ORIGINAL ARTICLE

Intelligent Diagnosis in Trauma: Exploring Machine Learning and Radiomics for Kidney Injury Assessment

Hanieh Alimiri Dehbaghi [1], Karim Khoshgard [2*], Samira Jafari [3]

[1] *Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran*

[2] *Department of Medical Physics, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran*

[3] *Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran*

*Corresponding Author: Karim Khoshgard

Email: khoshgardk@gmail.com

Received: 31 October 2024 / Accepted: 02 June 2025

Abstract

Purpose: The initial evaluation of trauma poses a formidable and time-intensive challenge. This study aims to scrutinize the diagnostic efficacy and utility of integrating machine learning models with radiomics features for the identification of blunt traumatic kidney injuries in abdominal CT images.

Materials and Methods: This investigation involved the collection of 600 CT scan images encompassing individuals with varying degrees of kidney damage resulting from trauma, as well as images from healthy subjects, sourced from the Kaggle dataset. An experienced radiologist performed the segmentation of axial images, and radiomics features were subsequently extracted from each region of interest. Initially, 30 machine learning models were deployed, with a final selection narrowed down to three models: Light Gradient-Boosting Machine (LGBM), Ridge Classifier, and Adaptive Boosting (AdaBoost). The performance of these chosen models was subjected to a more comprehensive examination.

Results: The AdaBoost model exhibited notable performance in diagnosing mild kidney injury, achieving accuracy and sensitivity rates of 93% and 94%, respectively. Furthermore, for severe kidney injury, the AdaBoost model demonstrated a remarkable sensitivity of 96% and an accuracy of 97%. The Area Under the Curve (AUC) values for this model were also calculated, yielding values of 92.91% and 97.04% for mild and severe renal injuries, respectively.

Conclusion: The artificial intelligence models employed in this study hold significant potential to enhance patient care by providing valuable assistance to radiologists and other medical professionals in the diagnosis and staging of trauma-related kidney injuries. These models offer the capability to prioritize positive studies, expedite evaluations, and accurately identify more severe injuries that may necessitate immediate intervention.

Keywords: Kidney; Blunt Trauma; Radiomics; Computed Tomography Scan; Machine Learning; Artificial Intelligence.

DOI:

1. Introduction

The kidney, as the primary filtration organ within the human body, is susceptible to trauma that may manifest as damage to the parenchyma or renal vasculature, leading to bleeding or injury to the collecting system with potential urine leakage. The kidney, while situated in a relatively safeguarded retroperitoneal area, is the organ most frequently harmed in the genitourinary system due to trauma [1]. Renal trauma constitutes a notable portion of overall trauma cases, ranging from 1% to 5%, with the majority attributed to blunt abdominal trauma, accounting for 80% to 90% of cases [2-4]. Renal trauma primarily impacts males, constituting 72% to 93% of reported cases. This type of injury is also more prevalent among younger individuals, with the average age ranging from 31 to 38 years [5]. Blunt renal trauma, more prevalent than penetrating trauma, is often caused by incidents such as motor vehicle collisions, falls, injuries from sports activities, and pedestrian accidents. The mechanisms of blunt renal trauma include a direct impact on the organ, compression against the paravertebral muscles, or rapid deceleration forces. Morbidity and mortality associated with renal trauma is closely tied to the grade of injury, concurrent injuries, and the effectiveness of management interventions [6].

Contrast-enhanced computed tomography is currently regarded as the preferred imaging technique for hemodynamically stable patients who have experienced blunt or penetrating renal trauma [5]. CT is frequently the primary diagnostic modality in patients with suspected or confirmed renal trauma, particularly in large trauma centers dealing with multiple-system injuries. In cases of acute trauma, CT plays a crucial role by offering comprehensive anatomical and physiological information that aids in distinguishing minor injuries from those necessitating intervention [7]. Despite its diagnostic utility, challenges arise in the timely interpretation of CT scans, particularly when dealing with abdominal CT examinations that cannot be promptly reviewed [8]. Additionally, not all imaging departments have round-the-clock on-site experienced radiologist coverage; although artificial intelligence techniques cannot replace radiologists, they can assist less experienced radiologists in order to reach the correct diagnosis [9]. The subsequent concern revolves around the formidable workload confronting radiologists within a department. In the daily operations of radiology, radiologists bear the responsibility of interpreting medical images sourced from diverse modalities. Their roles often necessitate thorough analyses and evaluations of these images within constrained timeframes. However, with the continuous evolution of modern medical technologies, the abundance of imaging data is on a swift upward trajectory. Notably, contemporary CT examinations feature thinner slices compared to historical practices [10]. Consequently, the development of an artificial intelligence model for the diagnosis of renal injuries aims to support radiologists in the identification, quantification, and analysis of changes in lesion size over time.

Artificial Intelligence (AI) has increasingly become a part of our everyday experiences, encompassing a range of technologies and methodologies rooted in computer science, statistics, and data science. Machine learning, a subset of AI, utilizes algorithms to autonomously detect patterns within data and make predictions or decisions based on these patterns, emphasizing reduced human intervention [11, 12].

In the field of medical image analysis, significant progress has been made in recent decades, enabling the extraction of quantitative features that may not be visible through visual inspection [13]. This methodology, known as radiomics, captures tissue and lesion characteristics, such as heterogeneity and shape, and utilizes these features to predict current and future variables, including disease presence or absence, treatment response, and time to recurrence [14, 15].

