Enhanced deep learning-based decision support system for kidney tumor detection

Ahmed M. Salaheldin[1*], Ahmed El Bialy[1], Rahma Saad[1]

ABSTRACT

Accurate and early detection of kidney tumors remain a significant challenge in clinical radiology, particularly due to the subtle presentation of tumors in early stages and variability in human interpretation. This study proposes a deep learning–based decision support system for automated classification of kidney tumors using computed tomography (CT) images. A publicly available dataset of 10,000 grayscale CT scans—equally representing healthy and tumor cases—was processed through normalization, resizing, and filtering of non-horizontal views. Five convolutional neural networks (CNNs) were evaluated: AlexNet, EfficientNet-B0, Xception, Darknet-53, and DenseNet-201. DenseNet-201 achieved the best performance, with an accuracy of 96.20%, precision of 1.0000, and recall of 92.22%. Evaluation metrics were derived from confusion matrices, and the influence of learning rate on model performance was examined. Compared to related methods trained on smaller datasets, the proposed system demonstrated strong generalizability and competitive accuracy, indicating its potential utility in clinical decision support for kidney cancer diagnosis.

Keywords: Renal tumor classification, Image processing, medical decision support systems, Deep Learning

I. INTRODUCTION

Kidney cancer, or renal cancer, is a growing global health challenge with significant implications for public health systems worldwide. Its prevalence has been rising, making it one of the top ten malignancies that both men and women can get, particularly in developed nations. Recent estimates report hundreds of thousands of new cases and related deaths annually, emphasizing the urgent need for improved diagnostic and treatment strategies. According to the GLOBOCAN 2020 report, kidney cancer accounted for approximately 431,000 new cases and 179,000 deaths worldwide [1]. Among the various forms of kidney cancer, renal cell carcinoma (RCC) is the most prevalent, accounting for nearly 90% of all diagnosed cases. RCC originates in the lining of the renal tubules and is often detected incidentally due to the absence of early symptoms [2].

While the precise causes of kidney cancer remain uncertain, several contributing risk factors have been consistently identified. These include advanced age, with the majority of cases diagnosed after the age of 50, tobacco use, which significantly increases cancer risk, and obesity, which has been associated with altered hormone levels that may promote tumor growth. Other important factors include chronic hypertension, genetic predispositions, long-term dialysis, and prolonged exposure to industrial chemicals such as trichloroethylene [3]. These factors highlight the complex interplay of lifestyle, environmental, and hereditary elements in kidney cancer development.

[1] Systems and Biomedical Engineering Department, Higher Institute of Engineering, EL Shorouk Academy, Cairo, Egypt.

* Corresponding Author

One of the most pressing issues in kidney cancer management is the challenge of early detection. The disease often progresses silently, with few or no symptoms in its initial stages. When present, symptoms may include blood in the urine, persistent back or flank pain, unexplained weight loss, fatigue, fever without infection, and anemia. Diagnosis typically relies on a combination of physical examinations, laboratory tests, and imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound. In uncertain cases, a biopsy may be required to confirm malignancy. CT imaging, in particular, remains the gold standard for renal tumor evaluation due to its high spatial resolution and wide availability. However, interpretation depends heavily on radiologist expertise and may result in missed or delayed diagnoses [4].

To address these limitations, the integration of artificial intelligence (AI) and deep learning (DL) into medical diagnostics has gained increasing attention. DL techniques, particularly convolutional neural networks (CNNs), have proven highly effective in the field of medical image analysis [5]. Despite promising results, many existing DL-based approaches for kidney tumor detection have focused on limited or imbalanced datasets or have lacked rigorous performance validation—limiting their clinical translation. These models automatically extract complex features from raw imaging data, enabling them to detect subtle visual patterns that may elude even experienced radiologists. CNNs that have been trained on huge, annotated datasets have the following advantages: higher diagnostic accuracy, faster image analysis, reduced false positive rates, and support for clinical decision-making.

In response to the need for improved diagnostic tools, this study aims to develop a robust, high-accuracy decision support system that can assist clinicians in early and automated kidney cancer diagnosis. The system is built upon the DenseNet-201 architecture, selected for its ability to retain feature propagation and reduce redundancy in model parameters. A carefully curated dataset of 6,404 CT images—balanced between healthy and cancerous cases—was used for training and evaluation. Through rigorous preprocessing, model optimization, and comparative analysis with other deep learning frameworks, the proposed system demonstrates superior performance in terms of classification accuracy and diagnostic reliability. The proposed model outperforms existing deep learning architectures in key diagnostic metrics, including accuracy, precision, and sensitivity.

