International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 4, August 2025, pp. 4148~4159
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i4.pp4148-4159

Explainable artificial intelligence and feature based technique for the classification of kidney ultrasound images

Fizhan Kausar, Ramamurthy Bojan
Department of Computer Science, School of Sciences, CHRIST (Deemed to be University), Bangalore, India

Article Info

Article history:

Received Aug 16, 2024
Revised Apr 22, 2025
Accepted May 24, 2025

Keywords:

Explainable AI
Kidney ultrasound (US) images
Machine learning
SHAP
Texture feature

ABSTRACT

Millions of people worldwide are affected by chronic kidney disease (CKD), which is one of the main causes of death. Using machine learning (ML) models, this study attempts to create a computer-aided diagnostic (CAD) system that can autonomously detect chronic kidney disease (CKD) with improved interpretability. An online medical database provided 340 ultrasound images used in this study, which included both normal and abnormal instances. 94 texture and intensity attributes were obtained from these images using Pyrandiomics. Six machine learning methods were used for classification: According to the evaluation results, support vector machine (SVM), decision tree (DT), random forest (RF), k-nearest neighbors (k-NN), XG-Boost, and naïve Bayes (NB) models were considered. Among these models, the random forest model demonstrated the highest accuracy. Explainable artificial intelligence (XAI) methods, namely Shapley additive explanation (SHAP), were utilized to improve model transparency. Clinicians could be assisted in comprehending the reasoning behind the predictions using SHAP analysis, which identifies the most important features impacting the ML model and visualizes the ranking of each individual feature.

Corresponding Author:

Fizhan Kausar
Department of Computer Science, School of Sciences, CHRIST (Deemed to be University)
Bangalore, India
Email: fizhan.kausar@res.christuniversity.in

1. INTRODUCTION

Chronic kidney disease (CKD) poses a significant global health threat, especially in low-income countries, due to its increasing prevalence and high mortality rates. Chronic diseases accounted for 60% of global deaths in 2005, increasing to 66.7% by 2020 (WHO). Early prediction and monitoring are essential to curtail CKD progression and prevent severe complications. CKD, marked by prolonged kidney damage or reduced function, incurs high medical costs and increases risks of stroke, heart disease, diabetes, and infections. Despite its widespread impact, CKD remains less recognized compared to other chronic illnesses [1].

Thus, this study considers CKD early prediction and monitoring focusing on pediatrics kidney ultrasound images. Ultrasound (US) is favored for identifying vascular irregularities due to its noninvasive approach and lack of ionizing radiation. Specifically, two-dimensional ultrasound (2-D US) is standard for measuring kidney dimensions and morphology of the kidney, despite challenges posed by image quality variations among low-cost and conventional machines [2].

The texture features in US scan images are used to describe the structural characteristics of tissues. These images can detect subtle structural abnormalities, such as cysts, scarring, and changes in tissue texture, which may indicate the early stages of CKD [3]. Several techniques are used to extract textural information from digital images, with the aim of capturing the structures, statistics, and model-based connections between image pixels. Only a few have compared the effectiveness of various known features using ultrasound dataset obtained from clinical settings. This research seeks to address this gap by evaluating the impact of certain US image characteristics on a pediatric dataset of kidney US images, with the objective of this study being to assess and evaluate the efficacy of machine learning techniques for identifying ultrasound images according to their texture and statistical features.

A study presented a decision support system that utilized texture features such as Haralick, shape, wavelet, Tamira, and histogram of oriented gradients (HOG), along with machine learning (ML) and convolutional neural network (CNN) approaches. The system employed a decision tree classifier to achieve an F1 score of 85.3% [4]. In a research study [5] aimed at improving image texture before segmentation, texture analysis was performed using the gray-level co-occurrence matrix (GLCM) and intensity histogram (IH). The results indicated that GLCM parameters were the most essential for analyzing texture in kidney ultrasound images.

In a separate study [6], a set of parameters was extracted from the regions of interest (ROI) of ultrasound images to diagnose chronic kidney disease. The GLCM and kidney size were used for diagnosis, and an artificial neural network (ANN) classifier yielded a result of 95.4%. In the following work [7], GLCM features with a fourteen-feature vector were extracted, and a linear discriminant analysis classifier was utilized for classification of different stages of CKD and could classify with 24% of higher accuracy when compared to the existing prosed work. In the following study [8] deep learning features and conventional imaging feature from US images were extracted, the conventional imaging feature included texture features obtained using Gabor filters, HOG features and geometrical features. Support vector machine (SVM) classifier was used to build on these features with AUC values of 0.92, 0.88, and 0.92 for left, right, and bilateral abnormal kidney scans, respectively and resulted in improved performance while combining traditional imaging data with deep learning-based features. A novel method for CKD screening using ultrasound images and a convolutional neural network (CNN) framework, named the texture branch network (TBN) [1]] was introduced, which integrates texture feature extraction methods such as GLCM and HOG into CNN which enhanced the models ability to classify CKD accurately.

The review above indicates that numerous investigations have focused on identifying and categorizing kidney ultrasound images using machine learning and deep learning approaches. Factors such as dataset characteristics, data quality, and extracted features significantly impact model performance. Earlier studies have primarily concentrated on GLCM and other statistical features in ultrasound and these studies have not elucidated which features contribute most to model effectiveness due to the lack of interpretability and the "black-box" [4] nature of their algorithms. Comprehending this "black-box" is crucial as it aids nephrologists in understanding the model's internal workings and decision-making processes. Moreover, many researchers have not used explainable (XAI) technique to determine the significance of individual texture feature in the classification of accuracy of pediatric ultrasound images.

