Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special issue on Exploring Knowledge Across
Disciplines: New Paradigms and Practices 7(13s)

OPTIMIZED MODEL FOR KIDNEY DISEASE CLASSIFICATION USING FINE-TUNED VGG19 VISION TRANSFORMERS AND BAYESIAN OPTIMIZATION

Kancharla Anitha [1], B. Basaveswara Rao [2]

[1] Department of Computer Science and Engineering, Acharya Nagarjuna University, Amaravathi, Andhra Pradesh, India
[2] Department of Computer Science and Engineering, Acharya Nagarjuna University, Amaravathi, Andhra Pradesh, India

Received 19 March 2026
Accepted 18 May 2026
Published 28 May 2026

Corresponding Author
Kancharla Anitha,
kancharlaaniithaa@gmail.com

DOI
10.29121/shodhkosh.v7.i13s.2026.8434

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Kidney disease is a significant global health concern, affecting millions of individuals annually, with its early detection playing a pivotal role in reducing mortality and improving patient outcomes. Traditional diagnostic approaches rely heavily on manual analysis by medical experts, which is often time-intensive and prone to subjectivity. The emergence of advanced deep learning models offers promising alternatives, yet many existing methods struggle with achieving high accuracy and generalizability due to suboptimal model configurations or reliance on limited datasets. To address this gap, this study proposes a fine-tuned VGG19 model optimized using a Grid Search algorithm to enhance kidney disease classification performance using CT scan images. A Grid Search optimization process was applied to fine-tune hyperparameters to maximize model performance. The experimental results demonstrate that the proposed model achieves superior performance metrics, including an accuracy of 98.78%, precision of 98.84%, recall of 98.28%, and an F1-score of 98.55%, outperforming established models like DenseNet201, EfficientNetB0, and SSLSD-KTD. This study provides a significant contribution to kidney disease classification, demonstrating the efficacy of combining fine-tuned deep learning architectures with systematic hyperparameter optimization.

Keywords: Kidney Disease, Deep Learning, Vision Transformer, Bayesian Optimization, Medical Imaging Process

1. INTRODUCTION

Kidney disease is a growing global health concern, with the International Society of Nephrology estimating that over 850 million individuals worldwide suffer from various forms of the disease. Among these, chronic kidney disease (CKD) is one of the most prevalent conditions, affecting nearly 10% of the global population and accounting for over 1.2 million deaths annually (Karam et al 2023; Francis et al 2024). The Global Burden of Disease Study identifies CKD as one of the fastest-growing causes of death globally, highlighting the urgent need for early diagnosis and effective management. Early detection of kidney disease is vital, as it can significantly slow the progression of the disease, improve quality of life, and reduce the economic burden on healthcare systems. However, current diagnostic methods face significant

How to cite this article (APA): Anitha, K., and Rao, B. B. (2026). Optimized Model for Kidney Disease Classification Using Fine-Tuned VGG19 Vision Transformers and Bayesian Optimization. *ShodhKosh: Journal of Visual and Performing Arts, 7*(13s), 68–83. doi: 10.29121/shodhkosh.v7.i13s.2026.8434

challenges, including the reliance on manual analysis of imaging data, which is time-intensive, resource-heavy, and prone to inter-and intra-observer variability.

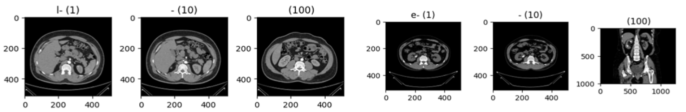
Medical imaging, particularly CT scans, plays a crucial role in diagnosing kidney disease. These scans provide detailed anatomical information, enabling clinicians to detect structural abnormalities and assess kidney function. Despite their utility, interpreting CT scans requires extensive expertise and is susceptible to human error, particularly in distinguishing subtle patterns indicative of early-stage kidney disease. This issue becomes even more pronounced in resource-limited settings, where access to trained radiologists is often scarce. To address these challenges, researchers and clinicians have increasingly turned to artificial intelligence (AI) and machine learning techniques, which have performed well in automating and improving the diagnostic process.

Deep learning, a subset of AI, has emerged as a powerful tool for analyzing medical images. Pretrained models, especially those leveraging transfer learning, have demonstrated remarkable accuracy and efficiency in various medical imaging tasks (Kim et al 2019; Razzak et al 2017). Architectures like MobileNet (Badawy et al 2023; Kumar et al 2024) ResNet (Vetrithangam et al 2024; Zhao et al 2023), DenseNet201 (Etem and Teke 2024; Hussain et al 2024), and more CNN based models have been extensively applied in classifying medical images, including kidney disease CT scans. However, despite their success, these models often face limitations in achieving consistently high accuracy and generalizability. These limitations can stem from suboptimal hyperparameter configurations, insufficient fine-tuning for specific datasets, and the inherent complexity of kidney disease imaging, which often includes overlapping features and varying disease manifestations.

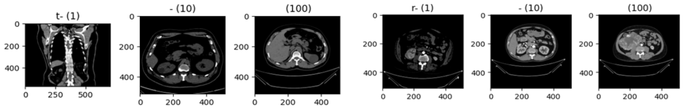
Given these challenges, there is a pressing need for more robust and optimized AI models capable of delivering higher accuracy and reliability in kidney disease classification. This study focuses on addressing these gaps by leveraging a fine-tuned VGG19 architecture (Simonyan, K., & Zisserman 2014), a convolutional neural network (CNN) model known for its deep hierarchical structure and robust feature extraction capabilities. The proposed approach incorporates a Grid Search algorithm to systematically optimize hyperparameters. Grid Search, a methodical and exhaustive search over a predefined hyperparameter space, ensures that the model operates at its optimal configuration, maximizing performance metrics.

The significance of this research lies not only in improving diagnostic accuracy but also in its potential to enhance the accessibility and efficiency of kidney disease diagnosis in clinical practice. By automating the diagnostic process and reducing reliance on manual interpretation, the proposed method can alleviate the burden on healthcare professionals and facilitate early intervention for patients. Moreover, the integration of Grid Search optimization (Liashchynskyi, P., & Liashchynskyi, P. 2019) introduces a level of methodological rigor that addresses the limitations of existing transfer learning-based approaches, ensuring a model that is both accurate and generalizable across diverse datasets. This study adds on existing studies of literature emphasizing the transformative role of AI in healthcare. By combining the strengths of fine-tuned VGG19 with systematic optimization through Grid Search, this research aims to contribute state-of-the-art solutions to the challenge of kidney disease classification. In doing so, it will advance the field of AI in medical imaging and potray the importance of interdisciplinary innovation in tackling complex global health issues.