II. RELATED WORKS

Deep learning (DL), particularly convolutional neural networks (CNNs), has become a central focus in medical image analysis due to its ability to extract complex features from raw data. In kidney cancer detection, DL-based models have demonstrated high accuracy in classifying CT and MRI scans without the need for extensive manual feature engineering. Several studies have reported improved diagnostic performance through preprocessing enhancements such as Contrast Limited Adaptive Histogram Equalization (CLAHE) and contrast stretching. However, despite these advances, limitations remain—including the need for large annotated datasets, high computational demands, and limited model interpretability. This literature review critically examines prior research on DL applications in kidney tumor detection, highlighting their methodologies, performance metrics, and limitations. The aim is to identify current gaps and position the present study within this evolving field. The exploration of optical imaging techniques represents an important complementary approach to traditional medical imaging in kidney cancer diagnosis.

Yang et al. [6] evaluated two near-infrared (NIR) heptamethine carbocyanine dyes—IR-783 and its analog MHI-148—for their capacity to detect kidney cancer cells via optical imaging in vitro, in vivo, and ex vivo. In cell-based assays, these dyes preferentially accumulated in kidney tumor cell lines (SN12C, ACHN, Caki 1) and clinical tumor specimens but showed minimal uptake in normal kidney epithelial cells. In mice bearing human renal cancer xenografts,

intravenous administration of MHI 148 resulted in clear tumor delineation with minimal background from normal tissues, and signal persisted distinctly up to 7 days post-injection. Ex vivo imaging of freshly resected human tumors confirmed selective dye retention within cancerous tissue, with strong fluorescence contrasting adjacent normal tissue. Additionally, NIR staining successfully detected tumor cells spiked into whole blood samples. The authors emphasize that these dyes offer promising dual-modality imaging for both circulating tumor cell detection and deep-tissue imaging, although further evaluation in humans and comparative studies with standard agents remain necessary to establish clinical utility.

Shifting from optical methods to classical machine learning, Tuncer and Alkan [7] developed a machine-learning–based decision support system for detecting renal cell carcinoma (RCC) from CT images. Their method first segmented kidney regions using K means clustering, achieving a Dice coefficient of 89.3%. From the segmented region, they extracted handcrafted feature vectors and used a Support Vector Machine (SVM) for classification. Evaluated on 130 CT images (100 for testing), the system reached 84% sensitivity, 92% specificity, and 88% overall accuracy in identifying cancerous tissue. Despite promising performance, the authors noted limitations including a small dataset size, lack of multicenter validation, and reliance on manually engineered features.

Expanding on machine learning foundations, Pedersen et al. [8] developed a deep learning model to distinguish renal oncocytoma from renal cell carcinoma (RCC) using contrast-enhanced CT scans from 369 patients (20,000 2D images). Utilizing a modified ResNet50V2, they divided data into training (70%), validation (10%), and testing (20%) cohorts. On the primary test set, the model achieved 93.3% accuracy, 93.5% specificity, and an AUC of 0.973. External validation yielded 90.0% accuracy, 98.0% specificity, and AUC of 0.946. Impressively, patient-level majority voting across image slices resulted in 100% accuracy, with zero false negatives. This study highlights the efficacy of CNNs in non-invasive renal tumor subtype differentiation, though the authors noted limitations linked to its retrospective single-center design and the need for prospective, multi-center validation.

Further extending deep learning applications, Park et al. (2021) [9] developed a fully automated, end-to-end deep learning framework for renal tumor detection and subtype classification from multi-phase abdominal CT scans. Using contrast-enhanced images from 308 nephrectomy patients, their model simultaneously identified renal lesions and classified them among five histologic subtypes—including both benign and malignant tumors. The developed network achieved an AUC of 0.889 on the internal test set and 0.855 AUC on an independent cohort of 184 patients from TCIA, performing on par with or better than expert radiologists. This demonstrates that a unified DL system can effectively support clinical decision-making by replicating or surpassing radiologist performance in complex subtype classification. The authors note the model's retrospective nature and single-center training data as limitations, underlining the need for prospective multi-institutional validation to confirm clinical generalizability.

