Integrating DenseNet and Ensemble Learning for Multimodal Prediction of Chronic Kidney Disease Stages

Utkarsha Rane[1] and Dr. Satish Ramesh Kolhe[2]

Abstract. Chronic Kidney Disease (CKD) is a major global health concern, and early detection is critical to preventing disease progression and associated complications. Conventional CKD diagnosis based on ultrasound image interpretation often suffers from inter-observer variability and limited diagnostic consistency. Moreover, many existing artificial intelligence–based studies rely on single-modality data, which restricts predictive performance and limits clinical applicability. To address these limitations, this study proposes a multimodal deep learning framework that integrates ultrasound imaging data with clinical tabular information for accurate CKD stage prediction. The proposed approach employs a DenseNet-based convolutional neural network to extract discriminative features from ultrasound images, while ensemble learning techniques are used to improve classification robustness and predictive accuracy. The model was evaluated using a dataset of 3,734 samples categorized into four classes: Normal, Cyst, Stone, and Tumor. Experimental results demonstrate strong predictive performance, achieving an overall accuracy of 99%, with precision, recall, and F1-score values of 0.99 across all classes. These findings indicate that the proposed multimodal framework significantly enhances diagnostic reliability and has strong potential to support clinicians in early CKD detection and improved decision-making in clinical practice.

Keywords: Chronic Kidney Disease, Kidney Ultrasound, DenseNet, Multimodal Learning, Ensemble Learning, Medical Image Classification.

Manuscript received March 20, 2026; revised April 13, 2026 and published on June 30, 2026
Utkarsha Rane, School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India. Email: utkarsharane05@gmail.com
Dr. Satish Ramesh Kolhe, School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India. Email: srkolhe2000@gmail.com

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a multifaceted global health issue that encompasses the deterioration of the renal system and requires early and precise detection to avoid further damage and enhance the outcome [1]. The conventional techniques for the detection of CKD are associated with the use of invasive techniques or manual inspection of the image, which is time-consuming and may exhibit variability in the outcome. Thus, the need for automated techniques with high precision is required [2]. The application of DL techniques has shown significant potential in the analysis of medical imaging by using the ability of the techniques to learn the features from the image and improve the discrimination of the disease state [3]. Specifically, the application of CNNs, including DenseNet models, is quite effective in the extraction of features from complicated ultrasound imaging using the effective gradient and the mechanism of feature reuse [4]. The integration of multi-modal representation learning with ensemble learning methodologies will enhance the predictive robustness since model bias can be mitigated and generalisation enhanced across heterogeneous data distributions-a feature common in CKD imaging datasets [5]. Recent multimodal ultrasound DL models have performed well in early detection of CKD and fibrosis assessment with superior performance against single-modality techniques, underpinning the potential clinical value of fusion in imaging features [6]. Ensemble frameworks combining multiple DL architectures have also demonstrated better predictive performances in different chronic disease prediction tasks by leveraging complementary strengths of individual models [7].

The focus of the current study will be to develop a multimodal framework for determining chronic kidney disease stages by integrating ultrasound imaging with ensemble learning of clinical data, which can improve diagnostic and early intervention accuracy. Due to the existence of challenges, such as variability in manual image interpretation and the limitation of single-modality models, this work is motivated to extract rich feature representations using DenseNet while taking advantage of ensemble strategies that could enhance the

10.32622/ijrat.1402202603

robustness. The objectives are the correct classification of renal conditions, assessment of model generalisation, and the clinical applicability of the proposed model. Therefore, the current paper discusses the methodological development, experimental validation, comparative analysis, and discussion of findings to provide a deeper understanding of the proposed framework.

II. LITERATURE REVIEW

Recent progress in deep learning has been influential in the diagnosis of kidney diseases using medical imaging, mostly ultrasound. [8]. Research proves that CNN-based methods and strategies related to multimodal fusion enhance the detection and staging of CKD and associated renal conditions more accurately than traditional approaches. [9]. Studies have focused on various techniques such as ultrasound radiomics, feature fusion, and ensemble learning to extract discriminative imaging features for improving classification performance. The literature review given summarises contemporary findings on deep learning-based renal imaging models, covering the trend, methodologies, and their clinical potential. [10].

