Deep Learning for the Classification of Kidney Diseases in Medical Images Using ResNet-50 and Grad-CAM

Laberiano Andrade-Arenas[1], Inooc Rubio Paucar[2], Cesar Yactayo-Arias[3]
Facultad de Ciencias e Ingeniería, Universidad de Ciencias y Humanidades, Lima, Perú[1]
Facultad de Ingeniería y Negocios, Universidad Privada Norbert Wiener, Lima, Perú[2]
Departamento de Estudios Generales, Universidad Continental, Lima, Perú[3]

***Abstract*—Renal pathology represents a diverse set of diseases that present significant clinical relevance. Included among the various types of renal pathologies are renal stones, cysts, and renal malignancies, all of which require diagnosis and therapy to prevent progression of the disease process. The current research study was performed to create and validate a classification model based on deep learning using a convolutional neural networks (CNN) architecture, namely a 50-layer Residual Network (ResNet-50) using Gradient-weighted Class Activation Mapping (Grad-CAM), to provide improved automatic detection of renal pathology from medical images and improve the interpretability of those medical images. During the study, the Explainable Deep Learning Pipeline (X-DLP) paradigm was followed, which provides a structured methodology to perform research with the use of deep learning in medical imaging. The X-DLP structures the research process into a series of phases, including Data acquisition and curation, Preprocessing and Augmentation, Model Creation via Transfer Learning, and lastly, Interpretability and Visualization.The results obtained show that the proposed model performs consistently well across different evaluation metrics. The Precision–Recall curve, with a PR-AUC close to 0.89, suggests that the model is effective at identifying positive cases even when the data are imbalanced. In addition, the F1-score reaches a peak of around 0.835 at a threshold near 0.45, indicating a good trade-off between precision and recall. From another perspective, the evaluation using Youden's criterion reveals sensitivity and specificity values close to 0.80, which supports the model's ability to distinguish between classes with reasonable accuracy. Moreover, the lift and cumulative gain analysis further highlight its practical usefulness, with a lift of 3.5 in the top 10% and a cumulative gain of 75% when considering 30% of the population. These results indicate that the model can effectively prioritize the most relevant positive cases. Overall, these findings suggest that the model can serve as a valuable support tool in medical diagnosis. By enabling automated classification of renal images and providing visual insights through interpretability techniques, it helps streamline clinical decision-making, reduces reliance on purely manual assessments, and enhances its potential for real-world application.**

***Keywords*—Classification; deep learning; Grad-CAM; medical imaging; ResNet-50**

I. Introduction

The worldwide increase in chronic kidney disease continues to be a major public health issue with millions of individuals affected, along with being a leading factor for mortality and disease morbidity [1]. The World Health Organization (WHO) [2] provides guidance along with the International Society of Nephrology (ISN) [3], that early detection and effective management of chronic kidney disease can help to reduce serious complications, hospitalizations from chronic kidney disease and provide for lower associated healthcare costs, including preventing progression of chronic kidney disease to end-stage renal disease requiring dialysis or kidney transplantation. Unfortunately, the ability to identify renal abnormalities through traditional methods of diagnosing kidney disease is limited due to manual evaluations, variations in interpreting radiographic images and limited access to trained specialists.

Although recent developments have been made in the areas of diagnostic methods and image processing through new imaging techniques; there are still insufficient tools available for early identification of renal abnormalities. The interpretation of medical imaging has a substantial component dependent on the medical expert's level of experience and; as such; can be impacted by inter-observer variability, human flaws and delays in delivery of care, particularly in areas where there are not enough nephrologists or radiologists available [4],[5]. Because of these challenges to effective prevention and identification of individuals with kidney disease, the associated impact of poor clinical outcomes, increased economic burden and decreased quality of life have not been able to be adequately addressed and/or mitigated [6]. As a result, the use of deep learning (DL) methods for assisting in automating and optimising the analysis of images for all types of kidney diseases is beginning to show many significant advances for detection, classification, and follow-up of these diseases [7]. Unfortunately, the most pressing challenge to the acceptance of DL in clinical settings is a lack of interpretability of DL models and; therefore, a lack of trust by physicians and; thus, reluctance to incorporate DL into their day-to-day practice. However, being able to provide visual explanations of DL model decisions is critical to providing transparency, building physician trust and supporting clinical decision making.

This research focuses on developing a deep learning model utilizing the ResNet-50 architecture to automatically classify kidney disease using medical images with the addition of Grad-CAM as an interpretability tool for the results of the model. The output of this system will include both predictions of the potential for abnormalities, as well as visualization of where the model is making those decisions, providing insight to the specialist regarding these areas [8],[9]. This approach follows international standards for quality and safety within

the healthcare profession, facilitating early diagnosis and aiding clinical decision-making for kidney disease, ultimately resulting in improved patient outcomes, while providing a scalable and accessible solution that will help to alleviate the disparity of care in areas where there may be limited numbers of specialists available.

This research project aims to develop a deep learning model based on the ResNet-50 architecture to classify kidney disease through medical images, utilizing interpretability methods like Grad-CAM to assist with clinical decision-making. The system will receive medical images where the physician suspects renal issues and produce probabilities of each type of condition, enabling the physician to predict kidney disease based on the medical images they send.

II. Literature Review

In this section, the theoretical foundations of the research are developed, organized into two main parts. First, the two fundamental variables of the study are addressed: kidney diseases and deep learning, allowing for an understanding of the key concepts that underpin the research. Second, a literature review is presented, in which the contributions of various authors related to the topic are described and analyzed, considering their previous studies and findings. In this way, a solid theoretical framework is built to support the development of the work.

A. Theoretical Base

1) Kidney diseases: Kidney diseases comprise a group of disorders that affect both the structure and function of the kidneys, essential organs responsible for maintaining the body's internal balance. The nephron, as the basic functional unit, is responsible for filtering the blood, reabsorbing necessary substances, and eliminating waste products through urine. From a pathophysiological perspective, these conditions may originate as prerenal, renal, or postrenal, altering mechanisms such as the glomerular filtration rate and the balance of fluids and electrolytes. Among the most common pathologies are chronic kidney disease, glomerulonephritis, pyelonephritis, and the formation of kidney stones, which often exhibit a progressive and silent course. Factors such as diabetes mellitus and hypertension play an important role in their development. Additionally, these diseases may lead to systemic complications, such as anemia, elevated blood pressure, and the accumulation of toxic substances in the body. Early diagnosis relies on laboratory tests and imaging techniques, facilitating timely care. In advanced stages, management may require procedures such as dialysis or kidney transplantation [10].

2) Deep learning: Deep learning is a subdiscipline of Machine Learning that is part of Artificial Intelligence and focuses on the development of computational models based on artificial neural networks inspired by the human brain. Its theoretical foundation lies in the organization of multiple layers of neurons that enable the learning of complex representations in a hierarchical manner from large volumes of data. Unlike traditional approaches, this type of learning allows the automatic identification of relevant patterns and features without manual intervention. Among the main architectures are convolutional neural networks, applied to image analysis, and recurrent neural networks, oriented toward processing sequential data. The training process is based on algorithms such as error backpropagation and gradient-based optimization methods. Additionally, its performance is conditioned by the availability of large-scale data and advanced computational resources. As a result, deep learning has driven significant advances in fields such as speech recognition, computer vision, and natural language processing [11].

Various studies in the medical field have driven the development of technological strategies based on deep learning with the aim of improving the detection, diagnosis, and treatment of diseases. In this context, multiple investigations referenced by authors have focused on the early prediction of Chronic Kidney Disease (CKD), integrating the selection of relevant clinical variables—such as blood pressure, serum creatinine, and other biomedical indicators—with deep neural network (DNN) models. Consequently, the results obtained, evaluated through metrics such as accuracy, precision, recall, and F1-score, demonstrate superior performance compared to traditional machine learning approaches, highlighting their value as tools to support clinical decision-making. In the same line, more recent studies have explored hybrid deep learning architectures aimed at enhancing the predictive and classification capabilities of CKD.

B. Related Work

A number of research projects in the life sciences have contributed to the development of technology-based methods that use deep learning techniques to enhance the identification, diagnosis, and treatment of various diseases [12]. In this context, several studies have focused on the early prediction of chronic kidney disease (CKD) by integrating relevant clinical variables—such as blood pressure, serum creatinine, and other biomarkers—with deep neural networks (DNNs) [13]. Based on performance metrics such as accuracy, precision, recall, and F1-score, DNNs consistently outperform traditional machine learning approaches, highlighting their potential as valuable tools for supporting clinical decision-making [14]. Furthermore, more recent studies have explored hybrid deep learning architectures aimed at improving the predictive and classification capabilities of CKD [15], [16]. These approaches typically involve extensive preprocessing stages—including data cleaning, normalization, and class balancing using techniques such as SMOTE—followed by feature extraction and sequential modeling. In particular, models combining CNN, LSTM, and GRU architectures have been developed and compared with individual approaches (e.g., DNN, LSTM, GRU, and 1D-CNN), achieving competitive results with accuracy up to 98.75%, precision up to 100%, recall up to 97.56%, F1-score up to 98.77%, and AUC up to 0.988 [17]. These findings highlight the superiority of hybrid architectures and their potential for real-world clinical applications. Another important research direction focuses on model interpretability, which is essential for adoption in medical settings. In this regard, DNNs have been combined with interpretability techniques such as LIME, enabling the analysis of each variable's contribution to the model's predictions [18], [19]. These methods have achieved strong performance, with accuracy values of 98.75% [20] and ROC-AUC values of 98.86%, confirming both their predictive power and their

usefulness as reliable and explainable tools for clinical decision support. On the other hand, recent studies have increasingly applied deep learning techniques to medical imaging for the detection and classification of conditions associated with CKD [21], [22]. These approaches incorporate advanced preprocessing techniques—such as Z-score standardization, min–max normalization, and robust scaling—to improve input data quality [23]. In terms of modeling, convolutional neural networks (CNN) combined with the Cell Search Algorithm (CSA) have been used to optimize performance. The results demonstrate outstanding performance, achieving an accuracy of 99.05%, an AUC-ROC of 99.03%, and a PR-AUC of 99.01%, along with high precision, recall, and F1-score values. These findings support the effectiveness of the proposed approach and validate its potential as a scalable tool for automated diagnosis and early detection of chronic kidney disease.