This research underscores the significance of gray-level texture and statistical features. These features capture spatial relationships between adjacent pixels and are comparatively straightforward to compute. This study uniquely examines a wide array of texture features, including gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level dependence matrix (GLDM), gray level size matrix (GLSZM), neighboring gray tone difference matrix (NGTDM), and first-order features. The proposed approach utilizes a broad spectrum of texture features alone in conjunction with efficient machine learning algorithms for image classification. Furthermore, it introduces the explainable artificial intelligence (XAI) algorithm Shapley additive explanation (SHAP) to determine the impact of each extracted gray-level feature in identifying the specific influence of extracted features (such as echogenicity patterns, texture, and size measurements) that contribute most significantly to the model's decision-making process and in classification of pediatric kidney ultrasound images.

The novel contribution of the proposed methodology:

a. Examining texture characteristics in ultrasound images of pediatric kidneys to identify abnormalities, showcasing how each feature contributes to classification accuracy
b. Incorporating XAI to reveal feature importance, enabling a clearer comprehension of how specific texture attributes affect classification precision.
c. Testing the proposed method using six machine learning algorithms to determine their efficacy in categorizing pediatric kidney images.

The paper is organized as follows: section 2 describes our methodology and the model we propose. Section 3 examines the experimental outcomes of our model. Section 4 provides an in-depth discussion of the proposed model, comparing it with prior research. Lastly, section 5 concludes the paper and suggests potential avenues for future research.

2. MATERIALS AND METHODS

This section provides a concise overview of the proposed methodology for classifying pediatric kidney ultrasound images using a machine learning model. As illustrated in Figure 1, the process consists of six distinct stages. The initial phase involves data collection, followed by data preprocessing. The third stage focuses on extracting features from the data. Subsequently, the model is trained and evaluated using various metrics. The final step incorporates an XAI approach to interpret and elucidate the model's predictions.

Figure 1. Representing the proposed workflow diagram

2.1. Data description

This research work uses kidney ultrasound images from a dataset restricted to infants under seven years old. This dataset was created from a publicly accessible medical database [9]. The scope of kidney-related illnesses is broad, and a subset concentrating on anomalies in pediatric kidneys has been carefully selected. There are 340 images in this subgroup of children less than seven, representing both normal and abnormal instances

2.2. Data pre-processing

During the data preprocessing phase, we optimized the performance of the proposed model by executing a series of steps. Initially, all images were standardized to a resolution of 224×224 pixels and removed any artifacts by focusing on the kidney region of interest through cropping. We then eliminated noise using technique, Gaussian noise with a median filter, the filter smoothened the image by averaging the pixel values with a gaussian and replaced each pixel value with the median of the neighboring pixel values, that effectively reduced the high-frequency noise by preserving the edges while removing noise and hence achieved optimal results. Contrast limited adaptive histogram equalization (CLAHE) [10] was employed to enhance the image quality, the adaptability ensured the region with even small and subtle details were enhanced. while the Sobel operator [11] was used for edge detection to accurately identify the kidney boundaries while enhancing the visibility of structure and boundaries with the ultrasound images. We employed the automated data augmentation approach known as synthetic minority oversampling technique (SMOTE), which was applied from the majority class to construct a new synthetic sample to increase the size of our dataset and avoid overfitting. The sample of the preprocessed image is shown in Figure 2.

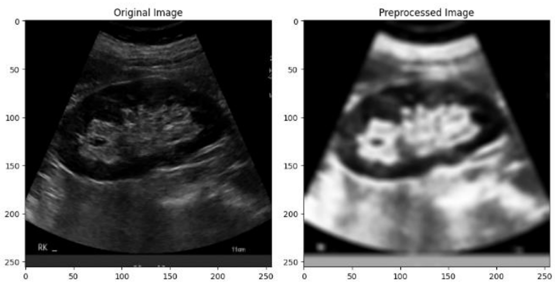

Figure 2. Sample of pediatric kidney ultrasound image

2.3. Feature extraction

The effectiveness of our proposed model relies heavily on image feature extraction, with the primary aim of this research being to identify and extract the most significant feature vectors. A valuable and efficient method for obtaining radiological features from medical images is PyRadiomics [12], which encompasses three categories: intensity-based, morphological, and textural features. Intensity-based (first-order) features utilize histograms to represent intensity distribution in ROIs without considering spatial relationships. Morphological traits describe the geometric characteristics of the area. Textural (second-order) features employ structures like GLCM, GLSZM, GLRLM, NGTDM, and GLDM to characterize the spatial arrangement of intensity levels based on gray levels. The PyRadiomics package provided a list of selected feature vectors. PyRadiomics is an open-source Python package designed for the processing and extraction of radiomics features from medical image data, utilizing a wide array of automated data characterization algorithms. This package was used to extract texture features from pediatric ultrasound images, with radiomics features being extracted for both normal and abnormal images. Figure 3 shows the texture feature extracted from the sample pediatric kidney image and Table 1 provides a detailed description of each texture feature type and its sub-features.

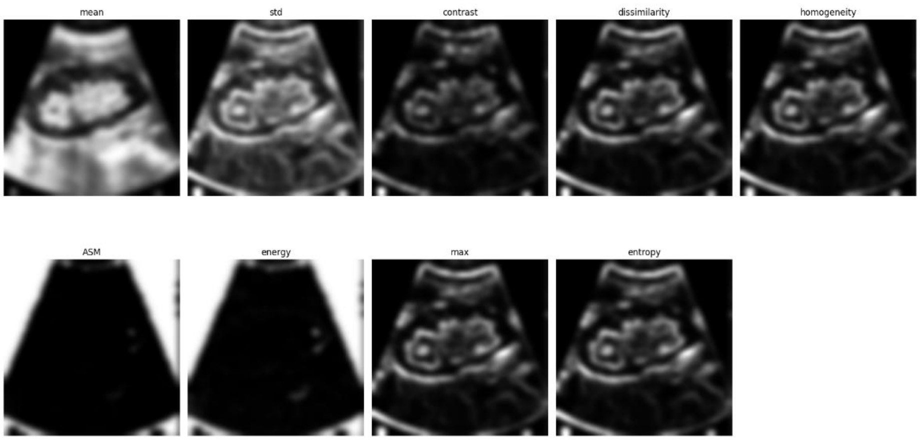

Figure 3. Visualization of the features extracted pediatric Kidney ultrasound image

