Hybrid Deep Learning Framework for Classification of Kidney CT Images: Diagnosis of Stones, Cysts, and Tumors

Kiran Sharma[a,b], Ziya Uddin[a,b], Adarsh Wadal[a] and Dhruv Gupta[a]

School of Engineering & Technology, BML Munjal University, Gurugram, Haryana-122413, India

Center for Advanced Data and Computational Science, BML Munjal University, Gurugram, Haryana-122413, India

ABSTRACT

Medical image classification is a vital research area that utilizes advanced computational techniques to improve disease diagnosis and treatment planning. Deep learning models, especially Convolutional Neural Networks (CNNs), have transformed this field by providing automated and precise analysis of complex medical images. This study introduces a hybrid deep learning model that integrates a pre-trained ResNet101 with a custom CNN to classify kidney CT images into four categories: normal, stone, cyst, and tumor. The proposed model leverages feature fusion to enhance classification accuracy, achieving 99.73% training accuracy and 100% testing accuracy. Using a dataset of 12,446 CT images and advanced feature mapping techniques, the hybrid CNN model outperforms standalone ResNet101. This architecture delivers a robust and efficient solution for automated kidney disease diagnosis, providing improved precision, recall, and reduced testing time, making it highly suitable for clinical applications.

1. Introduction

CT (computed tomography) imaging is an integral part of the diagnostic workup for renal diseases. The necessity of classifying kidney conditions as normal, stone, cyst, or tumor from CT images is to enable relevant medical intervention. This process is called multi-modal classification, where using CT imaging techniques allows for the effective differentiation of these conditions. Multi-modal classification is of primary importance due to its potential to enhance diagnostic accuracy and improve patient outcomes.

Correct distinction between normal kidney, stone, cyst, and tumor is crucial for treatment planning, as each disorder requires a tailored procedure. For instance, a simple cyst may need little or no treatment, while a malignant tumor often requires aggressive measures like surgery or chemotherapy. Early and accurate classification can have a profound effect on the treatment regimen and prognosis of renal diseases. Medical imaging and analysis is a highly progressive field, particularly because of advancements in machine learning.

Patro et al. (2023) proposed a novel variant of convolutional techniques, known as the Kronecker convolution method, for better feature extraction from CT images using their models for classification through improving classifiers. This decreases duplication of feature maps, thus bettering CNN performance for detecting kidney stones by capturing more salient features. Additionally, Wu and Yi (2020) have exhibited the ability to combine CNNs through feature fusion of more than one neural network, markedly enhancing the identification and categorization of renal pathologies. The pioneering study developed a multi-features CNN fusion model for detection of multiple kidney anomalies from ultrasound images.

In addition, deep learning algorithms have been developed to detect subtle differences in kidney tissues, allowing the identification of small kidney stones or early-stage tumors (Yildirim et al., 2021). For example, the ExDark19 model proposed by Baygin et al. (2022) applies transfer learning and feature selection techniques, achieving the model accuracy of 99% in kidney stone detection. Similarly, the model proposed by Çağlayan et al. (2022) achieved an accuracy rate of 99.1% to detect kidney stones. Hossain et al. (2023a) evaluated EANet with accuracy 83.65%, ResNet50 with accuracy 87.92%, and a CNN model with accuracy 98.66%. Furthermore, Gulhane et al. (2024) introduced an integrative approach that utilizes an improved deep neural network (DNN) architecture for the efficient detection of kidney stones. By combining traditional machine learning techniques with DNNs, the approach improves pattern

---

*Corresponding author1

**Corresponding author2

✉ kiran.sharma@bmu.edu.in (K. Sharma); ziya.uddin@bmu.edu.in (Z. Uddin)

ORCID(s):

recognition in datasets. Elhoseny et al. (2019) proposed a diagnostic prediction and classification system for chronic kidney disease (CKD) that integrates density-based feature selection (DFS) with ant colony optimization (ACO). By eliminating irrelevant data before classification, the system improves the accuracy and efficiency of CKD detection and management.

Despite significant advances, challenges remain in the multi-modal classification of kidney conditions. Variability in image quality and the complexity of kidney pathologies often result in inconsistent classification outcomes. Furthermore, there is a pressing need for larger and more diverse datasets to train models, ensuring robustness and generalizability across populations and clinical settings.

2. Related work

Recent advances in deep learning methods have greatly improved image classification in numerous domains. These advanced methods have also improved the classification of medical images in disease diagnosis such as kidney, retinal, cancer, etc. CNN models demonstrated high accuracy, precision, and recall rates, making them promising tools for the early detection and classification of such diseases. There is sufficient literature on multi-modal analysis of kidney abnormalities.

Rajora et al. (2024) explored CNNs for multi-modal classification of kidney diseases. They proposed a combined ensemble model, VGG16 and InceptionV3, achieving a model accuracy of 97.8%. Similarly, Ali et al. (2023) developed a custom CNN model that demonstrated high validation accuracy and precision in kidney disease classification, highlighting its robustness and efficiency in managing complex classifications. They introduced a hybrid CNN architecture with feature fusion and ensemble classification, achieving an accuracy of 94.5%.

The attention-based CNN model with dual attention mechanisms proposed by Chauhan et al. (2024) significantly enhanced feature extraction from CT images, thereby improving diagnostic performance. Along similar lines, Sri and Lakshmi (2023) proposed a deep ensemble model for kidney disease detection, achieving high precision and low loss, showcasing the effectiveness of combining multiple deep learning models for improved diagnostic results. Furthermore, Bindu Madavi and Sowjanya (2023) introduced the "DeepKidney" model, which utilizes CNNs for classifying kidney diseases, including stones, cysts, and tumors, demonstrating notable advancements in diagnostic accuracy and efficiency.

