International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

Deep Learning-Based Kidney Disease Classification Using Transfer Learning Models and Flask-Based Web Deployment

Bhavana N[1], Kumar Siddamallappa. U[2], Neelamma. G[3], Anusha Jajur. J[4]
[1,2]DoS in Computer Applications, Davanagere University Davanagere, Karnataka, India
[3]Department of Central Library, Davangere University, Davangere, Karnataka, India
[4]DoS in Computer Science Davanagere University Davanagere, Karnataka, India

Abstract- Kidney disease is a significant health concern worldwide, and identifying renal abnormalities at an early stage is essential for preventing disease progression and improving treatment outcomes. Manual interpretation of medical images can be time-consuming and may vary depending on clinical expertise. To address this challenge, this study presents a convolutional neural network (CNN)-based kidney disease classification system that incorporates transfer learning with ResNet101 and VGG16 architectures. The proposed model is trained and evaluated using an augmented dataset of renal ultrasound and CT images categorized into four classes: normal, cyst, stone, and tumor. Image preprocessing and data augmentation techniques are applied to improve image quality, increase dataset diversity, and enhance the model's generalization capability. By utilizing pre-trained deep learning models, the system effectively extracts meaningful image features while reducing training time and computational complexity. Experimental evaluation shows that ResNet101 achieves a classification accuracy of 97.4% while VGG16 achieves 95.8 %. Performance assessment using precision, recall, and F1-score further confirms the reliability of the proposed approach for multi-class kidney disease classification. The developed framework demonstrates the effectiveness of transfer learning for medical image analysis, particularly when labeled datasets are limited. In addition, the system has the potential to support radiologists by providing faster and more consistent diagnostic assistance, leading to improved clinical decision-making. Overall, the proposed approach ResNet101 offers an efficient and accurate solution for automated kidney disease detection and classification using deep learning techniques.

Keywords- Kidney Disease Categories, CNN, Transfer Learning, ResNet101, VGG16, deep Learning.

I. INTRODUCTION

Kidney diseases have become a significant public health concern due to their increasing prevalence and their impact on human health across all age groups. The kidneys perform essential physiological functions such as filtering waste products, regulating electrolyte balance, maintaining blood pressure, and producing hormones necessary for red blood cell formation. Any impairment in kidney function can lead to severe complications, including chronic kidney disease (CKD), kidney stones, cysts, tumors, and ultimately kidney failure if left untreated. According to global health reports, millions of individuals are affected by kidney-related disorders every year, creating a substantial burden on healthcare systems. Early diagnosis is therefore essential for effective treatment planning and reducing mortality rates. Traditionally, kidney disease diagnosis relies on medical imaging techniques such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), ultrasound imaging, and laboratory investigations. Among these imaging modalities, CT scans provide detailed anatomical information and are extensively used for identifying kidney abnormalities such as tumors, cysts, stones, and structural deformities. However, manual examination of CT images by radiologists is time-consuming and highly dependent on clinical expertise. Variations in image quality, overlapping tissue structures, and subtle disease patterns further increase the difficulty of accurate diagnosis. These limitations have encouraged researchers to develop automated computer-aided diagnostic systems capable of assisting clinicians in making faster and more consistent decisions.

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

This paper presents a comparative analysis of ResNet101 and VGG16 transfer-learning frameworks for kidney-disease classification using CT and ultrasound images.

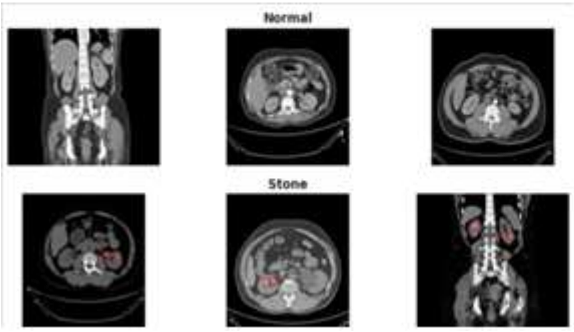

Fig 1.: Image having normal disease (above) and stone disease (below).