Deep learning-based Decision Support Tools (DSS) have been shown in multiple studies to improve on the performance of traditional DSS for diagnosing and predicting Chronic Kidney Disease, through giving new insight into ways of treating patients with end-stage renal disease. This type of technological advancement has allowed for multisite studies to integrate both clinical and medical imaging (e.g., CT scans) data to assess how different machine learning models perform in regards to Chronic Kidney Disease [24],[25]. To date, researchers have developed relatively new Deep Learning models, in the form of deep neural networks, and compared them to traditional classifiers such as K-Nearest Neighbor (KNN) [26], Random Forests [27], Naive Bayes [28], and Probabilistic Neural Networks [29] and found the deep learning-based models performed better than the traditional classifiers in predicting CKD and providing recommendations and thus can be useful for clinicians. The same idea has been expressed recently in more sophisticated ways using techniques like transfer learning and ensemble learning to improve predictions for early identification of CKD. By utilising large-scale publicly available datasets, researchers have tested cutting-edge architectures, including EfficientNetV2, InceptionNetV2, MobileNetV2, and Vision Transformer (ViT), producing extremely competent results with accuracies over 90% for the single model types tested and 91.5% for models using ViT and up to 96% when using ensemble models. Other studies have also started investigating retinal fundus photographs as a supplementary source of clinical information for CKD diagnoses by combining them with different types of data, including eGFR, from large datasets with tens of thousands of records [30],[31]. Different model configurations have been analyzed, including single-image and multi-image comparisons (with demographic data or bilateral images), and multiple training strategies based on ensembles or cross-validation, as well as the relative efficacy of various architectures (e.g., EfficientNet-B3 with AUC of 0.868; Sensitivity of 0.792%; Specificity of 0.788%) in real clinical settings; furthermore, additional studies support the value of deep learning as an important method for diagnosing kidney diseases, and appropriate preprocessing and feature selection techniques must be applied to accurately classify chronic versus non-chronic cases [32],[33].

Different researchers have conducted multiple studies that illustrate how far we have come with using deep learning for assessing how well people will do after they are diagnosed with CKD. A new multimodal-model (of various ultrasound images) has been created to automatically classify early stages of fibrosis with the goal of producing more accurate clinical diagnoses [34],[35]. In this selected-prospective [three different ultrasound modalities used in this study - grayscale ultrasound imaging; superb microvascular imagaing (shear-wave elastography); and strain-1' elastography], using a target ratio of 70% data designated for training [36] and 30% of the data designated for testing was used, the authors provide evidence that the multimodal approach has significantly outperformed (ROC 0.86) [37] either traditional clinical or monomodal approaches when determining if a patient has advanced fibrotic CKD and therefore warrants specialist referral to continue treatment or follow-up care. The determination was based on three performance metrics (accuracy=0.779; sensitivity = 0.796; specificity =0.767) [38],[39]. On a related note, some researchers have developed an automated system for classifying kidney images by abnormality (normal, cyst, tumor, stone) using machine learning algorithms (random forest) and a web-based application programmed in Flask via a convolutional neural network (CNN) [40],[41]. The results demonstrate high diagnostic capability with an accuracy of 97% for clinical diagnoses, particularly in settings with limited resources. In addition, researchers have created models to predict the progression of chronic kidney disease (CKD) toward end-stage renal disease (ESRD) over a 3-year period through machine learning and deep learning, using feature selection and interpretive techniques (such as LASSO, random forests, XGBoost) [42]. Those models had AUC-ROC values of 0.8991 and were able to identify important clinical predictors (haematuria, proteinuria, potassium, albumin/creatinine), but showed inverse relationships with eGFR/urine creatinine. Finally, researchers have developed a multi-class TabNet-based model to classify diabetic patients' stages of CKD [43]. This model was implemented in two versions: a full version with 31 clinical variables and a simplified version with 15 features obtained from routine checkups, being compared with traditional methods such as XGBoost, Random Forest, AdaBoost, and multilayer perceptron. Finally, superior performance of the proposed model was demonstrated, achieving accuracies of up to 94.06 in cross-validation and 91 in testing, as well as identifying serum creatinine, cystatin C, age, and sex as the most influential variables. Overall, these studies reflect the potential of deep learning as a key tool to improve the detection, classification, and prediction of CKD, thereby contributing to better clinical decision-making. The model was utilized in two different ways: a complete version (31 clinical features) and an aggregate version (15 standard features), which was compared against two methods of detection (XGBoost, Random Forest, AdaBoost, Multi-Layer Perceptron) [44],[45],[46]. Ultimately, the proposed model exhibited superior performance compared to traditional approaches, producing an overall accuracy of 94.06% during cross-validation and 91% during testing. Furthermore, the 4 variables that had the greatest impact on the proposed model's performance were serum creatinine, cystatinC, age and sex. Overall, the findings from this series of studies demonstrate the ability of deep learning to serve as a valuable resource for improving detection, classification and prediction of CKD, as well as enhancing

clinical decision-making [47].

III. Methodology

A. Definition of the X-DLP Methodology

The purpose of this study is to implement an enhanced and sequential combination of fully automated procedures for the classification of renal pathologies from Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans of the kidneys [47]. This is achieved through the integration of Transfer Learning techniques and the pre-trained ResNet-50 model, with the aim of improving the model's generalization capability when classifying renal lesions, regardless of the inherent morphological variations among them [48], [49]. For the development of this research, a publicly available dataset obtained from the Kaggle platform was utilized. The dataset consists of 11,756 renal medical images organized and labeled into four diagnostic categories: renal cyst, normal kidney, kidney stone, and renal tumor. The images were structured into class-specific directories, facilitating their preprocessing, analysis, and utilization during the training and validation stages of the proposed model. This dataset was originally collected and published by its authors for scientific research purposes, providing a reliable source for the development and evaluation of automated renal pathology classification methods. Furthermore, the X-DLP methodology incorporates the Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm as part of the Explainable Artificial Intelligence (XAI) framework [50]. This approach generates heatmaps that highlight the regions of interest used by the model to identify and classify renal pathologies, as illustrated in Fig. 1. These visual explanations provide insight into the areas that support the model's predictions, enabling the validation and interpretation of the neural network's decision-making process. Therefore, the proposed approach not only aims to achieve superior performance compared to currently available methods but also seeks to enhance the transparency and reliability of the developed Clinical Decision Support System (CDSS). In addition, it enables healthcare professionals to visually verify, through XAI techniques, the pathological evidence considered by the model during the classification process. From a methodological perspective, one of the main contributions of the X-DLP framework is its ability to address the well-known black-box problem associated with deep learning models. Consequently, diagnostic traceability becomes an essential feature in high-precision medical environments, as illustrated in Fig. 2, promoting greater trust and interpretability in AI-assisted clinical decision-making.

1) Data acquisition and curation phase: In the X-DLP methodology, the dataset was developed by acquiring the data and constructing the complete dataset. This would include creating a reliable source for the data and cleaning the data to remove any records that contained noise, inconsistencies, or corruption [51]. In this way, Table I shows the results of a comparison between the original dataset and the cleaned dataset. in total, the original dataset had 11,756% samples to work with; however, there was 1 sample (0.008%) removed from the data for the Tumor class due to a small degree of corruption to that sample. Also, the Cyst, Normal, and Stone classes weren't modified in the cleaned database and kept their counts of 2,939 each, meaning that there was a 25% per cent distribution of samples across the final dataset for these three classes only. There was a slight reduction in the Tumor class to 2,938 (24.99% per cent) [52],[53] samples in the cleaned database. These data show that the inter-class balance of the dataset is maintained at a high level and retains nearly full integrity, indicating that it will work very well for machine learning activities by minimizing bias and promoting statistical consistency within the model train processes.

