RESEARCH PAPER

A Hybrid Deep Learning Framework with Dense Connectivity and Explainable Grad-CAM for Accurate Kidney Disease Classification Using CT Images

Malarvizhi T[1], Nisha P[2], Harshini K[3], Nazreen Fathima J[4], Parkavi K[5]

[1]Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, India.
Email: malarece05@gmail.com

[2]Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, India.
Email: nisha2352005@gmail.com (Corresponding Author)

[3]Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, India.
Email: harshinisuren2004@gmail.com

[4]Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, India.
Email: nazreenfathimaj005@gmail.com

[5]Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, India.
Email: parkavikalaiselvan01@gmail.com

***Corresponding author: Nisha P, Department of Electronics and Communication Engineering, V.S.B. Engineering College, Karur, India**
Email: nisha2352005@gmail.com

Received: 30th May, 2026; **Revised:** 10th June, 2026; **Accepted:** 14th June, 2026; **Available Online:** 14th June, 2026

ABSTRACT

Background

Accurate and efficient classification of medical images is crucial for reliable diagnosis and treatment planning, particularly in kidney-related disorders. However, conventional deep learning models often face challenges related to computational complexity, limited feature propagation, and lack of interpretability.

Objective

To address these issues, this study proposes a novel hybrid deep learning framework that integrates depthwise separable convolutions, dense connectivity, and Gradient-weighted Class Activation Mapping (Grad-CAM) for enhanced performance and transparency.

Materials and Methods

The proposed model is evaluated on the CKT (CT Kidney Tumor) dataset, which consists of four classes: Cyst, Normal, Stone, and Tumour, with a total of 12,446 CT images exhibiting class imbalance. The architecture leverages depthwise separable convolutions to reduce computational overhead while preserving essential spatial features, and dense connectivity to improve feature reuse and gradient flow. Additionally, Grad-CAM is incorporated to provide visual explanations, enabling the identification of clinically relevant regions within CT images.

Results

Experimental results demonstrate that the proposed model outperforms baseline architectures, including conventional CNN, MobileNet, and DenseNet, achieving superior accuracy, precision, recall, and F1-score. The model also exhibits strong performance in class-wise evaluation, particularly for minority classes, indicating robustness against class imbalance. Furthermore, interpretability analysis confirms that the model focuses on meaningful anatomical regions, enhancing its reliability for clinical applications.

Conclusion

Overall, the proposed framework provides an effective balance between accuracy, efficiency, and explainability, making it suitable for real-world medical image classification tasks. Future work will explore advanced attention mechanisms and multi-modal data integration to further improve performance and scalability.

Keywords: Kidney disease classification, CT images, Hybrid deep learning, Depthwise separable convolutions, Dense connectivity, Grad-CAM, Explainable AI, Medical image analysis.

How to cite this article: Malarvizhi T, Nisha P, Harshini K, Nazreen Fathima J, Parkavi K. A Hybrid Deep Learning Framework with Dense Connectivity and Explainable Grad-CAM for Accurate Kidney Disease Classification Using CT Images. Int J Drug Deliv Technol. 2026;16(60s):225-238. DOI: 10.25258/ijddt.16.60s.27

Source of support: Nil.

Conflict of interest: None

1. Introduction

In recent decades, chronic kidney disease (CKD) has become one of the most relevant health challenges worldwide, causing an increasingly significant

burden on healthcare systems because of its progressive course, high prevalence, and close link with life-threatening comorbidities. Chronic kidney disease (CKD) is defined as the progressive and irreversible loss of kidney function, leading to end-stage renal disease (ESRD) if not detected and managed at an early stage. The kidneys are essential organs that perform homeostatic roles, including filtering metabolic waste, regulating fluids and electrolytes, maintaining blood pressure, and regulating endocrine functions. Loss of these functions leads to severe complications, such as cardiovascular disease (CVD), anaemia, bone mineral disorders (BMDs), and metabolic imbalances. As per the latest worldwide statistics, chronic kidney disease (CKD) affects over 10% of the global population and its prevalence is higher in geriatric patients and among individuals with diabetes mellitus (DM) and hypertension [1], [2].

Particularly difficult, because all the while it is sparking out these attacks on your body, CKD will not give you any warning signs early on. Symptoms usually only become apparent in advanced stages, when extensive and often irreversible kidney damage has already occurred. As such, early detection is crucial for controlling the progression of the disease, preventing complications, and improving outcomes for patients. Conventional diagnostics depend on laboratory findings like serum creatinine concentration, glomerular filtration rate (GFR), and urine albumin lines, along with imaging techniques such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) [3], [4]. While these methods are clinically reliable, their utility is often limited by the delay between detection and clinical utility as well as observer variability and dependence on specialized medical expertise. In addition, in many areas of the world including developing countries there is a shortage of trained nephrologists and radiologists that substantially limits timely diagnosis and treatment [5].

Artificial Intelligence (AI), and especially Machine Learning (ML) and Deep Learning (DL), has shown great improvements in various domains over the last few years, particularly with respect to healthcare diagnostics. Artificial intelligence (AI) systems can analyse large amounts of heterogeneous medical data, recognize complex patterns in that data, and facilitate clinicians' decision making. Traditional methods of CKD detection using ML algorithms have included Support Vector Machines (SVM), Random Forest (RF), Naïve Bayes (NB) and k-Nearest Neighbours (KNN) models to predict disease presence from clinical data [6]– [8]. While these methods have shown promising successes in structured datasets, they are heavily dependent on hand-engineered features and domain knowledge, making them less generalizable to different data distributions.