Yang *et al.* conducted a study focusing on the classification of small renal masses through machine learning techniques applied to CT scans. Their research involved data from 163 patients with small renal masses, which included 118 scans of Renal Cell Carcinoma (RCC) cases and 45 scans of renal angiomyolipoma without visible fat (AMLwvf). The researchers extracted textural features using the Pyradiomics package, applying it to the regions of interest generated with the ITK-SNAP tool. They employed eight different machine learning techniques in their analysis, with experimental results indicating that Support Vector Machines (SVM) utilizing t-score and relief methods achieved commendable

performance, reaching an AUC of 90% in the classification tasks [16]. Radiomics functions as an automated methodology for feature generation, extracting a myriad of quantitative phenotypes (radiomics features) from radiological images [15, 17]. Subsequently, machine learning algorithms can be trained to discern relationships between these radiomics features and patient diagnoses [18].

The purpose of this study is to investigate the efficiency of implementing machine learning models using radiomics features in the diagnosis of kidney injuries caused by trauma. In this study, the effectiveness of three machine learning models in diagnosing mild and severe injuries has been investigated using the criteria of accuracy, precision, F1 score, specificity, AUC, and sensitivity. If these artificial intelligence tools have an acceptable performance, they can serve as valuable assistants to less experienced radiologists. Especially in emergency departments, where due to the heavy workload of physicians and the need to quickly interpret the images, there is a greater possibility of mistakes.

2. Materials and Methods

The outline of the present research is shown in Figure 1.

2.1. Data Collection

In this investigation, a total of 600 axial CT slices sourced from the Kaggle database [https://kaggle.com/competitions/rsna-2023-abdominal-trauma-detection, 2023] were employed. These comprised 200 slices depicting a healthy kidney, another 200 slices exhibiting mild kidney damage, and the remaining 200 slices manifesting severe kidney damage resulting from blunt trauma. The images were in DICOM format with a matrix size of 512×512. The slice thickness ranged from 0.5 to 5 mm.

2.2. Segmentation

A fundamental step in quantitative image analysis, image segmentation involves the identification and delineation of regions of interest within an image. In this investigation, the region encompassing the kidney on each axial CT slice was identified and isolated. This segmentation process was executed by a skilled radiologist with more than 15 years of experience in interpreting trauma images, using 3D Slicer software.

2.3. Feature Extraction

In this study, this step refers to the concept of radiomics. In general, radiomics aims to extract quantitative and reproducible information from diagnostic images, including complex patterns that are

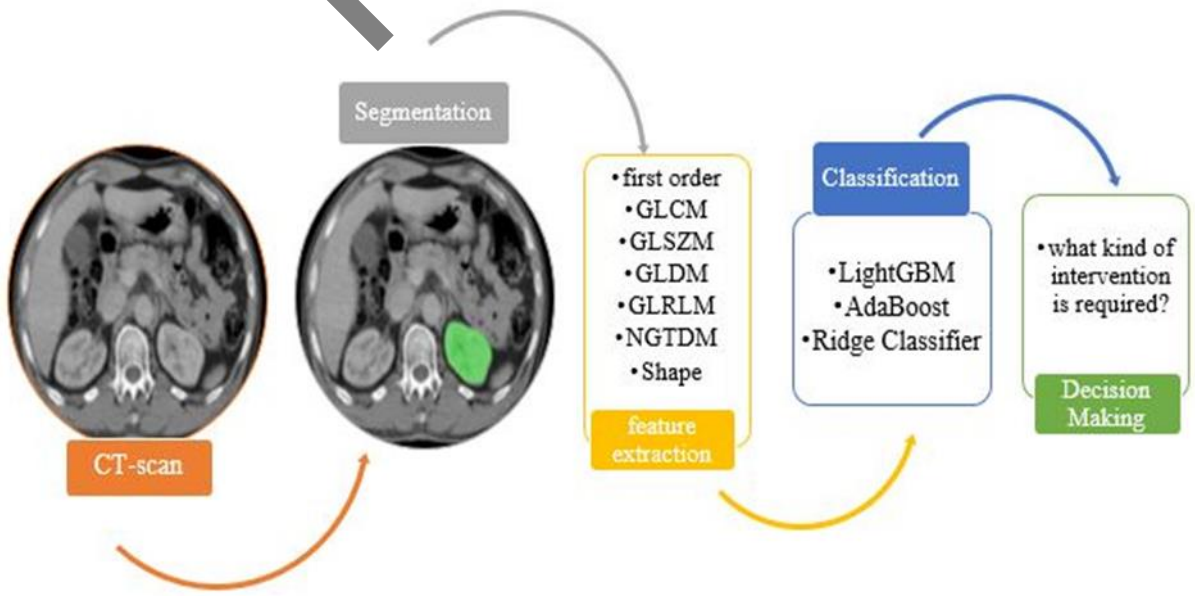

Figure 1. The outline of the research

difficult to detect or quantify by the human eye [15]. Employing the Radiomics toolbox within 3D Slicer software, this step involved the extraction of features, comprising first-order statistical features and texture features such as Gray Level Dependence Matrix (GLDM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM). Wavelet filters with various decompositions, including HHH, HHL, HLH, HLL, LHH, LHL, LLH, and LLL, were applied to obtain these features in all three dimensions. The results were recorded in an Excel file. Wavelet analysis of an image is possible using a pair of square mirror filters, a high-pass filter, and a low-pass filter [19]. The high-pass filter highlights the changes in the gray level and therefore emphasizes the details of the image, while the low-pass filter smoothes the image in terms of the gray level and removes the details of the image [15].

***First-order features*:** These features describe the intensity distribution of pixels or voxels in the image area defined by the mask through common and basic criteria [20].

Gray Level Dependence Matrix (GLDM): The features of this group define the gray level dependencies in an image. Gray-level dependency is defined as the number of related voxels within a certain distance that is dependent on the central voxel [20].

Gray Level Run Length Matrix (GLRLM): These features provide information about the spatial distribution of the run of consecutive pixels with the same gray level, in one or more directions, in 2 or 3 dimensions [15].