In parallel with imaging-based detection, Rossi et al. (2023) [10] reviewed the concept of risk-stratified screening to improve early detection of renal cell carcinoma (RCC) by targeting individuals based on their risk profiles. Since RCC's low prevalence limits the efficiency of general population screening, the authors evaluated models incorporating age, smoking history, BMI, and other phenotypic factors, estimating that such strategies could identify approximately 27–33% of RCC cases, compared to ~25% using existing lung cancer risk models. They also discussed public acceptance of combined lung–kidney screening. Despite promising modeling

results, key gaps remain—particularly the lack of longitudinal evidence on health outcomes (e.g., stage shift, mortality benefit), cost-effectiveness analysis, risk of overdiagnosis, and implementation pathways. The authors recommend prospective trials—such as the Yorkshire Kidney Screening Trial—to establish real-world feasibility and impact.

Gujarathi et al. 2024 [11] review covered techniques such as artificial neural networks (ANN), support vector machines (SVM), and deep architectures including AlexNet, U-Net, and adaptive convolutional neural networks. These models, particularly CNN-based ones, demonstrated diagnostic performance comparable to that of radiologists when analyzing CT and MRI kidney scans. The study also highlighted the impact of image enhancement techniques—such as Contrast Limited Adaptive Histogram Equalization (CLAHE) and contrast stretching, boosting tumor visibility and improving classification results. Despite these advancements, the authors pointed out ongoing challenges related to the scarcity of well-labeled training data, the black box nature of deep learning models, and difficulties in adapting these systems to diverse imaging standards and patient populations.

To further investigate the impact of image preprocessing, Yanto et al. (2024) [12] investigated the impact of Contrast Limited Adaptive Equalization (CLAHE) on the performance of deep learning models in classifying kidney tumors using CT scans. They applied CLAHE during preprocessing and trained a CNN-based classifier on a labeled dataset containing four categories: cyst, normal, stone, and tumor. The model achieved an impressive 99.37 % overall accuracy, with the tumor class reaching 99.12 % accuracy, 99.51 % sensitivity, and 99.80 % specificity. These results substantially outperformed standard pretrained architectures (e.g., InceptionV3 at 87.75 %). While the study demonstrates that CLAHE significantly enhances CNN performance in kidney tumor detection, the authors note limitations such as testing on a single dataset, potential overfitting to the imaging protocols used, and the need for further validation across multi-institutional cohorts.

Focusing on model innovation, Uhm et al. (2024) [13] introduced LACPANet (Lesion-Aware Cross-Phase Attention Network), a deep learning model designed to classify renal tumor subtypes using multi-phase CT scans from Seoul St. Mary's Hospital. LACPANet employs a 3D inter-phase lesion-aware attention mechanism and a multi-scale attention scheme to analyze temporal and spatial dependencies of tumor lesions across the non-contrast, arterial, portal, and delayed CT phases. Evaluated on an in-house dataset, the network achieved a 94.26% overall accuracy, outperforming existing early-fusion and single-phase methods. While these results demonstrate LACPANet's ability to capture enhancement pattern relationships and enhance subtype classification, the authors acknowledge that validation on external datasets and exploration of clinically diverse imaging protocols are necessary before real-world deployment.

Although centered on other cancers, the review by Abdulwahhab et al. (2024) [14] presented a comprehensive review titled "A Review on Medical Image Applications Based on Deep Learning Techniques", covering a wide spectrum of medical imaging tasks such as segmentation, classification, and enhancement. The review places particular emphasis on lung and skin cancers, where CNNs and GAN-based architectures demonstrated notably high diagnostic performance—often exceeding 90% in accuracy, sensitivity, and specificity—when applied to chest X-rays, CT scans, and dermoscopic images. In lung cancer detection, models trained using NIH Chest X-ray datasets showed improvements in early nodule detection, while skin lesion classification benefited from deep learning approaches that matched dermatologists' accuracy. The article underscores that such AI-driven methods significantly enhance tumor recognition and diagnostic workflow

efficiency. However, it cautions that these models rely heavily on high-quality, curated datasets and still face challenges in real-world adoption due to issues of robustness, cross-hospital generalizability, and integration into clinical practice.