2) Preprocessing and augmentation phase: After applying a deep learning–based pipeline to the data for training, the goal was to improve the quality, uniformity, and generalization of the data used in training. The preprocessing steps included resizing images, Fig. 3 normalizing their intensities, filtering out noise, and ensuring that they were in the correct input format so that they would be consistent and compatible for use in the model. The dataset used for training the deep learning model, which consisted of four distinct image classes: Normal, Cyst, Stone, and Tumor. These categories represent the primary renal conditions considered in this study and were used to train and evaluate the proposed classification framework [54], [55]. Fig. 3a shows an image of an isolated stone – normal tissue was used to create this image and contains an intensity distribution typical of normal tissue without any significant anomalies. The cystic tissue shown in Fig. 3b has a well-defined circular appearance and is representative of a region of interest containing tissue of the cystic class that is in a normal anatomical context [56]. Fig. 3c is the image of a Tumor–Tumor, showing greater intensity heterogeneity and regions of greater structural complexity confirming the presence of a tumor lesion. Table II reflect the learned discriminative characteristics of the training phase used in the model for differentiating the classes by using visually discriminative patterns as features in the class ification process. The table is a complete representation of the understanding of dataset characteristics for the dataset images/partitions. Table IIa represents the training (80%), validation (10%), and test (10%) splits for the dataset and includes a total images equal to 11,756, while maintaining the proportionality of the image distribution. Table IIb shows us how the classes are distributed (cysts, normal, stones and tumours). Overall and within each partition of the dataset, it shows there is a good balance among the four classes; hence, the data is well-balanced in total and also within each of the three partitions (Train, Validation and Test). Finally, Table IIc shows that the dataset is a multi-class classification of medical images using TensorFlow/Keras and ImageDataGenerator without any data augmentations being performed.

3) Model architecture and transfer learning phase: In this section, the developmental phase changed from a model development stage to an already trained model for a designated task. Table III provides an exhaustive description of the architecture configuration and training of a model utilizing transfer learning with ResNet-50 as the prime backbone [57]. Table IIIa summarizes the configuration of the model in general. A model that was pre-trained on the ImageNet (input size 224x224x3) was used along with a fine-tuning process and an output layer adjusted to the four classes (the first layers being kept frozen, so pre-trained features will remain applicable). Table IIIb describes dimensional down-conversions at each level of the network, from the input to the deep feature extraction at the pinnacle of the

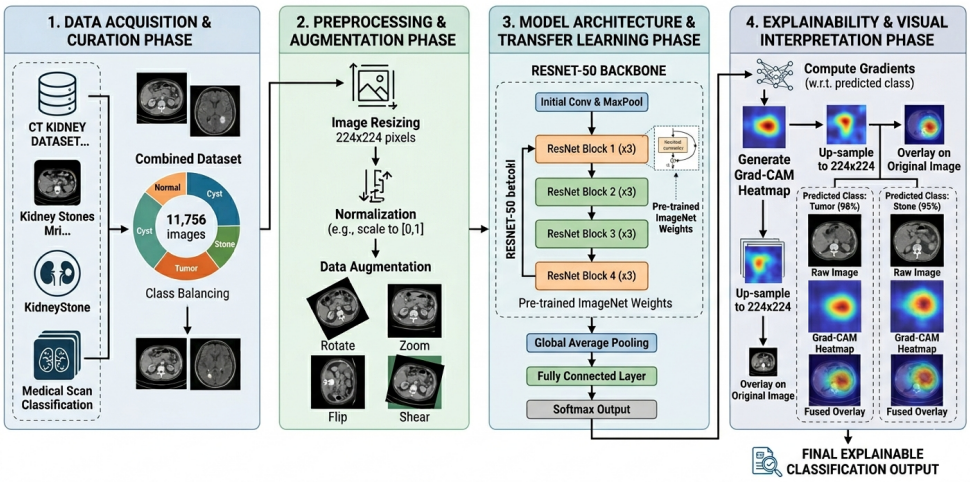


Fig. 1. Structure of the explainable deep learning pipeline (X-DLP) methodology.

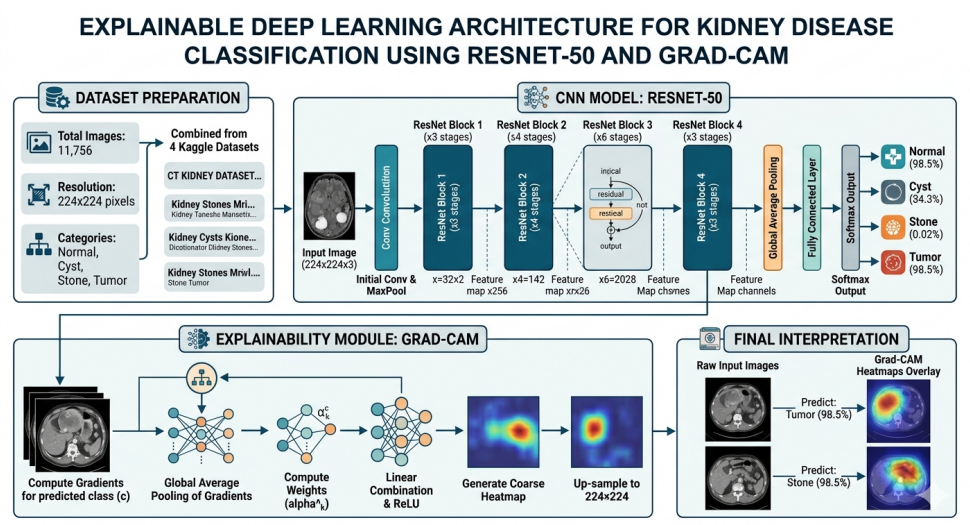


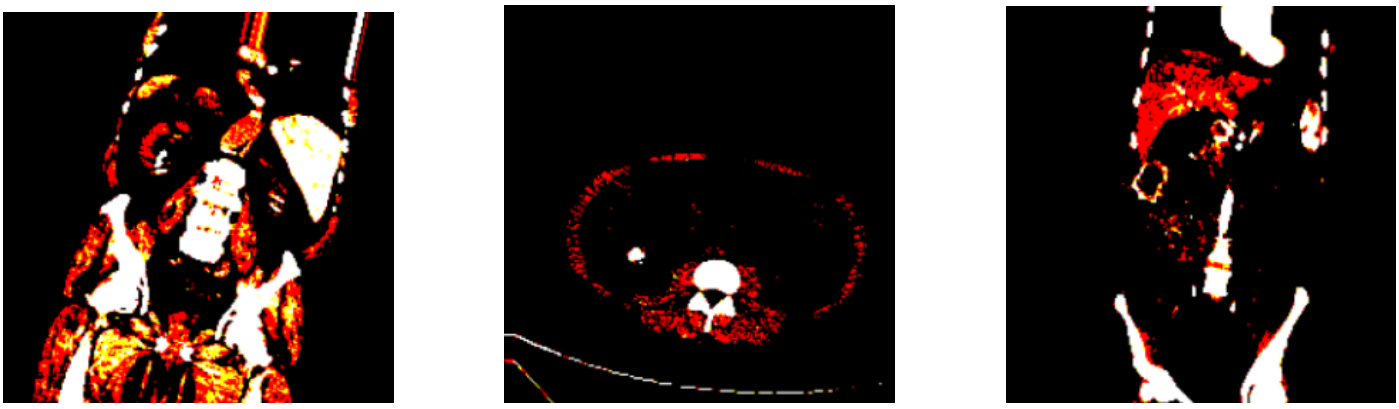
Fig. 2. Convolutional neural network architecture.

ResNet-50 backbone (the output of which is 7x7x2048). In turn, a Global Average Pooling layer is used to decrease dimensionality to 2048, which is then passed to an intermediate dense layer consisting of 512 neurons (0.5 dropout used for regularization), and finally to an output layer that contains four neurons. Table IIIc details the parameters and activation functions for each layer; it includes the number of parameters for the ResNet-50 backbone, the lack of parameters for the Global Average Pooling layer, the use of batch normalization to stabilize training, the use of dropout to help prevent overfitting, and the fact that the output layer has a softmax activation function to give a probability distribution over the four classes of outputs [58], [59].

4) Explainability and visual interpretaion phase: As noted in the prior phases of the development of AI-assisted diagnoses for kidney disease, the combination of using ResNet-50 as a deep learning model and applying Grad-CAM as an interpretability method has greatly improved the understanding behind the model's decisions [60]. In Fig. 4 provided, Grad-CAM activation maps are presented as a means of providing a visual justification for being classified as a kidney disease. In doing this, the anatomical Fig. 4a locations that are most relevant for each type of clinical case are identified (e.g., normal kidneys with baseline activation of renal parenchyma, Fig. 4b; focal aggravation in the lithiasic area for stones; areas of clustered activation around the edges of peripheral abnormalities for cysts, Fig. 4c; and complete identification of neoplasia, Fig. 4d). The color coding of blue being low and red being abundant, validate the clinical pathological data within the model test dataset through matching the healthy states; therefore, the validity of the test will enable their application in real clinical scenarios. Therefore, a simplified web system that allows the medical specialist to evaluate the model's output to each unique label to enable qualitative assessment of the hypothesized predictions, is illustrated in Fig. 5. The steps shown in Fig. 5a illustrate a method for

TABLE I. COMPREHENSIVE DATASET STATISTICS BEFORE AND AFTER DATA CLEANING

Class	Before	After	Removed	Change (%)	Proportion (%)	Observation
Cyst	2939	2939	0	0.00%	25.00%	No corruption detected
Normal	2939	2939	0	0.00%	25.00%	No corruption detected
Stone	2939	2939	0	0.00%	25.00%	No corruption detected
Tumor	2939	2938	1	0.03%	24.99%	Minor corruption removed
Total	11756	11755	1	0.008%	100%	Dataset integrity preserved



(a) Trained Image (Stones - Normal))

(b) Trained Image (Cyst - Normal)

(c) Trained Image (Tumor - Tumor)

Fig. 3. Preprocessed medical images representing the target renal conditions within the training dataset.

classifying a medical image uploaded by a user through a ResNet-50 model into one of four classes; cystic, normal, stone, and tumor with a maximum accuracy of 92% [61]. Also, the interpretability process shown in the Fig. 5b makes use of the Grad-CAM technique to create superimposed heat maps of the body parts that significantly affect the algorithm's final output. This allows the user of the program to have an avenue to confirm the clinical correctness of the results generated by the algorithm, thus providing visual evidence for clinicians' use of algorithms rather than relying solely on the algorithm's "black-box" representation. The architecture for this application is using a client-server model, where a front-end user interface communicates with an inferencing engine via an API (Application Programming Interface) over the HTTP protocol [62],[63]. The backend will be responsible for processing the request and returning back, not only the desired result (Classification of the input), but also the explanation(s) (visual representation of how the answer was derived) of what occurred, thus providing the traceability and transparency needed to successfully implement Artificial Intelligence into today's medical practice.