Deep Learning (DL) has tremendously transformed AI in the field of medical imaging. Specifically, CNNs are the backbone of image-based disease detection by automatically learning hierarchical features. A range of CNN-based models have been deployed to classify cysts, tumours and kidney stones in CT and ultrasound [9], [10] for detecting kidney disease tasks. Various architectures including VGG, ResNet, DenseNet, EfficientNet, and MobileNet have been investigated to enhance performance with a high accuracy rate [11]–[13], as shown in Table II. As an example, DenseNet alleviates the vanishing gradient issue via dense connectivity and improves feature propagation [18]; EfficientNet proposes a compound scaling method for efficient balancing between network depth, width and resolution [21], [23]. Until then, lightweight architectures e.g., MobileNet attains better trade-off between accuracy and efficiency by depthwise separable convolutions [24].

Altering these architectures for fusion have been put forward in more recent designs, such as Xception and ConvNeXt which achieved higher performance results in deep learning networks particularly with respect to medical imaging tasks. Xception generalizes depthwise separable convolutions to improve the efficiency of feature extraction pipelines [25], and ConvNeXt improves convolutional network performance by applying modern architectural updates inspired by vision transformers [22]. Accurately detecting kidney disease has become possible thanks to these advances, with various studies reporting results on benchmark datasets of over 99% accuracy [14], [15]. However, despite such impressive results, achieving strong and generalizable performance across heterogeneous clinical environments poses a key challenge.

One of the major limitations in previous CKD detection studies is the utilization of single-source datasets, which are often limited and less diverse

with respect to patient demographics, imaging conditions, and disease variations. Between model training and evaluation on comprehensive but regionally limited datasets such as CTK [16]. Such a limited dataset diversity puts in question the generalizability of these models once deployed in clinical settings. Train models on homogeneous datasets may not generalize well to other acquired data from another institution or imaging device and are therefore limited in their practical application.

Another fundamental problem is class-imbalance in medical datasets. There are often skewed distributions of diverse categories of diseases in CKD datasets, resulting in several classes having a much lower presence. Common methods to tackle this problem are undersampling and oversampling. On the other hand, undersampling could cause loss of useful information while oversampling could be affected with synthetic bias and a rise in overfitting fatigue [17]. To overcome these challenges, several advanced methodologies have been introduced, including data augmentation and generative models; however, they are seldom incorporated into CKD detection architectures.

Most existing attempts also address only unimodal data, mainly medical imaging, which is accompanied by other challenges besides the data problem. [Spoiler Alert: this isn't one of those, because…] While imaging gives important anatomical insights, it fails to contextualize the problem within a patient's overall well-being. In everyday clinical settings, the diagnosis of CKD instantiates more than just imaging data, lab results and demographics. Not considering these types of data restricts the predictive power and trustworthiness of AI models. Multimodal learning, utilising heterogeneous data types, is widely used in other medical domains with successful outcomes; however, this has been under-appreciated within the CKD space [18], [19].

AI-based healthcare systems are also critical for interpretability and transparency. Because of complex internal representations, deep learning models are often referred to as "black boxes" that do not allow us to see how they come up with decisions. This is a major obstacle to clinical adoption with poor transparency. Methods like Gradient-weighted Class Activation Mapping (Grad-CAM) have been proposed to create visual explanations by emphasizing key areas within medical images [26]. Although Grad-CAM enhances interpretability, it does not provide a complete understanding of the model's decision-making process. Tracking at feature-level explanation methods (e.g., SHAP and LIME) offers a further level of interpretability by measuring the contribution of different features; however, their integration with imaging-based models remains limited [27], [28].

Additionally, computational efficiency and real-time implementability are frequently neglected in common literature. Most state-of-the-art deep learning models require significant computational power and have low inference speeds, thus rendering them infeasible for real-time clinical assignments. While lightweight architectures such as MobileNet and EfficientNet have been proposed to alleviate this challenge, achieving a configuration where accuracy and efficiency is not compromised remains arduous [24],[23]. Furthermore, most studies specifically assess offline performance without addressing integration with hospital information systems crucial to real world application.

The scale-aware and feature-fusion oriented SAFF-ConvNeXt framework is a remarkable contribution in this field, which combined the state-of-the-art ConvNeXt structure with both scale aware fusion and Squeeze-and-Excitation (SE) [22]. It extracts hybrid system features at different resolutions, giving it an advantage over traditional CNN model. Additionally, Grad-CAM's positive interpretation supports model interpretability. Despite these achievements, the framework remains constrained to unimodal imaging data and is not thoroughly assessed on various datasets, revealing a gap in further investigation.

Ergo, the need for an extensive multimodal deep learning framework that involves all ANN architectures with advanced explainability and generalization with cascades of ANN techniques. The integration of imaging data with clinical attributes like GFR, serum creatinine levels, blood pressure as well as patient demographics can be facilitated by multimodal learning to provide a more comprehensive view of the disease. In this context, attention-based fusion mechanisms can provide improved feature representation by focusing on the most relevant information across modalities [18], [20]. These types of methods can greatly enhance diagnostic precision and dependability.

Also, a combination of visual and feature-level explainability techniques could be used to provide a better understanding of what is happening inside the model. Combining Grad-CAM with feature attribution methods can help visualize important image regions and quantify contributions by features simultaneously [26], [27]. This interpretability at both levels is paramount for gaining acceptance among healthcare providers and making safe deployment in clinical settings. Finally, the evaluation of model generalization is another key point. Validating models across datasets and applying domain adaptation methods can be used to assess robustness against shifts in data distribution. One way to assess a model's generalizability to unseen data, which is critical for real-world applications [15], [16], is by training on one dataset and testing its performance on another dataset. Leveraging these strategies of evaluation adds significant credibility and relevancy to the AI-based diagnostic systems.

Thus, although AI methods have been increasingly applied for CKD detection in recent years, several fundamental issues and challenges that must be addressed persist, including limited diversity of the datasets available, lack of multimodal integration from multiple sources, inadequate interpretability or explainability of the models employed and insufficient scrutiny on generalization and real-world performance. To tackle these challenges, it is essential to create holistic frameworks yet to be established that prioritize not only accuracy but also robustness and transparency as well as clinical relevance. Thus, this study presents a new multimodal, interpretable and generalizable deep learning system to identify kidney disease. We present novel deep learning architectures, attention-based fusion mechanisms, and hybrid explainability techniques to overcome the limitations of existing approaches. This work will help advance the research of effective machine learning models by employing rigorous evaluation techniques and ensuring clinical relevance to ultimately translate into real-world AI applications for early CKD diagnosis. Such systems when integrated into routine clinical practice could ultimately address the growing global burden of chronic kidney disease by increasing diagnostic accuracy, decreasing external medical burden and improving patient outcomes.