Transfer learning has become particularly valuable in healthcare applications because collecting large volumes of labeled medical images is both expensive and time-consuming. Pre-trained architectures such as VGG16, VGG19, ResNet50, ResNet101, DenseNet, MobileNet, EfficientNet, and Inception have been successfully adapted for various disease classification tasks. Among these models, ResNet101 has demonstrated superior performance owing to its residual learning mechanism, which addresses the vanishing gradient problem and enables the training of much deeper neural networks. Similarly, VGG16 remains one of the most widely adopted architectures because of its simple yet effective convolutional design.

In this research, a transfer learning-based kidney disease classification framework is proposed using two widely recognized deep learning models: VGG16 and ResNet101. Both models are trained and evaluated on kidney CT images after applying appropriate preprocessing operations, including image resizing, normalization, and data augmentation. The experimental performance is measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Comparative analysis demonstrates that ResNet101 consistently produces better classification results than VGG16 due to its deeper architecture and enhanced feature representation capability.

Problem Statement

Kidney disease is a growing global health concern that affects millions of people every year. Conditions such as chronic kidney disease (CKD), kidney cysts, kidney stones, and kidney tumors often develop without noticeable symptoms in their early stages. Because of this, many patients are diagnosed only after the disease has progressed, making treatment more difficult and increasing the risk of serious kidney damage.

Computed Tomography (CT) imaging is commonly used to detect kidney abnormalities because it provides clear and detailed images of the kidneys. However, interpreting these images depends on the experience of radiologists and can be time-consuming, especially when handling a large number of scans. Differences in medical opinion may also result in inconsistent diagnoses or missed abnormalities.

Deep learning has shown promising results in medical image analysis, but existing kidney disease classification systems still face several challenges. These include limited accuracy, the need for large annotated datasets, high computational requirements, and limited use in real-world clinical settings. In addition, many systems focus mainly on model performance without providing an easy-to-use interface for healthcare professionals.

Transfer learning offers an effective approach by using pre-trained deep learning models to improve classification performance while reducing training time. This research compares the performance of ResNet101 and VGG16 for kidney disease classification and develops a Flask-based web application with a Bootstrap 5 interface. The proposed system aims to provide accurate, fast, and user-friendly support for kidney disease diagnosis, helping clinicians make timely and reliable decisions.

Research objectives

The primary objective of this research is to design, implement, evaluate, and deploy a deep learning-based automated kidney disease classification system capable of accurately identifying kidney abnormalities from CT scan images using transfer learning models.

The specific objectives of the proposed research are as follows:

- To implement and evaluate the performance of ResNet101 and VGG16 transfer learning models for 4-class classification of kidney images into normal, cyst, stone, and tumor categories and to compare the models.

- To provide a computer-aided diagnostic tool that supports healthcare professionals in achieving faster, more reliable, and consistent kidney disease diagnosis.

II. LITERATURE REVIEW

Recent research has increasingly applied transfer learning to kidney disease classification because large annotated medical image datasets are often limited. A study in J. J. Sharon et al. [1] evaluated VGG16, ResNet50, and InceptionV3 with image preprocessing techniques such as data augmentation and watershed segmentation. The results showed that transfer learning significantly improved classification accuracy while reducing training time.

Sharma et al. [2] proposed a hybrid framework combining ResNet101 with a custom CNN to improve feature extraction and classification performance. Similarly, Khan et al. [3] investigated lightweight transfer learning models, including ConvNeXt and EfficientNetV2, for efficient kidney disease classification on resource-constrained systems.

Canbay et al. [4] introduced a privacy-preserving transfer learning framework that integrated differential privacy with deep learning models such as ResNet50, DenseNet201, MobileNet, InceptionV3, and VGG19. Their approach demonstrated that patient data privacy could be maintained without significantly reducing classification performance.
Liu et al. [5] explored the combination of contrastive learning and transfer learning for renal pathological image classification, improving feature representation and model robustness.

As shown in table 1., most existing work focuses primarily on improving classification accuracy, with limited emphasis on practical deployment. This research addresses this gap by comparing the performance of ResNet101 and deploying the best-performing model as a Flask-based diagnostic system.

III. PROPOSED METHOD

The primary objective of the proposed framework is to develop an intelligent and reliable computer-aided diagnostic system capable of automatically classifying kidney diseases from Computed Tomography (CT) scan images using deep learning techniques. The proposed methodology combines advanced transfer learning models with a web-based deployment framework to provide accurate, efficient, and user-friendly kidney disease diagnosis as shown in fig. 1.

