private NCCT (1,799 CT, Turkey)	DarkNet19 + INCA feature selection + KNN (ExDark19)	99.22%-99.71%
[21]	Bhandari et al.	2023	CT KIDNEY DATASET	Lightweight custom CNN + LIME + SHAP explainability	99.52%
[22]	Bingol et al.	2023	CT KIDNEY DATASET	29-layer CNN + Relief feature selection	99.37%
[15]	Sabuncu et al.	2023	Clinical axial CT (Near East Univ. Hospital, Cyprus)	InceptionV3	98.52%
[18]	Qian et al.	2024	CT KIDNEY DATASET	DenseNet121 (standard, no augmentation)	93.51%
[20]	Almuayqil et al.	2024	CT KIDNEY DATASET	KidneyNet (8-layer custom CNN)	99.88% / 99.92% precision

Ref	Author	Year	Dataset	Technique Used	Accuracy
[14]	Pimpalkar et al.	2025	CT KIDNEY DATASET	VGG16, ResNet50, InceptionV3 with hyperparameter optimisation	99.96% (InceptionV3)
[16]	Obaid et al.	2025	CT KIDNEY DATASET	DarkNet53 + multi-model ensemble fusion	99.69%
[19]	El-Hag et al.	2025	CT KIDNEY DATASET	EfficientNet-B2 + specular-free preprocessing	98.27%
[9]	Maçin et al.	2025	CT KIDNEY DATASET	KidneyNeXt (depthwise separable CNN, 224×224)	99.96%
[23]	Khan et al.	2025	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)

References

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

- tomography imaging. *Journal of Clinical Medicine*, 14(14), 4929. <https://doi.org/10.3390/jcm14144929>
- [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.combiomed.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>