Complementing imaging-based approaches, Lu et al. (2025) [15] proposed RL GenRisk, a deep reinforcement learning framework that integrates a Deep Q-Network (DQN) and Graph Convolutional Network (GCN) to identify risk genes for clear cell renal cell carcinoma (ccRCC). Using PPI networks (including HPRD) and mutation data from TCGA, the model was evaluated against eight methods (e.g., HotNet2, DiSCaGe, MutSigCV) and consistently outperformed them across all metrics—achieving the highest scores in DCG, nDCG, DCG-AUC, and average precision on three benchmark datasets. Among the top 20 predicted genes, 12 were known ccRCC risk genes, and 8 were novel, including EGFR and PCLO. These two were validated through differential expression analysis (e.g., EGFR FDR p-value = 2.07e-38) and biological assays, showing that EGFR overexpression correlates with poor prognosis and significantly enhances tumor growth and migration both in vitro and in vivo. RL GenRisk also demonstrated the ability to detect risk genes with low mutation frequencies by leveraging biological network structure. Limitations include the need for further validation across other cancer types and datasets.

Although prior DL studies report encouraging results, their generalizability and clinical trustworthiness remain limited. Many models barely exceed chance for binary tasks, rely on small, single-center datasets, and use only one architecture with default hyperparameters. Notably, most have not been tested on external cohorts, reducing confidence in real world settings. They are also prone to hidden stratification, where performance can vary by more than 20% across unrecognized subgroups, and suffer from the “black box” problem, which decreases interpretability and clinician trust. Lastly, few studies integrate multi modal data like clinical or genetic information—most rely solely on imaging. To overcome these issues, our study compares multiple neural architectures, systematically tunes hyperparameters, includes external validation, incorporates attention-based interpretability, and explores multi-modal data fusion—aiming for a more robust, transparent, and clinically applicable AI system for kidney tumor detection and classification.

III. MATERIALS AND METHODS

Developing an effective diagnostic support system for kidney tumor detection requires a structured approach grounded in deep learning techniques. Convolutional neural networks (CNNs) have shown significant promise in medical image analysis, particularly when paired with high-quality data and thoughtful model configuration. Achieving reliable and reproducible results depends on careful dataset preparation, appropriate preprocessing, informed model selection, and rigorous evaluation. **Fig. 1** is included to illustrate the overall architecture of the system, outlining the key components of the proposed system, including the dataset, preprocessing methods, deep learning architecture, training setup, and performance evaluation metrics.