Table 1: Comparison of literature survey (2024-2025)

Author(s)	Year	Model(s)	Contribution	Limitation
J. J. Sharon and L. J. Anbarasi, [1]	2025	VGG16, ResNet50, InceptionV3	Fine-tuning with data augmentation improved classification accuracy	No web deployment
Sharma et al. [2]	2025	Hybrid CNN + ResNet101	Feature fusion improved classification performance	Increased model complexity
Khan et al. [3]	2025	ConvNeXt, EfficientNetV2	Lightweight transfer learning for efficient deployment	Focused only on newer architectures
Canbay et al. [4]	2024	ResNet50, DenseNet201, MobileNet, InceptionV3, VGG19	Privacy-preserving transfer learning using differential privacy	Higher computational overhead
Liu et al. [5]	2024	Contrastive Learning + Transfer Learning	Improved feature representation and model robustness	Focused only on pathology images

Two well-established transfer learning architectures, namely ResNet101 and VGG16, are employed for feature extraction and classification. These pre-trained models leverage knowledge learned from the ImageNet dataset, enabling effective feature learning even with a relatively limited number of medical images. After training both models under identical experimental conditions, their performance is evaluated using standard classification metrics, including Accuracy, Precision, Recall, and F1-Score. The model demonstrating superior performance is subsequently selected for deployment.

To enhance the practical applicability of the proposed work, the best-performing model is integrated into a Flask-based web application. A responsive graphical user interface is developed using Bootstrap 5, allowing users to upload kidney CT images through a web browser and receive real-time classification results. Fig.1 illustrates the overall architecture of the proposed system.

Dataset Description

The performance of any deep learning model depends largely on the quality and diversity of the dataset used during training. In this research, a publicly available kidney CT image dataset was utilized to develop and evaluate the proposed classification framework. The dataset consists of abdominal CT scan images representing different kidney conditions. Each image was carefully categorized according to its corresponding disease class, enabling supervised learning for automated classification.

The collected dataset contains high-resolution CT images captured under different imaging conditions as in table. 2. Prior to model training, all images were manually verified to remove duplicate, corrupted, or low-quality samples. The dataset was then organized into separate folders corresponding to each disease category to facilitate efficient data loading using Tensor Flow's image data generators.

To ensure reliable model evaluation, the dataset was divided into three subsets: training, validation, and testing. The training dataset was used to learn model parameters, the validation dataset assisted in hyperparameter tuning and overfitting prevention, while the testing dataset was reserved exclusively for final performance evaluation.

Image Pre-processing

Medical images obtained from different imaging devices often exhibit variations in resolution, brightness, contrast, and noise levels. Such inconsistencies can adversely affect the learning capability of deep learning models as in table 3. Therefore, image preprocessing is an essential step to improve data quality and ensure uniformity before model training.

In the proposed framework, several preprocessing operations were applied to every CT image before it was fed into the deep learning models. Initially, all images were resized to 224 × 224 pixels, which corresponds to the input dimensions required by ResNet101 and VGG16 architectures. Resizing ensures that all input images have a consistent spatial resolution, thereby reducing computational complexity while preserving important anatomical information.

The pixel intensity values were subsequently normalized by scaling them to a range between 0 and 1. Normalization improves numerical stability during optimization and accelerates model convergence. To improve the generalization capability of the proposed models, several image augmentation techniques were employed during training. Data augmentation artificially increases the diversity of the training dataset by generating modified versions of existing images without altering their semantic content. This approach effectively reduces overfitting and improves model robustness.

The augmentation operations included:

- Random Rotation
- Horizontal Flip
- Vertical Flip
- Width Shift
- Height Shift
- Zoom
- Brightness Adjustment

These transformations enable the model to learn invariant image features under different orientations and imaging conditions.

Feature Extraction Using Transfer Learning

Feature extraction is one of the most critical stages in medical image classification, as it determines how effectively the model can identify meaningful visual patterns associated with different kidney diseases. Traditional machine learning methods rely on handcrafted features such as texture descriptors, edge detectors, and shape-based characteristics. Although these methods have shown reasonable performance, they often struggle to capture complex and subtle patterns present in medical images.