Table 1. List of feature vectors extracted from the feature type

Feature type	Feature
First Order: Based solely on the values of each pixel, without considering spatial relationships.	Energy, total energy, entropy, minimum, 10th percentile, 90th percentile, maximum, mean, median, inter quartile range, range, mean absolute deviation, robust mean absolute deviation, root mean squared, standard deviation, skewness, kurtosis, variance, uniformity.
GLCM: It calculates the frequency with which pairs of pixels in a picture with given values and spatial relationships occur.	Auto-correlation, joint average, cluster prominence, cluster shade, contrast, correlation, difference average, difference entropy, difference variance, joint energy, joint entropy, informational measure of correlation (IMC)1, informational measure of correlation (IMC)2, inverse difference (ID), inverse difference normalized (IDN), inverse variance, maximum probability, sum average, sum entropy, sum of squares.
GLSZM: Counts the number of interconnected zones or areas in an image of specified size and gray scale	Small area emphasis (SAE), large area emphasis (LAE), gray level non-uniformity (GLN), gray level non-uniformity normalized (GLNN), size-zone non-uniformity (SZN), size-zone non-uniformity normalized (SZNN), zone percentage (ZP), gray level variance (GLV), zone variance (ZV), zone entropy (ZE), low gray level zone emphasis (LGLZE), high gray level zone emphasis (HGLZE), small area low gray level emphasis (SALGLE), small area high gray level emphasis (SAHGLE), large area low gray level emphasis (LALGLE), large area high gray level emphasis (LAHGLE).
GLRLM: Measures the distance between consecutive groups of pixels that have the same grayscale.	Short run emphasis (SRE), long run emphasis (LRE), gray level non-uniformity (GLN), gray level non-uniformity normalized (GLNN), run length non-uniformity (RLN), run length non-uniformity normalized (RLNN), run percentage (RP), gray level variance (GLV), run variance (RV), run entropy (RE), low gray level run emphasis (LGLRE), high gray level run emphasis (HGLRE), short run low gray level emphasis (SRLGLE), short run high gray level emphasis (SRHGLE), long run low gray level emphasis (LRLGLE), long run high gray level emphasis (LRHGLE).
GLDM: Calculates the number of pixels at a specific distance that are influenced by a particular shade of gray.	Small dependence emphasis (SDE), large dependence emphasis (LDE), gray level non-uniformity (GLN), dependence non-uniformity (DN), dependence non-uniformity normalized (DNN), gray level variance (GLV), dependence variance (DV), dependence entropy (DE), low gray level emphasis (LGLE), high gray level emphasis (HGLE), small dependence low gray level emphasis (SDLGLE), small dependence high gray level emphasis (SDHGLE), large dependence low gray level emphasis (LDLGLE), large dependence high gray level emphasis (LDHGLE).
NGTDM: Determines the extent to which a pixel differs from the average gray tone of its neighbors within a specific radius.	Busyness, coarseness, complexity, contrast, strength.

2.4. Machine learning classifiers

After completing the feature extraction stage, the machine-learning model was trained using the extracted features to determine the algorithm that would provide better accuracy. In this study, we utilized some of the most widely used and reliable classifiers owing to their dependability: support vector machine (SVM), random forest (RF), k-nearest neighbor (KNN), XG-boost decision tree (DT), and XG-Boost, all of which are machine learning algorithms designed for classification. Prior research mentioned in this paper selected only a small number of texture features for the analysis of kidney images. Additionally, we incorporated first-order statistics into our analysis to evaluate the performance of both texture and statistical features using classifiers preferred in previous studies.

2.4.1. Support vector machine (SVM)

SVM have been extensively applied in areas such as cheminformatics, bioinformatics, and biometrics for tasks such as classification and regression. They rely on support vectors for binary classification and build models using training data, ensuring high precision using hyperplanes for linearly separable data. For nonlinear data, the SVM maps to higher-dimensional spaces using kernel functions, necessitating careful parameter selection [13].

2.4.2. Naive Bayes (NB)

NB is a straightforward yet effective algorithm that utilizes Bayes' theorem, assuming that features are independent. This approach is particularly advantageous in situations with limited training data and can sometimes surpass more sophisticated models like logistic regression in terms of how quickly they converge. Its efficiency and minimal computational demands make it a favoured choice across various disciplines [14].

2.4.3. Decision tree (DT)

DT is an essential classification method that operates by repeatedly splitting data according to particular feature values. This hierarchical arrangement enables the model to make choices by navigating a tree-like structure, where each node signifies a feature condition. As a crucial element of ensemble techniques such as RF, DT provide both interpretability and precision when handling intricate data sets [15].

2.4.4. Random forest (RF)

RF is a robust ensemble learning technique that improves prediction accuracy by integrating several DTs, each trained on different portions of the dataset. By consolidating the outcomes of individual trees through majority voting, [16], it effectively minimizes the likelihood of overfitting. This approach is particularly advantageous for tasks like medical image analysis, where precise disease classification is essential, especially when dealing with extensive and intricate datasets [17].

2.4.5. K-nearest neighbor (KNN)

KNN is a straightforward yet powerful classification technique that categorizes new data points by considering the predominant class among their closest neighbors in the training dataset. As a non-parametric approach, it does not rely on any assumptions regarding the data's distribution. The performance and precision of KNN are influenced by factors such as the value of k and the selected distance metric, which determine how the similarity between instances is assessed [18].