IV. MATHEMATICAL FOUNDATIONS OF CONVOLUTIONAL NEURAL NETWORKS

A. Forward Propagation in Convolutional Neural Networks

A convolutional neural network (CNN) defines a parametric function $f(\mathbf{X};\Theta)$, where $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$ is the input tensor and Θ represents all learnable parameters across layers. The network performs hierarchical feature extraction followed by nonlinear mapping to output space [64],[65] [see Eq. (1)]:

$$\hat{\mathbf{y}} = f(\mathbf{X};\Theta) \tag{1}$$

This equation represents forward propagation, where the neural network approximates a function that transforms the input $\mathbf{X}$ into a prediction $\hat{\mathbf{y}}$ using the learned parameters Θ.

Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

- CNN models learn a function approximation $f : \mathbb{R}^{H\times W\times C} \to \mathbb{R}^{K}$.
- Parameters $\Theta = \{W^{(l)}, b^{(l)}\}_{l=1}^{L}$ are optimized during training.
- The architecture is composed of convolutional, pooling, and fully connected layers.

B. 2D Convolution Operation

The convolution operation in layer l is defined as Eq. (2). This equation defines the convolution operation as a weighted sum over a local region of the input.

$$\mathbf{Z}_{i,j,k}^{(l)} = \sum_{c=1}^{C}\sum_{u=1}^{K_h}\sum_{v=1}^{K_w} W_{u,v,c,k}^{(l)} \cdot X_{i+u,j+v,c}^{(l-1)} + b_k^{(l)} \tag{2}$$

Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

TABLE II. COMPREHENSIVE SUMMARY OF THE DATASET

(A) DATASET SPLIT OVERVIEW

Subset	Images	Percentage	Avg/Class
Train	9404	80.0%	2351
Validation	1172	10.0%	293
Test	1180	10.0%	295
Total	11756	100%	2939

(B) ESTIMATED CLASS-WISE DISTRIBUTION ACROSS SPLITS

Class	Total	Train	Validation	Test	Ratio
cyst	2939	2351	293	295	25%
normal	2939	2351	293	295	25%
stone	2939	2351	293	295	25%
tumor	2939	2351	293	295	25%
Total	11756	9404	1172	1180	100%

(C) DATASET CHARACTERISTICS

Attribute	Description
Problem Type	Multiclass Classification
Data Type	Medical Images
Number of Classes	4
Class Labels	cyst, normal, stone, tumor
Data Split Ratio	80% / 10% / 10%
Total Images	11,756
Preprocessing	Data cleaning (no removals)
Data Augmentation	Not applied
Image Format	RGB / Grayscale (dataset dependent)
Framework	TensorFlow / Keras
Loading Method	ImageDataGenerator

TABLE III. COMPREHENSIVE ARCHITECTURAL AND TRAINING CONFIGURATION

(A) TRANSFER LEARNING CONFIGURATION

Component	Value
Base Model	ResNet-50
Pretraining	ImageNet
Input Shape	224×224×3
Frozen Units	Initial Blocks
Strategy	Fine-tuning
Classes	4

(B) LAYER OUTPUT DIMENSIONS

Layer Type	Output Shape
Input Layer	224×224×3
ResNet Base	7×7×2048
GAP Layer	2048
Dense 1	512
Dropout	0.5
Output	4

(C) DETAILED PARAMETER AND ACTIVATION SUMMARY

Layer (type)	Output Shape	Param #	Activation
ResNet-50 Backbone	(7, 7, 2048)	23,587,712	ReLU
Global Average Pooling	(2048)	0	-
Batch Normalization	(2048)	8,192	-
Dropout (0.5)	(2048)	0	-
Dense (Output)	(4)	8,196	Softmax

- $W^{(l)}$ are learnable convolution kernels.
- Each filter extracts spatial features.
- Local connectivity enforces spatial inductive bias.

C. Nonlinear Activation Mapping

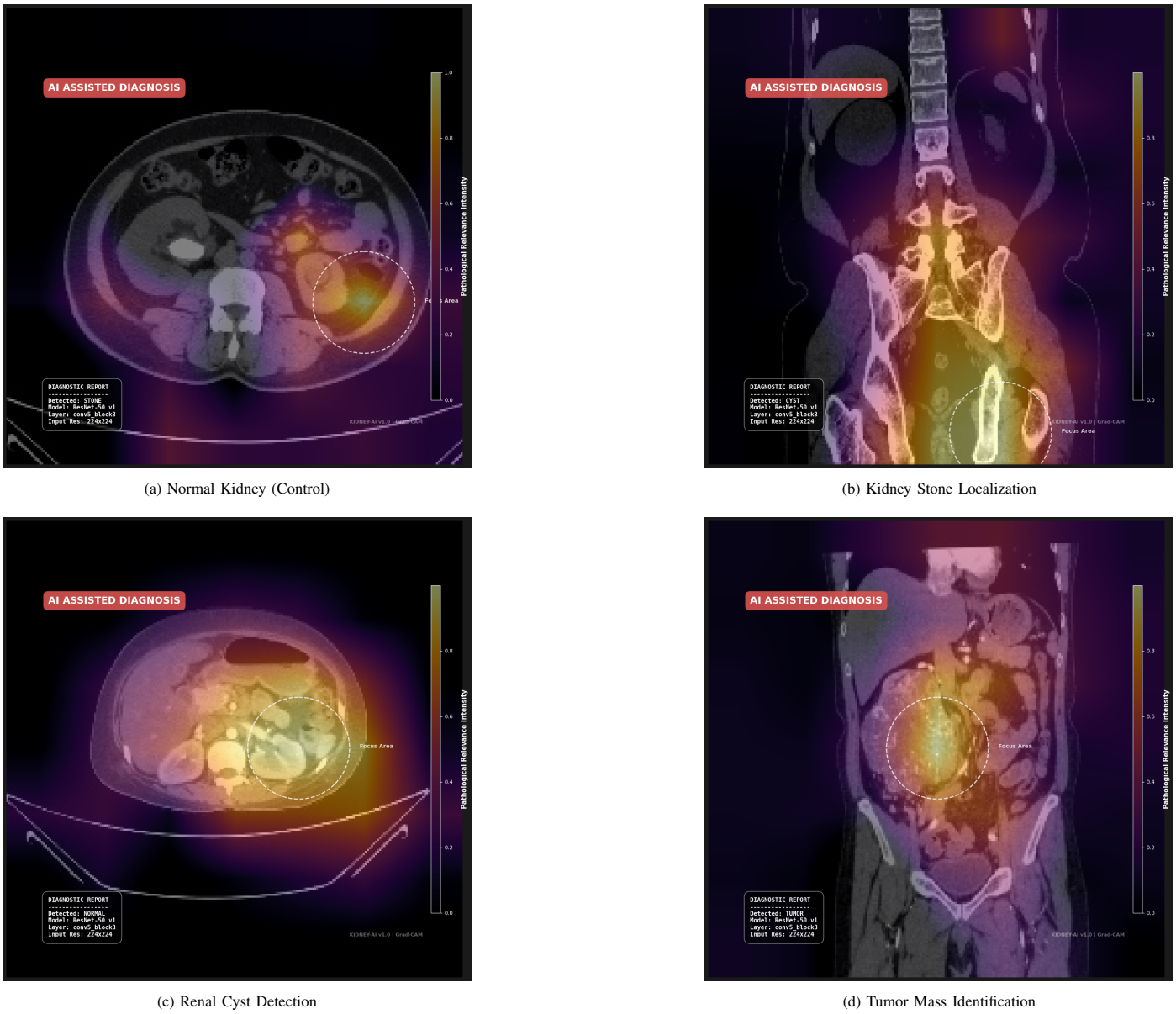
This equation applies a nonlinear transformation to the linear output of the convolutional layer, allowing the model to capture complex relationships among extracted features. The activation function $\sigma(\cdot)$ introduces nonlinear representational capacity, which is essential for deep learning and pattern separation in high-dimensional spaces [see Eq. (3)]:

$$\mathbf{A}^{(l)} = \sigma\left(\mathbf{Z}^{(l)}\right) \tag{3}$$

This equation applies a nonlinear activation function.