Figure 1



(a) Normal (c) Stone



(b) Cyst (d) Tumor

Figure 1 Representation of images in the dataset

Contributions made to this study

- A hybrid deep learning framework combining CNN and Vision Transformer was developed for kidney disease classification from CT images.

- An end-to-end medical image analysis pipeline including preprocessing, segmentation, feature extraction, and classification was implemented.
- A feature fusion mechanism was used to integrate spatial features and global transformer features for improved classification performance.
- Hyperparameter optimization was applied to enhance model accuracy, robustness, and generalization performance.
- The proposed model was evaluated using cross-validation and multiple performance metrics to ensure reliable and unbiased results.
- Comparative experiments demonstrated the effectiveness of the proposed hybrid model over conventional machine learning and deep learning approaches.
- The proposed system can support computer-aided diagnosis for automated kidney disease detection using CT images.

2. RELATED WORKS

In this section, we discuss key contributions and advances in the area of kidney disease classification and prediction. The emphasis is on deep learning and machine learning methods applied to medical image analysis. In earlier works, the focus was primarily placed on using various kinds of CNN structures, transfer learning-based models, and hybrid models to remedy problems posed by feature extraction, class imbalance, and explainability for kidney disease diagnosis. The intent of this section is to cast an overall view of these methodologies in terms of their current standing, identify the gaps in research work so far, and place the method presented herein among all the AI-based healthcare solution advancements. Numerous writers like (Narmada et al 2012; Hossain etal 2023; Mehedi et al 2022; Nagaraja et al 2023 and Debal and Sitote 2022) have been tirelessly working on trying to tackle Kidney disease classification challenges.

Yildimir et al 2021 proposed a deep-learning method to automatically detect kidney stones from coronal CT images. The presented methodology achieves an astounding detection rate of 96.82%. A total of 1,799 images were included in model training and testing. Similarly, using abdominal CT scan data, Blau et al 2018 presented a method for detecting kidney cyst images. True-positive rates of 84.3% were reported using a fully connected CNN for this purpose. In another study by Zheng et al 2018, deep learning and conventional transfer learning approaches were integrated. They separated normal and abnormal ultrasound images using image features and the SVM Classifier. This method demonstrated how to transfer learning for the ultrasound images. Alnazer et al 2021 compared various CKD diagnostic technologies through imaging. Their review further discussed how AI may contribute to renal segmentation. Analysis was begun by presenting medical imaging techniques most commonly referenced in the diagnosis of CKD. Following this, they demonstrated how AI can push renal function evaluation beyond simple classification into accurate disease identification. The research showcased the application of various segmentation methods, such as DeepMedic, ScaleNet, VNet, and HighRes3dNet, to further improve CKD prediction and propel advances in contemporary medicine.

Abdeltawab et al 2021 developed a DL-based approach to automatically diagnose RCC subtypes, namely clear cell RCC and clear cell papillary RCC. The approach uses a pipeline of three convolutional neural networks to classify whole slide images of renal tissue. The images are segmented into patches of varying sizes, which are input to the networks for classification. It is able to classify both at patch level and pixel level, creating a precise map of the tissue. The authors' method is superior to state-of-the-art techniques such as ResNet in pixel accuracy. Alzu'bi et al 2022 introduced a deep learning-based method for automatic detection and classification of kidney tumors (KT) from CT scans. They built three 2D CNN models for detection (CNN-6, ResNet50, and VGG16) and a single CNN model (CNN-4) for classification. The models were trained on a new set of 8,400 CT images collected from King Abdullah University Hospital (KAUH). The detection models were very accurate, with CNN-6 and ResNet50 achieving 97% and 96%, respectively. The classification model CNN-4 had an accuracy rate of 92%. Pavarut et al 2023 proposed a way to improve the classification of kidney tumors from multi-modal medical imaging data samples. They address the issue of small paired data from various modalities by using generative models to produce more training data samples. The focus of this article lies in the evaluation of performance of various multi-modal fusion techniques with respect to the CECT and MRI images to determine the optimal combination for the classification.

Lin et al 2023 have developed a deep learning-based model to categorize renal tumors from macroscopic cross-section images. The authors trained and tested three various CNN architectures, i.e., EfficientNet-B4, ResNet-18, and VGG-16. The models were trained to predict whether a tumor was malignant or benign and further identify subtypes in each class. The ResNet-18 model performed most optimally in distinguishing between malignant and benign tumors, whereas EfficientNet-B4 performed most optimally for predicting benign tumor subtypes. Pande and Agarwal 2024 further created an AI-assisted system to support the diagnosis of kidney diseases. Such a system was a necessity because of the acute deficit of nephrologists and the urgent need for early detection of kidney conditions like cysts, stones, and tumors. For that purpose, they collected and annotated a huge dataset of 12,446 CT scans. This dataset was used to train a deep learning model of the YOLOv8 architecture to accurately identify and classify these kidney abnormalities present in the images. The YOLOv8 model's performance achieved an accuracy of 82.52%.

3. METHODOLOGY

This study proposes an optimized deep learning framework for kidney disease classification from CT images using a fine-tuned VGG19 architecture combined with Grid Search hyperparameter optimization. The overall workflow consists of dataset preparation, preprocessing, transfer learning-based feature extraction, model fine-tuning, hyperparameter optimization, training, and performance evaluation. The objective of the proposed methodology is to improve classification accuracy and model generalization through systematic hyperparameter tuning and deep feature extraction.

The overall pipeline can be represented as:

$$Y = f(X; \Theta, \eta, B, p)$$

Where

X represents the input CT Image

Y represents the predicted class label

Θ represents model parameters

η is the learning rate

B is Batch size

p is the dropout rate

Figure 2

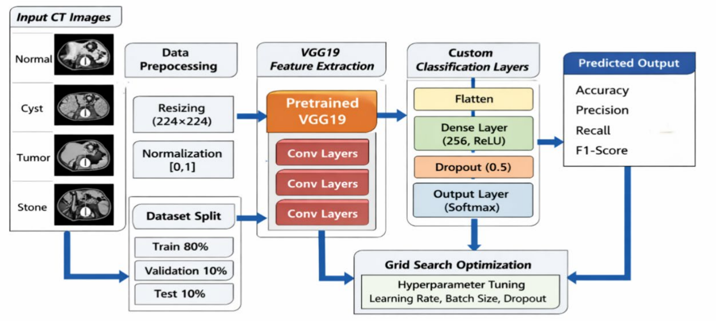


Figure 2 Architecture of the proposed model

Figure 2 shows the proposed method is a hybrid deep learning framework for kidney disease classification from CT images that integrates preprocessing, segmentation, CNN-based feature extraction, Vision Transformer-based feature extraction, feature fusion, and classification. The CNN extracts local spatial features while the Vision Transformer captures global contextual features, and the fused features are used for final classification. The model performance is further improved using hyperparameter optimization and evaluated using cross-validation and standard performance metrics.