Hossain et al. (2023b) introduced a CNN model which was a combination of customized CNN and ResNet50 and achieved an accuracy of 98.66%. This study underscores the effectiveness of CNN-specific methodologies combined with transfer learning to enhance classification performance. Similarly, Singh et al. (2023) used the EfficientNet-B3 model, fine-tuned on a large CT kidney image dataset, achieving a classification accuracy of 95.77%. Their approach highlights the efficiency of the model in processing medical images and delivering reliable diagnostic results. Sharma et al. (2024) designed a robust CNN-based framework that achieved an impressive precision of 99. 88% in the classification of multiclass kidney disease. This framework established a new standard for precision and reliability in detecting various kidney conditions.

Chowdhury et al. Chowdhury et al. (2023) utilized a mutual learning algorithm comprising multiple CNN and SVM models to enhance diagnostic accuracy while addressing data limitations and privacy concerns. This method improved learning efficiency through dynamic teacher-student interactions. Similarly, the YOLOv8 model was applied by Pande and Agarwal (2024) for the detection of kidney cysts, stones, and tumors with high accuracy, highlighting the potential of real-time detection systems in supporting rapid diagnosis and treatment planning in clinical settings.

Two studies emphasize the effectiveness of VGG16 and InceptionV3 models in detecting kidney diseases. Prasher et al. (2024) achieved a training and validation accuracy of 99.7% using the VGG16 model, demonstrating its strong ability to accurately classify kidney disorders. Meanwhile, Anand et al. (2024) modified the InceptionV3 model, achieving 96.52% accuracy in identifying conditions such as cysts, stones, and tumors. The ability of the VGG16 model to generalize from training to unseen data underscores its real-world applicability (Prasher et al., 2024). Both studies highlight that transfer learning models outperform traditional diagnostic methods, with their high precision showcasing their potential to aid in the early and accurate diagnosis of kidney diseases.

2.1. Motivation of study

Medical image classification is a key area of research, with numerous automated systems developed for accurate disease detection. Deep learning models have significantly contributed to the achievement of high performance in this field. Convolutional Neural Networks (CNNs) are widely used in biomedical computer vision tasks (Krizhevsky et al.,

Table 1

Comparative analysis of related literature.

Paper Refernce	Data Distribution	Methodology	Accuracy (%)	Limitations
Sharma et al. (2024)	Normal: 5,077, Stone: 1,377, tumor: 2283, and Cyst: 3709	Fine-tuned InceptionV3	96.52	Lower performance in tumor detection as compared to other classes.
Prasher et al. (2024)	Normal: 4500, Stone: 1000, tumor: 3000, and Cyst: 2000	Transfer learning with pre-trained VGG16 .	98.8	Over-reliance on pre-trained weights; lower performance on imbalanced datasets.
Hossain et al. (2023a)	Normal: 5,077, Stone: 1,377, tumor: 2283, and Cyst: 3709	custom CNN	99.92	Watershed method struggle with over-segmentation, noise sensitivity, and detecting thin or low-contrast areas.
Chowdhury et al. (2023)	Normal: 5,077, Stone: 1,377, tumor: 2283, and Cyst: 3709	CNN and SVM through teacher-student interaction	85.76; 74.04	The model's generalization to new datasets remains a challenge.
Bindu Madavi and Sowjanya (2023)	Specific counts not mentioned	CNN	98.91	Lower performance in tumor detection as compared to other classes.
Chauhan et al. (2024)	Includes pre-processed kidney ailment images	Attention-based DNN	95.5	Scaling to larger datasets poses computational challenges.
Singh et al. (2023)	1,500 images: 500 each for cyst, tumor, and stone	Fine-tuned EfficientNet-B3	95.77	Challenges in distinguishing overlapping features between tumor and stone cases.
Rajora et al. (2024)	Dataset of CT images	Ensemble model combining VGG16 and InceptionV3	97.8	High training time due to complex architecture; requires significant computational resources.
Sri and Lakshmi (2023)	Normal: 4000, Stone: 2500, tumor: 1500, and Cyst: 3000	Pre-trained VGG16 and fine tunned InceptionV3	95.7	Normalization dependencies limit adaptability to real-world noisy datasets
Ali et al. (2023)	Balanced dataset of 2,000 for each class	Hybrid CNN model integrating feature fusion and ensemble classification	94.5	Requires additional optimization for handling cross-modality image datasets.

2012). Most existing studies focus on customizing single pre-trained models, such as custom CNNs, attention-based DNN, fine-tuned EfficientNet-B3, SVM models,pre-trained VGG16, fine-tunned InceptionV3, etc. as summarized in Table 1. These proposed models suffers from certain limitations. To overcome some of the limitations, the proposed hybrid CNN model offers an effective feature-engineering approach, combining pre-trained ResNet101 and a custom CNN through feature fusion. This model achieves high accuracy in classifying kidney CT images. The classification process involves three key steps: first, feature extraction and classification using the pre-trained ResNet101; second, feature extraction with the custom CNN; and finally, feature fusion to classify images into categories such as normal, stone, cyst, and tumor.

2.2. Research objectives

The objectives of the study are two-folds:

- To develop a hybrid CNN model to classify the kidney as normal, stone, cyst, and tumor with improved accuracy.

- To compare the classification performance of the proposed hybrid CNN with base models: CNN and ResNet101.