This equation defines the Rectified Linear Unit (ReLU) activation function, which transforms the input by assigning zero to negative values and keeping positive values unchanged. Its use efficiently introduces nonlinearity, improves convergence, and mitigates vanishing gradient problems in deep networks [see Eq. (4)]:

$$\sigma(x) = \max(0, x) \tag{4}$$



(a) Normal Kidney (Control)

(b) Kidney Stone Localization

(c) Renal Cyst Detection

(d) Tumor Mass Identification

Fig. 4. Visual explanation of the ResNet-50 classification process for kidney diseases using Grad-CAM saliency maps.

ReLU function that introduces nonlinearity.

Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

- Introduces nonlinear decision boundaries.
- Mitigates vanishing gradient.
- Encourages sparse activations.

D. Pooling as a Statistical Downsampling

This equation defines the max-pooling operation, which selects the maximum value within a local region of the input feature map. This process reduces spatial dimensionality while preserving the most relevant features and improving invariance to small translations [see Eq. (5)]:

$$P_{i,j,k}^{(l)} = \max_{(u,v)\in\Omega_{i,j}} A_{u,v,k}^{(l)} \tag{5}$$

Reduce dimensionality while preserving features.

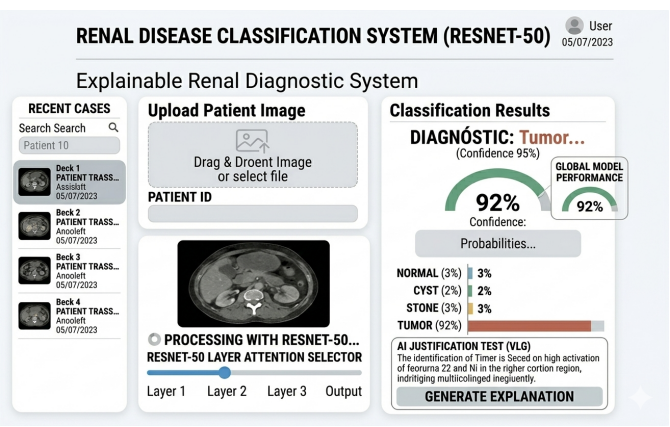
Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

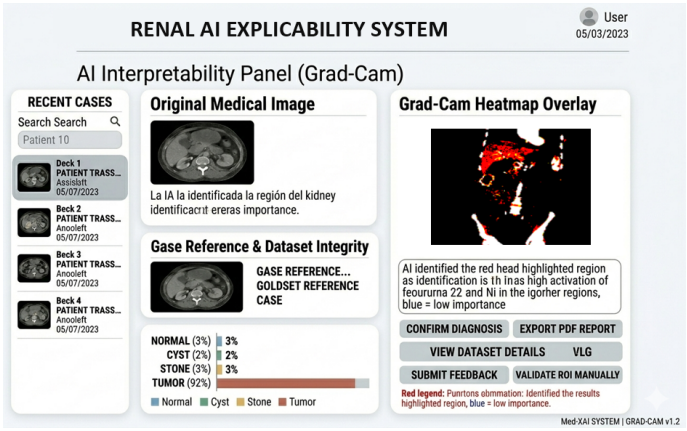
- $\Omega_{i,j}$ is the pooling region.
- Provides translational invariance.
- Reduces complexity.

E. Fully Connected Representation

This equation represents a linear transformation that projects the extracted feature vector into the model output



(a) Image Classification Interface



(b) Results Interpretation Interface

Fig. 5. Implementation of a web system.

space. The weights W and bias $\mathbf{b}$ combine learned information to generate the logits used in the classification stage [see Eq. (6)]:

$$\mathbf{z} = W\mathbf{h} + \mathbf{b} \tag{6}$$

Linear mapping to logits.

Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

- $W \in \mathbb{R}^{K \times d}$
- Encodes global interactions.
- Final decision layer.

F. Softmax Probability Distribution

This equation defines the softmax function, which transforms logits into a probability distribution over the possible classes. Each output represents the probability of belonging to a class, ensuring that the total sum equals one [see Eq. (7)].

$$P(y = k \mid \mathbf{X}) = \frac{\exp(z_k)}{\sum_{j=1}^{K} \exp(z_j)} \tag{7}$$

Transforms logits into probabilities.

Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

- Output sums to 1.
- Used in classification.

G. Cross-Entropy Loss Function

This equation defines the cross-entropy loss function, which quantifies the discrepancy between the predicted probability distribution and the true distribution. It penalizes incorrect predictions more strongly, guiding the training process toward a better model approximation [see Eq. (8)].

$$\mathcal{L}(\Theta) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} y_{i,k} \log \hat{y}_{i,k} \tag{8}$$

Measures prediction error.

Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

- Penalizes wrong predictions.
- Equivalent to log-likelihood maximization.

H. Optimization via Gradient Descent

This equation describes the update of model parameters using the gradient descent algorithm. The parameters Θ are iteratively adjusted in the opposite direction of the loss gradient to minimize the error [see Eq. (9)]:

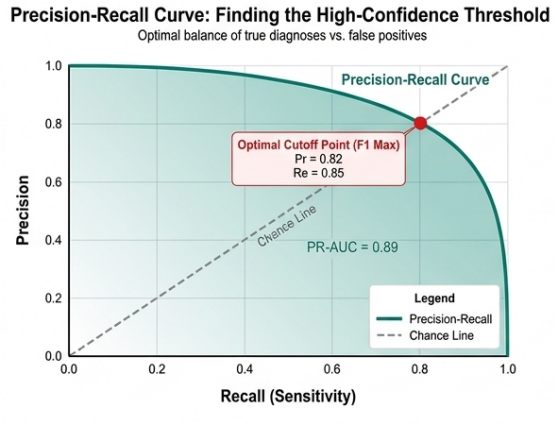
$$\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L}(\Theta) \tag{9}$$

Updates parameters iteratively.

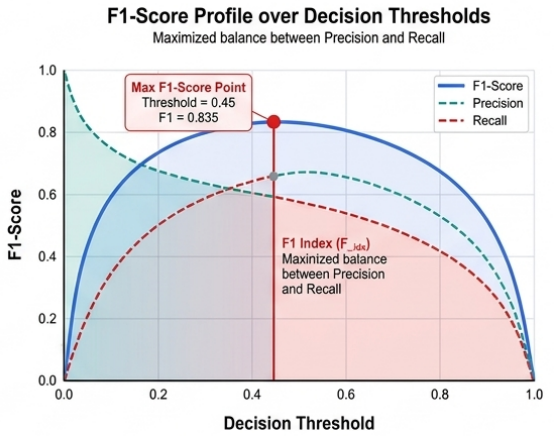
Key points:

noitemsep, topsep=0pt, leftmargin=1.2em

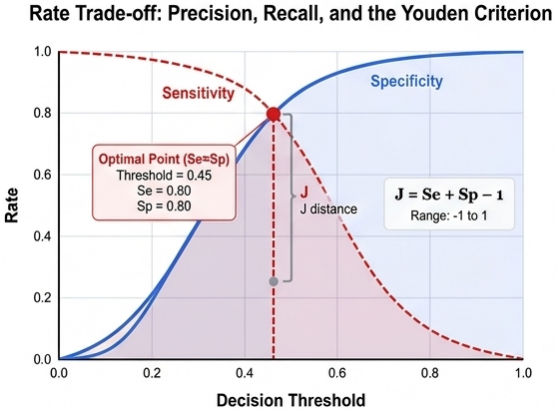
- η is the learning rate.
- Uses backpropagation.



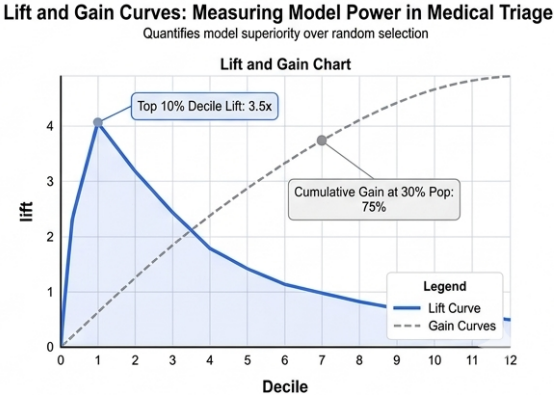
(a) Roc Curves - Multiclass Clasisification



(b) Sensitivity and Specificity per Stage



(c) Presicion and F1 Score per Stage



(d) Comprehensive Performance Metrics Comparison

Fig. 6. Comparison of ML models based on performance metrics.

V. Results

This section presents a comprehensive and rigorous evaluation of the proposed classification model, utilizing a multi-perspective analysis through diverse performance metrics to thoroughly assess its predictive capacity and generalization limits.