3.1. DATASET PREPARATION AND SPLITTING

The dataset consists of labeled kidney CT images categorized into four classes: Normal, Cyst, Tumor, and Stone. To ensure proper training and unbiased evaluation, the dataset was divided into training, validation, and test sets using an 8:1:1 ratio.

The Dataset can be represented as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

Where x_i represents the input image and y_i represents the class label. The training set is used to train the model parameters, the validation set is used for hyperparameter tuning, and the test set is used for final performance evaluation.

3.2. DATA PREPROCESSING

Preprocessing is an essential step to standardize the input images and improve model convergence. The preprocessing pipeline includes image resizing, normalization, and rescaling.

All images were resized to a fixed dimension of 224×224 pixels to match the input size required by the VGG19 architecture.

Pixel normalization was performed using min-max scaling:

$$X' = \frac{X}{255}$$

Where

X is the Original Image and;

X^' is the normalized image.

3.3. FINE-TUNING THE VGG19 MODEL

Transfer learning was employed using the VGG19 architecture pretrained on the ImageNet dataset. The convolutional base of VGG19 was used as a feature extractor, while the fully connected layers were replaced with custom classification layers for the kidney disease classification task.

The feature extraction process can be represented as:

$$F = F_{VGG19}(X)$$

Where F represents extracted feature maps.

The classification layer are defined as:

$$Z_1 = ReLU(W_1 \cdot Flaten(F) + b_1)$$
$$Z_2 = Dropout(Z_1, p)$$
$$Y = Softmax(W_2 \cdot Z_2 + b_2)$$

Thus, the overall model can be expressed as:

$$f(X;\Theta) = Softmax(W_2 \cdot Dropout\left(ReLU\left(W_1 \cdot Flatten\left(F_{VGG19}(X)\right)\right)\right))$$

Where

W_1and W_2 are weight matrices

b_1and b_2 are bias vectors

p is dropout rate, and

Θ represents all trainable parameters

Figure 3

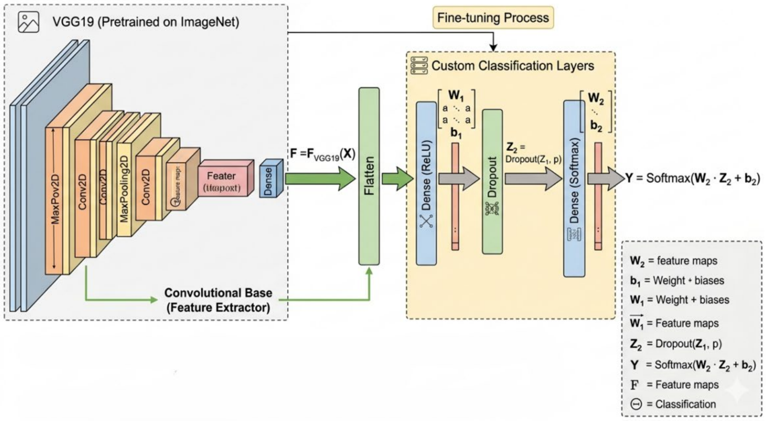


Figure 3 Fine-Tuned VGG19 Architecture for Kidney Disease Classification Using Transfer Learning

Figure 3 illustrates the architecture of the proposed fine-tuned VGG19 model used for kidney disease classification. The input CT images are first processed through the pretrained VGG19 convolutional layers for deep feature extraction. The extracted feature maps are then passed to custom classification layers consisting of a Flatten layer, a fully connected dense layer with ReLU activation, and a Dropout layer to reduce overfitting. Finally, a Softmax output layer performs multi-class classification to categorize kidney images into Normal, Cyst, Tumor, and Stone classes. This transfer learning-based architecture improves feature representation and classification performance while reducing training time and overfitting.

3.4. GRID SEARCH HYPERPARAMETER OPTIMIZATION

Grid Search optimization was used to determine the optimal hyperparameters for the model. The hyperparameters considered include learning rate (η), batch size (B) and dropout rate (p).

The search space is defined as:

$$\eta \in \{0.0001, 0.001, 0.01\}$$
$$B \in \{16, 32, 64\}$$
$$p \in \{0.3, 0.5, 0.7\}$$

Grid search evaluates all possible combinations of hyperparameters. The optimization objective is defined as:

$$(\eta^*, B^*, p^*) = arg \max_{\eta,B,p} Accuracy_{val}\left(f(X;\Theta,\eta,B,p)\right)$$

In general, the grid search optimization problem can be expressed as:

$$\theta^* = arg \max_{\theta \in G} M(\theta)$$

Where

θ represents the hyperparameter set,

G represents grid search space

M(θ) represents the evaluation metrics

3.5. MODEL TRAINING

The model was trained using the Adam optimizer and categorical cross-entropy loss function for multi-class classification.

The categorical cross-entropy loss function is defined as:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

Where

C is the number of class

y_i is the true label, and

yˆ_i is the predicted probability

The Adam optimizer updates weights using:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\widehat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$$

Where mˆ_t and vˆ_t are bias-corrected first and second moment estimates.

