Available online at http://sanad.iau.ir/journal/ijim/
Int. J. Industrial Mathematics (ISSN 2008-5621)
Vol.17, No.4, 2025 Article ID IJIM-1629, 10 pages
Research Article

Multi-Class Classification of Kidney Diseases in CT Images Using Deep Feature Extraction and Random Forest

Z. Amiri[1 *], S.M. Jalali[1], M. AlyanNezhadi[2], Z. Mortezaie[3], S.M.R. Hashemi[4]
[1]Department of Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran.
[2]Department of Science, University of Science and Technology of Mazandaran, Behshahr, Iran.
[3]Department of Mathematics and Computer Sciences, Hakim Sabzevari University, Sabzevar, Iran.
[4]Lecturer, Semnan University, Semnan, Iran.

Submission Date:2025/05/21, Revised Date:2025/08/21, Date of Acceptance:2025/09/04

Abstract
Kidney disease is among the leading causes of morbidity all over the world and requires early diagnosis to prevent severe complications. Intelligent diagnostic methods, such as machine learning, enhance the accuracy and efficiency of detection by analyzing advanced clinical and imaging information. This research investigates a novel approach for multi-class classification of kidney diseases in CT images. The classification includes four classes: cyst, stone, tumors, and Normal. The proposed method first extracts the key features from multiple pretrained convolutional neural network models. The selected features are then reduced using principal component analysis, and eventually, the extracted features are fed to the Random Forest classifier. The designed model achieved a high accuracy of 99.7%. The results indicate that the proposed model is more accurate than traditional methods and can assist physicians in making faster and more precise diagnoses. This research aids in improving the quality of healthcare and the precision of the diagnosis of kidney diseases.

Keywords: Convolutional neural network, Kidney disease, Image Processing, Multi-class Classification, Machine Learning.

* Corresponding author: Email: zahra.amiri@mazust.ac.ir

1. Introduction

Kidney disease is a global health problem affecting millions of individuals each year .Chronic kidney disease (CKD) is among the contributors to worldwide morbidity and mortality. It is more and more on the increase with more related conditions being on the upswing such as diabetes and hypertension [1]. CKD needs to be diagnosed early to prevent progression of the disease and minimize treatment costs. Early diagnosis, however, is hard as there are no symptoms of CKD during the initial phases.

Traditionally, diagnosis of kidney diseases from CT scans and clinical examinations relied greatly on human interpretation by physicians and radiologists. This process was time-consuming and prone to human error [2]. Therefore, with the enormous increase in medical imaging data, medical professionals have been increasingly required to interpret them in a correct and timely manner. This reality is towards creating intelligent, computerized systems that can diagnose kidney disease in time and accurately.

Deep learning (DL) and artificial intelligence (AI) have proved to be a highly promising technology for enhancing medical diagnosis [3-9]. Convolutional neural networks and deep learning algorithms, for example, are very capable of learning to obtain intricate patterns from medical images and thereby support better and accurate diagnosis [10]. They can capture the subtle details and abnormalities that may escape observation with the human eye and give a tremendous leap over traditional diagnostic methods. Despite these improvements, there are drawbacks such as computational processes being computationally intensive, time to train, and vulnerability to noisy data [11].

Traditional machine learning algorithms such as decision trees [12,13], Support Vector Machines (SVM) [14,15], K-Nearest Neighbors (KNN) [16,17], and Random Forests (RF) [18] were used in earlier studies to classify kidney disease. Although they produced good accuracy, they were limited by the quality of handcrafted features and sensitivity to imbalanced data. To surpass such limitations, deep CNNs have been used for automatic feature extraction and classification tasks, demonstrating better performance and robustness in handling complex visual patterns in medical images. By combining these two approaches, researchers can benefit from both, achieving optimal diagnostic accuracy and computational efficiency [19].