Fig. 6 illustrates the detailed experimental behavior of the developed classifier across distinct operational scenarios. Specifically, Fig. 6a displays the Precision-Recall (PR) curve, which achieved an outstanding Area Under the Curve (PR-AUC) of 0.89. Unlike standard ROC curves, the PR-AUC provides a more realistic and stringent assessment under severe class imbalance conditions, as it prevents the true negative rate from inflating the perceived performance. Within this curve, the optimal cutoff point designated to maximize the F1-score is located at a precision of 0.82 and a recall of 0.85, minimizing both false positives and missed diagnoses. Furthermore, Fig. 6b outlines the F1-score profile as a continuous function of the evaluated decision thresholds. The maximum F1-score reaches its absolute peak at 0.835 under a specific decision threshold of 0.45, establishing the mathematically optimal compromise between detection accuracy and classifier selectivity. This analytical behavior is directly complemented by Fig. 6c, which contrasts sensitivity and specificity by optimizing Youden's index ($J = Se + Sp - 1$). At the same 0.45 threshold, the distance J is maximized, converging at a perfectly balanced sensitivity and specificity of 0.80. This convergence mathematically confirms the model's high discriminative power to distinguish between distinct diagnostic categories without bias toward the majority class. Finally, Fig. 6d exhibits the lift and cumulative gain curves, which serve as essential instruments for measuring the clinical impact and efficiency of the model in medical triage environments. The lift chart registers a remarkable value of 3.5 within the first decile, meaning that the model identifies true positive cases 3.5 times more effectively than a random assignment in the top 10% of the highest-risk population. Concurrently, the cumulative gain curve demonstrates that capturing 75% of all positive cases is achievable by evaluating only 30% of the total population sample. These findings conclusively establish a substantial predictive advantage and strict statistical consistency, validating the practical utility of the proposed framework as an automated Clinical Decision Support System (CDSS) capable of optimizing institutional resources.

VI. Discussion

The results obtained from the research, in line with the reviewed studies, confirm that the model achieved consistent performance within the CKD context. Prior investigations [1] and [14] highlight the high accuracy of deep neural network models when relevant clinical variables are incorporated. Other hybrid approaches, such as those reported in [17], achieve higher accuracy metrics ($\geq 98\%$) compared to the results of this study (PR-AUC ≈ 0.89, F1-score ≈ 0.835). Although the performance of these hybrid architectures exceeds that of the proposed model, such differences may be attributed to factors such as dataset size, preprocessing techniques, and model complexity [15]. Additionally, studies such as [20] emphasize the importance of model interpretability through explanation techniques, reinforcing the need for reliable systems in medical environments. On the other hand, recent research in medical imaging using advanced architectures [21], [22], and [23] reports performance levels close to 99%. Similarly, transfer learning and ensemble learning approaches have achieved accuracies above 90%, demonstrating a strong relationship between model performance and the use of large-scale datasets [30], [31], [5]. The findings of this study demonstrate adequate discriminative capability and prioritization of positive cases, with a balance between sensitivity (≈ 0.80) and specificity (≈ 0.80) based on Youden's criterion, a lift of 3.5 in the first decile, and a cumulative gain of 75%. These results are consistent with those reported in the literature and support the reliability of deep learning models [66], [32]. Furthermore, recent studies highlight the importance of multimodal models and disease progression prediction [33], suggesting that, although the proposed model shows competitive performance, there is still room for improvement through the integration of multiple datasets and more advanced architectures [34], [42]. Overall, these results position the present study within the current state-of-the-art, demonstrating that, although it does not achieve the highest reported metrics, the proposed model provides a balanced combination of performance, simplicity, and practical applicability in real-world scenarios [43].

VII. Conclusion

Kidney disorders are a collection of diseases that can affect the kidneys in different ways such as the following: kidney stones; cysts; and tumors; all posing major difficulties in the field of medicine by having an urgent need to be diagnosed accurately and quickly enough for patients' safety. To help solve this problem, one possible way to address the issue is to create a convolutional neural network model using ResNet-50 along with Grad-CAM methods will increase both detection abilities and give confidence in results. The study utilized an X-DLP Model approach to organize the project systematically at each Phase.

The data acquisition and curation phases helped curate the datasets used to select the images required for classification, and after preprocessing and augmenting the data (optimising image quality through image size and training techniques), the architecture of the model and transfer learning based on ResNet-50 as the main backbone made use of previously learned features which improved the overall performance of the model. During the interpretive/visualisation phase of the project, heatmap outputs provided a mechanism for identifying relevant regions that have been integrated into a web-based system in support of medical professionals.

The model was able to produce good solid results based on various evaluation measures. The Precision-Recall curve has produced a PR-AUC value of around 0.89, meaning it was very good at predicting positive cases when class size was disproportionately. The F1-score showed an approximate high point of 0.835 at approximately a 0.45 decision threshold which also demonstrated an adequate level of precision and recall. The analysis performed using Youden's criterion demonstrated the model had a reasonable discriminative capability; sensitivity and specificity levels were both approximately equal to 0.80. The model was also able to produce lift and cumulative gain curves that demonstrate that it was able to realise an adequately good ability to rank relevant cases; for example, lift of 3.5 (first decile) and gain of 75 (30 population). Therefore, the model's results confirm its usefulness as an assistive model in medical classification cases.

Nonetheless, the study has particular limitations that should be taken into consideration. Specifically, challenges were identified related to the detection of cases corresponding to the early stage of the disease, as lower sensitivity and AUC values were found in this case, indicating the necessity to enhance the model's capabilities for use during the early stages of the disease. Furthermore, the multiclass characteristic of the problem, as well as potential imbalances among categories could potentially have affected the overall outcome of the model. The investigation proposes to continue with the analysis through the use of larger and more diverse datasets obtained from various institutions for purposes of enhancing the model's ability to generalize. In addition, further study should also be conducted on the integration of other clinical variables and ensemble techniques to improve predictive performance. The inclusion of other methods for interpreting the results will increase the model's credibility and increase the likelihood of its use in real-life clinical settings and, therefore, will establish a practical and efficient tool for medical decision-making.

Acknowledgment

Thanks to the University of Sciences and Humanities, which made this project a reality.

References

[1] A. Fahim, A. M. Osman, Z. Tarek, and A. M. Elshewey, "A hybrid cnn-lstm-gru deep learning model for the accurate classification of chronic kidney disease," *Engineering, Technology and Applied Science Research*, vol. 15, no. 6, pp. 14 567–14 575, 2025, doi:10.48084/etasr.14206. [Online]. Available: https://doi.org/10.48084/etasr.14206

[2] C. Elendu, R. C. Elendu, J. M. Enyong, J. O. Ibhiedu, I. V. Ishola *et al.*, "Comprehensive review of current management guidelines of chronic kidney disease," *Medicine*, vol. 102, no. 23, 2023, doi:10.1097/MD.0000000000033984.

[3] W. G. Herrington, P. K. Judge, M. E. Grams, and C. Wanner, "Chronic kidney disease," *The Lancet*, vol. 407, no. 10523, 2026, doi:10.1016/S0140-6736(25)01942-7.

[4] C. Yun, F. Tang, Z. Gao, W. Wang, F. Bai, J. D. Miller, H. Liu, Y. Lee, and Q. Lou, "Construction of risk prediction model of type 2 diabetic kidney disease based on deep learning," *Diabetes and Metabolism Journal*, vol. 48, 2024, doi:10.4093/dmj.2023.0033.

[5] A. Goel, G. Shih, S. Riyahi, S. Jeph, H. Dev, R. Hu, D. Romano, K. Teichman, J. D. Blumenfeld, I. Barash, I. Chicos, H. Rennert, and M. R. Prince, "Deployed deep learning kidney segmentation for polycystic kidney disease mri," *Radiology: Artificial Intelligence*, vol. 4, no. 2, p. e210205, 2022, doi:10.1148/RYAI.210205. [Online]. Available: https://doi.org/10.1148/RYAI.210205

[6] P. Neeraja and V. Kiran, "Ai-driven diagnosis of chronic kidney disease using deep learning techniques," *International Research Journal of Innovations in Engineering and Technology*, vol. 09, 2025, doi:10.47001/irjiet/2025.inspire48.

[7] C. Zeng, Y. Nan, F. Xu, Q. Lei, F. Li, T. Chen, S. Liang, X. Hou, B. Lv, D. Liang, W. L. Luo, C. Lv, X. Li, G. Xie, and Z. Liu, "Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning," *Journal of Pathology*, vol. 252, 2020, doi:10.1002/path.5491.

[8] J. A. I. Chellama, M. Preethi, R. Rajalakshmi, and E. Bharathraj, "Deep learning-based chronic kidney disease (ckd) prediction," 2024, doi:10.1201/9781003471059-47.

[9] S. M. A. Yousif, H. T. Halawani, G. Amoudi, F. M. O. Birkea, A. M. Almunajam, and A. A. Elhag, "Early detection of chronic kidney disease using eurygasters optimization algorithm with ensemble deep learning approach," *Alexandria Engineering Journal*, vol. 100, 2024, doi:10.1016/j.aej.2024.05.011.

[10] A. Francis, M. Harhay, A. C. M. Ong, S. Tummalapalli, A. Ortiz, A. B. Fogo, D. Fliser, P. Roy-Chaudhury, M. Fontana, M. Nangaku, C. Wanner, C. Malik, A. Hradsky, D. Adu, S. Bavanandan, A. Cusumano, L. Solá, I. Ulasi, and V. Jha, "Chronic kidney disease and the global public health agenda: an international consensus," *Nature Reviews Nephrology*, vol. 20, pp. 473–485, 2024.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi:10.1038/nature14539.

[12] N. Yogesh, P. Shrinivasacharya, N. Naik, and B. M. Vikranth, "Chronic kidney disease classification through hybrid feature selection and ensemble deep learning," *International Journal of Statistics in Medical Research*, vol. 14, 2025, doi:10.6000/1929-6029.2025.14.11.