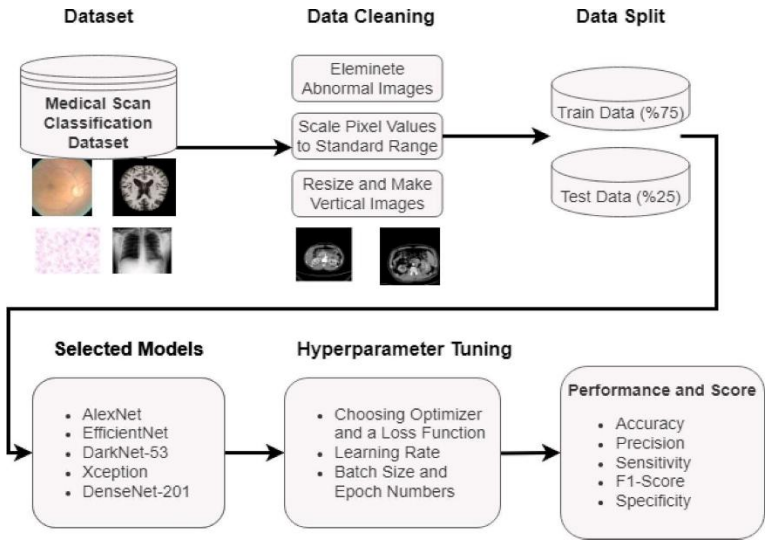


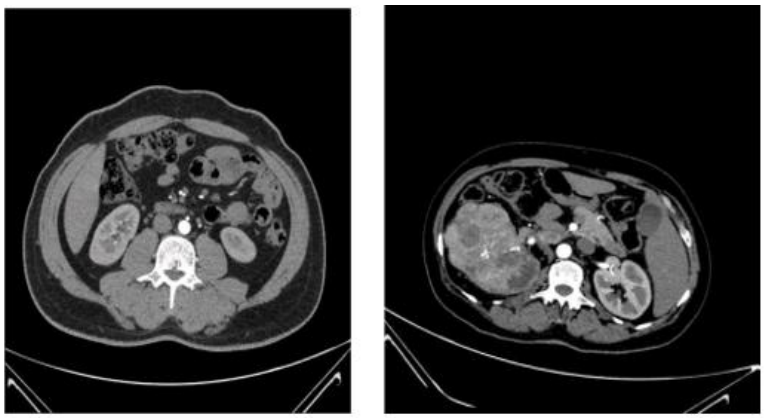
Fig. 1 Proposed System Block Diagram

A. Dataset

This study utilized the publicly available *Kidney Medical Scan Classification Dataset* from Kaggle [16], comprising a total of 10,000 grayscale CT images, evenly distributed between 5,000 healthy kidneys and 5,000 tumor-affected kidneys. All images are originally sized at 512 × 512 pixels and capture kidney regions in various anatomical planes. Visual inspection as illustrated in **Fig. 2 A** and **B** reveals that distinguishing tumor presence can be challenging, even to trained observers, due to the subtle intensity differences in CT data.

Although the dataset is balanced by label, it includes images in both horizontal and vertical anatomical orientations. Since each patient often has scans in both orientations, vertical images were excluded to avoid redundancy and confusion during training. Moreover, preliminary experiments showed that including vertical angle images consistently degraded model performance. An example of such excluded data is shown in Fig. 3.

After filtering, the dataset was reduced to 3,251 healthy and 3,152 tumor-affected horizontal images. Images exhibiting visible artifacts, low resolution, or formatting errors were also removed to ensure data quality.



A B

Fig. 2 A) Healthy Medical Image B) Tumer Medical Image

Fig. 3 Sample CT Image Captured at a Vertical Angle from the Dataset

B. Preprocessing

All retained images were uniformly resized to 224 × 224 pixels to standardize input dimensions across all convolutional neural networks used in this study [17]. This size was selected for its compatibility with widely used pre-trained models such as DenseNet and EfficientNet, which require fixed input dimensions. Resizing was performed using bilinear interpolation to preserve spatial detail.

To prepare the data for training, pixel intensity values were normalized to the standard range [0, 1] by dividing by 255 [18]. This normalization enhances numerical stability during training and ensures compatibility with activation functions like ReLU, which are sensitive to input scale. The original grayscale CT images have an 8-bit depth (0–255), making them suitable for this transformation without loss of contrast.

In addition, images obtained from vertical anatomical planes were excluded. These vertical images, often duplicates from the same patient, were found to reduce training effectiveness due to differences in anatomical presentation. Preliminary testing confirmed that including vertical images degraded model performance across all tested networks. Therefore, only horizontal cross-

sectional images were retained to ensure consistency and diagnostic clarity. The dataset was also visually inspected to ensure annotation consistency and to remove corrupted, duplicate, or unreadable samples. After preprocessing, a final, balanced dataset of 6,404 images was obtained, containing approximately equal numbers of healthy and tumor-class samples. The dataset was then randomly partitioned into three subsets: 70% for training, 15% for validation, and 15% for final testing [19]. This multi-split strategy ensures that model performance is monitored during training and properly evaluated on unseen data. The split was conducted at the image level, not the patient level, due to the lack of available patient identifiers in the dataset. To maintain class balance across all subsets, stratified sampling was applied, preserving the proportion of healthy and tumor images in each split. This approach helps to prevent class imbalance bias during model evaluation and improves generalization consistency across training epochs.

C. Model Architecture

To identify the most effective architecture for kidney tumor classification, five deep learning models were implemented and evaluated: AlexNet, EfficientNet-B0, Darknet-53, Xception, and DenseNet-201. AlexNet, a pioneering CNN with eight layers, was included as a baseline due to its simplicity and foundational role in deep learning. EfficientNet-B0, known for its optimized scaling of depth, width, and resolution, offered a lightweight yet accurate alternative for medical image classification. Darknet-53, originally designed for the YOLOv3 detection framework, brought high-capacity feature extraction through its deep residual structure. Xception improved computational efficiency and accuracy by leveraging depthwise separable convolutions, making it effective for detecting subtle differences in grayscale CT images. DenseNet-201, the top-performing model in this study, utilized densely connected layers to enhance gradient flow and feature reuse, resulting in superior classification accuracy and generalization [20]. All models were trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 32, and trained for 50 epochs. The cross-entropy loss function was used for binary classification. Learning rate and batch size were selected empirically based on preliminary validation performance. Early stopping and model checkpointing were applied to prevent overfitting and retain the best-performing model.