Within the context of applying deep learning techniques for kidney disease imaging, various research works have demonstrated their potential applications. Tian et al. [11] developed a radiomics model using ultrasound imaging, integrating ResNet34 features and textural analysis for screening CKD stages with significantly better diagnostic performance than experienced experts. Zhang et al. [12] offered a systematic summary of deep learning applications of renal ultrasound and stated that CNN-based models improve kidney structure analysis and disease diagnosis with emphasis on problems of data quality and generalizability. Qin et al. [13] proved that the multimodal ultrasound deep learning models comprising grey-scale, superb microvascular imaging, and elastography outperformed single-mode models in classifying early CKD fibrosis, therefore suggesting the value of feature fusion in renal pathology prediction. Svrcek et al.[14] explored the integration of conventional computer vision measurements with CNNs for CKD classification. It was shown to exhibit an increase in specificity and negative predictive value compared to raw image-only models. An application of a deep ensemble model technique, combining ResNet, DenseNet, and EfficientNet on 2D shear wave elastography images, for the evaluation of renal fibrosis in CKD was demonstrated by Ye et al. [15]. Thereby, high precision with robust performance showed the power of the ensemble on complex renal image processing tasks. Table 1 presents a comparative summary of these deep learning models and their key findings in CKD imaging.

Table 1. Comparative Summary of DL Models in CKD Imaging

Study	Method	Key Findings
[11]	ResNet34 + Ultrasound Radiomics	Outperformed senior physicians in CKD stage diagnosis, with a high AUC for early CKD detection.
[12]	Review of CNN and multimodal DL	Highlights CNN strengths and challenges in kidney disease imaging, and future outlook.
[13]	Multimodal ultrasound DL fusion	Multimodal DL improved detection accuracy over single-mode approaches.
[14]	DL combined with feature extraction	Improved CKD classification with conventional measurements and CNN integration.

Though there is much promise in diagnosing kidney diseases by means of deep learning techniques, there are certain challenges involved. For instance, many of these studies suffer from small or unbalanced datasets, a lack of standardisation in ultrasound imaging, and unimodality of inputs, among others. Multimodal fusion techniques have not yet been fully explored, and ensemble techniques have not yet been fully implemented in integrating imaging and clinical features simultaneously. This paper aims to improve these deficiencies by means of utilising a DenseNet-based CNN in extracting features from ultrasound images, as well as ensemble learning in dealing with tabular clinical data, in order to improve its accuracy, applicability, and usefulness in dealing with diverse CKD patient populations.

III. METHODOLOGY

This methodology is based on the development of a multimodal framework for predicting the stages of chronic kidney disease based on kidney ultrasound images and clinical characteristics. The methodology is based on the integration of convolutional neural networks based on DenseNet architecture for feature extraction from medical images and ensemble learning approaches to improve the robustness of the classification models. Data processing and data augmentation based on normalisation approaches are used to improve the generative capacity of the models, and the training processes are carried out to optimise the parameters of

10.32622/ijrat.1402202603

the models for precise prediction. The methodology was quite methodical in the development and integration of models to predict the stages of chronic kidney disease and was assessed based on various approaches to ascertain the credibility of the developed methodology. Fig. 1 shows the overall architecture of the proposed multimodal CKD prediction framework, illustrating the integration of ultrasound image processing, DenseNet-based feature extraction, and ensemble learning for classification.