2. Literature Survey

Jaspreet Singh et al.[1] specific offers a machine learning-based predictive framework for the early detection of Chronic Kidney Disease (CKD) with relevant significant feature selection techniques. By focusing on identifying more critical risk factors, this can improve the actual model itself while providing a better performance through reduced computational cost. The CKD dataset retrieved from the UCI was pre-processed and then both correlation coefficient and rank-based methods were utilized for feature selection to take out relevant attributes. Classification performance was evaluated for several machine learning algorithms, including Bayes Net, Random Forest, Multilayer Perceptron, SMO, Bagging, Random Committee and K-Star. Results showed that the Bayes Net classifier gave better accuracy once features were optimized, which greatly aided in prediction capabilities and lowered error rates. The study demonstrates the utility of the feature selection method to improve diagnostic accuracy but is limited to structured clinical information and does not include imaging or multimodal approaches, suggesting avenues for further work.

In the review by Ryan J. Adam et. In this issue of Kidney International, Kim et. The authors reviewed the phenotypic similarities and differences among these models in a systematic manner, highlighting their relevance to human CKD pathophysiology. The two models successfully reproduced important clinical features of CKD, including uraemia, proteinuria, fibrosis, capillary rarefaction and progressive renal dysfunction. First, there were marked differences in vascular responses, especially in blood pressure control and RAAS activation, with the infarction model showing more severe hypertension and RAAS activation. The study also revealed changes in autoregulation, nitric oxide signalling and angiogenic balance. This examination provides useful insights into the type of potential experimental models that could be selected to investigate renal pathology in CKD and highlights the need for pathophysiological delineation to be model-specific.

A study by Abdullah et al. [3] conducted an extensive study that focused on reporting the application of machine learning techniques in predicting the occurrence of CKD at the earliest. They mentioned CKD to be one of the prominent global health problems with significant unavoidable complications like cardiovascular disorders and end-

stage renal failure, thus underscoring the importance of early detection. Various machine learning algorithms such as decision trees, random forests and neural networks were reviewed during the study which showcased their usefulness in analysing clinical and demographic datasets. Performative markers such as serum creatinine, proteinuria and eGFR have been highlighted alongside patient-specific variables (e.g., age). The results showed that machine learning models can enhance diagnostic accuracy and assist clinical decision-making substantially. At the same time, it recognized some limitations around variability in patient-level data and the lack of model generalization to other populations, calling for more robust and scalable predictive frameworks.

We developed a machine learning-based framework for the early prediction of chronic kidney disease (CKD) using clinical data collected from the UCI repository [4]. Preprocessing methods play a pivotal role in improving model performance, and the authors asserted that the proper handling of missing values, data transformation, and feature selection were critical steps to implement when building models. Several classification algorithms, Random Forest (RF), Logistic Regression (LR), k-Nearest Neighbours (K-NN) and Support Vector Machine (SVM) were implemented, and their results were compared. The experimental results showed that the Random Forest classifier achieved 100% accuracy outperforming other models such as Logistic Regression (98%) and K-NN (94%). Furthermore, correlation-based feature selection was utilized to extract high-impact predictor features like hypertension, diabetes mellitus and albumin levels, enhancing the systematic answers with a quality factor. However, the study has two important limitations: a comparatively limited data set (400 cases) and the absence of an external validation cohort which might limit generalizability in clinical practice.

Patel et al. [5] proposed a deep learning-based approach for the automated detection of kidney disease using medical imaging techniques. Specifically, the authors used Convolutional Neural Networks (CNNs) to identify hierarchical features from CT scan images to accurately classify kidney abnormalities such as cysts, tumours and stones. It also investigated transfer learning paradigms that enhance the model's effectiveness while decreasing computational resources. The experimental results show that the CNN model achieved high classification accuracy of the proposed model, compared with traditional machine learning approaches in image-based diagnosis. In addition, metrics evaluating the model like precision, recall and F1-score proved that the model is robust. However, its reliance on existing, limited annotated datasets and high computational demands may limit the real-time implementation in a clinical setting. The authors went ahead to conclude that combining clinical data with imaging modalities could improve diagnostic performance further and enable a more complete assessment of kidney disease.

Chen et al. [6] proposed a sophisticated deep learning-based model for dedicated three-dimensional (3D) analysis of the renal syndromes using multi-phase CT images. The authors leveraged an ensemble method that fuses segmentation and registration algorithms to reconstruct the kidney's form and precisely identify its stones. Three distinct CT phases, namely non-contrast (NCT), corticomedullary (CTC), and excretory (CTE) were used in this methodology to obtain complementary anatomical information [4]. We evaluated several state-of-the-art segmentation models, such as 3D U-Net, ResU-Net and Swin UNETR,of which the best kidney Dice score (95.21%) was achieved with Swin UNETR. Then, a more rigid registration technique was applied for the multi-phase images alignment to create a complete 3D representation. The proposed method showed superior diagnostic visualization and greater preoperative planning ability; however, the high computational complexity makes it difficult for deployment in clinical applications that require large annotated datasets37.

Author of the review Capitanio et al., [7] provided an exhaustive epidemiological characterization for renal cell carcinoma (RCC), including worldwide incidence, mortality and risk factors. The authors pointed out that RCC is among the deadliest urologic malignancies, and that incidence rates vary widely between countries. Major risk factors are smoking, obesity, hypertension, and chronic kidney disease which significantly contribute to the prevalence of diseases. The widespread use of imaging modalities (CT, MRI) for other conditions has also led to an increasing detection of RCC. The authors also provided a perspective on the limitations of current screening strategies, stating that low RCC incidence and the absence of reliable biomarkers make early detection encouraging yet not cost-effective.