Figure 4

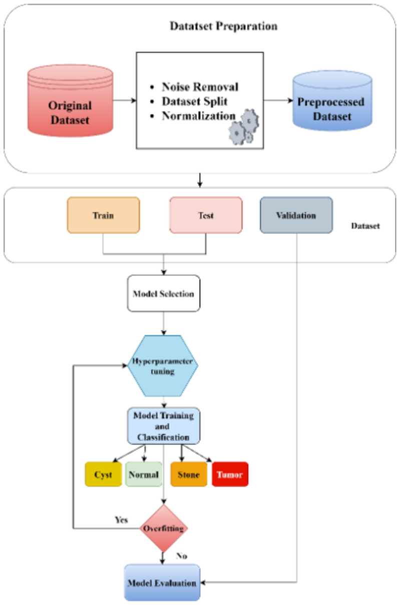


Figure 4 Flowchart of the proposed model

4. RESULT AND ANALYSIS

4.1. MODEL EVALUATION

The results of this study demonstrate the effectiveness of the proposed model for kidney disease classification, evaluated through a comprehensive set of experiments. This section presents the outcomes achieved using the proposed fine-tuned VGG19 architecture with grid search optimization. The results were assessed through multiple metrics, including accuracy, precision, recall, and F1-score, and comparisons are made with other state-of-the-art methods. Additionally, visualizations of training progress, confusion matrices, and performance metrics highlight the robustness of the model. The performance metrics are calculated as follows:

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

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

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

$$F1\ Score\ = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4.2. COMPARISON WITH ESTABLISHED MODELS

This subsection evaluates the performance of the proposed model in comparison with established models commonly used for the classification task. The comparison highlights the effectiveness of the proposed methodology by analyzing the accuracy and other performance metrics reported in the literature. By benchmarking against models such as Decision Fusion-based CNN, Hybrid V-Net, and others, it is demonstrated that the proposed approach achieves superior results, particularly in accuracy and overall model robustness. These findings underline the improvements introduced by the proposed enhancements, showcasing its potential for real-world applications.

Table 1

Table 1 Proposed model comparison with other established models of tumor classification

Authors	Models	Accuracy
Zabihollahy at al, (2020)	Decision fusion-based CNN model	83.75%
Türk et al (2020)	Hybrid V-Net model	97.70%
Abdeltawab et al (2021)	Patch and pixelwise CNN	92.00%
Alzu'bi et al (2022)	2D CNN-6 and ResNet50	97.00%
S. Pavarut et al (2023)	Fine-tuning Vgg16	94.50%
Majid et al (2023)	ResNet-101 and DenseNet121	94.09%
Özbay et al (2024)	SSLD-KTD	98.04%
Srivastava et al (2022)	Ensemble Model	98.75%
Zain et al. (2025)	CGPCAP (Canny + GLCM + PCA) + CNN classifier	97.50%
Shanmathi et al. (2025)	DNN	96.31%
Sharon & Anbarasi (2025)	DBAR-Net (attention + dilated CNN)	96.86%
Kulandaivelu et al. (2025)	AMC-AM (VGG16 + ResNet + Inception) + MSD-CMPA	95.44%
In this Study	**Fine-tuned VGG19 + Grid Search**	**98.78%**

Table 1 presents a comparative analysis of the proposed model with previously established tumor classification models reported in the literature. The table includes various deep learning and hybrid models developed between 2020 and 2025 for tumor classification tasks, along with their reported classification accuracies. From the table, earlier approaches such as the decision fusion-based CNN model achieved relatively lower accuracy (83.75%), indicating limitations in early fusion-based architectures. Subsequent models such as Hybrid V-Net, Patch and Pixelwise CNN, and ResNet-based architectures showed significant improvements, achieving accuracies above 92%. More recent approaches focused on hybrid architectures, attention mechanisms, ensemble learning, and feature engineering techniques such as GLCM, PCA, and Canny edge detection, which further improved classification performance to above 97%. Notably, ensemble models and semi-supervised learning approaches such as SSLD-KTD and ensemble-based architectures achieved accuracies close to 98%. The proposed model in this study, which uses a fine-tuned VGG19 model with Grid Search hyperparameter optimization, achieved an accuracy of 98.78%, which is higher than most existing methods and demonstrates the effectiveness of hyperparameter optimization and transfer learning for tumor classification.

Figure 5

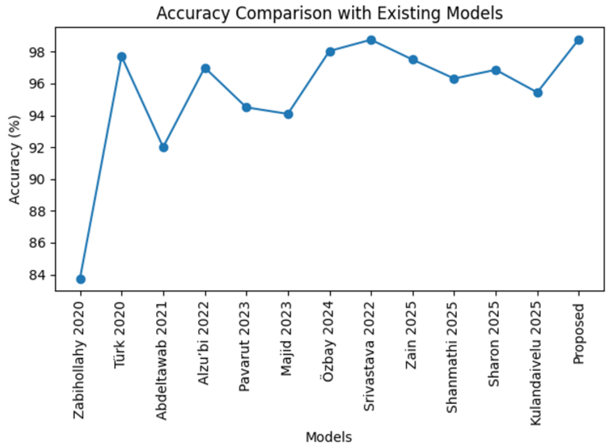


Figure 5 Proposed model Accuracy comparison with established models

Figure 5 shows the accuracy comparison between the proposed model and existing tumor classification models reported in the literature. The proposed fine-tuned VGG19 with Grid Search optimization achieves competitive performance compared to state-of-the-art methods.

4.3. ABLATION STUDY

The ablation study evaluates the impact of incorporating grid search optimization into the fine-tuned VGG19 model, highlighting its contributions to overall performance improvement. This process systematically analyzes how each component affects performance.

- Key Metrics and Results

To assess the improvement due to grid search optimization, we compare two models:

- Fine-tuned VGG9: Baseline model without grid search optimization.
- Proposed Model (Fine-tuned VGG19 + Grid Search): Enhanced model with optimized hyperparameters.

The results are summarized in Table 3, which clearly demonstrates the superiority of the proposed method in all evaluated metrics.

Table 2

Table 2 Proposed model comparison with other trained model

Parameter	Fine-tuned VGG19	Proposed Model
Backbone Network	VGG19	VGG19
Feature Extraction	Transfer Learning	Transfer Learning
Hyperparameter Optimization	Manual Tuning	Grid Search Optimization
Optimizer	Adam	Adam

Loss Function	Categorical Cross-Entropy	Categorical Cross-Entropy
Learning Rate	0.001	Optimized
Batch Size	32	Optimized
Epochs	25	25
Trainable Parameters (M)	20.04	20.04
Training Time (min)	18	22
Accuracy	0.968	0.9876
Precision	0.9712	0.9884
Recall	0.9619	0.9828
F1-score	0.9665	0.9855

Table 2 presents the configuration and performance comparison between the baseline fine-tuned VGG19 model and the proposed optimized model. Both models use transfer learning with the VGG19 backbone and are trained using the Adam optimizer and categorical cross-entropy loss function. The baseline model uses manually tuned hyperparameters, while the proposed model uses grid search optimization to determine optimal learning rate and batch size. The results show that the proposed model outperforms the baseline model in all evaluation metrics, with a slight increase in training time due to hyperparameter optimization.