3. Methodology

3.1. Data description and pre-processing

The ultrasound or CT scan data set of kidneys with kidney stones, tumor, or cyst was sourced from Kaggle (https://www.kaggle.com/datasets/srinivasbece/kindey-stone-dataset-splitted). The data set was intentionally designed to encompass various types of kidney disease, taking into account the variability in size, location, and composition. This extensive data set was designed to enhance the accuracy and robustness of the system, allowing effective identification and classification of kidney diseases with a comprehensive understanding of clinical manifestations. The data of the images of kidney diseases are classified into four categories, namely: Normal, Stone, Cyst, and Tumor. Table 2 displays the information about the count of each classification of images before and after augmentation. Original CT images are of size 640×640 and each image has been resized to 224×224 to improve the model performance. The data is decomposed into training and testing set as: training set contains (70%) and testing contains (30%) of data. Validation is performed on 20% of the testing data.

3.2. Data augmentation

Data enhancement improves the quality, robustness, and generalizability of machine learning models, especially when data are limited or unbalanced. It involves increasing the size and diversity of the data set by applying transformations such as rotations, translations, scaling, flipping, and cropping to existing data. In this study, flipping images, rotation by 30°, and median filtering are used for data augmentation. Figure1 shows the sample of data after augmentation, and the count of each category is given in Table 2.

Table 2
Data overview: Count of each class of kidney images before and after augmentation.

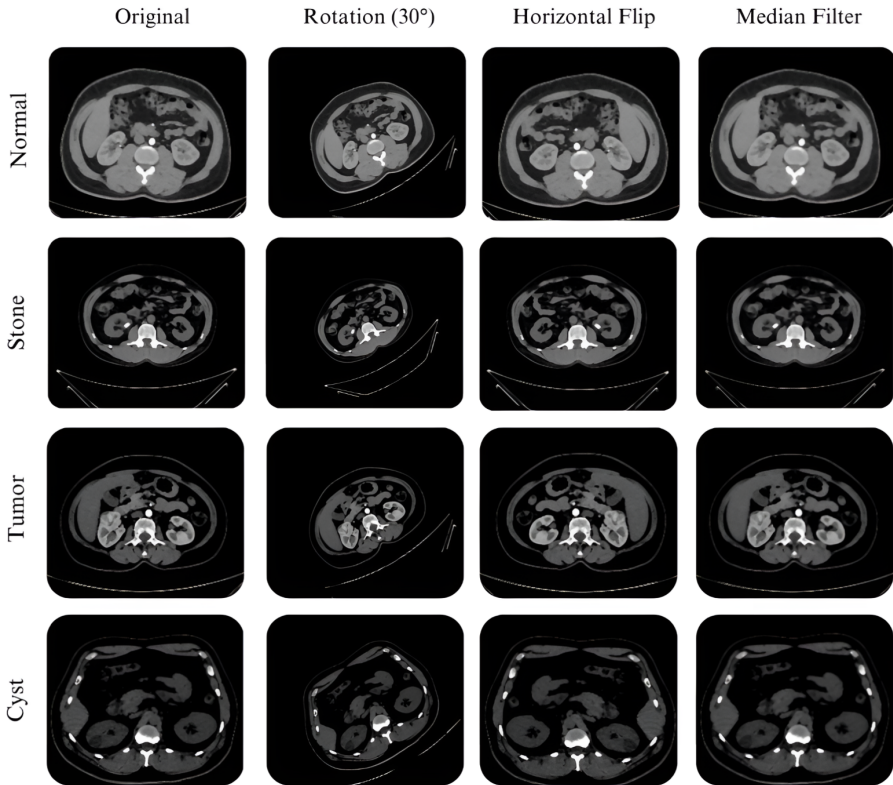| Kidney Type	Number of Kindly Images	
	Before Augmentation	After Augmentation
Normal	5,077	5,141
Stone	1,377	3,918
Cyst	3,709	11,127
Tumor	2,283	6,849
Total	12,446	27,035

Figure 1: An overview of the augmented data after rotation, flipping, and median filter.

4. Architecture of proposed model: Hybrid CNN

The proposed architecture of hybrid CNN shown in Figure 2 illustrates a deep learning architecture to classify kidney images into categories such as normal, stone, tumor and cyst. The architecture has two parts: one it passes input directly to the hybrid CNN model and the other to ResNet101. In Resnet101, the architecture is divided into two specialized branches to address different classification tasks. The first branch handles the classification between normal kidneys and stones, while the second focuses on distinguishing between cysts and tumors. Each branch consists of a series of convolutional layers that extract hierarchical features using various kernel sizes (e.g., 7x7, 3x3, 1x1) and filters, which gradually increase in number (e.g., 64, 256, 512) to capture complex spatial patterns. Batch normalization and max pooling are applied to stabilize training and down-sampling the feature maps, progressively reducing spatial dimensions from 56x56 to 7x7.

The outputs from each branch are passed through classification layers to predict normal vs. stone in one branch and cyst vs. tumor in the other. These features are then merged with the features extracted from hybrid CNN into a unified layer for the final classification. feature mapping filter out the common features with higher similarity. Fully connected layers flatten the feature maps, and a softmax activation function predicts probabilities for all four classes