[13] V. Arumugham, B. P. Sankaralingam, U. M. Jayachandran, K. V. S. S. R. Krishna, S. Sundarraj, and M. Mohammed, "An explainable deep learning model for prediction of early-stage chronic kidney disease," *Computational Intelligence*, vol. 39, no. 6, pp. 1423–1440, 2023, doi:10.1111/coin.12587. [Online]. Available: https://doi.org/10.1111/coin.12587

[14] B. Anoch and L. Parthiban, "Detection and classification of medical images using deep learning for chronic kidney disease," *International Urology and Nephrology*, pp. 1–15, 2025, doi:10.1007/s11255-025-04786-7. [Online]. Available: https://doi.org/10.1007/s11255-025-04786-7

[15] Akanksha and G. Suganeshwari, "An improved deep learning approach for prediction of the chronic kidney disease," *International Journal of Electrical and Electronics Research*, vol. 10, no. 4, pp. 321–326, 2022, doi:10.37391/IJEER.100414. [Online]. Available: https://doi.org/10.37391/IJEER.100414

[16] M. Arifuzzaman, I. Ahmed, M. J. U. Chowdhury, M. S. Rahman, M. E. Hossain, A. D. Nath, and S. Absar, "A novel ensemble-based deep learning model with explainable ai for accurate kidney disease diagnosis," *Advances in Artificial Intelligence and Machine Learning*, vol. 5, no. 1, pp. 1–12, 2025, doi:10.54364/AAIML.2025.51196. [Online]. Available: https://doi.org/10.54364/AAIML.2025.51196

[17] C. F. Hsu, T. M. Yu, Y. L. Wu, W. C. Wang, J. S. Wang, and S. S. Chang, "Prediction of advanced chronic kidney disease through retinal fundus images by deep learning," *Scientific Reports*, vol. 15, no. 1, p. Article 21366, 2025, doi:10.1038/s41598-025-21366-y. [Online]. Available: https://doi.org/10.1038/s41598-025-21366-y

[18] J. Rubia, S. Shibi, B. Lincy, J. P. Catherin, Vigneshwaran, and E. Nithila, "Automatic kidney disease prediction using deep learning techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 3, pp. 1798–1806, 2024, doi:10.11591/ijeecs.v36.i3.pp1798-1806. [Online]. Available: https://doi.org/10.11591/ijeecs.v36.i3.pp1798-1806

[19] P. C. Mo, H. Y. Hsu, C. F. Lin, Y. S. Cheng, I. T. Tu, L. C. Kuo, and F. C. Su, "Distinguishing sensorimotor performance between individuals with diabetes mellitus and chronic kidney disease through deep learning models," *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. Article 1351485, 2024, doi:10.3389/fbioe.2024.135148. [Online]. Available: https://doi.org/10.3389/fbioe.2024.1351485

[20] D. M. Alsekait, H. Saleh, L. A. Gabralla, K. Alnowaiser, S. El-Sappagh, R. Sahal, and N. El-Rashidy, "Toward comprehensive chronic kidney disease prediction based on ensemble deep learning models," *Applied Sciences*, vol. 13, no. 6, p. 3937, 2023, doi:10.3390/app13063937. [Online]. Available: https://doi.org/10.3390/app13063937

[21] A. A. Abdulmunem and A. J. Jabbar, "Implementation of clinical diagnosis system for chronic kidney disease using deep learning algorithms," *International Journal of Advanced Technology and Engineering Exploration*, vol. 11, no. 110, pp. 81–90, 2024, doi:10.19101/IJATEE.2023.10102081. [Online]. Available: https://doi.org/10.19101/IJATEE.2023.10102081

[22] X. Qin, X. Liu, L. Xia, Q. Luo, and C. Zhang, "Multimodal ultrasound deep learning to detect fibrosis in early chronic kidney disease," *Renal Failure*, vol. 46, no. 2, pp. 1–10, 2024, doi:10.1080/0886022X.2024.2417740. [Online]. Available: https://doi.org/10.1080/0886022X.2024.2417740

[23] S. B S, "Kidney disease classification using deep learning," *International Journal of Scientific Research in Engineering and Management*, vol. 9, no. 5, pp. 1–6, 2025, doi:10.55041/ijsrem48296. [Online]. Available: https://doi.org/10.55041/ijsrem48296

[24] P. Liang, J. Yang, W. Wang, G. Yuan, M. Han,

Q. Zhang, and Z. Li, "Deep learning identifies intelligible predictors of poor prognosis in chronic kidney disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3500–3510, 2023, doi:10.1109/JBHI.2023.3266587. [Online]. Available: https://doi.org/10.1109/JBHI.2023.3266587

[25] Y. Tan, Y. Ma, S. Rao, and X. Sun, "Performance of deep learning for detection of chronic kidney disease from retinal fundus photographs: A systematic review and meta-analysis," *European Journal of Ophthalmology*, vol. 34, no. 2, pp. 512–520, 2024, doi:10.1177/11206721231199848. [Online]. Available: https://doi.org/10.1177/11206721231199848

[26] M. N. H. Chowdhury, M. B. I. Reaz, S. H. M. Ali, M. L. Crespo, S. Ahmad, G. M. Salim, F. Haque, L. G. G. Ordóñez, M. J. Islam, T. M. Mahdee, K. S. Zaman, M. S. K. Hemel, and M. A. S. Bhuiyan, "Deep learning for early detection of chronic kidney disease stages in diabetes patients: A tabnet approach," *Artificial Intelligence in Medicine*, vol. 166, p. Article 103153, 2025, doi:10.1016/j.artmed.2025.103153. [Online]. Available: https://doi.org/10.1016/j.artmed.2025.103153

[27] L. Holmstrom, M. Christensen, N. Yuan, J. W. Hughes, J. Theurer, M. Jujjavarapu, P. Fatehi, A. Kwan, R. K. Sandhu, J. Ebinger, S. Cheng, J. Zou, S. S. Chugh, and D. Ouyang, "Deep learning-based electrocardiographic screening for chronic kidney disease," *Communications Medicine*, vol. 3, no. 1, p. Article 78, 2023, doi:10.1038/s43856-023-00278-w. [Online]. Available: https://doi.org/10.1038/s43856-023-00278-w

[28] R. H. Aswathy, P. Suresh, M. Y. Sikkandar, S. Abdel-Khalek, H. Alhumyani, R. A. Saeed, and R. F. Mansour, "Optimized tuned deep learning model for chronic kidney disease classification," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 2453–2468, 2022, doi:10.32604/cmc.2022.019790. [Online]. Available: https://doi.org/10.32604/cmc.2022.019790

[29] M. N. Hossain, E. Bhuiyan, M. B. A. Miah, T. A. Sifat, Z. Muhammad, and M. F. A. Masud, "Detection and classification of kidney disease from ct images: An automated deep learning approach," *Technologies*, vol. 13, no. 11, p. 508, 2025, doi:10.3390/technologies13110508. [Online]. Available: https://doi.org/10.3390/technologies13110508

[30] I. I. Ayogu, C. F. Daniel, B. A. Ayogu, J. N. Odii, C. L. Okpalla, and E. C. Nwokorie, "Investigation of ensembles of deep learning models for improved chronic kidney diseases detection in ct scan images," *Franklin Open*, vol. 11, p. Article 100298, 2025, doi:10.1016/j.fraope.2025.100298. [Online]. Available: https://doi.org/10.1016/j.fraope.2025.100298

[31] P. K. Rao, S. Chatterjee, K. Nagaraju, S. B. Khan, A. Almusharraf, and A. I. Alharbi, "Fusion of graph and tabular deep learning models for predicting chronic kidney disease," *Diagnostics*, vol. 13, no. 12, p. 1981, 2023, doi:10.3390/diagnostics13121981. [Online]. Available: https://doi.org/10.3390/diagnostics13121981

[32] A. S. Mahmoud, O. Lamouchi, and S. Belghith, "Advancements in machine learning and deep learning for early diagnosis of chronic kidney diseases: A comprehensive review," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 1–15, 2024, doi:10.58496/BJML/2024/015. [Online]. Available: https://doi.org/10.58496/BJML/2024/015

[33] A. Yumashev, P. Udayakumar, S. N. S. V. S. C. Ramesh, E. L. Lydia, and K. V. Kumar, "Role of rough neutrosophic attribute reduction with deep learning-based enhanced kidney disease diagnosis," *International Journal of Neutrosophic Science*, vol. 25, no. 1, 2025, doi:10.54216/IJNS.250126.

[34] K. Kumar, M. Pradeepa, M. Mahdal, S. Verma, M. V. RajaRao, and J. V. N. Ramesh, "A deep learning approach for kidney disease recognition and prediction through image processing," *Applied Sciences*, vol. 13, no. 6, p. 3621, 2023, doi:10.3390/app13063621. [Online]. Available: https://doi.org/10.3390/app13063621

[35] B. Dharaneesh, S. Sujan, and Hemalatha, "Deep learning architectures for kidney disease classification," *International Journal of Scientific Research in Engineering and Management*, vol. 8, no. 8, pp. 1–6, 2024, doi:10.55041/ijsrem37291. [Online]. Available: https://doi.org/10.55041/ijsrem37291

[36] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Deep-kidney: an effective deep learning framework for chronic kidney disease prediction," *Health Information Science and Systems*, vol. 12, no. 1, 2024, doi:10.1007/s13755-023-00261-8.