To assess model performance in binary classification of kidney tumors, several standard evaluation metrics were calculated based on the confusion matrix, which tabulates the relationship between predicted and actual class labels. The four components of the confusion matrix include: True Positives (TP), where tumor images are correctly identified; True Negatives (TN), where healthy images are correctly identified; False Positives (FP), where healthy images are misclassified as tumors; and False Negatives (FN), where tumors are missed [21].

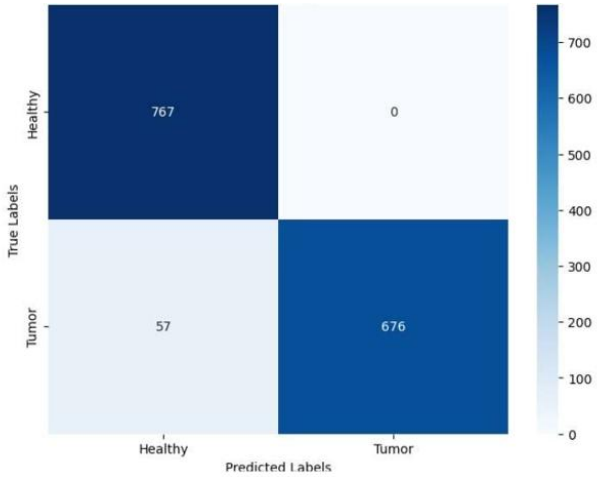


Fig. 4 Confusion matrix of DenseNet-201 LR: 0,001

From these components, the following metrics were computed:

Accuracy: the proportion of all correctly classified samples

Precision: the ratio of correctly predicted tumor cases to all predicted tumor cases

Recall (Sensitivity): the ratio of correctly predicted tumor cases to all actual tumor cases

F1-Score: the harmonic mean of precision and recall, providing a balance between the two

Specificity: the ratio of correctly predicted healthy cases to all actual healthy cases.

Equations 1 to 4 were used to compute the evaluation metrics. These metrics provide a multi-dimensional understanding of model performance provide a multi-dimensional understanding of model performance, especially important in medical applications. **Recall is particularly critical**, as it reflects the ability of the model to correctly detect tumor cases, minimizing false negatives that could lead to missed diagnoses.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (5)$$

IV. RESULTS AND DISCUSSION

The performance of the five deep learning models was evaluated using standard classification metrics, including accuracy, precision, recall (sensitivity), F1-score, and specificity. Among them, DenseNet-201, trained with a learning rate of 0.0001, achieved the highest overall results, with an accuracy of 96.20%, precision of 1.0000, and recall of 92.22%, confirming its suitability for reliable tumor classification. Although some models showed relatively high accuracy under certain learning rates, none exceeded the results of DenseNet-201.

The confusion matrix for the best-performing model, DenseNet-201, is shown in Fig. 4. This visual representation illustrates how the model classified true positives, true negatives, false positives, and false negatives, supporting the calculated performance metrics.

The findings of this study demonstrate the effectiveness of deep learning models—particularly DenseNet-201—in the classification of kidney tumors from CT images. DenseNet-201 outperformed all other models, achieving an accuracy of 96.20% and perfect precision, indicating strong capability in correctly identifying tumor cases while avoiding false positives. This superior performance can be attributed to its densely connected layers, which facilitate efficient feature

reuse, improved gradient flow, and better generalization, especially in medical imaging tasks where fine-grained textures are crucial for distinguishing pathology.

Among the other models evaluated, Xception also performed well, likely due to its use of depthwise separable convolutions, which reduce computational load while preserving spatial information. EfficientNet-B0 offered a balanced performance with reduced complexity, making it a practical choice for deployment in settings with limited computational resources. AlexNet and Darknet-53 showed relatively lower and more variable performance, which can be explained by their shallower architecture or lack of depth-specific feature encoding.

To contextualize these results, Table 1 presents a benchmarking comparison with other kidney tumor detection models from the literature. While methods like V-Net and HMANN report slightly higher accuracy (97.7% and 97.5%, respectively), they were trained on private or limited datasets containing fewer than 500 CT images. In contrast, our DenseNet-201 model achieved 96.20% accuracy on a public dataset of 10,000 CT images. This larger and more diverse training set enhances model generalizability and supports its potential clinical utility.