Gray Level Size Zone Matrix (GLSZM): GLSZM is based on a similar principle to GLRLM, but here, counting the number of groups (so-called regions) of contiguous adjacent pixels or voxels with the same gray level forms the basis of the matrix. A tissue with more homogeneity creates a wider and flatter matrix [20, 21].

***Neighboring Gray Tone Difference Matrix (NGTDM)*:** NGTDM calculates the sum of differences between the gray level of a pixel or voxel and the average gray level of pixels or voxels adjacent to it at a predetermined distance [15].

2.4. Training of Machine Learning Models

In this study, 75% of the data (slices) was used for training, and 25% of the data was also assigned to test the algorithms. These kinds of issues are supervised learning issues; they need labels. In supervised learning, each sample contains two parts: one is input observations or features, and the other is output observations or labels [22, 23]. In this study, the input observations are radiomics features, and the output observations are the presence or absence of kidney injuries. The purpose of supervised learning is to conclude a functional relationship from training data that generalizes well to testing data [10].

A brief description of each model is provided below:

LGBM is a high-speed, distributed, high-performance machine learning framework based on a decision tree algorithm. This framework can be used in various tasks such as sorting, classification, regression, and other machine learning tasks. By maintaining accuracy, the speed of this framework increases about ten times, and the amount of occupied memory is about three times less. This framework has advantages such as high training efficiency, low memory occupancy, high precision, and support for parallelization, and it can also be implemented using a GPU to process large data [24, 25].

Ridge Classifier is a linear classification algorithm that is based on the Ridge regression algorithm. It is used to classify data into two or more classes based on features. In Ridge regression, the objective is to minimize the sum of squared errors between the predicted values and the actual values while also incorporating a penalty term to prevent overfitting. This penalty term is based on the L2 norm of the coefficients, which helps to shrink the coefficients of less important features towards zero [26].

Among the various ensemble learning algorithms, ***adaptive boosting (AdaBoost)*** [27] is one of the most prominent. The fundamental concept behind ensemble learning is to train multiple weak learners using training data and then combine these weak learners to create a strong learner. Weak learners are typically based on individual learning algorithms, such as decision stumps, though other algorithms can also be used [28].

During preliminary experiments, these three models consistently achieved the highest accuracy, sensitivity, and specificity on our validation data, outperforming other algorithms in terms of classification metrics. LGBM and AdaBoost are ensemble methods known for their ability to handle high-dimensional, heterogeneous data and reduce overfitting challenges in radiomics. Ridge Classifier, with its regularization capability, performed well with correlated features, which are common in radiomics datasets.
These models demonstrated efficient training and prediction times, making them suitable for potential clinical application where rapid decision support is valuable [24, 26]. All three models provide mechanisms for feature importance ranking, which was essential for our goal of identifying clinically relevant radiomics features.

Hyperparameter tuning is the process of systematically selecting the best set of configuration values (hyperparameters) for a machine learning model to optimize its performance. Unlike model parameters, which are learned during training, hyperparameters are set before training and control how the model learns or how complex it can become [29].

Below are brief definitions of some commonly tuned hyperparameters:

- **learning_rate:** Controls how much the model's weights are updated during training.
- **max_depth:** Specifies the maximum depth of each decision tree in ensemble models.
- **min_child_weight:** Sets the minimum sum of instance weights (or number of samples) required in a child node.
- **n_estimators:** The number of trees (for ensemble methods like boosting or bagging) or iterations used in the model.
- **penalty:** Used in linear models (e.g., logistic regression), this hyperparameter determines the type of regularization applied to prevent overfitting.
- **solver:** Specifies the algorithm used to optimize the model's parameters during training (e.g., 'liblinear', 'saga', 'lbfgs' in logistic regression) [29-31].

The results of hyperparameter tuning for all three models are given in Tables 1 and 2.

Table 1. The results of Hyperparameter tuning (Mild renal injury)

LGBM	learning_rate = 1.0 max_depth = 5 min_child_weight = 1.0 n_estimators = 200
AdaBoost	n_estimators = 30 learning_rate = 0.97
Ridge Classifier	penalty = l2 solver = "liblinear"

Table 2. The results of Hyperparameter tuning (Severe renal injury)

LGBM	learning_rate = 0.1 max_depth = 5 min_child_weight = 3.0 n_estimators = 25
AdaBoost	n_estimators = 30 learning_rate = 0.99
Ridge Classifier	penalty = l2 solver = "liblinear"

In the course of this study, a dedicated feature selection stage was not incorporated into our methodology. An intrinsic advantage of the employed machine learning models lies in their capacity to perform feature selection seamlessly during algorithm execution. This eliminates the necessity for a discrete step labeled 'feature selection'. The most important features selected by all three models are shown in Table 3.

Since these features have shown their importance in the diagnosis of trauma-related kidney injuries, they can be considered good candidates for biomarkers of kidney complications [32], so it is suggested that in future studies, the correlation between these features and clinical parameters can be investigated.

Radiomics features identified as most important by the models are clinically relevant because they quantitatively capture underlying pathological

Table 3. The most important features selected by all three models in CT scan images