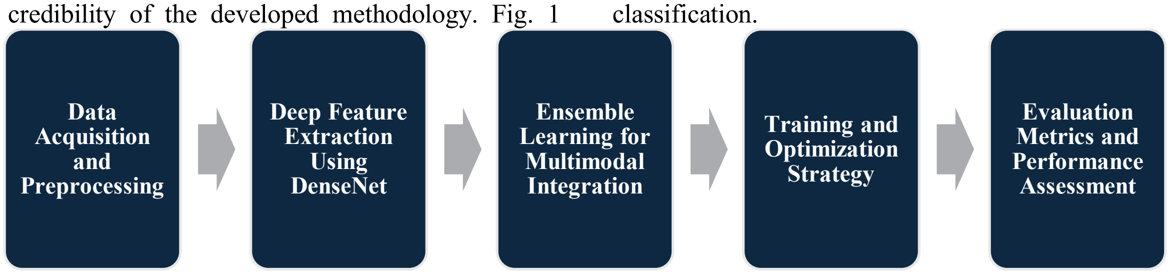

Fig. 1. Multimodal CKD Prediction Framework.

A. Data Acquisition and Preprocessing

This work uses a dataset of medical images comprising four classes: Normal, Cyst, Stone, and Tumour, which describe kidney ultrasound images. Images are preprocessed by resizing, normalisation, and noise reduction to enhance feature extraction performance. In this work, various forms of data augmentation have been applied, such as image rotation, flipping, and scaling, in order to avoid the problem of class imbalance and enhance model generalisation. The current study split its data into training and validation data so that any model assessment would not be biased. Training data was used for the optimisation of the DenseNet parameters. Feature standardisation across the images ensures that all images are distributed with the same intensity. This study prepares an input tensor for the DenseNet CNN that represents the hierarchical features to be used for further classification. Therefore, this work has integrated preprocessing, augmentation, and normalisation systematically to make sure that input images are in the appropriate format for deep learning training; this enhances predictive accuracy and robustness against diverse conditions of the renal organs.

B. Deep Feature Extraction Using DenseNet

This work utilizes a type of convolutional neural network known as DenseNet for the extraction of imaging features at a high level from kidney ultrasound data. DenseNet connects every layer to every other layer in a feed-forward manner, hence enabling flowing gradients efficiently and reusing features across layers. The investigation makes use of several convolutional layers with the addition of batch normalization and ReLU activation in the hope of capturing the texture and structural patterns from the renal images. During the training of the network, the investigation tries to minimize the loss function, which is categorical cross-entropy, in the hope of achieving the optimization of accuracy in classification. Representation of feature extraction theoretically:

Equation 1 – Convolution Operation:

$$F_l = \sigma(W_l * F_{l-1} + b_l) \qquad (1)$$

Where F_l is the feature map at the layer l,W_l are the convolutional weights, b_l is the bias, and σ represent the ReLU activation function.

Equation 2 – Dense Connectivity:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \qquad (2)$$

Where x_l is the output of the layer l and H_l defines the composite operations of convolution, normalisation, and activation on the concatenated feature maps of previous layers. This configuration enables the study to retain fine-grained spatial information while allowing for deeper network training.

C. Ensemble Learning for Multimodal Integration

It is based on this work that combines tabular data with the imaging features to better predict the CKD stages. In the concatenation operation, the DenseNet features are concatenated with the various attributes of the patients to form a multimodal representation. For this purpose, to make accurate predictions, a form of ensemble learning could be employed. This could be Gradient Boosting or Random Forest, perhaps to avoid overfitting. The different outputs of multiple base learners are combined using the ensemble method, each giving a different perspective on the combined data. Training is performed based on the task by minimizing the mean squared error or cross-entropy.

10.32622/ijrat.1402202603

Equation 3 – Weighted Voting in Ensemble:

$$\hat{y} = \sum_{i=1}^{N} w_i y_i \qquad (3)$$

Where $\hat{y}$ is the ensemble prediction, y_i is the prediction from the ith model, and w_i represents its weight.

Equation 4 – Loss Function for Ensemble Training:

$$L = -\sum_{c=1}^{C} y_c \log(p_c) \qquad (4)$$

Where y_c is the true label, p_c is the predicted probability for class c and C is the number of classes.

This combination ensures that the study utilises complementary information from both images and clinical features for the robust classification of CKD.