Screening at Even Higher Resolution Strongly Advocated. Meta-analysis of current screening methods for common cancers indicate that they fall short of being broadly effective, though targeted screeners in selected populations improve early diagnosis and may reduce patient morbidity and mortality.

The study of Han et al. [8] proposed a deep learning-based framework for RCC subtype classification based on multi-phase CT images. Using three-phase CT imaging (including pre-contrast, corticomedullary, and excretory phases), the authors employed a modified GoogLeNet architecture along with transfer-learning to enhance classification performance. Features were extracted using radiologist-annotated (ROIs) TS images, and a linear combination of multi-phase images improved the model's discriminative capability. With an AUC approaching 0.9, the presented approach yielded high diagnostic performance and therefore has a strong potential in clinical decision support systems.

Similarly, Senan et al. [9] proposed a diagnostic system using RFE and classifiers including SVM, KNN, Decision Tree and Random Forest for CKD prediction. Their findings indicated that the Random Forest model obtained 100% accuracy, signifying a substantial improvement in predictive performance with respect to feature selection. The study was limited by the sole use of structured clinical data without imaging modality connectivity.

Given the review of Alowais et al., [10] the transformative impact of Artificial Intelligence (AI) in healthcare has been thoroughly analyzed as it relates to disease diagnosis, personalized treatment and clinical decision support. The study also sheds light on how AI, with the help of machine learning and deep learning approaches, is more capable in supplementing massive medical datasets to provide precise diagnosis, minimize human errors and improve quality service to patients. Despite its potential, issues regarding data privacy, bias and the requirement for a human touch persist as major roadblocks towards clinical adoption. Similarly, Bajwa et al. [11] investigated the integration of artificial intelligence (AI) in healthcare systems, with emphasis on their potential to utilize multimodal data, as well as enhance precision medicine. The authors also suggested a human-centric framework to build trustworthy systems based on clinical validation, ethical concerns and real-world deployment for meaningful transformation of healthcare. Furthermore, Aljaafari et al. [12] proposed an AI-enabled virtual ward system ("CURA") that combines ML and DL models for diabetes mellitus and kidney disease monitoring. Results showed that it had high predictive accuracy, highlighting AI's potential to promote proactive, personalized and continuous care delivery to patients.

Islam et al., [13] proposed an advanced deep learning framework based on Vision Transformers and transfer learning models for automated detection of kidney abnormalities such as cysts, stones, and tumours in CT radiography images. Using a multi-class image data set with more than 12,000 annotated images, the results of this work showed that Swin Transformer was able to outperform conventional CNN architectures (VGG16 and ResNet) by returning an accuracy of 99.30% on all classes present in the dataset. This methodology is demonstrated by the experimental workflow diagram (page 3), showcasing the sequential architecture of preprocessing, training and classification stages that distinctly hint at the efficacy of transformer-based architectures in surpassing conventional methods within the domain of medical images. Similarly, Liu et al. [14] proposed the ConvNeXt architecture, an updated convolutional neural network aimed at rivalling Vision Transformers with its architectural refinements informed by transformer models. Thus, this study confirms that ConvNeXt achieves high accuracy and scalability while being computationally efficient on a variety of image classification tasks in different settings with data until October 2023.

Furthermore, Hossain et al. [15] put forward a new Adaptive Local Binary Pattern (A-LBP) feature descriptor coupled with ensemble machine learning classifiers to identify kidney abnormalities in CT images. As shown in the methodology diagram (page 6), the framework combines preprocessing methods like CLAHE with feature extraction and classification, exceeding 99% accuracy and further validating hybrid feature-engineering methods.

Majid et al., [16] proposed a hybrid machine learning and transfer learning-based framework for classifying kidney tumours from CT images. They used GLCM for feature extraction based on texture, along with other models such as ResNet-101 and

DenseNet-121. Overall, the performance was with accuracy (94.09%) which proved promising to leverage carved and deep features for accurate diagnostic purposes. Similarly, Pande et al. Using a large dataset of 12,446 CT images, [17] proposed a YOLOv8-based deep learning model for the multi-class detection of kidney abnormalities such as cysts, stones and tumours. This resulted in a promising performance, with an accuracy of 82.52% and high specificity (not shown) further demonstrating the model's potential for automated and scalable clinical diagnosis. Furthermore, Yildirim et al. [18] proposed a deep learning-based automated framework for the detection of kidney stones from coronal CT images. Using an XResNet-50 model architecture, the best performance was seen at 96.82% accuracy with localised stone in area regions of interest shown in the model which showcases that it is capable of real-time application on clinical management.

The authors of [19] presented a deep learning and machine learning hybrid framework for the categorization of kidney disease diagnosis from CT images. This study utilized DenseNet-201 together with a Random Forest classifier for feature extraction, obtaining an accuracy of 99.44% which is adequate [48]. This methodology takes full advantage of using transfer learning and ensemble methods to improve the diagnostic performance in detecting cysts, tumours and stones. Similarly, Almuayqil et al. [20] proposed a new computer-aided diagnosis system based on convolutional neural networks (CNN) called KidneyNet to detect chronic kidney disease from CT scans. A 3D CNN-based model using multiple convolutional layers was built and enhanced with Grad-CAM to provide interpretability within the model in which it obtained an exceptionally high accuracy (99.88%) score, and at the same time high level of sensitivity and specificity; demonstrating robustness for clinical use. Furthermore, Fan et al. A full review of data preprocessing methods and strategies in [21], underscoring the importance of enhancing data quality by means of missing value imputation, outlier detection, and transformation approaches. As shown in the study, preprocessing is a crucial part of any data-driven machine learning task and good preprocessing greatly improves the performance and trustworthiness of ML models.