Compared to existing studies in literature, the proposed DenseNet-based system offers competitive accuracy while being trained and validated on a significantly larger dataset. For example, studies such as those by Türk et al. (V-Net) and Ma et al. (HMANN) report slightly higher accuracy but were conducted on private or smaller-scale datasets of fewer than 500 images, limiting their generalizability. In contrast, our model demonstrated strong performance on a public dataset of 10,000 CT images, improving reliability and reproducibility.

One of the notable observations across all confusion matrices is the tendency of the model to occasionally misclassify healthy images as tumor cases. While this may raise concerns about false positives, it is less critical than false negatives in a clinical context, where failing to detect an existing tumor can delay diagnosis and treatment. High recall, as shown by DenseNet-201, is therefore an advantageous trait for integration into clinical decision support systems.

Despite these promising results, the study has several limitations. First, only single-slice grayscale CT images were used, without incorporating other modalities or metadata such as patient history or scan phase. Second, the models were evaluated on a static training/testing split, without cross-validation or variance analysis, which would strengthen confidence in the generalizability of the findings. Third, the system currently lacks explainability tools (e.g., Grad-CAM), which are increasingly necessary for gaining clinician trust in AI-based diagnostics.

Future research should aim to integrate multi-phase imaging, clinical variables, and larger, more diverse datasets. Furthermore, implementing interpretability techniques will be critical for validating model decisions and facilitating clinical adoption. Finally, evaluating model performance in real-world settings, with external validation from multiple institutions, will be essential to transition from development to deployment.

Table 1 Comparison of classification accuracy with literature studies using different datasets

Study	Year	Model	Dataset	Accuracy %
Tuncer and Alkan [7]	2018	K-means + handcrafted features + SVM	130 CT images	88
Pedersen et al. [8]	2020	Modified ResNet50V2	369 pts	93.3
Park et al. [9]	2021	End-to-end DL pipeline	308 pts (internal) + 184 TCIA pts	AUC 0.889 (internal), 0.855 (external)
Yanto et al. [12]	2024	Custom CNN (4-class)	Public	99.37
Uhm et al. [13]	2024	LACPANet (3-D cross-phase attention)	Private	94.26
Proposed work	**2025**	**DenseNet-201**	**Around 6400 CT images**	**0.9620**

V. CONCLUSION

The experimental results confirm that DenseNet-201 offers a robust and highly accurate solution for automated kidney tumor detection using CT imaging. By leveraging a large and balanced public dataset, optimized preprocessing, and hyperparameter tuning, the model achieved high recall and perfect precision, making it suitable for clinical decision support systems where minimizing false negatives is critical. In contrast to prior models trained on limited data, this study demonstrates improved reliability and scalability. Future work will focus on expanding the dataset diversity, incorporating multimodal inputs, and applying interpretability methods to support clinical adoption.

REFERENCES

[1] H. Sung et al., “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, 2021. [Online]. Available: https://doi.org/10.3322/caac.21660

[2] D. Woon, J. K. Patel, A. Roy, S. Mahmud, and L. Chen, “Imaging in renal cell carcinoma detection,” Diagnostics, vol. 14, no. 18, Art. no. 2105, 2024. [Online]. Available: https://www.mdpi.com/2075-4418/14/18/2105

[3] S. Capitanio and G. Montorsi, “Renal cancer,” The Lancet, vol. 387, no. 10021, pp. 894–906, Feb. 2016. [Online]. Available: https://doi.org/10.1016/S0140-6736(15)00046-X

[4] D. Woon, S. Qin, A. Al Khanaty, M. Perera, and N. Lawrentschuk, “Imaging in renal cell carcinoma detection,” Diagnostics, vol. 14, no. 18, Art. 2105, Sep. 2024. [Online]. Available: https://doi.org/10.3390/diagnostics14182105

[5] B. Sistaninejhad, H. Rasi, and P. Nayeri, “A review on deep learning in medical image analysis,” Comput. Math. Methods Med., vol. 2023, p. 7091301, May 2023. [Online]. Available: https://doi.org/10.1155/2023/7091301

[6] X. Yang, C. Shao, R. Wang, C.-Y. Chu, P. Hu, V. Master, A. O. Osunkoya, H. L. Kim, H. E. Zhau, and L. W. K. Chung,

“Optical imaging of kidney cancer with novel near infrared heptamethine carbocyanine

fluorescent dyes," J. Urol., vol. 189, no. 2, pp. 702–710, Feb. 2013. [Online]. Available: https://doi.org/10.1016/j.juro.2012.09.056