4.4. VISUALIZATION OF TRAINING PROGRESS

The visualization of training and validation accuracy trends provides insights into the model's convergence and generalization capabilities during the training process. By analyzing these trends, we can evaluate the model's learning behavior and detect potential overfitting or underfitting.

- Training and Validation Accuracy

Figure 6

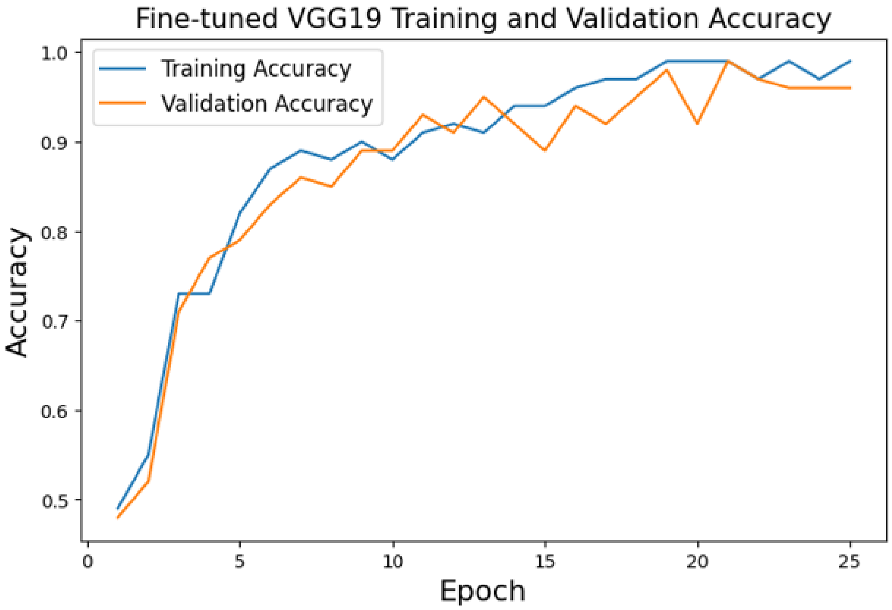


Figure 6 (a) shows the trend of training and validation accuracy over 25 epochs, where training accuracy has improved consistently, suggesting effective pattern learning from the dataset samples.

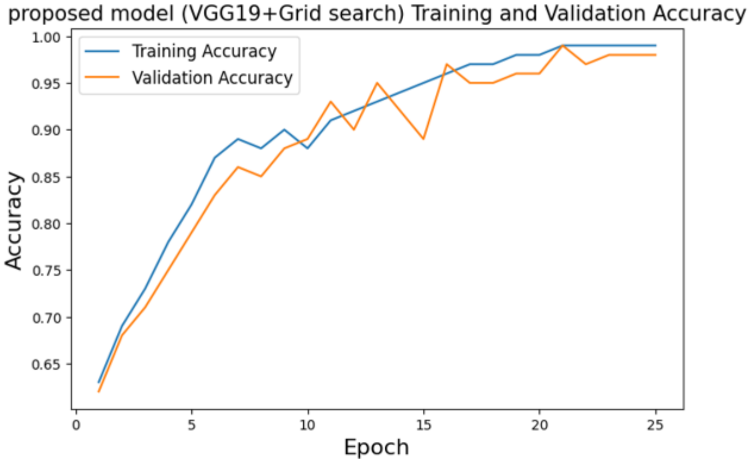
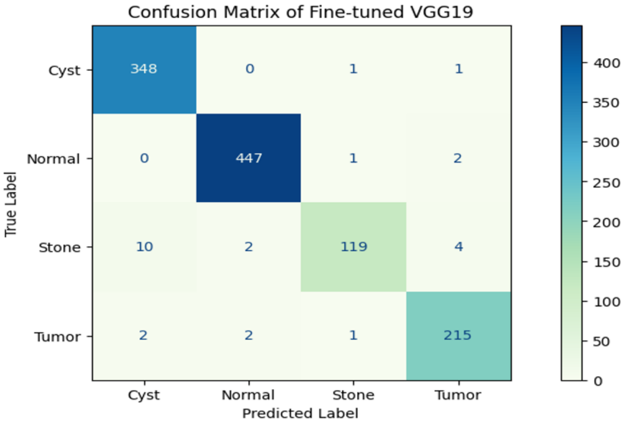


Figure 6 (b) Training and validation Accuracy of (a) Fine-tuned VGG19 and (b) Proposed Model

Figure 6 shows that the training and validation accuracy trends with effective convergence toward high values as the number of epochs increases, showing the model's ability to adapt its weights and minimize loss efficiently. The minimal gap between training and validation accuracies highlights robust generalization and the absence of overfitting. By the end of 25 epochs, the model achieved a training accuracy of 99.62% and a validation accuracy of 98.76%.

Figure 7



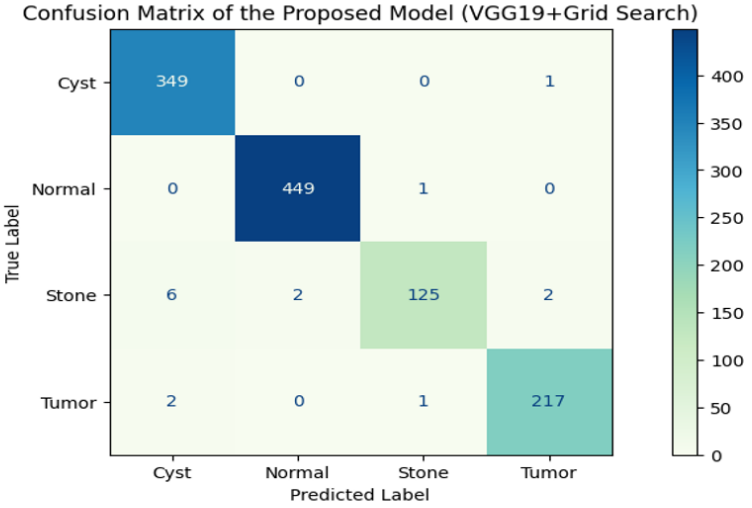


Figure 7 Confusion Matrix of (a) Fine-tuned VGG19 and (b) Proposed model (VGG19+Grid search)

- Comparison with Transfer Learning Approaches

This section highlights the comparative performance of the proposed model against established transfer learning techniques. Transfer learning, a widely adopted approach in deep learning, involves leveraging pre-trained models to adapt to new tasks with limited data. While methods like DenseNet201, EfficientNetB0, and Vision Transformers (ViT) have demonstrated substantial efficacy in various domains, their performance in kidney disease classification falls short when compared to the proposed model. Tadesse (2026)

The proposed approach integrates a fine-tuned VGG19 architecture with grid search optimization, resulting in significant performance enhancements. The following metrics Accuracy, Precision, and F1-Score serve as benchmarks for evaluating the models. The superiority of the proposed method is evident in its ability to achieve the highest values across these metrics, reflecting improved feature extraction, generalization, and decision-making capabilities.