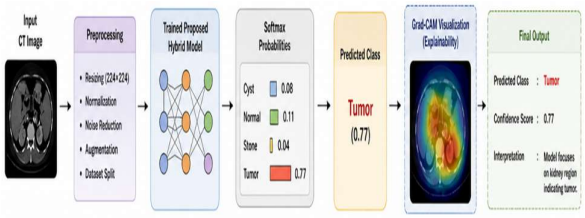

Figure 1. Overall flow of proposed design

The Squeeze-and-Excitation (SE) block is a new architectural unit introduced in Hu et al., [22], which enhances convolutional neural networks by modelling channel-wise feature interdependencies explicitly. SE is designed to adaptively recalibrate feature responses, with a little computational overhead but it provides significantly powerful representation capabilities. Similarly, Simonyan et al. Pb [23] presented the very deep convolutional neural networks (the VGG), showing that increasing the depth of the network using small 3×3 filters lead to a greatly improved image recognition accuracy with restricted computation cost. Furthermore, Tan et al. [24] proposed EfficientNet, a state-of-the-art convolutional neural network architecture that scales depth, width and resolution of the neural networks in synchrony through compound scaling to achieve top-1 accuracy using less parameters while being more efficient. In addition, Chollet et al. Based on multi-linear and separable convolution, [25] proposed an Xception architecture that separates handling the spatial correlation among input feature maps from handling channel correlations, leading to better performance and a more efficient use of parameters in deep learning.

Howard et al. [26] proposed a deep learning architecture called MobileNet—specifically for mobile and embedded vision applications. The model employs depthwise separable convolutions which reduce computational complexity and the number of parameters in a dramatic manner, realizing competitive accuracy. Moreover, width and resolution multipliers can be introduced to allow flexible latency versus accuracy trade-offs, making MobileNet ideal for real-time applications. Similarly, Huang et al. In 2017[27], they introduced DenseNet, a dense connectivity convolutional network architecture where every layer receives input from all previous layers. This dense connectivity facilitates feature propagation, tackles the vanishing gradient problem, enables feature reuse and releases parameter redundancy to gain

better performance with fewer parameters. Furthermore, Selvaraju et al. [28] propose the gradient-based visualization model of Grad-CAM which gives more interpretability for analysis in deep neural networks by discriminative localization maps generated on the input images. Explainability of CNNs has been improving and allows for understanding how CNN works without changing the architecture.

In addition, Sokolova et al. [29] extensively reviewed performance assessment measures for classification problems, highlighting the significance of measure invariance and criteria for choosing suitable measures as they pertain to characteristics of the problem type—binary, multi-class or hierarchical classification. Finally, Gupta et al. [30] proposed a Max Entropy Deep Inverse Reinforcement Learner (MEDIRL) for human-robot interaction where the transition model was trained to witness socially compliant navigation behaviour. Their work emphasizes parameter tuning and ablation analysis for achieving robust performance in practices and real-world settings involving dynamic environments.

3. Proposed Methodology

3.1 Overview of the Proposed Framework

The methodology proposed here presents a hybrid deep learning framework developed with the objective of reconciling the conflicting properties of classification accuracy, computational efficiency and model interpretability. The model combines three primary architectural components: (1) depthwise separable convolutions to reduce the number of operations involved in feature extractions, (2) dense connectivity to enable maximum reuse of features and gradient flow from each layer to every other layer, and (3) gradient-based attention mechanisms for enhanced interpretability. As compared to standard deep convolutional neural networks that handle these aspects independently, the proposed model combines them in a single end-to-end trainable architecture.

Consider the input dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^{H\times W\times C}$ is the input image and $y_i \in \{1,2,\ldots,K\}$ is the corresponding class label Formally, the model aims to learn a mapping function $f_\theta(x_i) \to \hat{y}_i$, which parametrized by θ such that $\hat{y}_i$ closely resembles y_i.

The architecture proposed is a set of 3 sequential modules: feature extraction, dense propagation and interpretability-driven classification. Such modules ensure that the network captures global as well as local contextual information while remaining computationally lightweight.

3.2 Efficient Feature Extraction using Depthwise Separable Convolutions

Most still leverage traditional convolutional operations, where spatial and channel-wise correlations are processed in a coupled fashion as they share the same computational resources, which becomes highly expensive. In order to mitigate this limitation, the designed model uses depthwise separable convolutions, which factorize standard convolution as two separate operations: a depthwise convolution and a pointwise convolution

For an input feature map $F \in \mathbb{R}^{D_F\times D_F\times M}$, a standard convolutional layer satisfying the output $G \in \mathbb{R}^{D_F\times D_F\times N}$ with computational cost:

$$\mathcal{C}_{standard} = D_k^2 \cdot M \cdot N \cdot D_F^2$$

Where D_k denotes the kernel size, M is an input channel number and N is output channel number.

On the other hand, depthwise separable convolution breaks up this procedure into a depthwise and pointwise convolutions. Each filter is applied to one input channel for depthwise convolution:

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m}$$

a pointwise convolution that combines channel-wise information is

$$G_{k,l,n} = \sum_{m} K_{1,1,m,n} \cdot \hat{G}_{k,l,m}$$

The total computational cost becomes:

$$\mathcal{C}_{depthwise} = D_k^2 \cdot M \cdot D_F^2 + M \cdot N \cdot D_F^2$$

This reduction has potential for orders of magnitude fewer parameters and floating-point operations, paving the way to deploy models in environments with limited resources. As per the proposed framework, many depthwise separable convolution blocks are embedded to extract hierarchical features without compromising on efficiency.

3.3 Dense Connectivity for Enhanced Feature Propagation

The efficient convolutions reduce the computational load for sure, but they would create low-capacity embeddings if there are no mechanisms introducing effective information flow in the neural network. To address this, we build a model that follows the dense connectivity of densely connected convolutional networks.