Ridge Classifier	AdaBoost	LGBM
First order		
90Percentile (LLL)		
10Percentile		
90Percentile		
Maximum		
Mean		
Median		
Minimum	**First order**	
Range	Entropy (LHH)	
Variance	Mean (LHH)	**First order**
Variance (LLH)	Mean (HLL)	Mean (HLL)
10Percentile (LLL)	Median (HLH)	10Percentile
GLCM	Kurtosis (HHL)	10Percentile (LLH)
Cluster Prominence (LLH)	Mean Absolute Deviation (HHH)	Mean (LLH)
GLDM	Robust Mean Absolute Deviation	Minimum (LLH)
Large Dependence High Gray Level Emphasis (LLL)	Skewness	Mean Absolute Deviation (LHL)
	Mean (HLL)	Entropy (LHH)
Large Dependence High Gray Level Emphasis	Skewness (HLL)	Kurtosis (LLH)
GLRLM	Maximum (LLL)	Mean (LHH)
High Gray Level Run Emphasis (LLL)	**GLCM**	**GLCM**
Long Run High Gray Level Emphasis (LLL)	Contrast (HLL)	Idn
	Inverse Variance (HLH)	Imc1 (HHL)
Run Length Non Uniformity (LLL)	Correlation (HHL)	Cluster Shade (LHL)
Short Run High Gray Level Emphasis (LLL)	Imc1 (HHL)	Cluster Shade (HLL)
GLSZM	MCC (HHH)	Difference Variance (HHL)
High Gray Level Zone Emphasis (LLL)	MCC (LHL)	**GLDM**
Large Area High Gray Level Emphasis (LLL)	Cluster Shade (HLL)	Gray Level Non Uniformity
	Imc2 (HLH)	Dependence Non Uniformity (LLH)
Small Area High Gray Level Emphasis (LLL)	Imc2 (HHL)	Dependence Non Uniformity Normalized (LHL)
Zone Variance (LLL)	**GLDM**	
Large Area Emphasis	Dependence Variance (HLL)	Large Dependence High Gray Level Emphasis
Large Area High Gray Level Emphasis	Dependence Variance	
Zone Variance	Small Dependence Low Gray Level Emphasis (LLH)	Small Dependence Low Gray Level Emphasis (LLH)
Large Area High Gray Level Emphasis (LLH)	Dependence Non Uniformity Normalized (LHL)	Dependence Non Uniformity Normalized (HLL)
Large Area Emphasis (LHL)	Dependence Non Uniformity Normalized (HLH)	Dependence Variance (HHL)
Large Area High Gray Level Emphasis (LHL)	Large Dependence Emphasis (HHH)	**GLRLM**
	GLRLM	Run Entropy (LLL)
Zone Variance (LHL)	Short Run Low Gray Level Emphasis (LLH)	Gray Level Variance
Large Area Emphasis (LHH)	Run Length Non Uniformity	Long Run Emphasis (HHL)
Large Area High Gray Level Emphasis (LHH)	Run Percentage (HLL)	Run Variance (HHH)
	Run Variance (HLH)	**GLSZM**
Large Area Low Gray Level Emphasis (LHH)	Short Run Low Gray Level Emphasis (LLH)	Gray Level Non Uniformity Normalized
Zone Variance (LHH)	Long Run Emphasis (HHL)	Small Area Emphasis (LHL)
Large Area Emphasis (HLL)	Run Entropy (HHL)	Low Gray Level Zone Emphasis (HLL)
Large Area High Gray Level Emphasis (HLL)	**GLSZM**	Zone Entropy (HLL)
	Size Zone Non Uniformity Normalized (LLH)	Size Zone Non Uniformity Normalized (LHL)
Large Area Low Gray Level Emphasis (HLL)	Size Zone Non Uniformity Normalized (LHL)	Small Area Emphasis (LHL)
Zone Variance (HLL)	Zone Entropy (LHL)	Gray Level Variance (HHH)
Large Area Low Gray Level Emphasis (HHL)	Zone Percentage (HHL)	**NGTDM**
Large Area Emphasis (HHH)	**NGTDM**	Busyness (LLL)
	Busyness (LHL)	Busyness
Large Area Low Gray Level Emphasis (HHH)		Contrast
NGTDM		
strength		
strength (LLL)		
Busyness (LHH)		
Complexity (LLL)		

changes in kidney tissue. For example, texture features like GLCM, GLDM, etc. reflect tissue heterogeneity, which increases with fibrosis, scarring, or hemorrhage following trauma. Intensity-based (first-order) features may reveal changes in tissue density from bleeding, edema, or necrosis. These quantitative markers have been shown to correlate with kidney function, fibrosis severity, and overall disease progression, making them valuable for non-invasive detection and assessment of traumatic kidney injuries [33].

2.5. Evaluation of Machine Learning Models

The confusion matrix is a valuable tool for evaluating the performance of classification models. It provides a detailed breakdown of the model's predictions compared to the actual outcomes, allowing for a comprehensive assessment of its effectiveness. It provides a more detailed understanding of how well the model is performing across different classes. In this research, in order to evaluate the performance of machine learning models, the evaluation criteria of the confusion matrix, including accuracy, precision, F1 score, specificity, sensitivity, AUC, and misclassification rate, were used. Relations 1 to 6 show how to calculate these criteria [34, 35].

In our study:

- TP (True Positive): Cases correctly classified as having renal injury.
- TN (True Negative): Cases correctly classified as not having renal injury.
- FP (False Positive): Cases incorrectly classified as having renal injury, despite the absence of injury.
- FN (False Negative): Cases incorrectly classified as not having renal injury, despite the presence of injury.

$$Accuracy: \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision}: \frac{TP}{TP+FP} \quad (2)$$

$$\text{Sensitivity}: \frac{TP}{TP+FN} \quad (3)$$

$$\text{Specificity}: \frac{TN}{TN+FP} \quad (4)$$

$$\text{F1 score}: \frac{2TP}{2TP+FP+FN} \quad (5)$$

$$\text{Misclassification: 1-accuracy} \quad (6)$$

3. Results

The confusion matrix for all three models is shown in Table 4.

According to Table 5, in terms of accuracy, the AdaBoost and LGBM models have the best performance with values equal to 93% and 92%, respectively, in diagnosing mild kidney damage and 97% for severe kidney damage. In terms of precision criteria, the AdaBoost model, with values equal to 92.59% and 98.07%, respectively, for detecting mild and severe kidney damage, showed a stronger performance than the two other models. Of course, the LGBM model has the same performance as the AdaBoost model, with a precision of 92.45% in detecting mild kidney damage. The AUC value for the AdaBoost model in diagnosing mild kidney injuries was 92.91% and for severe injuries was 97.04%, which indicates a stronger performance of this model in comparison with other models.