[37] M. Zhang, Z. Ye, E. Yuan, X. Lv, Y. Zhang, Y. Tan, C. Xia, J. Tang, J. Huang, and Z. Li, "Imaging-based deep learning in kidney diseases: Recent progress and future prospects," *Insights into Imaging*, vol. 15, no. 1, p. Article 36, 2024, doi:10.1186/s13244-024-01636-5. [Online]. Available: https://doi.org/10.1186/s13244-024-01636-5

[38] F. Mallamaci and G. Tripepi, "Risk factors of chronic kidney disease progression: Between old and new concepts," *Journal of Clinical Medicine*, vol. 13, no. 3, p. 678, 2024, doi:10.3390/jcm13030678. [Online]. Available: https://doi.org/10.3390/jcm13030678

[39] Ákos Géza Pethő, M. Tapolyai, Éva Csongrádi, and P. Orosz, "Management of chronic kidney disease: The current novel and forgotten therapies," *Journal of Clinical and Translational Endocrinology*, vol. 36, p. 100354, 2024, doi:10.1016/j.jcte.2024.100354.

[40] F. G. Scurt, M. J. Ganz, C. Herzog, K. Bose, P. R. Mertens, and C. Chatzikyrkou, "Association of metabolic syndrome and chronic kidney disease," *Obesity Reviews*, vol. 25, no. 1, p. N/A, 2024, doi:10.1111/obr.13649.

[41] K. Tsuji, N. Uchida, H. Nakanoh, K. Fukushima, S. Haraguchi, S. Kitamura, and J. Wada, "The gut–kidney axis in chronic kidney diseases," *Diagnostics*, vol. 15, no. 1, p. 21, 2025, doi:10.3390/diagnostics15010021.

[42] P. Tao, C. W. Chien, C. Liu, J. Zheng, D. Sun, J. Zeng, Q. Song, Y. Liu, T. H. Tung, and L. Kang, "Diabetes mellitus is a risk factor for incident chronic kidney disease: A nationwide cohort study," *Heliyon*, vol. 10, no. 7, 2024, doi:10.1016/j.heliyon.2024.e28780.

[43] N. Pradhan, J. Kerner, L. A. Campos, and M. Dobre, "Personalized nutrition in chronic kidney disease," *Biomedicines*, vol. 13, no. 3, p. 647, 2025, doi:10.3390/biomedicines13030647.

[44] P. I. Georgianos and R. Agarwal, "Hypertension in chronic kidney disease - treatment standard 2023,"

Nephrology Dialysis Transplantation, vol. 38, no. 12, p. N/A, 2023, doi:10.1093/ndt/gfad118.

[45] L. Cirillo, L. D. Chiara, S. Innocenti, C. Errichiello, P. Romagnani, and F. Becherucci, "Chronic kidney disease in children: an update," *Clinical Kidney Journal*, vol. 16, no. 10, p. N/A, 2023, doi:10.1093/ckj/sfad097.

[46] K. Badura, J. Janc, J. Wąsik, S. Gnitecki, S. Skwira, E. Młynarska, J. Rysz, and B. Franczyk, "Anemia of chronic kidney disease—a narrative review of its pathophysiology, diagnosis, and management," *Biomedicines*, vol. 12, no. 6, p. 1191, 2024, doi:10.3390/biomedicines12061191.

[47] S. Maringhini and C. Zoccali, "Chronic kidney disease progression—a challenge," *Biomedicines*, vol. 12, no. 10, p. 2203, 2024, doi:10.3390/biomedicines12102203.

[48] Y. Li and R. Padman, "Enhancing end-stage renal disease outcome prediction: a multisourced data-driven approach," *Journal of the American Medical Informatics Association*, vol. 33, no. 1, pp. 1–12, 2026, doi:10.1093/jamia/ocaf118.

[49] M. S. A. Huda, E. Kanon, M. S. K. Pappo, M. A. Ali, and N. Ahmed, "Nefroai: An explainable and real-time framework for predicting chronic kidney disease using diverse machine learning models and different feature selection techniques," *IEEE Access*, vol. 14, pp. 1–15, 2026, doi: 10.1109/ACCESS.2025.3649006.

[50] Y. Zhang, Y. Hou, T. Qiu, Y. Zhuang, K. Chen, W. Ling, Y. Luo, and J. Lin, "Deep learning in renal ultrasound: applications, challenges, and future outlook," *Frontiers in Oncology*, vol. 15, p. 1730628, 2026, doi:10.3389/fonc.2025.1730628.

[51] S. Sudharson and P. Kokil, "Automated diagnosis tool for clinical ultrasound image analysis in kidney disease identification," *Biomedical Signal Processing and Control*, vol. 111, p. 108345, 2026, doi:10.1016/j.bspc.2025.108345.

[52] V. P. Samuel and G. Annasamy, "Ultrasensitive detection of creatinine using deep learning-integrated graphene oxide gold nanocomposites sers sensor," *IEEE Sensors Letters*, vol. 10, no. 1, pp. 1–4, 2026, doi:10.1109/LSENS.2025.3642276.

[53] P. I. Nainggolan, F. Lubis, M. F. SyahPutra, T. G. M.R, A. Lumbangaol, and R. Sulaiman, "Advancing kidney disease diagnosis with yolov9 deep learning: A comprehensive classification approach," pp. 1–6, 2026, doi:10.1109/iccai65301.2025.11279427.

[54] S. S and V. A, "A hybrid deep learning approach based on optimized feature selection on environmental multi-disease predictive models," *International Journal of Cognitive Computing in Engineering*, vol. 7, pp. 100 150–100 165, 2026, doi:10.1016/j.ijcce.2025.11.006.

[55] S. Samenezhad and D. Rafighi, "The role of artificial intelligence in advancing urologic care: From diagnostics to therapeutics," *Surgery in Practice and Science*, vol. 24, p. 100322, 2026, doi:10.1016/j.sipas.2025.100322.

[56] W. Cheungpasitporn, A. Athavale, L. Ghazi, K. B. Kashani, T. Colicchio, J. L. Koyner, J. Chen, J. H. Ix, G. Nadkarni, and J. A. Neyra, "Transforming nephrology through artificial intelligence: A state-of-the-art roadmap for clinical integration," *Clinical Kidney Journal*, vol. 19, no. 1, pp. 1–15, 2026, doi:10.1093/ckj/sfag004.

[57] M. N. Hossain, E. Bhuiyan, M. B. A. Miah, S. Awang, M. S. Hossain, and T. A. Sifat, "Classification of kidney disease from ct images: A deep learning-based comparative analysis," pp. 1–6, 2026, doi:10.1109/icsecs65227.2025.11279151.

[58] R. O. Ogundokun, P. A. Owolawi, and E. A. van Wyk, "A comparative analysis of deep transfer learning and traditional handcrafted feature approaches for kidney disease classification using random forests," vol. 2667, pp. 205–218, 2026.

[59] S. M, T. M, E. R. G, and A. M, "A chronic kidney disease prediction system based on internet of things using walrus optimized deep learning technique," *Informatics for Health and Social Care*, pp. 1–15, 2026, doi:10.1080/17538157.2025.2610695.

[60] I. Sargunan, S. K. Krishnamoorthy, K. Govindharaju, and V. Cheemalapati, "Topk-kiu-net and squeeze residual fraction network for kidney tumor segmentation and detection," *Biomedical Signal Processing and Control*, vol. 113, p. 109116, 2026, doi:10.1016/j.bspc.2025.109116.

[61] M. Abniki, M. Amirdosara, and M. Zangi, "Ai-driven innovations in intensive care nephrology; bridging intensive care and kidney diseases," *Journal of Nephropharmacology*, vol. 15, no. 1, 2026, doi:10.34172/npj.2025.12820.

[62] Y. Chen, C. Chang, Y. Yang, X. Zhang, P. Chao, C. Chen, and C. Lu, "Rapid diagnosis of chronic kidney disease via raman spectroscopy combined with a lightweight mixture of experts method," *Journal of Raman Spectroscopy*, vol. 57, no. 1, 2026, doi:10.1002/jrs.70028.

[63] J. Eda, T. Mohaidat, S. I. Niloy, Z. Niu, M. R. K. Khan, and K. Khalil, "Artificial intelligence in healthcare: A comprehensive survey on ai-driven diagnosis and patient monitoring," *Applied Soft Computing*, vol. 190, p. 114429, 2026 doi:10.1016/j.asoc.2025.114429.

[64] S. C. Kunjumohamad and S. Govindaraj, "Nlp-driven federated learning framework for multi-organ rare disease prediction," *Iran Journal of Computer Science*, vol. 9, no. 1, 2026, doi:10.1007/s42044-025-00375-0.

[65] W. Dai, R. Liu, Z. Wu, T. Wu, M. Wang, J. Zhou, Y. Yuan, and J. Liu, "Exploiting scale-variant attention for segmenting small medical objects," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2026, doi:10.1109/TNNLS.2025.3645355.

[66] Z. A. Khan, M. Waqar, H. U. Khan, N. I. Chaudhary, A. T. Khan, I. Ishtiaq, F. A. Khan, and M. A. Z. Raja, "Fine-tuned deep transfer learning: An effective strategy for the accurate chronic kidney disease classification," *PeerJ Computer Science*, vol. 11, p. e2800, 2025, doi:10.7717/peerj-cs.2800. [Online]. Available: https://doi.org/10.7717/peerj-cs.2800