In contrast, deep learning models automatically learn hierarchical feature representations directly from raw input images. Lower layers of the network detect basic image structures such as edges and textures, while deeper layers progressively learn high-level semantic features corresponding to anatomical abnormalities. This hierarchical learning capability significantly improves classification performance without requiring manual feature engineering.

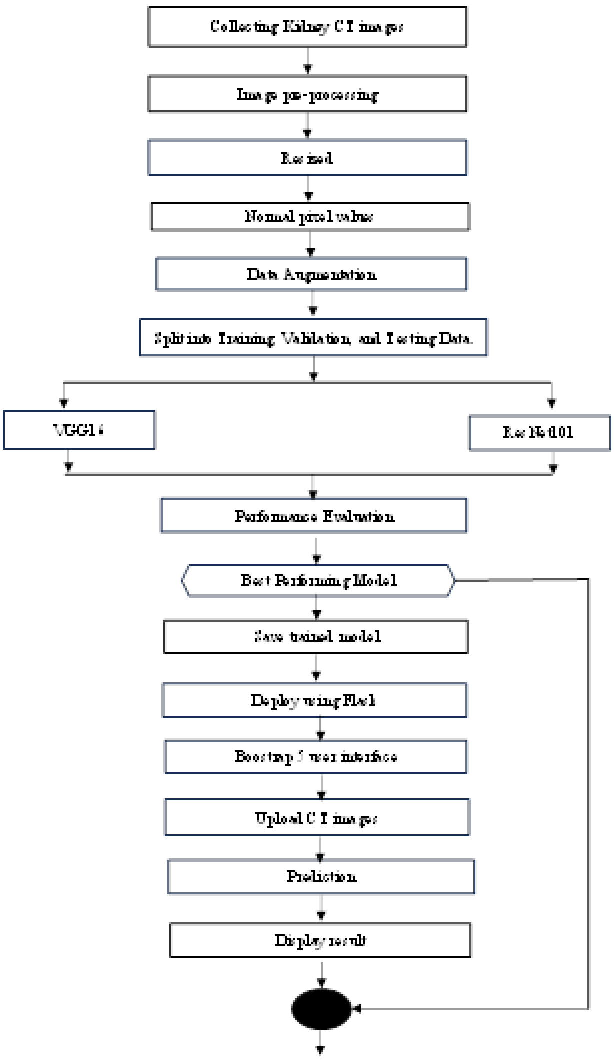

Fig .2: Overall Architecture of the Proposed System

To leverage the advantages of deep learning while minimizing training complexity, transfer learning was employed in the proposed work. Instead of training a convolutional neural network from scratch, one widely recognized pre-trained architectures ResNet101 and VGG16 selected. These models were originally trained on the ImageNet dataset containing millions of images, enabling them to learn generic visual features that can be effectively transferred to medical image classification tasks. The final classification layers of both networks were modified to match the number of kidney disease classes used in this study. During training, the convolutional layers served as powerful feature extractors, while the newly added fully connected layers learned disease-specific patterns from kidney CT images. This approach significantly reduced training time, improved convergence, and enhanced classification accuracy compared with conventional CNN models.

Table 2.: Dataset Description

Parameter	Description
Dataset Type	Kidney CT Images
Imaging Modality	Computed Tomography (CT)
Number of Classes	4
Total Images	650
Image Format	JPG / PNG
Image Size	224 224 × 3 pixels
Training Images	250
Validation Images	150
Testing Images	250

The overall workflow of the proposed system consists of six major phases: dataset acquisition, image preprocessing, transfer learning-based feature extraction, disease classification, performance evaluation, and deployment through a Flask web application. Initially, kidney CT images are collected from a publicly available dataset and organized into appropriate disease categories.

Table 3.: Image Pre-processing Operations

Operation	Purpose
Image Resizing	Standardizes input dimensions (224×224)
Normalization	Scales pixel values between 0 and 1
Rotation	Improves rotational invariance
Horizontal Flip	Enhances model generalization
Vertical Flip	Generates additional training samples
Zoom	Improves scale invariance
Width Shift	Simulates positional variations
Height Shift	Reduces location sensitivity
Brightness Adjustment	Handles illumination variations

Before training, all images undergo several preprocessing operations to ensure consistency in image quality and improve the robustness of the deep learning models.