The ROC curves for all three models are shown in Figures 2 and 3.

Regarding the sensitivity criterion, the AdaBoost model has surpassed the other two models with values of 94% and 96%, respectively, in detecting mild and severe injuries. This model achieved 91% specificity in detecting mild kidney injuries and 98% for severe injuries. The F1-score, which is a combination of precision and recall criteria, was calculated as 93% and 97% for the AdaBoost model in detecting mild and severe kidney damage, respectively. Misclassification indicates the number of samples that have been incorrectly classified, and in this sense, the LGBM and AdaBoost models have a better performance.

Table 4. Confusion matrix for mild and severe renal injuries

Mild renal injury			
LGBM		Predicted Negative	Predicted Positive
	Actual Negative	43 (TN)	4 (FP)
	Actual Positive	4 (FN)	49 (TP)
AdaBoost		Predicted Negative	Predicted Positive
	Actual Negative	43 (TN)	4 (FP)
	Actual Positive	3 (FN)	50 (TP)
Ridge Classifier		Predicted Negative	Predicted Positive
	Actual Negative	42 (TN)	5 (FP)
	Actual Positive	12 (FN)	41 (TP)
Severe renal injury			
LGBM		Predicted Negative	Predicted Positive
	Actual Negative	45 (TN)	2 (FP)
	Actual Positive	1 (FN)	52 (TP)
AdaBoost		Predicted Negative	Predicted Positive
	Actual Negative	43 (TN)	4 (FP)
	Actual Positive	12 (FN)	41 (TP)
Ridge Classifier		Predicted Negative	Predicted Positive
	Actual Negative	46 (TN)	1 (FP)
	Actual Positive	2 (FN)	51 (TP)

Table 5. The results of implementing the machine learning models

Injury	Models	Accuracy	Precision	AUC	Mis-Classification	Sensitivity	Specificity	F_1 Score
Mild	LGBM	92%	92.45%	91.97%	8%	92%	91%	92%
	AdaBoost	93%	92.59%	92.91%	7%	94%	91%	93%
	Ridge Classifier	83%	89.13%	83.36%	17%	77%	89%	83%
Severe	LGBM	97%	96.29%	96.92%	3%	98%	96%	97%
	AdaBoost	97%	98.07%	97.04%	3%	96%	98%	97%
	Ridge Classifier	84%	91.11%	84.42%	16%	77%	91%	84%

4. Discussion

In this study, we used machine learning models and radiomics features to facilitate and accelerate the diagnosis of renal injuries caused by blunt trauma in CT scan images. Using artificial intelligence to analyze medical images in emergency departments helps doctors quickly diagnose and treat patients who need urgent care. The proposed framework allows for an objective and quantitative evaluation of kidney trauma, contrasting with the often subjective assessments currently used in clinical practice. This advancement has the potential to improve real-time diagnostics for kidney trauma and serve as an effective triage tool.

Renal injuries, while uncommon, are not rare occurrences. The diagnosis and treatment of such