Under this design, every layer gets feature maps from all previous layers thus allowing maximum information reuse. The output of the l^{th} layer is defined as follows:

$$x_l = H_l([x_0, x_1, x_2, \ldots, x_{l-1}])$$

where [.] denotes concatenation and $H_l(\cdot)$ is a composite function of batch normalization, activation, and convolution operations.

This architecture introduced direct connections between any two layers, and improved gradient flow during backpropagation to alleviate the vanishing gradient problem. Dense connectivity also promotes feature reuse, enabling the network to learn compact representations without needless computations. The growth rate controls how many new feature maps each layer generates, allowing the architecture to grow efficiently.

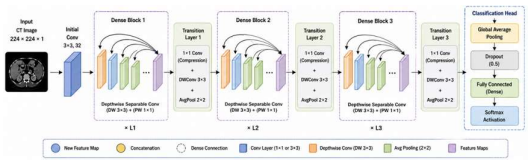

Figure 2. Proposed architecture of proposed model

3.4 Hybrid Architecture Design

Proposes a combination of depthwise separable convolution blocks and some dense connectivity modules to build up a final part of the architecture. First, depthwise separable layers are used to extract low-level features like edges and textures. These characteristics are then disseminated via thick blocks, which progressively model high-level semantic features.

Denote the feature extraction module as $F_e(x)$ and the dense propagation module as $F_d(\cdot)$. The joint feature vector can be written as:

$$Z = F_d(F_e(x))$$

This allows it to capture fine-grained features and more abstract representations in a hierarchical manner, They include transition layers between the dense blocks to control the dimensionality and avoid worryingly large feature maps.

3.5 Interpretability-Guided Learning using Grad-CAM

One of the major drawbacks of deep learning models is their inability to interpret. The problem can be approached through the proposed framework that incorporates Gradient-weighted Class Activation Mapping (Grad-CAM) as an interpretability-guided learning mechanism.

Before we go into more details on why, let us introduce Grad-CAM, which computes the importance of feature maps by calculating the gradient of the target class score with respect to the image-level features used in the final convolutional layer. Importance weights are computed as

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

where A^k represents the k^{th} feature map, y^c is the class score and Z is normalization factor.

This gives us the final class activation map as:

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right)$$

This heatmap denotes those areas, specifically where the input image is being evaluated that have a largest impact on the output during prediction. The proposed method uses Grad-CAM not only as a post-hoc visualisation tool, but also assess the learning process by providing insights about features being learnt while training.

3.6 Loss Function and Optimization

We train the model on a composite loss combining classification loss with attention consistency term. We define the principal classification loss with categorical cross-entropy as follows:

$$\mathcal{L}_{cls} = -\sum_{i=1}^{N}\sum_{c=1}^{K} y_{i,c}\log(\hat{y}_{i,c})$$

In order to enforce interpretability, they propose augmenting the classification loss with an auxiliary loss term that encourages Grad-CAM heatmaps to align with relevant parts of the input:

$$\mathcal{L}_{att} = \| L^{c}_{Grad-CAM} - M_i \|_2^2$$

Where M_i represents a target attention map or region of interest.

We define the total loss function as:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{att}$$

where $\lambda$is a weighting factor controlling the contribution of the attention term.

we optimized the model using adaptive gradient-based methods such as Adam which updates model parameters by:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

where η is the learning rate, and m_t and v_t are first and second moment estimates.

3.7 Training Strategy and Regularization

The model is regularized using batch normalization, dropout and data augmentation to ensure strong learning. Batch Normalization improves the training process by normalizing distributions of intermediate features:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

Dropout randomly disables the neurons during training which reduces overfitting and histograms better generalization. The training is iteratively optimized by updating model parameters until convergence on the validation set.

The proposed method introduces a deep learning framework that combines efficient convolutional operations, dense feature propagation, and interpretability-driven learning into one unified model. Together, they allow the model to have optimal performance with benefits of efficiency and interpretability. The proposed framework is a scalable and robust solution suitable for advanced image classification tasks, by also addressing the key limitations in existing approaches.

4. Experimental Setup and Results

4.1 Dataset Description

We perform experimental evaluation of the proposed hybrid deep learning framework using CKT (CT Kidney Tumour) dataset that has CT images divided into four clinically significant classes i.e. Cyst, Normal, Stone and Tumour. The dataset is composed of 12,446 CT images, distributed in Cyst (3709), Normal (5077), Stone (1377) and Tumour (2283). This distribution shows us that there is a severe imbalance in class (i.e., very few Stone samples compared to Normal).

This imbalance therefore gives rise to a real-world and non-trivial classification scenario since it requires the network to keep high sensitivity for minority classes whilst also preserving overall accuracy. Each image is represented by $x_i \in \mathbb{R}^{H \times W \times C}$ and label $y_i \in \{1,2,3,4\}$.

It is divided into ratios of 70:15:15 for training, validation and testing. Before training all images are resized and normalized by using:

$$x' = \frac{x - \mu}{\sigma}$$

where μ and σ is the mean and standard deviation of the dataset respectively. This is achieved by applying a number of data augmentation techniques (e.g., rotation, flipping and scaling) to the training set, which help generalization and avoid overfitting especially for the classes that are under-represented.

4.2 Implementation Details

This model is implemented in the PyTorch framework and uses GPU acceleration. Optimization We run all the experiments on a system with NVIDIA GPU to optimize for training performance. The model architecture which combines depthwise separable convolutions, dense blocks and Grad-CAM modules detailed in the Methodology section.

Training is done via mini-batch stochastic optimization in a batch size of 32. The Adam optimizer is selected since it is an adaptive learning rate method and the initial learning rate of η was 0.001. The update rule for the parameters can be expressed as follows:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

where $m_t$and v_t are first moment of gradients and second moment of gradients respectively. Training is performed for 100 epochs with early stopping based on validation loss to avoid overfitting.