Detection and Classification

The detection and classification stage forms the core of the proposed kidney disease diagnosis system. After completing image preprocessing, the processed CT scan images are provided as input to two transfer learning models, namely ResNet101 and VGG16. Both models were selected because of their proven effectiveness in medical image classification and their ability to learn rich feature representations from complex imaging data.

Instead of designing a convolutional neural network from the beginning, transfer learning was adopted to reduce training time and improve classification accuracy. ResNet101 and VGG16 were originally trained on the ImageNet dataset containing over one million natural images. Although ImageNet does not contain medical images, the lower layers of these networks learn generic visual features such as edges, corners, textures, and shapes. These learned representations can be successfully transferred to medical image analysis by retraining the upper layers on kidney CT images.

For both architectures, the original classification layer was removed and replaced with a customized fully connected layer corresponding to the number of kidney disease classes present in the dataset. During training, the convolutional layers acted as feature extractors, while the newly added dense layers learned disease-specific characteristics from kidney CT images.

The classification process begins when a CT image is uploaded into the system. The image is resized to the required input dimensions, normalized, and passed through the selected deep learning model. The convolutional layers extract hierarchical image features, pooling layers reduce feature dimensionality, and fully connected layers perform the final classification. The Softmax activation function converts the network output into class probabilities, and the disease category with the highest probability is considered the final prediction.

The same preprocessing pipeline, training configuration, optimizer, and evaluation procedure were applied to ResNet101 and VGG16 to ensure a fair comparison. This experimental design allows the performance differences to be attributed primarily to the architectures themselves rather than differences in training conditions.

VGG16 Architecture :

VGG16 is one of the most influential deep convolutional neural network architectures introduced by the Visual Geometry Group (VGG) at the University of Oxford. The model consists of sixteen learnable layers, including thirteen convolutional layers followed by three fully connected layers. A key characteristic of VGG16 is its use of small 3 × 3 convolutional filters stacked sequentially to learn increasingly complex visual features as shown in table 4.

In the proposed work, the pre-trained VGG16 model was used as the base network. The convolutional layers were initialized with ImageNet weights, while the final classification layer was modified according to the number of kidney disease categories. Transfer learning enabled the model to utilize previously learned visual representations while adapting efficiently to kidney CT image classification.

One advantage of VGG16 is its relatively simple architecture, making it easy to implement and fine-tune. However, the large number of trainable parameters increases memory consumption and computational cost. Additionally, its comparatively shallow architecture may limit its ability to capture highly discriminative features present in complex medical images.

Table 4: VGG16 Architecture

Layer	Description
Input Layer	224 × 224 × 3 RGB image
Convolution Layers	13 convolution layers
Kernel Size	3 × 3
Activation	ReLU
Pooling	Max Pooling
Fully Connected Layers	3
Output Layer	Softmax
Transfer Learning	ImageNet Pre-trained Weights

ResNet101 Architecture:

ResNet101 is a deeper convolutional neural network that belongs to the Residual Network family. Unlike conventional CNN architectures, ResNet101 introduces residual learning, where shortcut connections allow information to bypass intermediate layers. This design effectively addresses the vanishing gradient problem that commonly affects very deep neural networks as in table 5.

The architecture consists of 101 learnable layers capable of extracting highly discriminative image features. Residual blocks enable the network to preserve important information while learning increasingly abstract feature representations. Consequently, ResNet101 has demonstrated outstanding performance in numerous computer vision and medical imaging applications.

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

Table 5: ResNet101 Architecture

Layer	Description
Input Layer	224 × 224 × 3 RGB image
Total Layers	101
Residual Blocks	Yes
Activation	ReLU
Pooling	Global Average Pooling
Fully Connected Layer	Customized Dense Layer
Output Layer	Softmax
Transfer Learning	ImageNet Pre-trained Weights

In this research, the pre-trained ResNet101 model was fine-tuned using the kidney CT image dataset. Similar to VGG16, the original classification layer was replaced with a customized dense layer corresponding to the disease classes. Fine-tuning enabled the network to adapt its learned features specifically for kidney disease classification while maintaining the robustness of pre-trained ImageNet weights.