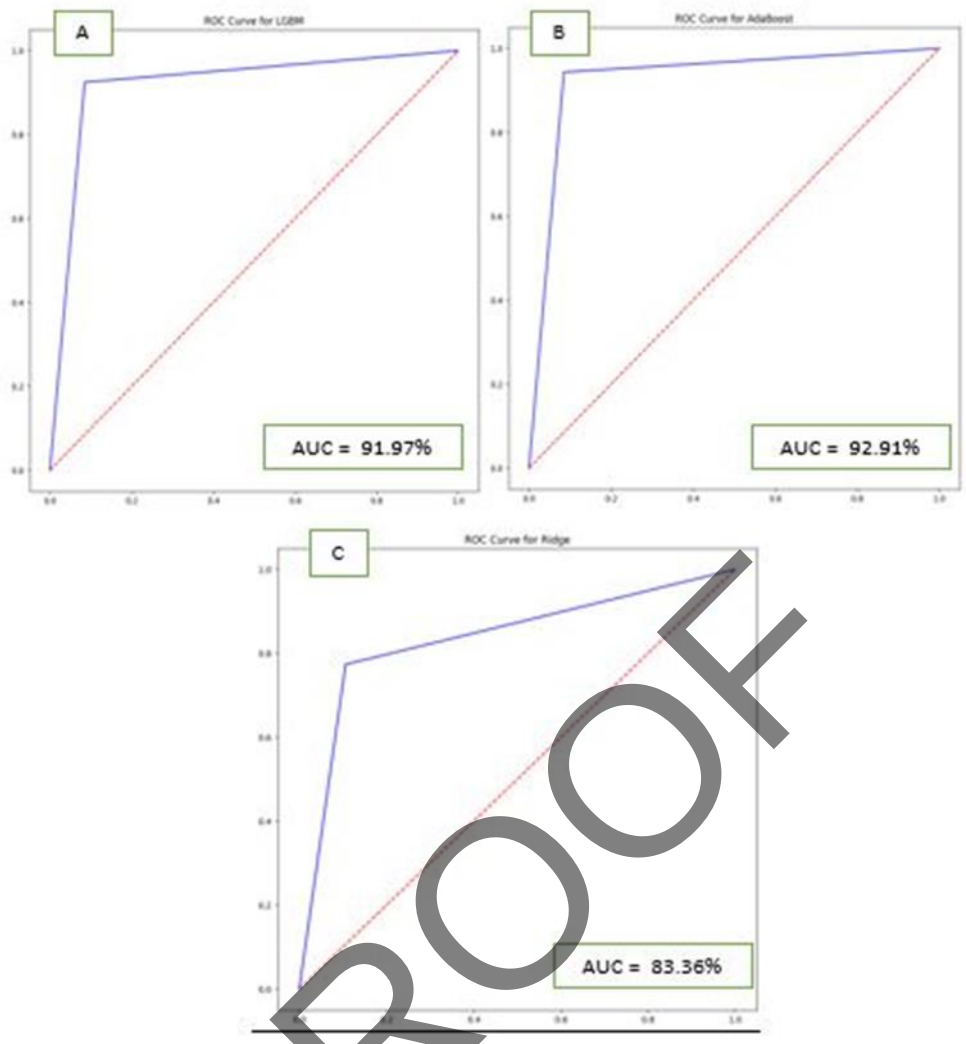

Figure 2. ROC curves for all three models in the detection of mild kidney injuries: A: LGBM, B: AdaBoost, and C: Ridge Classifier

injuries demand a comprehensive understanding of the retroperitoneal region. These injuries can manifest with diverse patterns, often necessitating intricate diagnostic and therapeutic evaluations [36]. The workload burden on radiologists in emergency departments may contribute to delays in diagnosing these complications. Studies have indicated that, in certain instances, general radiologists must interpret an image every three to four seconds throughout an eight-hour workday to meet the demands of their workload [37]. The proliferation of AI in medical imaging is primarily motivated by the need to enhance the efficiency and efficacy of clinical care. The volume of radiological imaging data is expanding at an unprecedented rate in comparison to the available number of skilled readers. Concurrently, diminishing reimbursement for imaging procedures is compelling healthcare providers to address the challenge through heightened productivity [38].

Kate *et al.* conducted a study utilizing logistic regression (LR), support vector machines (SVM), decision trees, and naïve Bayes to detect undiagnosed Acute Kidney Injury (AKI) in a large population of hospitalized elderly patients aged over 60 years. The study reported AUC ranging from 0.66 to 0.74, indicating moderate performance in identifying undiagnosed AKI within this demographic [39]. Yap *et al.* conducted a study on the classification of renal masses using machine learning techniques applied to CT scans. The dataset included CT scans from 735 patients, comprising 196 scans of benign masses and 539 scans of malignant cases. The researchers manually segmented the scans using the 3D Synapse

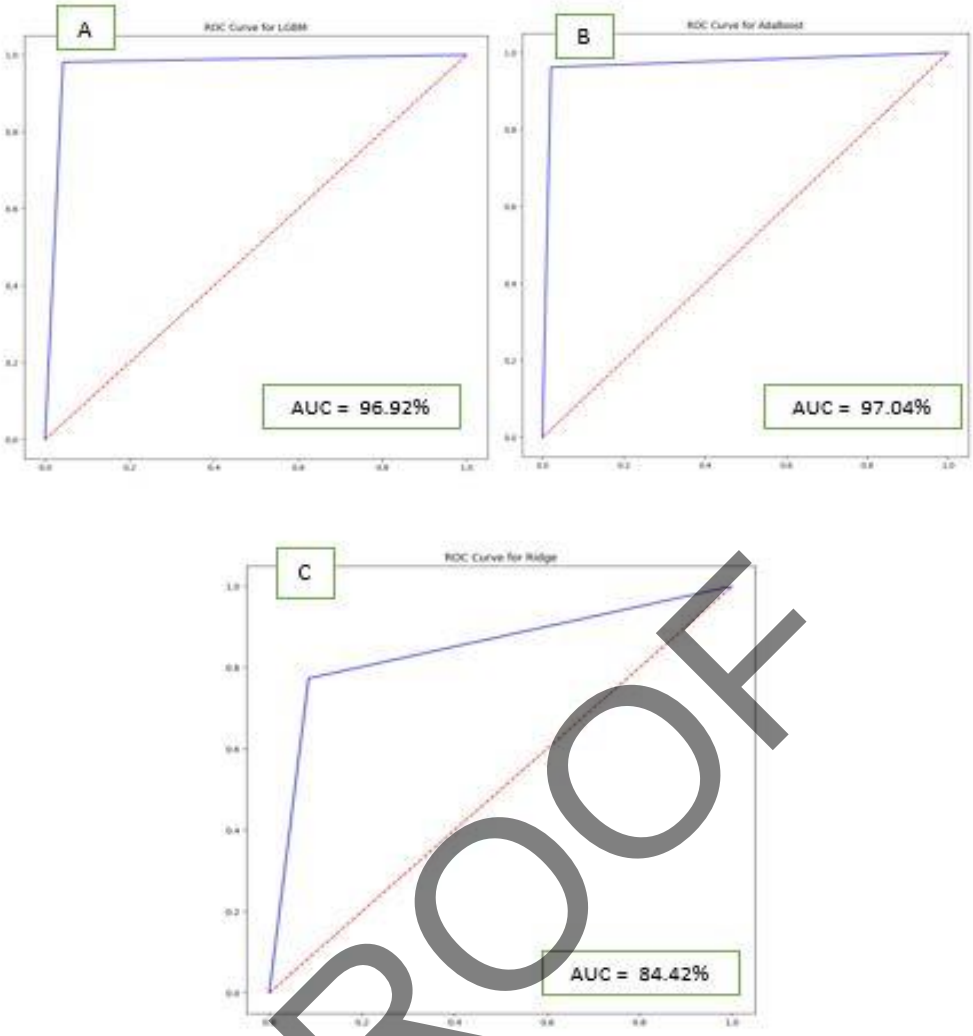

Figure 3. ROC curves for all three models in the detection of severe kidney injuries: A: LGBM, B: AdaBoost, and C: Ridge Classifier

3D tool, collaborating with two expert radiologists to extract features based on shape and texture matrices. The study utilized two machine learning algorithms: AdaBoost and random forest. The experimental results indicated that the random forest model achieved high performance, with AUC values ranging from 68% to 75% for classifying renal masses [40]. Uhlig *et al.* conducted a study focused on classifying renal tumor subtypes using machine learning techniques applied to CT scans. In their approach, features were extracted through manual segmentation with the aid of 3D Slicer and PyRadiomics, which generated ROI features from the axial slices of the renal scans. The proposed model utilized a random forest algorithm for the multi-class classification of renal tumors, achieving a notable performance with an AUC of 78% after excluding the oncocytoma subtype [41]. There is an increasing interest in the role of artificial intelligence in emergency abdominal imaging. However, abdominal imaging poses more intricate and specific challenges compared to other areas, such as skeletal fracture detection. This complexity arises from the diverse anatomy of the abdomen and the intricate imaging characteristics involved [42].