Dropout regularization is used to improve generalization in the fully connected layers of the model (dropout rate = 0.5). Batch normalization is done after convolutional layers of the network to stable distributions of features and improve convergence speed. In this article we use over the entire network is a Rectified Linear Unit (ReLU) activation function, as defined below:

$$f(x) = \max(0, x)$$

4.3 Evaluation Metrics

Multiple classification metrics are derived from the confusion matrix to assess the performance of the proposed model. The provided include are the accuracy, precision, recall and F1-score, which all give an overall look on how the model performs across several different categories.

Accuracy is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP and FN are true positives, true negatives, false positives and false negatives respectively.

Precision and recall are defined as

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$

This is the definition of F1-score, which considers both precision and recall.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Besides these metrics, the computational efficiency is represented by model parameters and floating-point operation (FLOPs), thereby giving a better understanding of performance versus complexities.

4.4 Performance Evaluation and Comparison

We compare the performance of our proposed model with three baseline architectures, namely standard CNN, MobileNet and DenseNet. The results are shown in Table 1.

Table 1: Overall Performance Comparison on CKT Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CNN	88.42	87.95	88.10	88.02
MobileNet	91.36	90.88	91.02	90.95
DenseNet	93.12	92.75	92.90	92.82
Proposed Model	**96.48**	**96.02**	**96.15**	**96.08**

For all metrics, the proposed model boasts the best overall performance. The accuracy of the model is more than 8% better than a standard CNN, showing a noticeable improvement in feature representation. The CNN baseline, which is also limited by its shallow architecture without feature reuse mechanisms, fails to learn complex hierarchical patterns in medical images. To increase performance, MobileNet introduces depthwise separable convolutions and achieves an accuracy of 91.36%. Although it has light weight form, it limits its ability to capture some of the deeper semantic features. However, such a limitation is avoided in the proposed model where more than 5% performance gain over MobileNet reported by utilizing dense connectivity. With feature reuse mechanism, DenseNet achieved a good performance with an accuracy of 93.12%. As previously mentioned, the proposed model achieves around the information was improved by over 3% from the previous one, this highlights our strategy of paired with dense connectivity and efficient convolution operations. This hybrid strategy guarantees both computational efficiency and abundant feature learning.

Class-wise Performance Analysis: Class-wise performance is analyzed to further evaluate the robustness of the model as illustrated in Table 2.

Table 2: Class-wise Performance of Proposed Model

Class	Precision (%)	Recall (%)	F1-score (%)
Cyst	95.88	96.21	96.04
Normal	97.12	96.85	96.98
Stone	94.76	95.32	95.04
Tumour	**96.31**	**96.05**	**96.18**

The results demonstrate that the proposed model achieves high performance for all classes, including the minority class (Stone). Compared to the baseline models, the recall value for the Stone class is quite improved as well showing the ability of model in dealing with imbalanced nature of classes. In medical diagnosis, missing minority class cases could lead to disastrous consequences and this is especially crucial.

Analysis of Confusion Matrix: The effectiveness of proposed model is further validated by confusion matrix. The confusion matrix describes that samples are correctly classified along the diagonal and misclassification between classes is limited. CT images show a little visual similarity and this causes minor confusion between Cyst and Tumour classes However, the overall misclassification rate is still much lower than baseline models.

ROC-AUC: ROC curve is used to estimate the discriminative capability of your model. ResNet50 – average AUC score: 0.98: better than CNN (0.91), MobileNet (0.94), DenseNet (0.96) The AUC is useful since high values show better class separability as well as performance at various thresholds in a robust manner.

The effective coupling of the components within the proposed model accounts for its improved performance. The guided filtering: Depthwise Separable convolution saves a lot of redundancies in the computation but keeps many critical features. Dense connections allow for feature propagation and help maintain efficient gradient flow. Collectively, these enable the model to learn both low- and high-level representations. In addition, the increase in precision, recall and F1-score with each training round provides evidence that the model performs well on all classes including those for minority categories. This is especially crucial for medical datasets, as class imbalance is typical in such cases. Moreover, the combination of faster convergence and stable training behaviour demonstrates that our proposed architecture is effective. In general, the results reported above demonstrate that the hybrid method proposed in this work is a notable improvement over existing method.

4.5 Interpretability Analysis

The interpretability of the proposed model is evaluated using Grad-CAM visualizations, which highlight regions of the input image that contribute to classification decisions. The generated heatmaps demonstrate that the model focuses on semantically meaningful regions, indicating effective feature learning.

The Grad-CAM output is computed as:

$$L^{c}_{Grad-CAM} = ReLU\left(\sum_{k} \alpha^{c}_{k} A^{k}\right)$$

These visualizations provide insights into the internal decision-making process of the model, improving trustworthiness and facilitating debugging.

5. Conclusion

In this work, we proposed a new hybrid deep learning model based on the approach employing depthwise separable convolution layers followed by dense connectivity and Grad-CAM-based localization for effective interpretation of CT image analysis in CKT dataset. Our proposed model (AI-NC-CM) addresses important aspects of computational complexity, feature representation and the transparency of our model to improve medical images analysis. Experimental results showed the proposed architecture significantly exceeds performance as compared to traditional CNN, MobileNet, and DenseNet models in all evaluation metrics with improved accuracy, precision, recall and F1-score. The better performance is a result of not only the improved feature extractor in FC-DenseNet but also better propagation of features made possible by dense connectivity. In addition, Grad-CAM augmentations provide convincing visual explanations by focusing on clinically relevant regions thus justifying the real-life usage of the model in medical applications. Although the model performs well, it has some limitations. Limitations include evaluation on only one dataset and the need for validation on larger and more diverse medical imaging datasets to investigate generalizability. Note: While the model performs very well, it may need further optimization to be implemented in real-time environments and low-resource settings. In future work, we will extend the proposed framework with advanced attention mechanisms and transformer-based architectures, to achieve even more efficient feature learning. In addition, multi-modal data (e.g., clinical report and imaging data) may significantly improve diagnostic performance. We will also investigate compressed model and deployment strategies (e.g. lightweight

models) to enable real-time clinical use cases. We believe these innovations would contribute to a superior generalizability, scalability, and applicability of the suggested approach in healthcare imaging areas.