D. Training and Optimization Strategy

The proposed methodology involves an iterative training strategy for optimizing the weights of the DenseNet classifiers and models. The proposed solution involves mini-batch gradient descent along with adaptive learning rate scheduling. Techniques such as dropout and weight decay are involved in regularisation. The proposed solution stops overfitting of training data. There is an unbiased adjustment of gradients along with early stopping techniques based on validation loss values. The proposed solution involves performance metrics such as accuracy, precision, and recall values on the validation dataset. The proposed solution tunes the hyperparameters such as the learning rate, mini-batch sizes, and number of layers of the DenseNet models. This allows the solution to solely concentrate on generalised feature extraction and robustness across different kidney types. This is because there is smooth operation of both image and table data for multimodal training.

E. Evaluation Metrics and Performance Assessment

The performance of the models is assessed based on a range of metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve. Confusion matrices are drawn on the performance of the individual classes of normal, cyst, stone, and tumor. The time and cost of inference, computational complexity, and use of graphics processing units are also assessed to check the feasibility of use. Cross-validation is performed in order to validate the generality of the performance metrics. Performance curves based on the evolution of loss and improvement of accuracy are plotted during the training phase. The above criteria, which involve statistical, computational, and clinical performance, ensure that the reliability and interpretability of the model in predicting the stages of CKD using multi-modal data are properly validated.

IV. RESULTS

The result has illustrated the efficiency of the DenseNet-based multimodal system in properly classifying kidney images from an ultrasound procedure into four classes, viz. Normal, Cyst, Stone, and Tumour. This has been measured in terms of accuracy for both training and validation phases, conducted for various numbers of epochs. This has also focused on class-specific precision, recall, and F1-measure. This experiment has also identified the efficiency of renal image processing through deep feature extraction, helping in the classification and prediction of chronic kidney disease staging.

Table 2. Training–Validation Accuracy Across Epochs

Epoch	Train Accuracy	Validation Accuracy
1	77.7%	91.9%
2	88.3%	93.7%
3	90.8%	96.5%
4	92.9%	97.1%
5	94.4%	98.2%
6	94.4%	98.2%
7	95.5%	97.8%
8	95.6%	97.5%
9	96.4%	98.7%
10	96.6%	98.5%

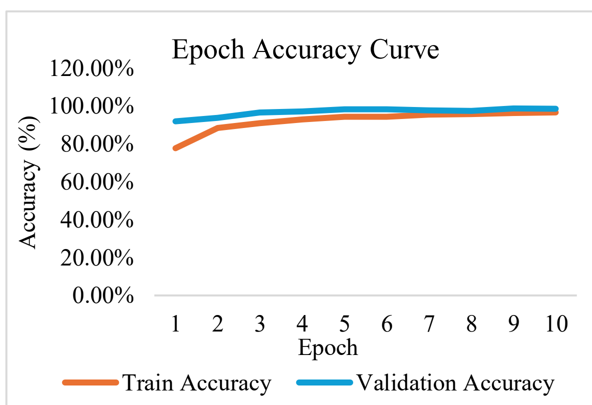

Fig. 2. Epoch Accuracy Curve

Table 2 and Fig. 2 show, in a comparative manner, the progressive improvement in the results obtained from the DenseNet-based model after 10 training epochs: the

10.32622/ijrat.1402202603

training accuracy starts from 77.7% and gradually increases to 96.6% on the last epoch, which signifies that there is good learning of image features. On the other hand, the validation accuracy starts higher at 91.9%, signifying that the model generalises well even in early epochs. Furthermore, the highest validation accuracy is at 98.7%, as is clear from the low overfitting. Also, the stagnation region in the above graph at points 5 and 6 indicates convergence. Thus, based on the convergence, the following results can be considered constant. All the above phenomena reflect the training process's validity, the effectiveness of the preprocessing and augmentation techniques, and lastly, the learning capacity of the model regarding the existing patterns in the renal ultrasound images.