2.4.6. XG-Boost

XGBoost is a robust algorithm for gradient-boosted trees, specifically crafted for high efficiency and performance in managing structured data. It constructs models in a sequential manner, with each subsequent tree aiming to rectify the mistakes of its predecessors, thus enhancing the overall predictive accuracy. By optimizing a regularized objective function, XGBoost adeptly balances the complexity and performance of the model, making it particularly well-suited for large-scale classification tasks [19].

2.5. Explainable artificial intelligence (XAI)

AI addresses issues with ethics, justice, privacy, bias, and transparency by making AI decisions easier for humans to understand. XAI is a technique that is used to decipher AI systems [20]. XAI seeks to define AI behaviors, strengths, limitations, and future behavior through better decision making, time savings, and increased trust in the medical sector. XAI provides developers with choices to strike a compromise between explainability and speed. Expert systems must explain how they work and why they are necessary to guarantee that people comprehend their applications and designs correctly. Shapley additive explanation (SHAP) and local interpretable model-agnostic explanation (LIME) are the two most well-known approaches in XAI [21].

The SHAP method, which is a transparent machine learning strategy, demonstrates the relevance of input variables in making predictions. This helps discern useful prediction techniques, such as categorizing peptides or images. SHAP combines local and global interpretations to assign a value to each feature based on its contribution to the predictions [22]. Although it may occasionally produce contradictory explanations, SHAP's solid theoretical foundation in game theory guarantees unbiased forecasts. It improves comprehension by working in harmony with other techniques such as LIME [23]. TreeSHAP is used for conditional predictions because it can be time-consuming for large samples and may disregard feature dependency.

By assessing the influence of extracted features on predictions, SHAP ranks of features in machine learning models, improving interpretability by highlighting important contributors of each feature to model outputs. SHAP is adaptable because, unlike standard ranking techniques, it can handle non-linear linkages and it reveals which combination of features influencing the prediction and ensures a fairer and more transparent evaluation and more balanced analyses of the model’s performance. In the end, using SHAP enhances accuracy, increases model refinement.

2.6. Performance metrics

Kidney abnormality images are classified as true positive (TP), true negative (TN), false positive (FP), or false negative (FN) based on diagnosis accuracy, crucial for evaluating classification models. Evaluation metrics include accuracy, sensitivity, specificity, precision, recall, F1 score, and AUC curve. Below, we provide a brief description of these metrics [24].

$$Accuracy = TP + TN/TP + TN + FP + FN \tag{1}$$

$$Sensitivity = TP/(TP + FN) \tag{2}$$

$$Specificty = TN/TN + FP \tag{3}$$

$$Recall = TP/TP + FN \tag{4}$$

$$F1 - SCORE = 2 * TP/2 * TP + FN + FP \tag{5}$$

$$AUC = TPR - FPR + 1/2 \tag{6}$$

3. EXPERIMENTAL RESULTS

The experiment was carried out on a Windows 10 system with an Intel i5 processor. The development environment used was Jupyter Notebook. This research proposes a ML approach for effectively categorizing pediatric kidney ultrasound images. The dataset, comprising both normal and abnormal pediatric kidney scans, was collected from online medical ultrasound image repositories. The primary objective was to classify these images based on texture characteristics. Image preprocessing involved applying a Gaussian median filter to enhance quality and reduce noise, utilizing CLAHE for image improvement, and implementing the Sobel operator for edge detection. Due to a notable imbalance between normal and abnormal images in the dataset, SMOTE was employed to generate synthetic images, creating a more suitable foundation for ML training. Following data stabilization, texture features were extracted, including first-order derivatives, GLCM, GLRLM, GLSZM, GLDM, and NGTDM. The extracted features were divided in a 70:30 ratio and input into various ML classifiers, including RF, DT, NB, SVM, KNN and XG-Boost. The classifiers' performance was assessed using a standard evaluation approach, incorporating metrics like precision, recall, accuracy, sensitivity, specificity, F1 score, and AUC curve. Table 2 displays the performance outcomes for each classifier. To comprehend the influence of individual features on the classification outcomes, XAI techniques were employed. The test results are elaborated on in the following section.

The RF classifier achieved the highest accuracy of 98% among the evaluated models. This superior performance can be attributed to its method of creating multiple trees using various data subsets and averaging their predictions, enabling it to rank feature importance. Other classifiers' accuracy ranged from 60% to 95%. The results table indicates that while random forest and XG-Boost excelled, SVM demonstrated a strong ability to differentiate between classes, making it more reliable in terms of AUC. NB showed poor performance, whereas KNN performed well across multiple thresholds for class distinction. DT required a balance between sensitivity and specificity. Figure 4 provides a graphical comparison of the classifiers. Furthermore, the Random Forest algorithm's feature importance ranking is displayed in Figure 5. In this graph, the x-axis represents individual feature values calculated based on entropy, while the y-axis shows the feature ranking. The top five most influential features contributing to the classification of pediatric ultrasound images, as ranked by RF, are $FirstOrder\_skewness$, $glcm\_IMC2$, $gldm\_lowgraylevelempasis$, $gldm\_smalldependencelowgraylevelempasis$, and $glszm\_smallareaempasis$.