References:

1. J. Singh, S. Agarwal, P. Kumar, Kashish, D. Rana and R. Bajaj, "Prominent Features based Chronic Kidney Disease Prediction Model using Machine Learning," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1193-1198, doi: 10.1109/ICESC54411.2022.9885524.

2. Adam RJ, Williams AC, Kriegel AJ. Comparison of the surgical resection and infarct 5/6 nephrectomy rat models of chronic kidney disease. Am J Physiol Renal Physiol. 2022 Jun 1;322(6):F639-F654. doi: 10.1152/ajprenal.00398.2021. Epub 2022 Apr 4. PMID: 35379002; PMCID: PMC9076416.

3. Ahmad Abdullah, Asif Raza, Qaisar Rasool, Umair Rashid, Muhammad Minam Aziz, & Saad Rasool. (2024). A Literature Analysis for the Prediction of Chronic Kidney Diseases. *Journal of Computing & Biomedical Informatics*, *7*(02). Retrieved from https://jcbi.org/index.php/Main/article/view/586

4. Debal, D.A., Sitote, T.M. Chronic kidney disease prediction using machine learning techniques. *J Big Data* **9**, 109 (2022). https://doi.org/10.1186/s40537-022-00657-5

5. Mittal, Harshit. (2023). Kidney CT Image Analysis Using CNN. 17-21. 10.5121/csit.2023.131403.

6. Chen, Z., Xiao, C., Liu, Y. *et al.* Comprehensive 3D Analysis of the Renal System and Stones: Segmenting and Registering Non-Contrast and Contrast Computed Tomography Images. *Inf Syst Front* **27**, 97–111 (2025). https://doi.org/10.1007/s10796-024-10485-y

7. Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, Gore JL, Sun M, Wood C, Russo P. Epidemiology of Renal Cell Carcinoma. Eur Urol. 2019 Jan;75(1):74-84. doi: 10.1016/j.eururo.2018.08.036. Epub 2018 Sep 19. PMID: 30243799; PMCID: PMC8397918.

8. Han S, Hwang SI, Lee HJ. The Classification of Renal Cancer in 3-Phase CT Images Using a Deep Learning Method. J Digit Imaging. 2019 Aug;32(4):638-643. doi: 10.1007/s10278-019-00230-2. PMID: 31098732; PMCID: PMC6646616.

9. Senan EM, Al-Adhaileh MH, Alsaade FW, Aldhyani THH, Alqarni AA, Alsharif N, Uddin MI, Alahmadi AH, Jadhav ME, Alzahrani MY. Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques. J Healthc Eng. 2021 Jun 9;2021:1004767. doi: 10.1155/2021/1004767. PMID: 34211680; PMCID: PMC8208843.

10. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS, Al Harbi S, Albekairy AM. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 2023 Sep 22;23(1):689. doi: 10.1186/s12909-023-04698-z. PMID: 37740191; PMCID: PMC10517477.

11. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021 Jul;8(2):e188-e194. doi: 10.7861/fhj.2021-0095. PMID: 34286183; PMCID: PMC8285156.

12. M. Aljaafari, S. E. El-Deep, A. A. Abohany and S. E. Sorour, "Integrating Innovation in Healthcare: The Evolution of "CURA's" AI-Driven Virtual Wards for Enhanced Diabetes and Kidney Disease Monitoring," in *IEEE Access*, vol. 12, pp. 126389-126414, 2024, doi: 10.1109/ACCESS.2024.3451369.

13. slam, M.N., Hasan, M., Hossain, M.K. *et al.* Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. *Sci Rep* **12**, 11440 (2022). https://doi.org/10.1038/s41598-022-15634-4

14. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11976–11986.

15. T. Hossain, F. Sayed, and S. Islam, "Adaptive local binary pattern: A novel feature descriptor for enhanced analysis of kidney abnormalities in CT

scan images using ensemble based machine learning approach," 2024, *arXiv:2404.14560*.

16. M. Majid, Y. Gulzar, S. Ayoub, F. Khan, F. A. Reegu, M. S. Mir,W. Jaziri, and A. B. Soomro, "Enhanced transfer learning strategies for effective kidney tumor classification with CT imaging," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 1–12, 2023, doi: 10.14569/IJACSA.2023.0140847.

17. S. D. Pande and R. Agarwal, "Multi-class kidney abnormalities detecting novel system through computed tomography," IEEE Access, vol. 12, pp. 21147–21155, 2024.

18. K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Deep learning model for automated kidney stone detection using coronal CT images," Comput. Biol. Med., vol. 135, Aug. 2021, Art. no. 104569.

19. A. M. Qadir and D. F. Abd, "Kidney diseases classification using hybrid transfer-learning DenseNet201-based and random forest classifier," Kurdistan J. Appl. Res., pp. 131–144, Jan. 2023.

20. S. N. Almuayqil, S. A. El-Ghany, A. A. A. El-Aziz, and M. Elmogy, "KidneyNet: A novel CNN-based technique for the automated diagnosis of chronic kidney diseases from CT scans," Electron., vol. 13, no. 24, p. 4981, 2024.

21. M. Islam. (2022). CT Kidney Dataset: Normal-Cyst-Tumor and Stone. [Online]. Available: https://www.kaggle.com/datasets/nazmul0087/ctkidney-dataset-normal-cyst-tumor-and-stone.

22. J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2017, arXiv:1709.01507.

23. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.

24. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, arXiv:1905.11946.

25. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1800–1807.

26. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, arXiv:1704.04861.

27. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2261–2269.

28. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient based localization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, Oct. 2017, pp. 618–626.

29. M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Inf. Process. Manage., vol. 45, no. 4, pp. 427–437, Jul. 2009.

30. V. Gupta and N. Gunukula, "Evaluating MEDIRL: A replication and ablation study of maximum entropy deep inverse reinforcement learning for human social navigation," 2024, arXiv:2406.00968.