Recently, several studies have used deep learning to diagnose kidney diseases. Blau et al. [20] employed a fully connected CNN to identify kidney cysts in abdominal CT images, achieving a true positive rate of 84.3%. In a different study, Yildirim et al. [21] introduced a deep learning approach for identifying kidney stones in kidney CT scans, reaching an accuracy of 96.82%. Uhm et al. [22] adjusted ResNet-101 by adding 1×1×1 convolutions for classifying five types of kidney cancer, achieving an AUC of 0.88 and an accuracy of 72%. Wagih et al. [23] applied PCA for feature selection from ultrasound images and trained a neural network, reaching 97% classification accuracy in five kidney-related categories.

Pande and Agarwal [4] proposed a YOLOv8-based deep learning model for classifying kidney abnormalities in CT images into four classes: cyst, tumor, stone, and normal. Their approach focused on real-time detection using object localization and single-pass classification. This method is compared with our proposed model in the results section, where our approach demonstrates improved performance.

In this paper, a hybrid approach is proposed for computer-aided detection of kidney disease from CT scans. Our approach uses three strong pre-trained CNN models—ResNet50, InceptionV3, and DenseNet121—to exhaustive feature extraction. To avoid the risk of overfitting and feature dimensionality, we apply Principal Component Analysis (PCA), followed by random forest classification. The hybrid approach blends the strong feature extraction power of deep learning with the strength and power of conventional classifiers to produce an overwhelmingly well-balanced and effective solution.

The greatest contribution of this research is to integrate deep learning and conventional machine learning algorithms in an overall fusion in to outsmart the inherent challenges posed in the diagnosis of kidney disease. The integration not only improves diagnostic precision and reduces computational load but also resolves class imbalances in the database very efficiently. The proposed system will have an ideal and realistic solution to be implemented in clinical settings in real life to facilitate early diagnosis and improved decision-making in the clinical setting.

2. Proposed method

This section provides the overall methodology used to design an effective automated kidney disease classifier system from CT scan images. The proposed framework integrates feature extraction using different pre-trained CNNs, dimensionality reduction using PCA, elimination of class imbalance using dynamic weighting, and ultimate classification using a RF algorithm. Process flow is presented in Fig. 1 and discussed in the following subsections.

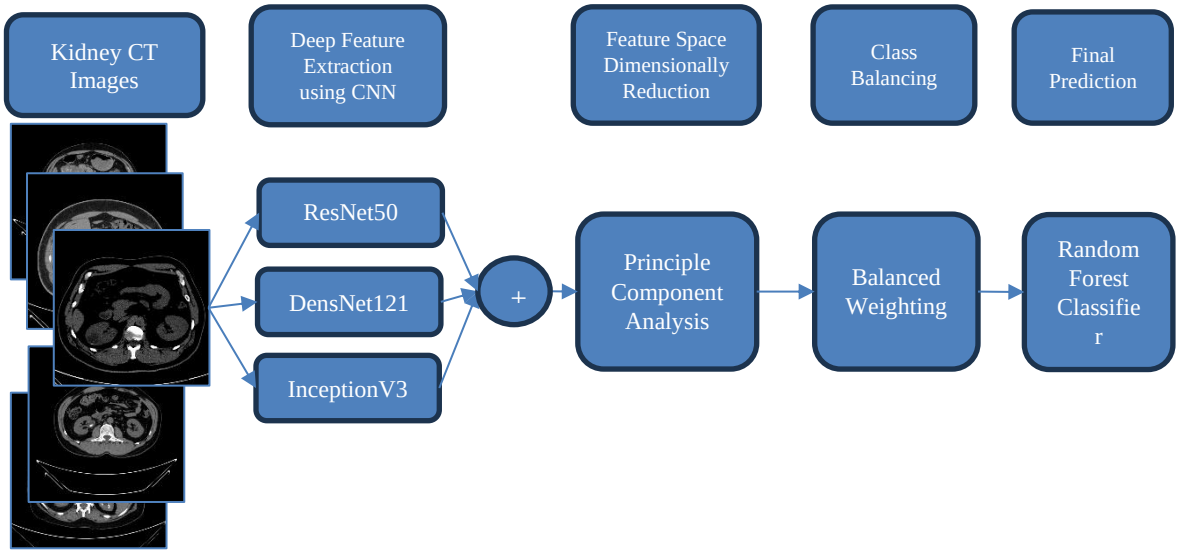

Figure 1: The flowchart of the proposed method.