At the beginning of the study, 30 machine learning models were implemented on the data, and after examining these models, it was found that three models, LGBM, AdaBoost, and Ridge Classifier, perform better than other models. Another advantage of these models is that they do not need a separate feature selection algorithm.

In recent years, deep learning-a specialized area within machine learning-has gained substantial attention in the field of medical imaging because of its capacity to transform how diseases are diagnosed and

risks are predicted. By automatically identifying and categorizing radiographic features, deep learning models help minimize observer bias, enhance the reliability of diagnoses, and overcome many challenges associated with traditional manual assessments [43, 44]. But this technology has limitations. Deep learning is resource-intensive, requiring powerful Graphics Processing Units (GPUs), high-performance computing resources, and substantial storage capacity; In this sense, the machine learning techniques used in our research are simpler and less expensive. Unlike traditional machine learning methods, deep learning models are often considered 'black boxes' because it is difficult to understand how they make decisions. While we can see the input and the resulting output, it is usually unclear which features the model used or how it arrived at its final classification of an image as healthy or diseased [45, 46]. The use of other clinical information of patients along with radiomics features in the training of machine learning models can lead to a more accurate evaluation of the performance of these models. Lack of access to this information is one of the limitations of this study.

5. Conclusion

The artificial intelligence models presented in this research have significant potential to diagnose and grade renal injuries caused by trauma and can help radiologists and other doctors to speed up and facilitate the diagnosis of these complications in emergency departments. The use of artificial intelligence models in this way can help to prioritize positive studies for faster diagnosis and identify more severe complications that necessitate prompt intervention.

Acknowledgment

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Kermanshah University Medical Science (IR.KUMS.MED.REC.1402.281). The authors are grateful to the Student Research Committee of Kermanshah University of Medical Sciences for supporting this project with code 4020530.

References

1- Javad Salimi and MR NIKOUBAKHT, "Epidemiologic study of 284 patients with urogenital trauma in three trauma centers in Tehran." (2004).

2- M. McPhee, N. Arumainayagam, M. Clark, N. Burfitt, and R. DasGupta, "Renal injury management in an urban trauma centre and implications for urological training." (in eng), *Ann R Coll Surg Engl,* Vol. 97 (No. 3), pp. 194-7, Apr (2015).

3- M. T. Heller and N. Schnor, "MDCT of renal trauma: correlation to AAST organ injury scale." (in eng), *Clin Imaging,* Vol. 38 (No. 4), pp. 410-17, Jul-Aug (2014).

4- T. Ząbkowski, R. Skiba, M. Saracyn, and H. Zieliński, "Analysis of Renal Trauma in Adult Patients: A 6-Year Own Experiences of Trauma Center." (in eng), *Urol J,* Vol. 12 (No. 4), pp. 2276-9, Sep 4 (2015).

5- T. Erlich and N. D. Kitrey, "Renal trauma: the current best practice." (in eng), *Ther Adv Urol,* Vol. 10 (No. 10), pp. 295-303, Oct (2018).

6- B. B. Voelzke and L. Leddy, "The epidemiology of renal trauma." (in eng), *Transl Androl Urol,* Vol. 3 (No. 2), pp. 143-9, Jun (2014).

7- W. Szmigielski, R. Kumar, S. Al Hilli, and M. Ismail, "Renal trauma imaging: Diagnosis and management. A pictorial review." (in eng), *Pol J Radiol,* Vol. 78 (No. 4), pp. 27-35, Oct (2013).

8- Xiang Li *et al.*, "Deep learning-enabled system for rapid pneumothorax screening on chest CT." *European journal of radiology,* Vol. 120p. 108692, (2019).

9- H. P. Chan, R. K. Samala, and L. M. Hadjiiski, "CAD and AI for breast cancer-recent development and challenges." (in eng), *Br J Radiol,* Vol. 93 (No. 1108), p. 20190580, Apr (2020).

10- S. Wang and R. M. Summers, "Machine learning and radiology." *Med Image Anal,* Vol. 16 (No. 5), pp. 933-51, Jul (2012).

11- Sheetal Chaudhuri *et al.*, "Artificial intelligence enabled applications in kidney disease." in *Seminars in dialysis*, (2021), Vol. 34 (No. 1): *Wiley Online Library*, pp. 5-16.

12- Nafiseh Ghaffar Nia, Erkan Kaplanoglu, and Ahad Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction." *Discover Artificial Intelligence,* Vol. 3 (No. 1), p. 5, 2023/01/30 (2023).

13- H. Alimiri Dehbaghi, K. Khoshgard, H. Sharini, and S. J. Khairabadi, "Diagnosis of traumatic liver injury on computed tomography using machine learning algorithms and radiomics features: The role of artificial intelligence for rapid diagnosis in emergency rooms." (in eng), *J Res Med Sci,* Vol. 29p. 77, (2024).

14- Robert J Gillies, Paul E Kinahan, and Hedvig Hricak, "Radiomics: images are more than pictures, they are data." *Radiology,* Vol. 278 (No. 2), pp. 563-77, (2016).