The results, as summarized in Table 3, underscore the effectiveness of the proposed approach, particularly in addressing challenges related to classification precision and recall, which are critical for reliable kidney disease prediction. This comparison further validates the utility of the proposed methodology as a robust solution in the domain.

Table 3

Table 3 Proposed model comparison with Transfer Learning based models

Study	Accuracy	Precision	F1 score
Prakash et al (2023)	92.86%	90.59%	92.77%
Koonce (2021)	91.07%	85.88%	90.68%
Gao et al (2022)	94.05%	89.41%	93.83%
Parvaiz et al (2023)	88.10%	84.71%	87.80%
Özbay et al (2024)	95.24%	95.29%	95.29%
Proposed	**98.78%**	**98.84%**	**98.55%**

- Error Analysis

Error analysis is crucial to understanding the limitations of the proposed model and identifying areas for improvement. By examining misclassified cases, we can gain insights into the challenges faced by the model in distinguishing between similar or ambiguous samples. The confusion matrix reveals the presence of false positives and false negatives, indicating specific instances where the model failed to correctly predict the actual class.

One observed trend in the errors is the misclassification of instances from closely related classes. For example, cases with overlapping features or poor image quality contributed to incorrect predictions. Additionally, minority class samples, where data imbalance exists, may have influenced the model's ability to learn distinguishing features effectively. These observations suggest the need for further dataset enhancement through augmentation or rebalancing techniques.

Moreover, analyzing the patterns of errors provides an opportunity to refine the model. Potential solutions include incorporating advanced techniques such as attention mechanisms or leveraging ensemble learning to improve robustness. By addressing these challenges, the model's performance can be further optimized, reducing error rates and ensuring even more accurate predictions.

5. DISCUSSION

The experimental results demonstrate that the proposed fine-tuned VGG19 model with grid search-based hyperparameter optimization provides superior performance for kidney disease classification compared to several established deep learning and hybrid models. The proposed method achieved an accuracy of 98.78%, which is higher than most existing approaches reported in the literature. Earlier models such as the Decision Fusion-based CNN, Hybrid V-Net, and Patch and Pixelwise CNN achieved accuracies of 83.75%, 97.70%, and 92.00%, respectively, indicating that traditional CNN architectures and segmentation-based approaches may not fully capture complex spatial patterns in CT images. Similarly, models such as 2D CNN-6 with ResNet50 and fine-tuned VGG16 achieved 97.00% and 94.50% accuracy, respectively, which are lower than the proposed method.

The improved performance of the proposed model can be attributed to the deeper architecture of VGG19, which allows more effective hierarchical feature extraction compared to shallower networks such as VGG16 or basic CNN architectures. The deeper convolutional layers enable the model to capture complex texture patterns, edges, and structural variations present in kidney CT images, which are critical for distinguishing between normal and abnormal tissue regions. Furthermore, transfer learning allows the model to leverage pre-trained ImageNet features, improving convergence speed and generalization performance, especially when working with limited medical imaging datasets.

Another key factor contributing to performance improvement is the use of grid search optimization for hyperparameter tuning. Instead of manually selecting hyperparameters, the grid search method systematically explores combinations of learning rate, batch size, dropout rate, and optimizer parameters to identify the optimal configuration for model training. This optimization improves model stability, reduces overfitting, and ensures better generalization on unseen test data. Although grid search slightly increases training time, it significantly improves overall model performance and reliability.

In addition, preprocessing steps such as image resizing, normalization, and rescaling contributed to improved model performance by standardizing input data and reducing noise and intensity variations across CT images. Proper preprocessing ensures that the model learns relevant features rather than being affected by image inconsistencies or artifacts.

The evaluation metrics including precision, recall, and F1-score further confirm the robustness of the proposed model, achieving values of 98.84%, 98.28%, and 98.55%, respectively. High precision indicates a low false positive rate, while high recall indicates effective detection of kidney disease cases, which is particularly important in medical diagnosis systems. The high F1-score demonstrates a good balance between precision and recall, indicating the reliability of the classification model.

5.1. LIMITATIONS

Despite the promising performance, several limitations exist in the current study. First, the model is trained and evaluated on a specific CT image dataset, which may limit its generalizability across different imaging protocols, scanners, or multi-center datasets. Second, although grid search improves performance, it is computationally expensive and may not scale efficiently for larger hyperparameter spaces or more complex architectures. Third, the proposed approach primarily relies on 2D image slices, which may not fully capture three-dimensional spatial dependencies present in volumetric CT data. Additionally, the absence of multimodal data (e.g., clinical or biochemical parameters) restricts the model's ability to incorporate complementary diagnostic information. Finally, while the model achieves high accuracy, further statistical validation and testing on external datasets are necessary to ensure robustness in real-world deployment scenarios.

6. FUTURE WORK

Future research will focus on addressing the identified limitations and further enhancing the model's performance and generalization capability. One important direction is the incorporation of larger and more diverse multi-institutional datasets to improve robustness against domain variations and imaging inconsistencies. Advanced data augmentation techniques and class balancing strategies will be explored to reduce misclassification and improve model reliability.

In addition, future work will investigate the integration of hybrid architectures combining convolutional neural networks with transformer-based models to capture both local and global feature representations more effectively. Ensemble learning strategies, such as model stacking and boosting, will also be explored to further enhance classification accuracy and stability. Another promising direction is the extension of the current 2D framework to 3D volumetric analysis, enabling the model to exploit spatial continuity across CT slices for improved feature representation. Furthermore, incorporating multimodal data, including patient clinical records and laboratory findings, may significantly enhance predictive performance and clinical relevance. Efforts will be made to optimize computational efficiency and reduce training complexity through advanced optimization techniques such as Bayesian optimization or adaptive learning strategies, facilitating real-time deployment in clinical decision support systems.