Table 2. Results of classification of kidney images

Classifier	Accuracy	Sensitivity	Specificity	Precision	Recall	F-1 Score	AUC
Support vector machine	0.82	0.74	0.91	0.91	0.74	0.82	0.91
Naïve Bayes	0.75	0.46	0.56	0.71	0.46	0.56	0.78
Decision Tree	0.84	0.54	0.94	0.92	0.54	0.70	0.85
Random Forest	0.98	0.97	0.92	0.93	0.97	0.98	0.78
KNN	0.67	0.51	0.86	0.82	0.51	0.62	0.85
XG Boost	0.95	0.91	0.88	0.96	0.91	0.95	0.79

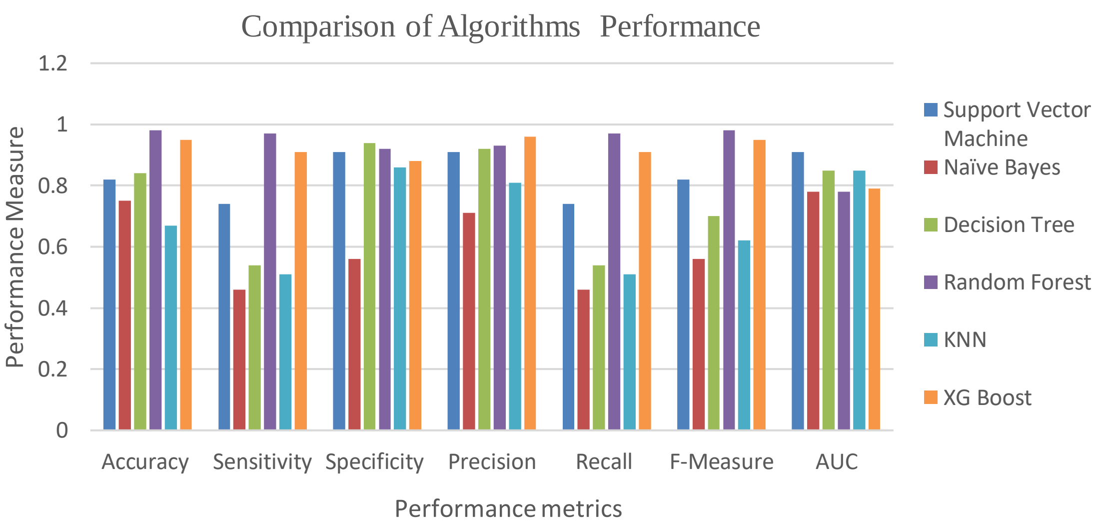

Figure 4. Performance comparison of ML algorithms

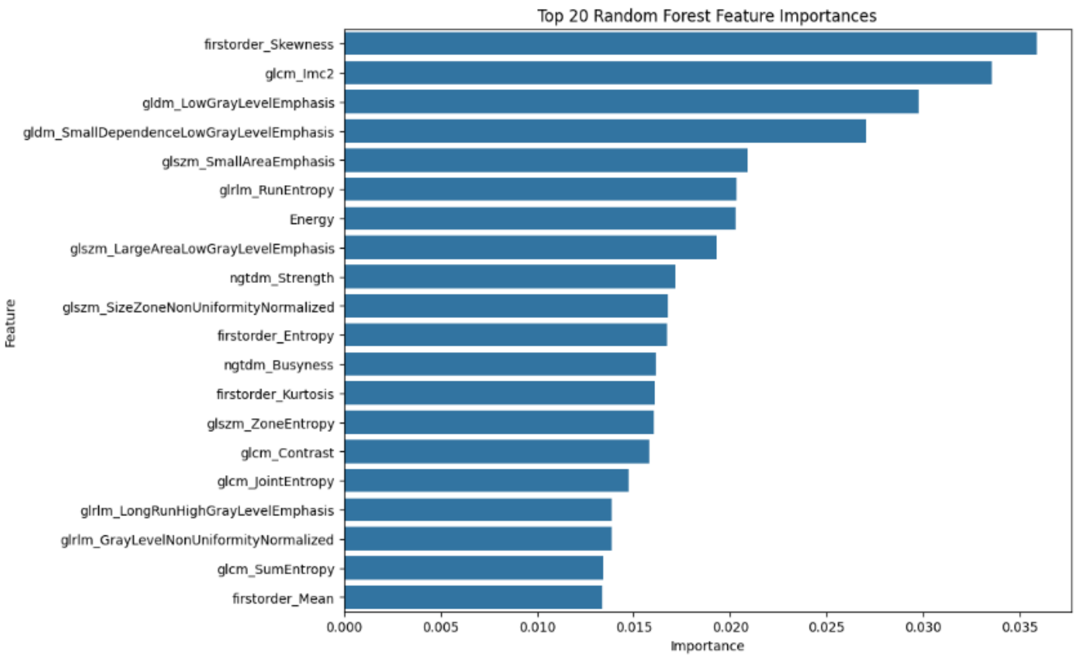

Figure 5. Random forest feature ranking on the extracted features form ultrasound images

Furthermore, our study utilizes SHAP to gain insight into our ML models prediction, Figure 6 presents the feature importance ranking with the SHAP summary plot for the Random Forest. The top three most important features that contribute to the predictive model are $energy$, $GLCM\_Imc2$, and $FirstOrder\_Skweness$. The graph presents the SHAP values on the x-axis and the rankings of the features on the y-axis. The features that have a higher SHAP value are considered to be more influential in detecting the classes based on texture features and are arranged in ascending order based on their SHAP scores, with the highest scores at the top of the figure. Figure 7 SHAP beeswarm graphic offers a detailed explanation of the impact of feature characteristics on the final outcome, including both global explanations and interpretations (comprehensive explanations).

This graph offers significant insights into the distinct roles of textual and first-order derivative features, enabling a comprehensive grasp of their influence on model predictions. The x-axis displays the SHAP values, with each line representing an individual feature. Blue dots signify lower feature values, while red dots represent higher ones. Each point on the graph corresponds to a single observation, and the feature's effect on the model's output is indicated by its position along the x-axis.