Table 3. Performance Metrics for Kidney Ultrasound Classes

Class	Precision	Recall	F1-Score	Support
Cyst	0.99	0.99	0.99	1118
Normal	1.00	0.99	1.00	1534
Stone	0.96	0.99	0.97	428
Tumor	1.00	1.00	1.00	654
Overall	0.99	0.99	0.99	3734

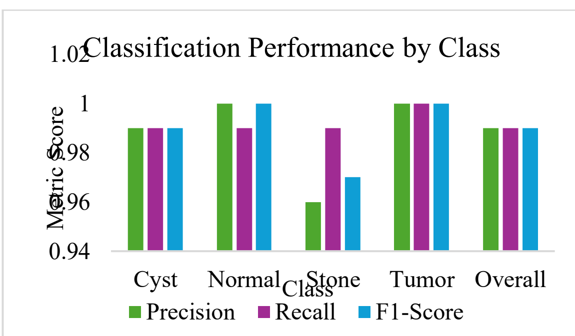

Fig. 3. Classification Performance by Class

The performance analysis for all four classes, which are Normal, Cyst, Stone, and Tumor, has been provided through a table and a graph in Table 3 and Fig. 3, respectively, through the summarisation of the precision, recall, F1 score, and support for the respective classes. The performance analysis for the classes such as Normal, Cyst, and Tumor has come out to be almost perfect, with all the scores being almost equal to 0.99, thus ensuring the efficiency of the predictions made and hence the efficiency in the performance for the differentiation between the respective classes. The class Stone has a little lower precision value, 0.96, whereas its recall remains at a very healthy value, 0.99, thus pointing to minute mispredictions that do not hamper overall dependability. Taken altogether, the model results in an overall accuracy of 0.99, precision of 0.99, recall of 0.99, and F1 score of 0.99 over 3,734 samples, establishing that the step of feature extraction using DenseNet has efficiently identified key differences in renal images through ultrasound, and that the multi-modal fusion step has ensured the soundness of the predictions for all four classes.

V. DISCUSSION

This work demonstrates the relevance and effectiveness of applying DenseNet to the multimodal framework in efficiently predicting the stages of chronic kidney diseases (CKD) through Kidney Ultrasound Images and Clinical Data. The significant results obtained identify the high level of accuracy achieved in classifications pertaining to Normal, Cyst, Stone, and Tumor classes with very few errors and equal efficiencies in all classes. The results clearly indicate the importance of efficiently achieving classifications pertaining to renal diseases based on the efficient extraction of deep features from images, which typically require minute details in image structures and textures to achieve classifications differentiating renal pathologies. Also significant is the complementarity achieved through the integration of the tabulated clinical data in images, to achieve robustness through the presence of flaws in both the sources of data, to justify the efficiency achieved in classifications through efficient application in healthcare settings, which may typically require variation in image quality and patient characteristics, posing challenges to classifications through original methodologies.

A comparison of a similar deep learning study suggests that the current framework outperforms in terms of predictive accuracy, particularly in classes that generally exhibit overlapping properties in terms of classification, such as classes Stone and Cyst. This data also suggests that the integration of a DenseNet function in feature extraction and multimodal integration not only works to increase accuracy in the prediction system but also increases its efficacy in terms of interpretability. This study suggests a great potential in terms of using such a predictive system in a variety of diagnostic models, such that decisions can be quickly and efficiently made within a predictive module. Future studies, therefore, must investigate in terms of size and diversity of data as well as theability to integrate such predictive models in a real-time predictive module.

10.32622/ijrat.1402202603

VI. CONCLUSION

This study introduced a multimodal framework for the prediction of the stages of chronic kidney disease (CKD) using the combination of kidney ultrasound images and clinical information through the application of the DenseNet-based deep learning approach and the ensemble learning strategy. The evaluation results of the proposed framework using the dataset of 3,734 samples showed promising classification results, obtaining an accuracy of 99%, precision, recall, and F1-score of 0.99 for the Normal, Cyst, Stone, and Tumor classes. The results showed that the proposed framework has consistent prediction results, except for the precision value of the Stone class, which was slightly lower at 0.96. The results showed that multimodal learning plays an important role in improving the reliability of the diagnosis of the stages of chronic kidney disease compared to the application of single-modality approaches. Although the study has some limitations, such as the application of the single-center approach and the variability of the ultrasound images, the proposed framework can provide an effective solution for the automated prediction of the stages of chronic kidney disease.