7. CONCLUSION

In this study, an optimized deep learning framework based on a fine-tuned VGG19 architecture combined with grid search-based hyperparameter optimization was proposed for kidney disease classification using CT images. The proposed approach leverages transfer learning to extract discriminative spatial features while systematically optimizing key hyperparameters such as learning rate, batch size, and dropout rate to enhance model performance. Experimental results demonstrate that the proposed model achieves a classification accuracy of 98.78%, outperforming several existing deep learning and hybrid approaches reported in the literature.

The model exhibited strong convergence behavior, achieving a training accuracy of 99.62% and validation accuracy of 98.76% after 25 epochs, indicating stable learning and minimal overfitting. The consistency between training and validation performance highlights the model's ability to generalize effectively to unseen data. Furthermore, evaluation using precision, recall, and F1-score confirms the robustness of the proposed approach, ensuring a balanced performance that is critical for medical diagnostic applications where both false positives and false negatives must be minimized.

The contribution of grid search optimization was validated through performance improvements over manually tuned configurations, demonstrating its effectiveness in identifying optimal hyperparameter settings. Additionally, confusion matrix analysis provided insights into classification behavior, highlighting the model's ability to accurately distinguish between different kidney disease classes while identifying specific areas of misclassification.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Abdeltawab, H., Khalifa, F., Ghazal, M., Cheng, L., Gondim, D., & El-Baz, A. (2021). A pyramidal deep learning pipeline for kidney whole-slide histology images classification. Scientific Reports, 11(1), 20189. https://doi.org/10.1038/s41598-021-99735-6

Alnazer, I., Bourdon, P., Urruty, T., Falou, O., Khalil, M., Shahin, A., & Fernandez-Maloigne, C. (2021). Recent advances in medical image processing for the evaluation of chronic kidney disease. Medical Image Analysis, 69, 101960. https://doi.org/10.1016/j.media.2021.101960

Alzu'bi, D., Abdullah, M., Hmeidi, I., AlAzab, R., Gharaibeh, M., El-Heis, M., Almotairi, K. H., Forestiero, A., Hussein, A. M., & Abualigah, L. (2022). Kidney tumor detection and classification based on deep learning approaches: a new dataset in CT scans. Journal of Healthcare Engineering, 2022, 1–22. https://doi.org/10.1155/2022/3861161

Badawy, M., Almars, A. M., Balaha, H. M., Shehata, M., Qaraad, M., & Elhosseini, M. (2023). A two-stage renal disease classification based on transfer learning with hyperparameters optimization. Frontiers in Medicine, 10. https://doi.org/10.3389/fmed.2023.1106717

Blau, N., Klang, E., Kiryati, N., Amitai, M., Portnoy, O., & Mayer, A. (2018). Fully automatic detection of renal cysts in abdominal CT scans. International Journal of Computer Assisted Radiology and Surgery, 13(7), 957–966. https://doi.org/10.1007/s11548-018-1726-6

CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone. (2021, November 1). https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone

Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. Journal of Big Data, 9(1). https://doi.org/10.1186/s40537-022-00657-5

Etem, T., & Teke, M. (2024). Enhanced deep learning based decision support system for kidney tumour detection. BenchCouncil Transactions on Benchmarks Standards and Evaluations, 4(2), 100174. https://doi.org/10.1016/j.tbench.2024.100174

Etem, T., & Teke, M. (2024). Enhanced deep learning based decision support system for kidney tumour detection. BenchCouncil Transactions on Benchmarks Standards and Evaluations, 4(2), 100174. https://doi.org/10.1016/j.tbench.2024.100174

Francis, A., Harhay, M. N., Ong, A. C. M., Tummalapalli, S. L., Ortiz, A., Fogo, A. B., Fliser, D., Roy-Chaudhury, P., Fontana, M., Nangaku, M., Wanner, C., Malik, C., Hradsky, A., Adu, D., Bavanandan, S., Cusumano, A., Sola, L., Ulasi, I., Jha, V., . . . Nephrology, I. S. O. (2024). Chronic kidney disease and the global public health agenda: an international consensus. Nature Reviews Nephrology, 20(7), 473–485. https://doi.org/10.1038/s41581-024-00820-6

Gao, P., Ma, T., Li, H., Lin, Z., Dai, J., & Qiao, Y. (2022). CoNVMAE: Masked Convolution meets Masked Autoencoders. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2205.03892

Hussain, S., Songhua, X., Aslam, M., Waqas, M., & Hussain, S. (2024). Quantum Deep Learning for Automatic Chronic Kidney Disease Identification and Classification with CT images. Research Square. https://doi.org/10.21203/rs.3.rs-4743771/v1

K, K. V. Nagaraja, T. Singh and B. P. KN, "Optimization Study of Renal CT Image Classification Using Convolution Neural Networks," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-8, doi: 10.1109/ICCCNT56998.2023.10307772.

Karam, S., Amouzegar, A., Alshamsi, I. R., Ghamdi, S. M. A., Anwar, S., Ghnaimat, M., Saeed, B., Arruebo, S., Bello, A. K., Caskey, F. J., Damster, S., Donner, J., Jha, V., Johnson, D. W., Levin, A., Malik, C., Nangaku, M., Okpechi, I. G., Tonelli, M., . . . Zaidi, D. (2024). Capacity for the management of kidney failure in the International Society of Nephrology Middle East region: report from the 2023 ISN Global Kidney Health Atlas (ISN-GKHA). Kidney International Supplements, 13(1), 57–70. https://doi.org/10.1016/j.kisu.2024.01.009

Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H., & Kim, N. (2019). Deep learning in medical imaging. Neurospine, 16(4), 657–668. https://doi.org/10.14245/ns.1938396.198

Koonce, B. (2021). Convolutional Neural Networks with Swift for Tensorflow. In Apress eBooks. https://doi.org/10.1007/978-1-4842-6168-2

Kulandaivelu, G., Suchitra, M., Pugalenthi, R., & Lalit, R. (2025). An implementation of adaptive Multi-CNN feature fusion model with attention mechanism with improved heuristic algorithm for kidney stone detection. Computational Intelligence, 41(1). https://doi.org/10.1111/coin.70028

Kumar, Y., Brar, T. P. S., Kaur, C., & Singh, C. (2024). A comprehensive study of deep learning methods for kidney tumor, cyst, and stone diagnostics and detection using CT images. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-024-10112-8

Liashchynskyi, P., & Liashchynskyi, P. (2019, December 12). Grid search, random search, genetic algorithm: A big comparison for NAS. arXiv.org. https://arxiv.org/abs/1912.06059

Lin, Z., Yang, W., Zhang, W., Jiang, C., Chu, J., Yang, J., & Yuan, X. (2023). Recognizing pathology of renal tumor from macroscopic cross-section image by deep learning. BioMedical Engineering OnLine, 22(1), 3. https://doi.org/10.1186/s12938-023-01064-4

M. H. K. Mehedi, E. Haque, S. Y. Radin, M. A. Ur Rahman, M. T. Reza and M. G. R. Alam, "Kidney Tumor Segmentation and Classification using Deep Neural Network on CT Images," 2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, Australia, 2022, pp. 1-7, doi: 10.1109/DICTA56598.2022.10034638.