(normal, stone, cyst, tumor). This dual-branch design enables specialized feature extraction for specific abnormalities while providing a comprehensive output. The inclusion of pre-processing, task-specific branches, and hierarchical feature learning makes this architecture efficient for the detection of automated kidney abnormalities in CT images. The proposed model is explained stepwise in the algorithm 1.
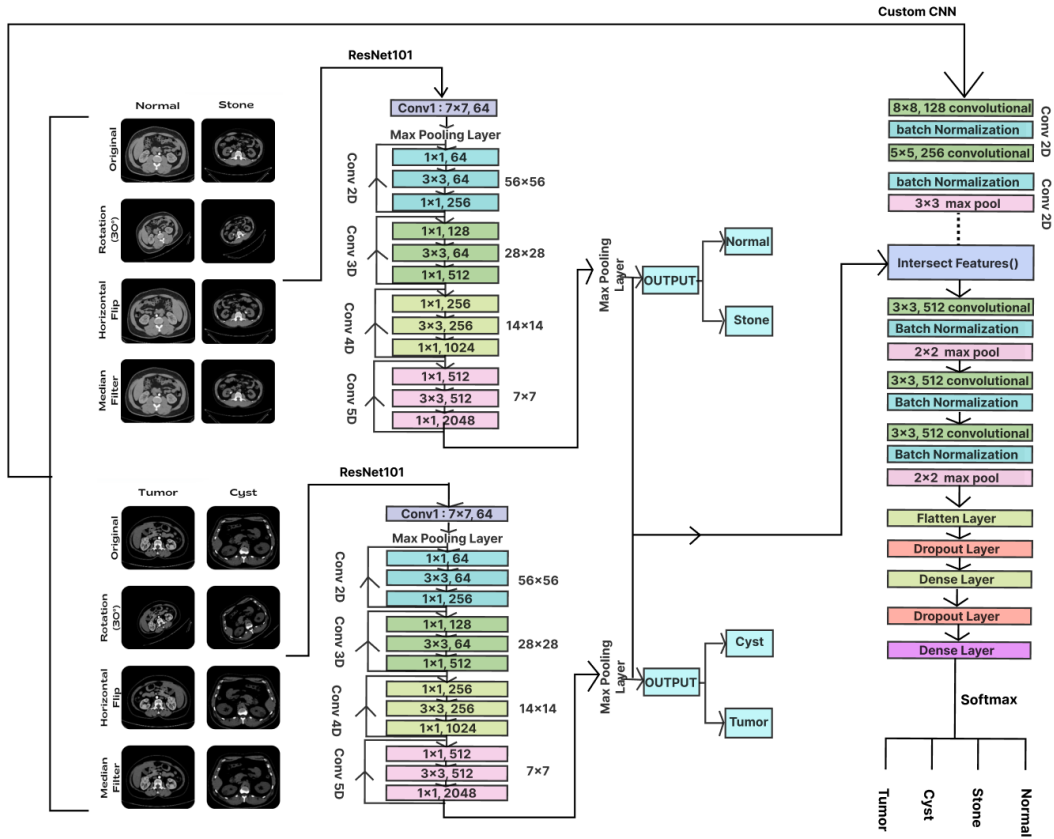

Figure 2: Architecture of the proposed hybrid CNN model.

4.1. Architecture description

Table 3 provides a detailed description of the proposed hybrid CNN architecture, illustrating its layer types, output shapes, and parameter counts. The model begins with a series of Convolutional Layers (Conv2D) to extract spatial features from the input images. These layers progressively increase the number of filters (from 128 to 512) to capture finer and more complex details. Batch Normalization follows each convolutional layer to stabilize and speed up the training process. MaxPooling2D layers are interspersed to downsample the feature maps, reducing their spatial dimensions while retaining essential features.

A specialized Intersect Features function combines relevant features extracted from different models, enhancing the model's ability to discern fine-grained details. The architecture concludes with Fully Connected (Dense) layers, where the feature maps are flattened into a 1D vector and passed through dense layers to perform high-level representation learning. Dropout layers are included to prevent overfitting by randomly deactivating neurons during training. The final dense layer outputs probabilities for the four target classes: normal, stone, cyst, and tumor. With over 15 million trainable parameters, this architecture effectively balances complexity and efficiency to achieve high classification accuracy.

Algorithm 1: Hybrid CNN: Feature Extraction and Classification of Kidney CT Images

1: **Data Preprocessing**

- Dataset: labeled kidney images
- Normalize pixel values and resize CT images
- Apply augmentations (rotation, flipping, and median filter)

2: **Feature Extraction**

1. **ResNet101**
   - Divide the labeled data in two sets: *Normal-Stone* and *Tumor-Cyst*
   - Use ResNet101 on individual classes to extract features
2. **Custom CNN**
   - Input kidney images to custom CNN
   - Extract features from images

3: **Training: Hybrid CNN**

- Pass *Normal-Stone* and *Tumor-Cyst* features extracted from ResNet101 to custom CNN
- Perform feature mapping on features extracted from both ResNet101 and custom CNN
- Keep the features extracted from both ResNet101 and custom CNN with higher similarity
- Train Hybrid CNN by using cross-entropy loss, batch normalization, max pooling, and Adam optimizer

4: **Testing**

- Pass any kidney CT image to Hybrid CNN model and extract features
- Classify the Kidney CT image

5: **Output**

- Predicted class of kidney: Normal, Stone, Cyst, and Tumor
- Evaluate model accuracy: accuracy, precision, recall, and F1 score

4.2. Testing of the model

Figure 3 illustrates the workflow of a hybrid CNN-based model designed for the classification of kidney CT images. Initially, the CT image of the input kidney is processed by the proposed custom CNN to extract relevant features. These features are then refined and passed through a feature extraction process to generate meaningful data representations. The extracted features are subsequently combined using a feature mapping module within the hybrid CNN architecture. This fusion process ensures the integration of diverse set of features for robust classification. Finally, fused features are classified into one of four categories: normal, stone, cyst, or tumor.

5. Experimental setup

This section elaborates the experimental setup that includes the configuration and implementations of the proposed hybrid CNN along with the evaluation metrics.

5.1. Configuration

Table 4 shows all the pre-initialized parameters used for hybrid CNN. It specifies that the model is trained using 5 epochs and a batch size of 32. The learning rate is set to 0.001, which dictates the speed of weight updates during training. The stride, which controlled the size of the step during convolution, is set to 1, ensuring fine-grained feature extraction. The first layer of the model comprises 128 kernels, each applying a convolution operation to extract features, and uses the ReLU activation function for non-linearity. The dropout rate of 0.5 is applied to prevent overfitting by randomly deactivating half the neurons during training. The model has 15,605,124 trainable parameters and 6,400 non-trainable parameters, emphasizing its complexity. On average, it required 132 seconds per epoch for training. These details highlight the design choices aimed at optimizing the hybrid CNN performance.