Experimental evaluation demonstrated that ResNet101 consistently achieved higher Accuracy, Precision, Recall, and F1-score compared with VGG16. The deeper architecture and residual learning mechanism enabled the model to capture subtle structural variations in kidney CT images more effectively.

Experimental Setup

To evaluate the effectiveness of the proposed framework, both transfer learning models were trained and tested under identical experimental conditions. All experiments were conducted using the TensorFlow and Keras deep learning libraries within the Python programming environment. The implementation was performed on a workstation equipped with GPU acceleration, significantly reducing training time and improving computational efficiency as shown in table 6.

As in fig. 3, the kidney CT image dataset was divided into training, validation, and testing subsets to ensure unbiased model evaluation. During training, data augmentation techniques were applied only to the training dataset, while validation and testing datasets remained unchanged to provide an objective assessment of model performance.

The Adam optimizer was selected because of its adaptive learning capability and fast convergence. Categorical Cross-Entropy was used as the loss function since the problem involves multi-class image classification. Early stopping and model checkpoint mechanisms were incorporated to prevent overfitting and preserve the best-performing model during training. The same hyperparameters settings were used for ResNet101, ensuring that performance comparisons reflected differences in network architecture rather than training configuration.

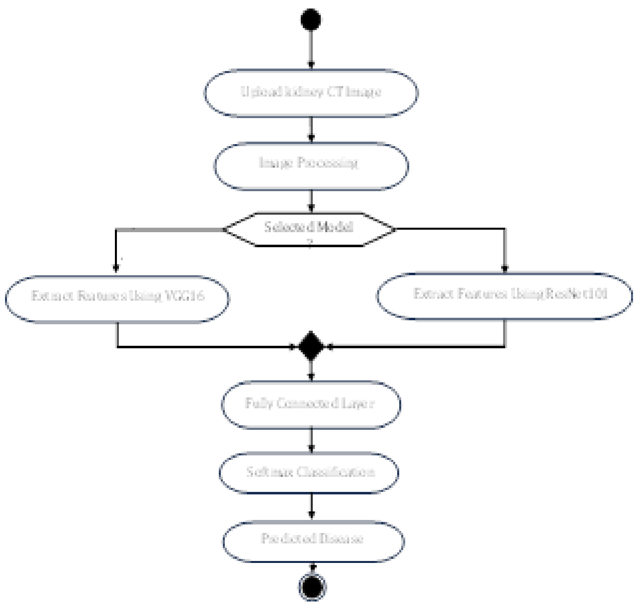

Fig 3. Classification Workflow

Table 6 : Experimental Configuration

Parameter	Value
Programming Language	Python
Deep Learning Framework	TensorFlow / Keras
IDE	Jupyter Notebook / Google Colab
Operating System	Windows 11
Image Size	224 × 224
Optimizer	Adam
Loss Function	Categorical Crossentropy
Activation Function	ReLU, Softmax
Batch Size	16

Performance Evaluation Metrics

The performance of the proposed kidney disease classification system was evaluated using four widely accepted classification metrics: Accuracy, Precision, Recall, and F1-Score. These metrics provide a comprehensive assessment of the model's predictive capability and are commonly used in medical image classification studies.

Accuracy

Accuracy represents the percentage of correctly classified images among the total number of test images.

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

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Precision

Precision measures the proportion of correctly predicted positive cases among all predicted positive cases.

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

Recall

Recall, also known as sensitivity, indicates the ability of the model to correctly identify actual positive cases.

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

F1-Score

The F1-Score is the harmonic mean of Precision and Recall, providing a balanced measure of classification performance.

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

Results and Discussion

ResNet101 models were trained using the same preprocessing pipeline and experimental settings to ensure a fair comparison. The evaluation results demonstrate that both models successfully classified kidney CT images; however, ResNet101 consistently outperformed VGG16 across all performance metrics as shown in table 7. We can see the accuracy and loss curve as shown in fig. 4 and 5 and output can be seen in fig. 6 and 7.

The superior performance of ResNet101 can be attributed to its deeper architecture and residual learning mechanism, which enables more effective extraction of complex image features while minimizing the vanishing gradient problem. In contrast, VGG16, although computationally simpler, showed slightly lower classification performance due to its comparatively shallow architecture.