M. S. Hossain, S. M. Nazmul Hassan, M. Al-Amin, M. N. Rahaman, R. Hossain and M. I. Hossain, "Kidney Disease Detection from CT Images using a customized CNN model and Deep Learning," 2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS), Kochi, India, 2023, pp. 1-6, doi: 10.1109/AICAPS57044.2023.10074314.

Majid, M., Gulzar, Y., Ayoub, S., Khan, F., Reegu, F. A., Mir, M. S., Jaziri, W., & Soomro, A. B. (2023). Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging. International Journal of Advanced Computer Science and Applications, 14(8). https://doi.org/10.14569/ijacsa.2023.0140847

N. Narmada, V. Shekhar and T. Singh, "Classification of Kidney Ailments using CNN in CT Images," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-5, doi: 10.1109/ICCCNT54827.2022.9984256.

Özbay, E., Özbay, F. A., & Gharehchopogh, F. S. (2024). Kidney Tumor Classification on CT images using Self-supervised Learning. Computers in Biology and Medicine, 176, 108554. https://doi.org/10.1016/j.compbiomed.2024.108554

Özbay, E., Özbay, F. A., & Gharehchopogh, F. S. (2024). Kidney Tumor Classification on CT images using Self-supervised Learning. Computers in Biology and Medicine, 176, 108554. https://doi.org/10.1016/j.compbiomed.2024.108554

Parvaiz, A., Khalid, M. A., Zafar, R., Ameer, H., Ali, M., & Fraz, M. M. (2023). Vision Transformers in medical computer vision—A contemplative retrospection. Engineering Applications of Artificial Intelligence, 122, 106126. https://doi.org/10.1016/j.engappai.2023.106126

Prakash, N. N., Rajesh, V., Namakhwa, D. L., Pande, S. D., & Ahammad, S. H. (2023). A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis. Scientific African, 20, e01629. https://doi.org/10.1016/j.sciaf.2023.e01629

Razzak, M. I., Naz, S., & Zaib, A. (2017). Deep learning for medical Image Processing: Overview, challenges and the future. In Lecture notes in computational vision and biomechanics (pp. 323–350). https://doi.org/10.1007/978-3-319-65981-7_12

S. D. Pande and R. Agarwal, "Multi-Class Kidney Abnormalities Detecting Novel System Through Computed Tomography," in IEEE Access, vol. 12, pp. 21147-21155, 2024, doi: 10.1109/ACCESS.2024.3351181.

S. Pavarut et al., "Improving Kidney Tumor Classification With Multi-Modal Medical Images Recovered Partially by Conditional CycleGAN," in IEEE Access, vol. 11, pp. 146250-146261, 2023, doi: 10.1109/ACCESS.2023.3345648.

Shanmathi, K., Varshini, S. S., Varsha, M., Indrakumar, S., & PraveenKumar, G. D. (2025). Deep neural network models for comprehensive kidney stone prediction. In Information systems engineering and management (pp. 304–319). https://doi.org/10.1007/978-3-031-90478-3_26

Sharon, J. J., & Anbarasi, L. J. (2025). An attention enhanced dilated bottleneck network for kidney disease classification. Scientific Reports, 15(1), 9865. https://doi.org/10.1038/s41598-025-90519-w

Simonyan, K., & Zisserman, A. (2014, September 4). Very deep convolutional networks for Large-Scale image recognition. arXiv.org. https://arxiv.org/abs/1409.1556

Srivastava, S. S. . R. K. Y., . V. N. ,. P. K. (2022). An Ensemble Learning Approach for Chronic Kidney Disease Classification. www.pnrjournal.com. https://doi.org/10.47750/pnr.2022.13.S10.279

Tadesse, Y. (2026). AI-POWERED SECURITY STRATEGIES FOR THE OSI MODEL. ShodhAI: Journal of Artificial Intelligence, 3(1), 9–19. https://doi.org/10.29121/shodhai.v3.i1.2026.62

Türk, F., Lüy, M., & Barışçı, N. (2020). Kidney and renal tumor segmentation using a hybrid V-Net-Based model. Mathematics, 8(10), 1772. https://doi.org/10.3390/math8101772

Vetrithangam, D., Himabindu, Saranya, Neha, S., Naresh Kumar, P., Fathima, A., Ashok, B., & Akanksha, K. (2024). Improved RESNET models for chronic kidney disease prediction. Journal of Electrical Systems, 20(2s), 165–183. https://doi.org/10.52783/jes.1121

Yildirim, K., Bozdag, P. G., Talo, M., Yildirim, O., Karabatak, M., & Acharya, U. (2021). Deep learning model for automated kidney stone detection using coronal CT images. Computers in Biology and Medicine, 135, 104569. https://doi.org/10.1016/j.compbiomed.2021.104569

Zabihollahy, F., Schieda, N., Krishna, S., & Ukwatta, E. (2020). Automated classification of solid renal masses on contrast-enhanced computed tomography images using convolutional neural network with decision fusion. European Radiology, 30(9), 5183–5190. https://doi.org/10.1007/s00330-020-06787-9

Zain, R. H., Sumijan, & Defit, S. (2025). Development of an ultrasound image extraction method for detection and classification of kidney abnormalities using a convolutional neural network. International Journal of Online and Biomedical Engineering (iJOE), 21(04), 150–170. https://doi.org/10.3991/ijoe.v21i04.53289

Zhao, T., Sun, Z., Guo, Y., Sun, Y., Zhang, Y., & Wang, X. (2023). Automatic renal mass segmentation and classification on CT images based on 3D U-Net and ResNet algorithms. Frontiers in Oncology, 13, 1169922. https://doi.org/10.3389/fonc.2023.1169922

Zheng, Q., Furth, S., Tasian, G., & Fan, Y. (2018). Computer-aided diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data by integrating texture image features and deep transfer learning image features. Journal of Pediatric Urology, 15(1), 75.e1-75.e7. https://doi.org/10.1016/j.jpurol.2018.10.020