Table 3

The architecture description of the proposed hybrid CNN model.

| Layer (type)	Output Shape	#Param
conv2d (Conv2D)	(None, 73, 73, 128)	24,704
batch_normalization (BatchNormalization)	(None, 73, 73, 128)	512
conv2d_1 (Conv2D)	(None, 73, 73, 256)	819,456
batch_normalization_1 (BatchNormalization)	(None, 73, 73, 256)	1,024
max_pooling2d (MaxPooling2D)	(None, 24, 24, 256)	0
conv2d_2 (Conv2D)	(None, 24, 24, 256)	590,080
batch_normalization_2 (BatchNormalization)	(None, 24, 24, 256)	1,024
conv2d_3 (Conv2D)	(None, 24, 24, 256)	65,792
batch_normalization_3 (BatchNormalization)	(None, 24, 24, 256)	1,024
conv2d_4 (Conv2D)	(None, 24, 24, 256)	65,792
batch_normalization_4 (BatchNormalization)	(None, 24, 24, 256)	1,024
Intersect Features()		
conv2d_5 (Conv2D)	(None, 24, 24, 512)	1,180,160
batch_normalization_5 (BatchNormalization)	(None, 24, 24, 512)	2,048
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 512)	0
conv2d_6 (Conv2D)	(None, 12, 12, 512)	2,359,808
batch_normalization_6 (BatchNormalization)	(None, 12, 12, 512)	2,048
conv2d_7 (Conv2D)	(None, 12, 12, 512)	2,359,808
batch_normalization_7 (BatchNormalization)	(None, 12, 12, 512)	2,048
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 512)	0
conv2d_8 (Conv2D)	(None, 6, 6, 512)	2,359,808
batch_normalization_8 (BatchNormalization)	(None, 6, 6, 512)	2,048
max_pooling2d_3 (MaxPooling2D)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 1024)	4,719,616
dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 1024)	1,049,600
dropout_1 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 4)	4,100
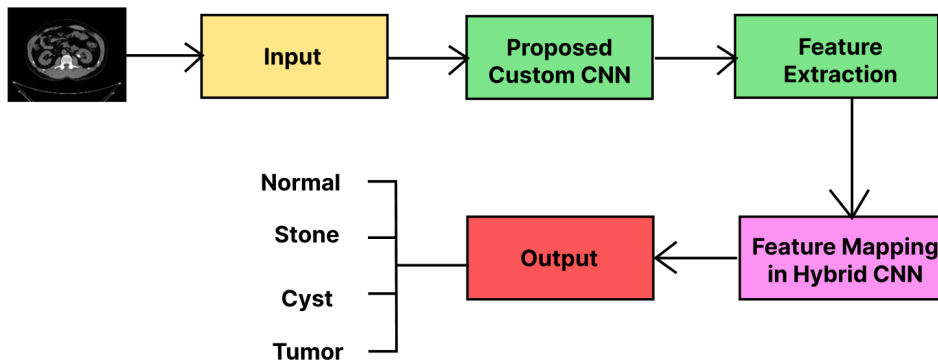
Figure 3: Schematic outline of hybrid CNN testing.

5.2. Computational setup

The model operates on a system equipped with a T4 GPU, 12.7 GB of system RAM, 15.0 GB of GPU RAM, and 112.6 GB of disk space, providing a robust computational environment. The model runs efficiently for up to 10 epochs, using these resources optimally. Beyond this threshold, it fully utilizes the available system and GPU resources, highlighting the system's computational capacity and the model's demand for memory and processing power during

Table 4

Hyperparameters used in the proposed model.

Hybrid CNN Hyperparameters	Details
Number of epochs	5
Batch size	32
Learning rate	0.001
Stride	1
Number of kernels in different layers	128, 256 and 512
Activation	ReLu
Dropout rate	0.5
Number of trainable parameters	15,605,124
Number of non trainable parameters	6,400

Table 5

Training hybrid CNN: Performance metrics of the model for classification of kidney as normal, stone, cyst, and tumor for each epoch.

Epoch	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Std Dev
1	74.83	96.71	96.59	96.48	0.0942
2	94.78	99.84	99.84	99.84	0.0219
3	98.58	100	100	100	0.0061
4	99.45	100	100	100	0.0023
5	99.73	100	100	100	0.0012

extended training.

5.3. Evaluation metrics

The following properties are evaluated to measure the performance of the model.

1. **Precision:** Precision measures the accuracy of positive predictions and is caluclated as

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

where TP is true positive and FP is false positive.

2. **Recall:** Recall, or sensitivity, measures the ability of a model to find all true positives cases and is calculated as

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where FN is false negative.

3. **F1-Score:** The F1 score is the harmonic mean of precision and recall and is calculated as

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

4. **Accuracy:** Accuracy measures the overall correctness of the model and is calculated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TN is true negative.