2.1. Feature Extraction

In this step of the proposed method, features are extracted from three well-established CNN architectures pre-trained on the ImageNet dataset, containing ResNet50, InceptionV3 and

DenseNet121. Each network is truncated at the global average pooling layer to obtain high-level features without performing full network retraining. Specifically, ResNet50 and InceptionV3 produce 2048-dimensional feature vectors from their avg_pool and global_average_pooling2d layers, respectively, whereas DenseNet121 produces a 1024-dimensional feature vector from its global_average_pooling2d layer. The concatenated feature vectors form a single 5120-dimensional representation for each image. The aforementioned feature fusion process takes advantage of the complementary nature of the selected CNN models and, in doing so, enhances the features and class discriminability.

2.2. Dimensionality Reduction by PCA

To fight the curse of dimensionality and reduce computational expense, Principal Component Analysis was applied to the 5,120-dimensional feature vectors. PCA finds orthogonal directions that capture the most variance of the data and projects the features onto a lower-dimensional space. In this case, the number of principal components was selected such that it captures 99.5% of the cumulative variance, which resulted in a final vector size of 334 dimensions. This step has the effect of eliminating redundancy while preserving the most discriminative information.

2.3. Class Imbalance Handling

In most existing datasets for medical images, and especially for kidney diseases, the data is imbalanced across classes, and the tumor and stone classes have significantly fewer samples. To counteract model bias towards dominant classes, class weighting is employed during training. The class weight w_i for class i is computed as:

$$W_i = \frac{N}{n_i + k} \qquad (1)$$

where N is the total number of training instances, n_i is the number of samples in the ith class and k is the number of classes. The weighting causes minority classes to contribute to the model's learning objective proportionally and increases the model’s efficiency in recognizing data across all classes.

2.4. Classification using Random Forest

After dimensionality reduction and reweighting, the 334-dimensional feature vectors are classified using an RF classifier. RF is an ensemble learning technique that averages the output of many decision trees trained on random data and feature subsets. The hyperparameters used in the RF classifier are reported in Table 1. RF is chosen because of its high interpretability, resistance to overfitting, and good performance on multi-class and imbalanced data.

Table 1-The parameters used in the RF classifier

Hyperparameter	Value
Number of trees	100
Splitting criterion	Gini impurity
Maximum depth	Unlimited

Minimum samples to split an internal node	2
Minimum samples per leaf node	1
Random seed (for reproducibility)	42

3. Experiments and Results

In this section, a comprehensive analysis of the proposed kidney disease classification system with CT scan images is presented. Details regarding the dataset, normalization process, evaluation metrics, experimental setup, and result analysis are presented in the following sections.

3.1. Dataset Description

We used the publicly available CT Kidney Dataset: Normal-Cyst-Tumor and Stone, consisting of 12,446 annotated abdominal CT scans in four clinical classes. The detail of the dataset is given in below table. As can be seen, the classes are imbalanced, with some classes, such as Stone, having a small number of samples. Therefore, the proposed model must be robust in handling these imbalances.

Table 2- The detail of the CT Kidney dataset.

Class	Number of Images	Percentage
Normal	5,077	40.8%
Cyst	3,709	29.8%
Tumor	2,283	18.3%
Stone	1,377	11.1%

3.2. Evaluation Metric

In this subsection, the performance metrics for evaluating the performance of the model to classify kidney images of CT-Scan into four classes (stone, cyst, tumor, and normal) are created. Precision, Recall (Sensitivity), F1-Measure, and Accuracy are the metrics that are used here, providing a better understanding of the strengths and weaknesses of the model. Precision shows the ratio of correctly predicted positive samples (TP) to the total predicted positives (i.e., TP + FP). For a specific class i:

$$\Pr ecision_i = \frac{TP_i}{TP_i + FP_i} \qquad (2)$$

Recall (or sensitivity) measures the ratio of correctly predicted positive samples (TP) to the actual positive samples (i.e., TP + FN, where FN are False Negatives). For a specific class i:

$$\mathrm{Re}\, call_i = \frac{TP_i}{TP_i + FN_i} \qquad (3)$$

The F1-Measure combines precision and recall into a single metric using their harmonic mean. It is particularly useful in cases where the class distribution is imbalanced. For a specific class $\boldsymbol{i}$:

$$F1-Measure_i = 2 \cdot \frac{\Pr ecision_i \cdot \mathrm{Re}\, call_i}{\Pr ecision_i + \mathrm{Re}\, call_i} \qquad (4)$$