When comparing the random forest algorithm's feature importance results with those shown in Figure 3, we notice that they don't perfectly match the SHAP influence analysis results presented in Figure 4. However, upon closer examination, several features appear in both features ranking analyses, including $gldm\_smalldependencelowgraylevelempasis$, $glcm\_IMC2$, $glrlm\_runentropy$, $ngtdm\_strength$, $firstorder\_skweness$, $gldm\_lowgraylevelemphasis$, $glszm\_smallareaempasis$, $ngtdm\_busyness$, $energy$, $largearealowgratlevelemphasis$, $glszm\_sizezonenonuniformitynormalized$, and $glszm\_zoneentropy$. Although their rankings differ based on the scale used, we can infer that both approaches are relatively consistent in identifying the most influential features for analyzing ultrasound images. Additionally, we have compared our research to existing studies, as shown in Table 3. This comparison reveals that our method, which employs texture features for classification, has attained higher accuracy than previous investigations in this area. While prior studies primarily focused on GLCM or various image features and achieved good accuracy, our approach concentrates solely on different types of texture features, as kidney scans are primarily based on tissue characteristics

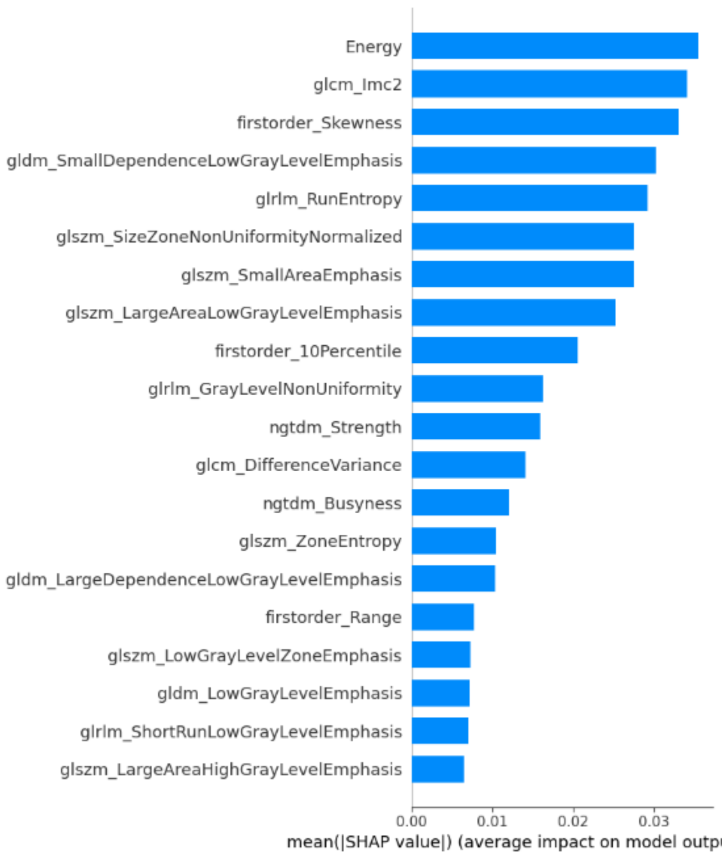

Figure 6. The SHAP algorithm's ranking of the importance of the features extracted from the US images

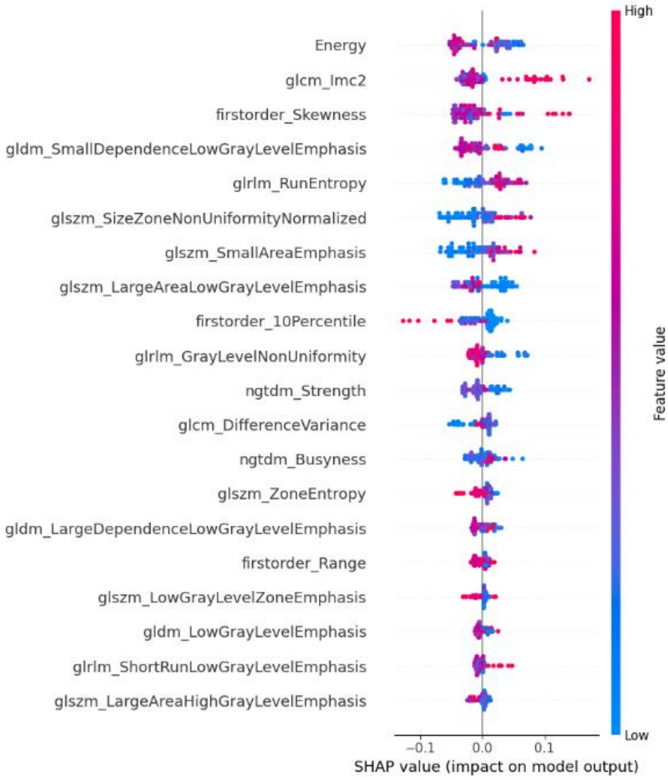

Figure 7. Beeswarm summary plot illustrates the influence of extracted feature on the performance of random forest using SHAP values

Table 3. Comparison of proposed model with existing work

Ref no	Year	Features Extracted	Total feature count	Accuracy	Classifier	XAI
[25]	2017	HOG, Haar as feature extraction USING SIFT, SURF	324	HOG- 96.67% Haar - 93.3% LBP -90.0%, SURF -89.3%	AdaBoost, SVM	-
[26]	2019	GLCM	21	84.61%	ANN	-
[27]	2020	GLCM	44	77.80%	ANN	-
[1]	2020	GLCM	16	88.48%	CNN	-
[7]	2021	GLCM	21	97.46	LDA	-
[6]	2021	GLCM	57	95.40%	ANN	-
[4]	2022	Haralick, Shape, Wavelet, Tamura, HOG	42	96.68%	k-NN, fuzzy k-NN, and SVM	
	2024	**Proposed- GLCM, GLRL, GLSZM, GLDM, NGTDM, First-Order**	**94**	**Highest accuracy Random Forest- 98%**	**SVM, NB, DT, XG-Boost, KNN, RF**	**SHAP**

4. DISCUSSION

CKD is a progressive health issue that often goes unnoticed until advanced stages, emphasizing the need for early detection to ensure effective treatment. This study explores the development of an interpretable ML model for CKD prediction using texture features extracted from pediatric ultrasound images. Our research addresses key challenges in ML-based medical diagnostics, including data imbalance, comprehensive texture analysis, and model interpretability.