Table 5 provides a classification report for the performance of a model over five epochs, detailing various evaluation metrics. The precision increased from 74.83% in Epoch 1 to 99.73% in Epoch 5, with precision, recall, and F1 scores reaching perfect values (100%) from Epoch 3 onward. The average standard deviation decreased significantly, from 0.0942 in Epoch 1 to 0.0012 in Epoch 5, reflecting increased stability in predictions.
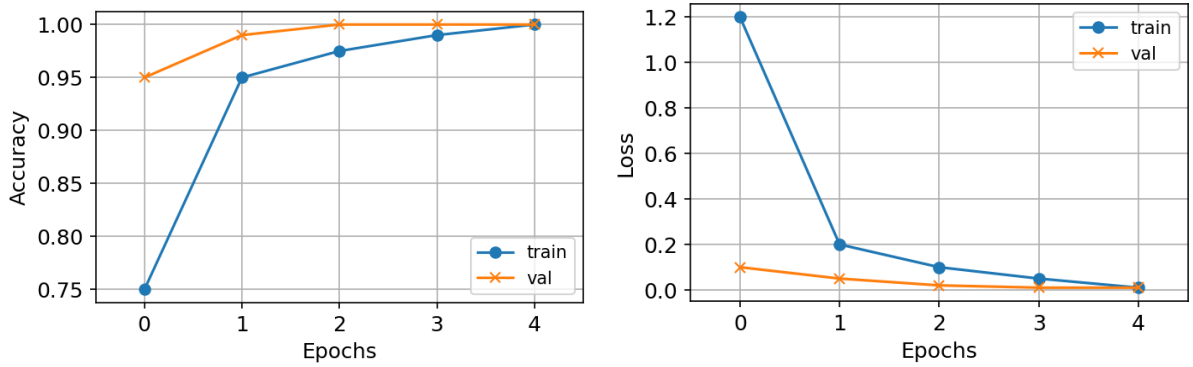
Figure 4: (*left*) Model accuracy trend. (*right*) Model loss trend.

Table 6
Testing the performance of hybrid CNN with base CNN and pre-trained ResNet101.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (hrs)	Testing Time (sec)
ResNet101	99.81	99.75	99.50	99.75	0:47:55	31
Hybrid CNN	100	100	100	100	1:25:02	23

6. Results and discussion

6.1. Accuracy and loss

Figure 4 (left) illustrates the accuracy trends for the training and validation datasets in five epochs during model training. Training accuracy starts around 74% in Epoch 1 and rapidly increases, reaching nearly 100% by Epoch 3, maintaining this high performance through subsequent epochs. This demonstrates that the model quickly learned the training data with high precision. However, validation accuracy begins higher than training accuracy at approximately 95% in Epoch 1 and achieves 100% by Epoch 3, remaining stable through Epochs 4 and 5. The steady and close alignment with the training accuracy indicates that the model generalizes well and avoids overfitting. The overlapped accuracy values after Epoch 3 signify a well-trained model with balanced performance on both datasets.

Similarly, Figure 4 (right) shows the loss trends for both the training and validation data sets in five epochs during model training. Training loss starts high at around 0.65 in Epoch 1 and decreases sharply to near 0 by Epoch 3. This indicates that the model effectively minimizes error in the training dataset as the training progresses. However, validation loss starts at around 0.2 in Epoch 1 and also decreases steadily to near 0 by Epoch 3, maintaining this minimal loss in subsequent epochs. The low and stable validation loss suggests that the model generalizes well to unseen data and avoids overfitting. Rapid convergence of both training and validation losses to near-zero values highlights efficient learning of the model and excellent performance on both datasets at the end of training.

6.2. Comparative analysis

To test the performance of the model, a test sample of 3,734 CT images is used. Table 6 shows the comparison of hybrid CNN with pre-trained ResNet101 and base CNN models. Among the three, hybrid CNN outperforms the others, achieving 100% accuracy, precision, recall, and F1 score, demonstrating its ability to classify all images perfectly without false positives or negatives. ResNet101 also performs strongly, with an accuracy of 99.81%, precision of 99.75%, recall of 99.50%, and an F1 score of 99.75%, while the base CNN is significantly short, with an accuracy of 77.39%, precision of 78.10%, recall of 71.72%, and an F1 score of 72.80%. The training time is longest for the hybrid CNN due to its complexity, compared to ResNet101 and base CNN. However, the hybrid CNN compensates for this with the fastest testing time compared to other two. This indicates that the hybrid CNN effectively combines the strengths of pre-trained ResNet101 and the custom CNN, resulting in exceptional classification performance with practical efficiency.

6.2.1. Confusion matrix

We assessed the performance of the hybrid CNN by analyzing its confusion matrix in comparison with ResNet101. The confusion matrix provides a two-dimensional representation where rows represent the actual class labels (Cyst, Normal, Stone, Tumor) and columns represent the predicted labels by the model. The ResNet101 model demonstrated high precision across all four classes, accurately classifying 1,017 out of 1,018 cysts, 1,531 out of 1,534 normals, 423 out of 428 stones, and 648 out of 654 tumors. The highest misclassification occurred with four stones being incorrectly predicted as cysts, but overall, the true positive rates were exceptional, showcasing robust feature extraction and classification capabilities. When comparing the hybrid CNN's performance to ResNet101, the hybrid CNN achieved perfect classification accuracy. It correctly identified 1,118 cases of Cyst, 1,534 cases of Normal, 428 cases of Stone, and 654 cases of Tumor, with zero misclassifications.
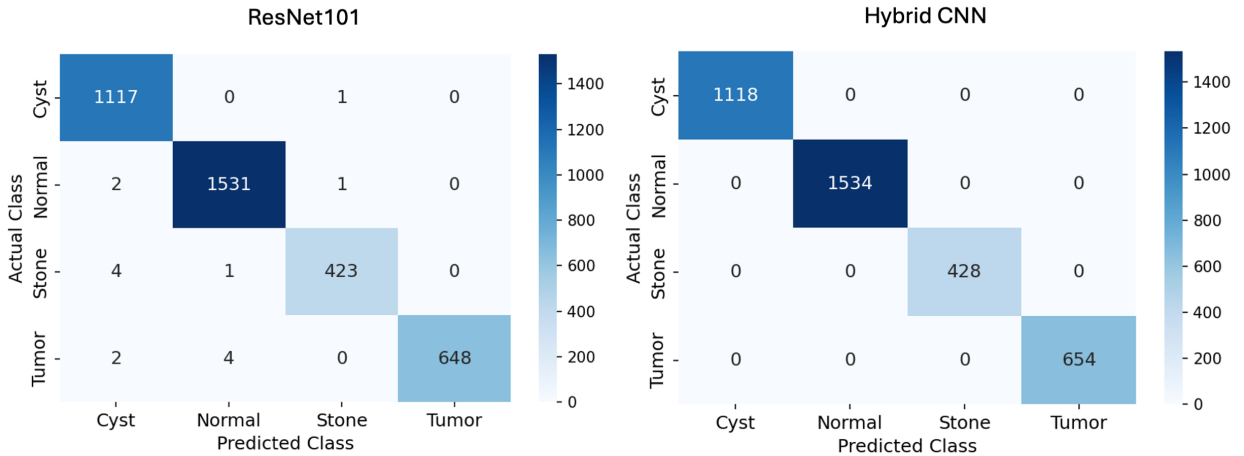