A higher F1-Score represents a good balance between sensitivity and precision, which is particularly important when the dataset is imbalanced across the four classes. In multi-class classification, macro-averaged value (unweighted mean across all classes) or weighted value (weighted mean based on the number of samples in each class) can be used.

Accuracy is the simplest metric and is defined as the ratio of correctly predicted samples (both positive and negative) to the total number of samples.

$$Accuracy_i = \frac{TP_i + TN_i}{Total\ Samples} \qquad (5)$$

where k is the number of classes. The value of the above metrics is in range 0 to 1. In all of above metrics, higher values are better, as they indicate better model performance.

3.3. Results and Discussion

This section evaluates the suggested approach by utilizing the introduced evaluation metrics. In every experiment, the dataset is split into training and testing data. Eighty percent of the data is allocated for training, while the other twenty percent is designated for testing. Table 3 displays the confusion matrix obtained by applying the suggested method to the test dataset. The approach is also assessed using the evaluation metric, and the findings are presented in

Table **4**.

Table 3- Confusion Matrix

True \ Predicted	Normal	Cyst	Stone	Tumor
Normal	741	1	0	0
Cyst	1	1014	0	1
Stone	4	4	267	0
Tumor	2	1	0	454

Table 4- Results of model evaluation on the test dataset

Class /Metric	Accuracy	Precision	Recall	F1-score
Cyst	0.997	0.994	0.998	0.996
Normal	0.997	0.991	0.999	0.995
Stone	0.997	1.000	0.971	0.985
Tumor	0.998	0.998	0.993	0.996
Weighted Average	0.997	0.994	0.994	0.994

As can be seen from the above results, the system also attains superb classification performance with a weighted average accuracy of 99.7% in all metrics. Quite notably, even for minority classes such as "Stone" and "Tumor," the model achieves high accuracy, unlike classical classifiers that underperform due to class imbalance. This stable and quality

performance indicates the strength and capability of generalization of the designed system and provides it with usability in real clinical applications despite inherent data imbalance.

In the next experiment the impact of dimensionality reduction is evaluated. To this aim the PCA step was removed from introduced method. Table 5 indicates the classification performance of the suggested approach without and with PCA usage. As shown, both configurations achieved identical results for all main evaluation metrics, including Accuracy (0.997), Precision (0.994), Recall (0.994), and F1-Score (0.994). The results confirm that PCA usage retained the full discriminative power of the features but potentially reduced the computational complexity and size. The fact that dimensionality reduction could do so without sacrificing performance demonstrates that redundant or noisy features were effectively removed, resulting in a compact but equally informative feature space. Therefore, PCA-augmented pipeline is considered more efficient and scalable, particularly in the case of high-dimensional medical imaging data.

Table 5- Evaluating the impact of the feature reduction step on the model

Method	Accuracy	Precision	Recall	F1-Score
Proposed Method without PCA	0.997	0.994	0.994	0.994
Proposed Method with PCA	0.997	0.994	0.994	0.994

To further validate the efficiency of our system, A comparative analysis was performed with the approach presented by Pande et al. [4]. As shown in Table IV, the proposed method outperforms the baseline model in all categories, including accuracy, precision, recall, and F1-score.