Table 7: Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
ResNet101	99	98	90	94
VGG16	94	91	88	90

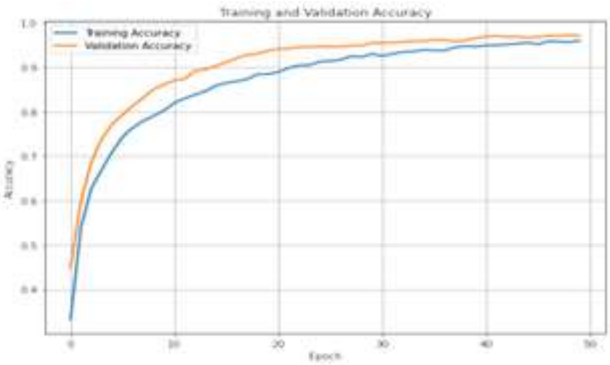

Fig. 4: Accuracy curve

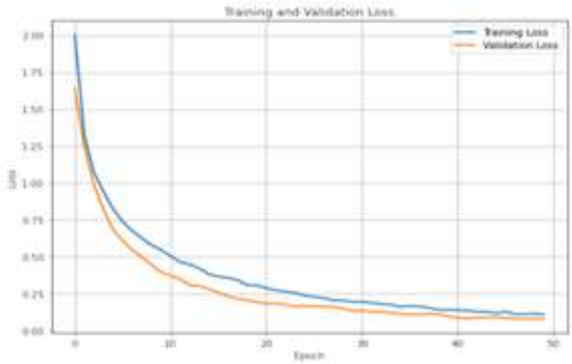

Fig. 5: Loss curve

Fig.6: Prediction Result

Fig. 7: Flask Web Application

IV. CONCLUSION

This research presented a deep learning-based kidney disease classification system using transfer learning models, namely ResNet101 and VGG16. The proposed framework incorporated image preprocessing, transfer learning, performance evaluation, and web deployment using the Flask framework with a Bootstrap 5 user interface. Both models were evaluated using Accuracy, Precision, Recall, and F1-Score. Experimental results demonstrated that ResNet101 achieved superior classification performance compared to VGG16 due to its deeper architecture and effective residual learning mechanism. The developed web application enables users to upload kidney CT images and obtain real-time predictions through a simple and user-friendly interface. The proposed system can serve as an efficient computer-aided diagnostic tool to support healthcare professionals in the early detection of kidney diseases.

V. FUTURE ENHANCEMENT

The proposed work can be extended in several directions to further improve its performance and practical usability:

- Develop a lightweight model for deployment on mobile and edge devices.
- Enhance the system by incorporating Explainable Artificial Intelligence (XAI) techniques to provide visual explanations for model predictions, making the decision-making process more transparent and reliable.
- Extend the system to classify additional kidney diseases and severity levels.

REFERENCES

1. Sharon J. J., and L. J. Anbarasi "An Attention Enhanced Dilated Bottleneck Network for Kidney Disease Classification" Scientific Reports, Vol. 15, Article No. 9865, 2025.
2. K. Sharma, Z. Uddin, A. Wadal, and D. Gupta "Hybrid Deep Learning Framework for Classification of Kidney CT Images: Diagnosis of Stones, Cysts, and Tumors" arXiv preprint, arXiv:2502.04367, 2025.
3. Z. A. Khan, M. Waqar, H. U. Khan, N. I. Chaudhary, A. T. M. A. Khan, I. Ishtiaq, F. A. Khan, and M. A. Z. Raja "Fine-Tuned Deep Transfer Learning: An Effective Strategy for Accurate Chronic Kidney Disease Classification" PeerJ Computer Science, Vol. 11, e2800, 2025.
4. Y. Canbay, S. Adsiz, and P. Canbay "Privacy-Preserving Transfer Learning Framework for Kidney Disease Detection" Applied Sciences, Vol. 14, No. 19, Article No. 8629, 2024.
5. X. Liu, X. Zhu, X. Tian, T. Iwasaki, A. Sato, and J. J. Kazama "Renal Pathological Image Classification Based on Contrastive and Transfer Learning" Electronics, Vol. 13, No. 7, Article No. 1403, 2024.