The study involved the development and evaluation of six ML models. These models were trained and tested using a dataset derived from an online medical database, with 94 texture features extracted from the images. The primary objective was to classify kidney images based on these texture features. The RF model demonstrated superior performance compared to DT, NB, XG-BOOST, KNN, and SVM models. To address ongoing concerns about the opacity of ML model predictions, we incorporated explainability and interpretability features into the RF model using the SHAP technique. This approach provides a detailed analysis of specific texture features and their impact on image classification as normal or abnormal.

Previous research has employed various ML algorithms to classify kidney ultrasound images, typically using a limited number of texture features. There is a growing trend in exploring ML applications for CKD detection and classification. However, studies have not yet investigated external validation and explainability in the use of texture features alone for classifying pediatric kidney ultrasound images. Our research seeks to address this gap by examining the potential of machine learning algorithms in classifying chronic kidney disease. We employed SHAP to enhance the clarity and interpretability of our method, offering insights into the model's assumptions and highlighting the significance of particular features. Among the evaluated texture features, $gldm\_smalldependencelowgraylevelempasis$, $glcm\_IMC2$, $glrlm\_runentropy$, $ngtdm\_strength$, $firstorder\_skweness$, $gldm\_lowgraylevelemphasis$, $glszm\_smallareaempasis$, $ngtdm\_busyness$, $energy$, $largearealowgratlevelemphasis$, $glszm\_sizezonenonuniformitynormalized$, and $glszm\_zoneentropy$ emerged as the most influential for classification accuracy. These features have the potential to assist physicians in image classification. Future research should explore different combinations of texture features from GLCM, GLRLM, GLDM, NGTDM, and GLSZM, along with urine test reports, to provide more accurate and timely assessments of patient health, potentially leading to improved management of at-risk individuals.

5. CONCLUSION

This research introduces a sophisticated machine learning (ML) framework for categorizing pediatric kidney ultrasound images and determining the texture feature alone provide significant classification accuracy. The study employed various ML models, including RF, DT, KNN, NB, SVM and XG-Boost. RF emerged as the top performer, achieving 98% accuracy in distinguishing normal and abnormal pediatric ultrasound images. The investigation also utilized XAI technique like SHAP to provide insights into model behavior at both local and global levels. These methods offered coherent and logical explanations about $energy$, $GLCM\_Imc2$, and $FirstOrder\_Skewness$ texture features played a crucial role in determining classification outcome, thereby enhancing model transparency. The finding emphasizes the importance of tissue-level texture variation in diagnosing pediatric kidney abnormalities. The next phase of this research will focus on feature selection technique to identify the most influential features, thereby inproving classification accuracy and model efficiency. The insight gained from this study can contribute to the development of AI-assisted decision support system for early detection and diagnosis of kidney abnormalities in children.

FUNDING INFORMATION

Authors state no funding involved.

AUTHORS CONTRIBUTION

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Fizhan Kausar	✓	✓	✓	✓			✓	✓	✓					
Ramamurthy Bojan		✓			✓	✓				✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABLITY

The data that support the findings of this study will be available in *https://www.ultrasoundcases.info/*

REFERENCES

[1] P. yi Hao *et al.*, "Texture branch network for chronic kidney disease screening based on ultrasound images," *Frontiers of Information Technology and Electronic Engineering*, vol. 21, no. 8, pp. 1161–1170, Oct. 2020, doi: 10.1631/FITEE.1900210.

[2] S. Jeevitha and N. Priya, "Identifying and classifying an ovarian cyst using SCBOD (size and count-based ovarian detection) algorithm in ultrasound image," *International Journal of Electrical and Computer Engineering Systems*, vol. 13, no. 9, pp. 799–806, Dec. 2022, doi: 10.32985/ijeces.13.9.8.

[3] M. Bhandari, P. Yogarajah, M. S. Kavitha, and J. Condell, "Exploring the capabilities of a lightweight CNN model in accurately identifying renal abnormalities: Cysts, stones, and tumors, using LIME and SHAP," *Applied Sciences (Switzerland)*, vol. 13, no. 5, p. 3125, Feb. 2023, doi: 10.3390/app13053125.

[4] Sunanda Biradar, P. T. Akkasaligar, and S. Biradar, "Feature extraction and classification of digital kidney ultrasound images: A hybrid approach," *Pattern Recognition and Image Analysis*, vol. 32, no. 2, pp. 363–372, Jun. 2022, doi: 10.1134/S1054661822020043.

[5] N. A. Shaharuddin and W. M. H. W. Mahmud, "Feature analysis of kidney ultrasound image in four different ultrasound using gray level co-occurrence matrix (GLCM) and intensity histogram (IH)," *International Journal of Integrated Engineering*, vol. 10, no. 3, pp. 42–47, Dec. 2018, doi: 10.30880/ijie.2018.10.03.007.

[6] D. H. Kim and S. Y. Ye, "Classification of chronic kidney disease in sonography using the glcm and artificial neural network," *Diagnostics*, vol. 11, no. 5, p. 864, May 2021, doi: 10.3390/diagnostics11050864.

[7] R. Ahmad and B. K. Mohanty, "Chronic kidney disease stage identification using texture analysis of ultrasound images," *Biomedical Signal Processing and Control*, vol. 69, p. 102695, Aug. 2021, doi: 10.1016/j.bspc.2021.102695.

[8] Q. Zheng, S. L. Furth, G. E. Tasian, and Y. Fan, "Computer-aided diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data by integrating texture image features and deep transfer learning image features," *Journal of Pediatric Urology*, vol. 15, no. 1, pp. 75.e1-75.e7, Feb. 2019, doi: 10.1016/j.jpurol.2018.10.020.

[9] D. Joe, "A gallery of high-resolution, ultrasound, color doppler & 3D images," *Kidneys*, 2018. https://www.ultrasound-images.com/breast/ (accessed Jul. 18, 2024).