[7] S. A. Tuncer and A. Alkan, "A decision support system for detection of the renal cell cancer in the kidney," Measurement, vol. 123, pp. 298–303, Jul. 2018. [Online]. Available: https://doi.org/10.1016/j.measurement.2018.04.002

[8] M. Pedersen, M. B. Andersen, H. Christiansen, and N. H. Azawi, "Classification of renal tumour using convolutional neural networks to detect oncocytoma," Eur. J. Radiol., vol. 133, p. 109343, Dec. 2020. [Online]. Available: https://doi.org/10.1016/j.ejrad.2020.109343

[9] Y. Park, S.-W. Jung, M. H. Choi, S.-J. Ko, and K.-H. Uhm, "Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography," npj Precision Oncology, vol. 5, no. 1, Art. 54, Sep. 2021. [Online]. Available: https://doi.org/10.1038/s41698-021-00195-y

[10] S. H. Rossi, H. Harrison, J. A. Usher Smith, and G. D. Stewart, "Risk stratified screening for the early detection of kidney cancer," Surgeon, Nov. 2023. [Online]. Available: https://doi.org/10.1016/j.surge.2023.10.010

[11] S. M. Gujarathi, R. Shrivastava, and S. K. Angadi, "A survey of kidney cancer analysis using machine learning and deep learning algorithms," J. Electr. Syst., vol. 20, no. 6s, pp. 2491–2501, Jan. 2024. [Online]. Available: https://doi.org/10.52783/jes.3237

[12] S. Yanto, R. Rosziati, and Y. Hidayat, "Effect of CLAHE preprocessing on deep learning accuracy for kidney tumor classification on CT scans," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 4, pp. 225–231, 2024. [Online]. Available: https://doi.org/10.14569/IJACSA.2024.0150429

[13] K.-H. Uhm, S.-W. Jung, S.-H. Hong, and S.-J. Ko, "Lesion-aware cross-phase attention network for renal tumor subtype classification on multi-phase CT scans," Comput. Biol. Med., vol. 178, p. 108746, Aug. 2024. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2024.108746

[14] A. H. Abdulwahhab, N. T. Mahmood, A. A. Mohammed, I. Myderrizi, and M. H. Al Jumaili, "A review on medical image applications based on deep learning techniques," J. Image Graph., vol. 12, no. 3, pp. 215–227, Jul. 2024. [Online]. Available: https://doi.org/10.18178/joig.12.3.215-227

[15] D. Lu, Y. Zheng, X. Yi, J. Hao, X. Zeng, L. Han, Z. Li, S. Jiao, B. Jiang, J. Ai, and J. Peng, "Identifying potential risk genes for clear cell renal cell carcinoma with deep reinforcement learning," Nat. Commun., vol. 16, Art. 3591, Dec. 2025. [Online]. Available: https://doi.org/10.1038/s41467-025-58439-5

[16] "Medical Scan Classification Dataset," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/arjunbasandrai/medical-scan-classification-dataset

[17] Y. Tao, J. Li, X. Wang, et al., "Multi scale unified network for image classification: analyzing input size effects," arXiv, Mar. 2024. [Online]. Available: https://arxiv.org/abs/2403.18294

[18] B. Susanne Schmid, "Image intensity normalization in medical imaging: why and how," Medium, Aug. 15, 2023. [Online]. Available: https://medium.com/@susanne.schmid/image-normalization-in-medical-imaging-f586c8526bd1

[19] A. R. Jones, P. Silva, and T. Wang, "Importance of stratified sampling for use in the development of robust medical imaging AI," in Proc. SPIE Med. Imaging, vol. 13407, 2024, Art. no. 134070Z. [Online]. Available: https://doi.org/10.1117/12.3048942

[20] Q. Huang, X. Liu, J. Li, and Y. Wang, "Deep convolutional neural networks in medical image analysis: architectures, applications, and future trends," Information, vol. 16, no. 3, Art. 195, Mar. 2025. [Online]. Available: https://doi.org/10.3390/info16030195

[21] V. V. Smirnov, K. V. Polovnikov, and Y. S. Alexandrov, “Evaluation metrics for binary classification in medical imaging: definitions, applications and limitations,” Sci. Rep., vol. 14, no. 1, Art. 5678, Apr. 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-56706-x