Figure 5: Comparative analysis of confusion matrix. (*left*) ResNet101 and (*right*) Hybrid CNN.

6.2.2. Principal component analysis (PCA)

Figure 6 shows the confusion matrix of the ResNet101 model.

Principal Component Analysis (PCA) is used to reduce the high-dimensional feature space from CT images to two dimensions, while preserving the maximum variance. PCA computes the eigenvectors of the covariance matrix, which represent directions of maximum variance in the data. This allows visualization of the distribution of four classes: healthy kidney, kidney stone, cyst, and tumor. The first two principal components, PC1 and PC2 for ResNet101, and proposed hybrid CNN, are shown in Figure 6.

In both PCA's, each point represents a CT image, colored according to its class: Normal (orange), Stone (green), Cyst (blue), and Tumor (red). The plot reveals a clear separation between classes. Normal kidney samples primarily cluster in the lower region, cyst samples form distinct clusters in the upper regions, and Stone and Tumor samples overlap in the intermediate regions. Kidney stone samples and tumor samples overlap, probably due to shared features such as irregular shapes or intensity patterns.

PCA for ResNet1010 shows a mirrored pattern along the PC1 axis. The left half ($PC1 < 0$) likely represents one subgroup, while the right half ($PC1 > 0$) represents another. Almost mirror images in scatter plots might occur because PCA relies on eigenvectors and eigenvalues that define directions in space, but do not have inherent orientations. Eigenvectors are normalized to have unit length, but their signs are arbitrary. For example, if $[v, v, v]$ is an eigenvector, $[-v, -v, -v]$ is also valid because both represent the same direction. This mirroring could also be due to the following reasons:

1. *Non-linear or Multi-modal Data Structure*: Variations in CT images or differences within each class (e.g., cyst size, stone composition, tumor location) could cause this split.
2. *Data Augmentation:* Techniques like rotation, flipping, mirroring up/down, etc. may create separate groups, as the augmented features are treated as distinct by PCA.

The mirroring does not affect the interpretation of PCA because it does not change the variance explained or the relative distances between points in the reduced space. The PCA plots 1 and 2 demonstrates ResNet101 ability to separate

healthy kidneys from pathological conditions. However, the overlap between Tumor and Stone classes highlights challenges in differentiating these conditions based on the current feature set. This suggests that additional features or a more refined model could improve the classification accuracy for overlapping classes. The proposed hybrid CNN model (Figure 6(right)) shows a clearer separation of the four classes without the mirroring effect, indicating that the hybrid CNN model resolves the subgroup issue by filtering common features.
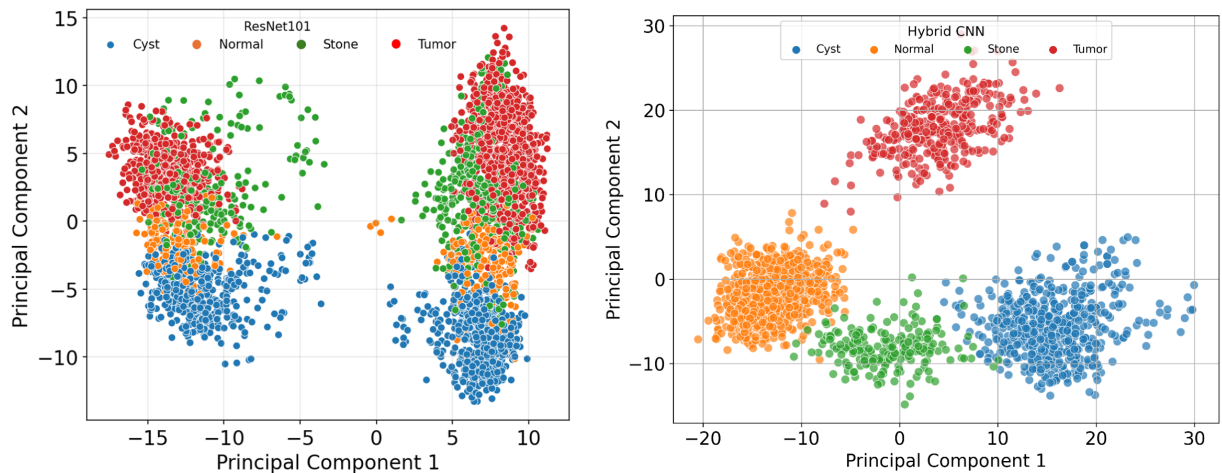

Figure 6: Comparative analysis of confusion matrix. (*left*) ResNet101 and (*right*) Hybrid CNN.