Table 6- Comparison of the proposed method with existing methods

Class	Metric	Year	Accuracy	Precision	Recall	F1-Score
Cyst	Proposed Method		**0.997**	**0.994**	**0.998**	**0.996**
	Pande et al.[4]	2024	0.972	0.984	0.922	0.960
	Sowjanya et al.[24]	2023	0.980	-	-	-
	Sharon and Anbarasi[3]	2025	-	0.98	0.98	0.98
Normal	Proposed Method		**0.997**	**0.991**	**0.999**	**0.995**
	Pande et al., 2024[4]	2024	0.863	0.749	1.000	0.857
	Sowjanya et al.[24]	2023	0.996	-	-	-
	Sharon and Anbarasi[3]	2025	-	0.96	0.97	0.96
Stone	Proposed Method		**0.997**	**1.000**	**0.971**	**0.985**
	Pande et al., 2024[4]	2024	0.945	0.732	0.786	0.756
	Sowjanya et al.[24]	2023	0.984	-	-	-
	Sharon and Anbarasi[3]	2025	-	0.98	0.94	0.96
Tumor	Proposed Method		**0.998**	**0.998**	**0.993**	**0.996**
	Pande et al., 2024[4]	2024	0.870	0.965	0.304	0.463
	Sowjanya et al.[24]	2023	0.993	-	-	-
	Sharon and Anbarasi[3]	2025	-	0.97	0.98	0.97
Weighted Average	Proposed Method		**0.997**	**0.994**	**0.994**	**0.994**
	Pande et al., 2024[4]	2024	0.825	0.858	0.753	0.757
	Sowjanya et al.[24]	2023	0.989	-	0.980	0.980
	Sharon and Anbarasi[3]	2025	0.989	-	-	-

The improved performance of our proposed method over other methods can be attributed to three main design choices. First, instead of relying on a single deep model, we extracted features from three complementary CNN architectures (ResNet50, InceptionV3, DenseNet121). This fusion increased the discriminative power of the representation, particularly for visually similar classes such as cysts and tumors. Second, by reducing the fused 5,120-dimensional feature space to 334 dimensions while preserving 99.5% of the variance, PCA effectively removed redundancy and noise. This not only reduced computational cost but also stabilized classification results, as shown in Table 5 where performance was maintained even after dimensionality reduction. Third, the dataset was highly imbalanced (e.g., stone class only 11%). Our class-weighting approach ensured that minority classes contributed proportionally during training, leading to consistently high precision and recall across all four categories, unlike previous works where minority classes were underrepresented.

Together, these strategies produced a more robust and balanced model, as evidenced by the high accuracy (99.7%) and stable performance across all classes (**Error! Reference source not found.**). While a full ablation study is beyond the current scope of this manuscript, the provided analyses and comparisons qualitatively support the contribution of each component to the overall performance gains.

4. Conclusion

In this paper, a hybrid method for multi-class classification of kidney disease in CT images based on deep CNN-based feature extraction, PCA-based dimensionality reduction, and Random Forest classification was presented. The designed model achieved a high accuracy of 99.7% and outperformed existing methods even when the data were class-imbalanced. PCA effectively reduced computational complexity without compromising classification performance. The results demonstrate the efficacy of the integration of deep learning and traditional machine learning methods in developing a balanced and efficient diagnostic tool. With its potency and high precision in all types of kidney diseases, the proposed framework is extremely promising to be applied in real clinical environments to assist radiologists in early and reliable diagnosis.

Despite the promising results and superior performance compared to existing methods, some limitations must be considered. First, the evaluation was carried out on a single publicly available dataset due to the limited availability of other large-scale annotated kidney CT databases. As a result, further testing on independent datasets would be necessary to confirm generalizability across different imaging conditions and populations.

References

[1] K. J. Jager, C. Kovesdy, R. Langham, M. Rosenberg, V. Jha, and C. Zoccali, "A single number for advocacy and communication—worldwide more than 850 million individuals have kidney diseases," vol. 34, ed: Oxford University Press, 2019, pp. 1803–1805.

[2] M. Zhang *et al.*, "Imaging-based deep learning in kidney diseases: recent progress and future prospects," *Insights into imaging,* vol. 15, no. 1, p. 50, 2024.

[3] J. J. Sharon and L. J. Anbarasi, "An attention enhanced dilated bottleneck network for kidney disease classification," *Scientific Reports,* vol. 15, no. 1, p. 9865, 2025.