REFERENCES

1. Rao I.R., Bangera A., Nagaraju S.P., et al.: Chronic kidney disease of unknown aetiology: A comprehensive review of a global public health problem. Tropical Medicine & International Health (2023). 10.1111/tmi.13913
2. Hossain M.N., Bhuiyan E., Miah M.B.A., Sifat T.A., Muhammad Z., Al Masud M.F.: Detection and classification of kidney disease from CT images: an automated deep learning approach. Technologies 13(11) (2025). 10.3390/technologies13110508
3. Jain A., Dhasal P., Bhardwaj S.M.: Multimodal deep learning system for early detection of chronic diseases using medical images and EHR data. Vascular and Endovascular Review 8, 128–136 (2025). 10.64149/J.Ver.8.18s.128-136
4. Hamza A., Mezl M.: Deep learning-enhanced ultrasound analysis: classifying breast tumors using segmentation and feature extraction. IEEE Access 13, 83528–83541 (2025). 10.1109/ACCESS.2025.3568588
5. Iftikhar H., Hashem A.F., Qureshi M., Rodrigues P.C.: Clinical application of machine learning models for early-stage chronic kidney disease detection. Diagnostics 15(20) (2025). 10.3390/diagnostics15202610
6. Xie W., Wu G., Qi X., et al.: An innovative approach for predicting prostate cancer Gleason grading: machine learning-based fusion of multimodal ultrasound, clinical, and laboratory indicators. European Journal of Medical Research 30(1), 1051 (2025). 10.1186/s40001-025-03426-1
7. Chen H., Huang Y., Chen L.: Ensemble machine learning for predicting renal function decline in chronic kidney disease: development and external validation. Frontiers in Medicine 12, 1598065 (2025). 10.3389/fmed.2025.1598065
8. Zhao D., Wang W., Tang T., Zhang Y.Y., Yu C.: Current progress in artificial intelligence-assisted medical image analysis for chronic kidney disease: a literature review. Computational and Structural Biotechnology Journal 21, 3315–3326 (2023). 10.1016/j.csbj.2023.05.029
9. Chowdari D.L., Reddy K.V.K., Shoba L.K., Pongiannan R.K.: Optimizing chronic kidney disease detection through multi-modal data fusion using CNN–MLP architectures. In: Proceedings of the 8th International Conference on Inventive Computation Technologies (ICICT), pp. 1240–1247 (2025). 10.1109/ICICT64420.2025.11005275
10. Wang Z., Zhang X., Wang X., et al.: Deep learning techniques for imaging diagnosis of renal cell carcinoma: current and emerging trends. Frontiers in Oncology (2023). 10.3389/fonc.2023.1152622
11. Tian S., Yu Y, Shi K, et al.: Deep learning radiomics based on ultrasound images for the assisted diagnosis of chronic kidney disease. Nephrology 29(11), 748–757 (2024). 10.1111/nep.14376
12. Zhang M., Ye Z, Yuan E, et al.: Imaging-based deep learning in kidney diseases: recent progress and future prospects. Insights into Imaging (2024). 10.1186/s13244-024-01636-5
13. Qin X., Liu X., Xia L., Luo Q., Zhang C.: Multimodal ultrasound deep learning to detect fibrosis in early chronic kidney disease. Renal Failure 46(2), 2417740 (2024). 10.1080/0886022X.2024.2417740
14. Svrcek P.T., Jang J., Ge C., Lee H., Kim Y.H.: Combined application of deep learning and conventional computer vision for kidney ultrasound image classification in chronic kidney disease: preliminary study. Ultrasonography 44(5), 346–353 (2025). 10.14366/usg.25074
15. Ye D., Ou Z., Ye F, et al.: Ensemble deep learning model for accurate assessment of renal fibrosis in chronic kidney disease using two-dimensional shear wave elastography images. BMC Medical Imaging 25(1) (2025). 10.1186/s12880-025-01964-y

10.32622/ijrat.1402202603