[10] P. R. Eveline, R. S. Vinod Kumar, and S. C. Ebbie, "FOE-NET: Segmentation of fetal in ultrasound images using V-NET," *International Journal of Electrical and Computer Engineering Systems*, vol. 14, no. 10, pp. 1141–1149, Dec. 2023, doi: 10.32985/ijeces.14.10.7.

[11] D. S. Korti, Z. Slimane, and K. Lakhdari, "Enhancing dynamic hand gesture recognition using feature concatenation via multi-input hybrid model," *International Journal of Electrical and Computer Engineering Systems*, vol. 14, no. 5, pp. 535–546, Jun. 2023, doi: 10.32985/ijeces.14.5.5.

[12] J. J. M. Van Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, pp. e104–e107, Oct. 2017, doi: 10.1158/0008-5472.CAN-17-0339.

[13] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, "Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review," *Expert Systems with Applications*, vol. 167, p. 114161, Apr. 2021, doi: 10.1016/j.eswa.2020.114161.

[14] V. R. Balaji, S. T. Suganthi, R. Rajadevi, V. Krishna Kumar, B. Saravana Balaji, and S. Pandiyan, "Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier," *Measurement: Journal of the International Measurement Confederation*, vol. 163, Oct. 2020, doi: 10.1016/j.measurement.2020.107922.

[15] K. Hema, K. Meena, and R. Pandian, "Analyze the impact of feature selection techniques in the early prediction of CKD," *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 66–77, 2024, doi: 10.1016/j.ijcce.2023.12.002.

[16] N. Amini and A. Shalbaf, "Automatic classification of severity of COVID-19 patients using texture feature and random forest based on computed tomography images," *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 102–110, Nov. 2022, doi: 10.1002/ima.22679.

[17] A. Subudhi, M. Dash, and S. Sabut, "Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 277–289, Jan. 2020, doi: 10.1016/j.bbe.2019.04.004.

[18] O. A. Jongbo, A. O. Adetunmbi, R. B. Ogunrinde, and B. Badeji-Ajisafe, "Development of an ensemble approach to chronic kidney disease diagnosis," *Scientific African*, vol. 8, p. e00456, Jul. 2020, doi: 10.1016/j.sciaf.2020.e00456.

[19] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," *Journal of Pathology Informatics*, vol. 14, p. 100189, 2023, doi: 10.1016/j.jpi.2023.100189.

[20] G. Dharmarathne, M. Bogahawaththa, M. McAfee, U. Rathnayake, and D. P. P. Meddage, "On the diagnosis of chronic kidney disease using a machine learning-based interface with explainable artificial intelligence," *Intelligent Systems with Applications*, vol. 22, p. 200397, Jun. 2024, doi: 10.1016/j.iswa.2024.200397.

[21] M. J. Raihan, M. A. M. Khan, S. H. Kee, and A. Al Nahid, "Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP," *Scientific Reports*, vol. 13, no. 1, Apr. 2023, doi: 10.1038/s41598-023-33525-0.

[22] P. R. Magesh, R. D. Myloth, and R. J. Tom, "An explainable machine learning model for early detection of Parkinson's Disease using LIME on DaTSCAN imagery," *Computers in Biology and Medicine*, vol. 126, p. 104041, Nov. 2020, doi: 10.1016/j.compbiomed.2020.104041.

[23] S. S Band *et al.*, "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods," *Informatics in Medicine Unlocked*, vol. 40, p. 101286, 2023, doi: 10.1016/j.imu.2023.101286.

[24] S. Janthakal and G. Hosalli, "An automatic feature extraction technique from the images of granular parakeratosis disease," *International Journal of Electrical and Computer Engineering Systems*, vol. 13, no. 8, pp. 611–619, Oct. 2022, doi: 10.32985/ijeces.13.8.1.

[25] R. Vasanthselvakumar, M. Balasubramaniam, and S. Palanivel, "Pattern analysis of kidney diseases for detection and classification using ultrasound B-mode images," *International Journal of Pure and Applied Mathematics*, vol. 117, no. 15, pp. 635–653, 2018.

[26] A. Nithya, A. Appathurai, N. Venkatadri, D. R. Ramji, and C. Anna Palagan, "Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images," *Measurement: Journal of the International Measurement Confederation*, vol. 149, p. 106952, Jan. 2020, doi: 10.1016/j.measurement.2019.106952.

[27] Priyanka and D. Kumar, "Feature extraction and selection of kidney ultrasound images using GLCM and PCA," *Procedia Computer Science*, vol. 167, pp. 1722–1731, 2020, doi: 10.1016/j.procs.2020.03.382.

BIOGRAPHIES OF AUTHORS

Fizhan Kausar received her MCA degree from Kristu Jayanti College, Bengaluru, India, in 2019, and she is currently pursuing her Ph.D. degree in Data Science from CHRIST (Deemed to be University), Bengaluru, as a full-time scholar. She has 3 years of teaching experience, her research interests are medical image processing using artificial intelligence, deep learning, machine learning, soft computing, genetics algorithms, computational intelligence. She is also an active member of the IEEE and ACM student chapter. She can be contacted at email: fizhan.kausar@res.christuniversity.in.

Ramamurthy Bojan is currently working as associate professor in the Department of Computer Science at CHRIST (Deemed to be University), Bangalore, Karnataka, India. He has completed his Ph.D. Programme in Computer Applications in the area of Image Processing at Anna University, Chennai. He did his M.Phil. Programme in the area of Genetic Algorithms for optimization problems at Bharathiar University, Coimbatore. He has more than 25 years of teaching experience and 20 years of research experience. Also, He has worked with different Optimization problems using genetic algorithms and also active of ISTE and Senior member of IEEE. He can be contacted at email: ramamurthy.b@christuniversity.in.