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

[5] M. Forghani, A. Firstkov, M. Alyannezhadi, D. Danilenko, and A. Komissarov, "Reduced amino acid alphabet-based encoding and its impact on modeling influenza antigenic evolution," *Инфекция и иммунитет,* vol. 12, no. 5, pp. 837–849, 2022.

[6] M. Forghani, A. Firstkov, M. M. AlyanNezhadi, and K. Forghani, "Evaluating Keyphrase Extraction Methods for Clustering Influenza-Related Scientific Papers," in *2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2022: IEEE, pp. 1–7.

[7] M. Forghani, M. Khachay, and M. M. AlyanNezhadi, "The impact of amino acid encoding on the prediction of antigenic variants," in *2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2020: IEEE, pp. 1–5.

[8] Z. Amiri, H. Hassanpour, and A. Beghdadi, "Abnormalities detection in wireless capsule endoscopy images using EM algorithm," *The Visual Computer,* vol. 39, no. 7, pp. 2999–3010, 2023.

[9] Z. Amiri, H. Hassanpour, and A. Beghdadi, "Feature extraction for abnormality detection in capsule endoscopy images," *Biomedical Signal Processing and Control,* vol. 71, p. 103219, 2022.

[10] Z. Amiri, H. Hassanpour, and A. Beghdadi, "Combining deep features and hand-crafted features for abnormality detection in WCE images," *Multimedia Tools and Applications,* vol. 83, no. 2, pp. 5837–5870, 2024.

[11] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis,* vol. 42, pp. 60–88, 2017.

[12] H. Ilyas *et al.*, "Chronic kidney disease diagnosis using decision tree algorithms," *BMC nephrology,* vol. 22, no. 1, p. 273, 2021.

[13] I. Pasadana *et al.*, "Chronic kidney disease prediction by using different decision tree techniques," in *Journal of Physics: Conference Series*, 2019, vol. 1255, no. 1: IOP Publishing, p. 012024.

[14] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *Journal of medical systems,* vol. 41, pp. 1–11, 2017.

[15] N. A. Almansour *et al.*, "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," *Computers in biology and medicine,* vol. 109, pp. 101–111, 2019.

[16] P. Sinha and P. Sinha, "Comparative study of chronic kidney disease prediction using KNN and SVM," *International Journal of Engineering Research and Technology,* vol. 4, no. 12, pp. 608–12, 2015.

[17] C. Priyadharshini, K. Sanjeev, M. Vignesh, N. Saravanan, and M. Somu, "KNN based detection and diagnosis of chronic kidney disease," *Annals of the Romanian Society for Cell Biology,* vol. 25, no. 4, pp. 2870–2877, 2021.

[18] N. Chumuang *et al.*, "An efficiency random forest algorithm for classification of patients with kidney dysfunction," in *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2020: IEEE, pp. 1–6.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[20] N. Blau, E. Klang, N. Kiryati, M. Amitai, O. Portnoy, and A. Mayer, "Fully automatic detection of renal cysts in abdominal CT scans," *International journal of computer assisted radiology and surgery,* vol. 13, pp. 957–966, 2018.

[21] K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Deep learning model for automated kidney stone detection using coronal CT images," *Computers in biology and medicine,* vol. 135, p. 104569, 2021.

[22] K.-H. Uhm *et al.*, "Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography," *NPJ precision oncology,* vol. 5, no. 1, p. 54, 2021.

[23] M. W. Attia, F. Abou-Chadi, H. E.-D. Moustafa, and N. Mekky, "Classification of ultrasound kidney images using PCA and neural networks," *International Journal of Advanced Computer Science and Applications,* vol. 6, no. 4, pp. 53–57, 2015.

[24] B. M. KP and N. Patwari, "DeepKidney: Multiclass Classification of Kidney Stones, Cysts, Tumors, and Normal Cases Using Convolutional Neural Networks," in *2023 4th International Conference on Communication, Computing and Industry 6.0 (C216)*, 2023: IEEE, pp. 1–5.