7. Conclusion

To conclude, the hybrid CNN model excelled in classifying kidney CT images, achieving 100% accuracy, precision, recall, and F1 score. By combining ResNet101 features and a custom CNN via feature fusion, the model captured diverse feature representations, significantly outperforming standalone models. This dual-branch architecture facilitated robust feature extraction and mapping, enabling precise differentiation across all kidney conditions. The study underscores the potential of the model to optimize clinical workflows, improve diagnostic accuracy, and support early intervention strategies, thus significantly improving patient care and outcomes.

7.1. Limitations

Despite its high performance, the proposed model has some limitations. First, the training time is longer compared to individual models due to the complexity of the hybrid architecture. Second, the model relies on a relatively large, well-annotated dataset, which may not always be available in real-world scenarios.

References

Ali, M.M., Agrawal, J., Mishra, T., Raj, M., 2023. An automated deep learning approach for kidney disease detection, in: 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), IEEE. pp. 1–8.

Anand, V., Chauhan, R., Sharma, G., Pokhariya, H.S., Gupta, S., Sunil, G., 2024. Transfer learning empowered multi-class classification of kidney diseases: A deep learning approach, in: 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), IEEE. pp. 1–8.

Baygin, M., et al., 2022. Exemplar darknet19 feature generation technique for automated kidney stone detection with coronal ct images. Artificial Intelligence in Medicine 127, 102274.

Bindu Madavi, K.P., Sowjanya, K., 2023. Deepkidney: Multiclass classification of kidney stones, cysts, tumors, and normal cases using convolutional neural networks, in: 4th International Conference on Communication, Computing and Industry 6.0 (C216), IEEE. pp. 1–8.

Chauhan, R., Karnati, M., Singh, P., 2024. Attention based deep neural network for classification of kidney ailments using ct images, in: 15th ICCCNT IEEE Conference, IEEE. pp. 1–8.

Chowdhury, S.T., Mukhopadhyay, S., Narendra, K.S., 2023. Mutual learning algorithm for kidney cyst, kidney tumor, and kidney stone diagnosis, in: 18th Conference on Computer Science and Intelligence Systems, IEEE. pp. 401–410.

Elhoseny, M., Shankar, K., Uthayakumar, J., 2019. Intelligent diagnostic prediction and classification system for chronic kidney disease. Scientific Reports 9, 9583. doi:10.1038/s41598-019-46074-2.

Gulhane, M., Kumar, S., et al., 2024. Integrative approach for efficient detection of kidney stones based on improved deep neural network architecture. SLAS Technology 29, 100159.

Hossain, M.S., Hassan, S.M.N., Al-Amin, M., Rahaman, M.N., Hossain, R., Hossain, M.I., 2023a. 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) , 1–6doi:10.1109/AICAPS57044.2023.10074314.

Hossain, M.S., Hassan, S.N., Al-Amin, M., Rahaman, M.N., Hossain, R., Hossain, M.I., 2023b. Kidney disease detection from ct images using a customized cnn model and deep learning, in: IEEE Conference on Advances in Intelligent Computing and Applications, IEEE. pp. 1–8.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25.

Pande, S.D., Agarwal, R., 2024. Multi-class kidney abnormalities detecting novel system through computed tomography. IEEE Access 12, 1–10.

Patro, K.K., Allam, J.P., et al., 2023. Application of kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal ct images. Information Sciences 640, 119005.

Prasher, S., Nelson, L., Sandhya, V., 2024. Vgg16 transfer learning model for diagnosing multi-class kidney disorder, in: 2024 5th International Conference for Emerging Technology (INCET), IEEE. pp. 1–8.

Rajora, R., Gupta, H., Malhotra, S., Devliyal, S., Sunil, G., 2024. Advancing renal health: Unleashing the power of cnns in multi-class classification for precise kidney condition diagnosis, in: IEEE 9th International Conference for Convergence in Technology (I2CT), IEEE. pp. 1–8.

Sharma, G., Pokhariya, H.S., Anand, V., Gupta, S., Chauhan, R., Sunil, G., 2024. Revolutionizing kidney disease diagnosis: A comprehensive cnn-based framework for multi-class ct classification, in: IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems, IEEE. pp. 1–8.

Singh, R., Sharma, N., Chauhan, R., Choudhary, A., Gupta, R., 2023. Precision kidney disease classification using efficientnet-b3 and ct imaging, in: 3rd International Conference on Smart Generation Computing, Communication and Networking, IEEE. pp. 1–8.

Sri, V.S., Lakshmi, G.R.J., 2023. Deep learning technique to detect and diagnose the anomalous in kidney, in: International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), IEEE. pp. 1–8.

Wu, Y., Yi, Z., 2020. Automated detection of kidney abnormalities using multi-feature fusion convolutional neural networks. Knowledge-Based Systems 200, 105873.

Yildirim, K., et al., 2021. Deep learning model for automated kidney stone detection using coronal ct images. Computers in Biology and Medicine 135, 104569.

Çağlayan, A., Horsanalı, M., Kocadurdu, K., İsmailoğlu, E., Guneyli, S., 2022. Deep learning model-assisted detection of kidney stones on computed tomography. International Brazilian Journal of Urology : Official Journal of the Brazilian Society of Urology 48, 830 – 839. doi:10.1590/S1677-5538.IBJU.2022.0132.

Ethical Statement

The authors declare that the study did not involve human participants or animals.

Data Availability Statement

The data used in this study is available on https://www.kaggle.com/datasets/srinivasbece/kindey-stone-dataset-splitted. The data is also available through the corresponding author on reasonable request.

Author Contributions

KS conceptualized the research problem, conducted the analysis, wrote and reviewed the manuscript. ZU conceptualized the research problem, performed analysis, wrote and reviewed the manuscript. AW collected the data and carried out the analysis. DG designed the analysis framework and contributed to figure preparation.

Funding

No funding is available.

Competing Interests

The authors declare that they have no conflict of interest.