15- Marius E Mayerhoefer *et al.*, "Introduction to radiomics." *Journal of Nuclear Medicine,* Vol. 61 (No. 4), pp. 488-95, (2020).

16- Ruimeng Yang *et al.*, "Radiomics of small renal masses on multiphasic CT: accuracy of machine learning–based classification models for the differentiation of renal cell carcinoma and angiomyolipoma without visible fat." *European radiology,* Vol. 30pp. 1254-63, (2020).

17- Philippe Lambin *et al.*, "Radiomics: extracting more information from medical images using advanced feature analysis." *European journal of cancer,* Vol. 48 (No. 4), pp. 441-46, (2012).

18- Hossein Naseri *et al.*, "Radiomics-based machine learning models to distinguish between metastatic and healthy bone using lesion-center-based geometric regions of interest." *Scientific Reports,* Vol. 12 (No. 1), p. 9866, 2022/06/14 (2022).

19- Andrew Laine and Jian Fan, "Texture classification by wavelet packet signatures." *IEEE Transactions on pattern analysis and machine intelligence,* Vol. 15 (No. 11), pp. 1186-91, (1993).

20- Alex Zwanenburg *et al.*, "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping." *Radiology,* Vol. 295 (No. 2), pp. 328-38, (2020).

21- Guillaume Thibault, Jesus Angulo, and Fernand Meyer, "Advanced statistical matrices for texture characterization: application to cell classification." *IEEE Transactions on Biomedical Engineering,* Vol. 61 (No. 3), pp. 630-37, (2013).

22- Ethem Alpaydin, Introduction to machine learning. *MIT press*, (2020).

23- Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics). (2009).

24- Baoyang Cui, Zhonglin Ye, Haixing Zhao, Zhuome Renqing, Lei Meng, and Yanlin Yang, "Used Car Price Prediction Based on the Iterative Framework of XGBoost+LightGBM." *Electronics,* Vol. 11 (No. 18), p. 2932, (2022).

25- H. Alimiri Dehbaghi, K. Khoshgard, H. Sharini, S. Jafari Khairabadi, and F. Naleini, "Radiomics-based machine learning for automated detection of Pneumothorax in CT scans." (in eng), *Plos one,* Vol. 19 (No. 12), p. e0314988, (2024).

26- Rakesh Katuwal and P. N. Suganthan, "An Ensemble of Kernel Ridge Regression for Multi-class Classification." *Procedia Computer Science,* Vol. 108pp. 375-83, (2017).

27- Xindong Wu *et al.*, "Top 10 algorithms in data mining." *Knowledge and Information Systems,* Vol. 14 (No. 1), pp. 1-37, 2008/01/01 (2008).

28- Feng De-Cheng *et al.*, "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach." *Construction and Building Materials,* Vol. 230p. 117000, (2020).

29- Warut Pannakkong, Kwanluck Thiwa-Anont, Kasidit Singthong, Parthana Parthanadee, and Jirachai Buddhakulsomsiri, "Hyperparameter Tuning of Machine Learning Algorithms Using Response Surface Methodology: A Case Study of ANN, SVM, and DBN." *Mathematical Problems in Engineering,* Vol. 2022 (No. 1), p. 8513719, (2022).

30- Justus A Ilemobayo *et al.*, "Hyperparameter Tuning in Machine Learning: A Comprehensive Review." *Journal of Engineering Research and Reports,* Vol. 26 (No. 6), pp. 388-95, 06/07 (2024).

31- Christian Arnold, Luka Biedebach, Andreas Küpfer, and Marcel Neunhoeffer, "The role of hyperparameters in machine learning models and how to tune them." *Political Science Research and Methods,* pp. 1-8, (2023).

32- Seongho Jo and Kipyo Kim, "CT-Based Radiomic Feature Analysis for Identifying Baseline Kidney Function in Patients With Underlying CKD: SA-PO915." *Journal of the American Society of Nephrology,* Vol. 33 (No. 11S), pp. 856-57, (2022).

33- Y. H. Choi *et al.*, "Histopathological correlations of CT-based radiomics imaging biomarkers in native kidney biopsy." (in eng), *BMC Med Imaging,* Vol. 24 (No. 1), p. 256, Sep 27 (2024).

34- Samira Jafari, Afshin Almasi, Hamid Sharini, Sajad Heydari, and Nader Salari, "Diagnosis of borderline personality disorder based on Cyberball social exclusion task and resting-state fMRI: using machine learning approach as an auxiliary tool." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* pp. 1-11, (2022).

35- Y. Baştanlar and M. Ozuysal, "Introduction to machine learning." (in eng), *Methods Mol Biol,* Vol. 1107pp. 105-28, (2014).

36- Petrone Patrizio, Perez-Calvo Javier, E. M. Brathwaite Collin, Islam Shahidul, and K. Joseph D'Andrea, "Traumatic kidney injuries: A systematic review and meta-analysis." *International Journal of Surgery,* Vol. 74pp. 13-21, (2020).

37- Robert J McDonald *et al.*, "The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload." *Academic Radiology,* Vol. 22 (No. 9), pp. 1191-98, (2015).

38- Giles WL Boland, Alexander S Guimaraes, and Peter R Mueller, "The radiologist's conundrum: benefits and costs of increasing CT capacity and utilization." *European radiology,* Vol. 19pp. 9-11, (2009).

39- R. J. Kate, R. M. Perez, D. Mazumdar, K. S. Pasupathy, and V. Nilakantan, "Prediction and detection models for acute kidney injury in hospitalized older adults." (in eng), *BMC Med Inform Decis Mak,* Vol. 16p. 39, Mar 29 (2016).

40- Felix Y Yap *et al.*, "Shape and texture-based radiomics signature on CT effectively discriminates benign from malignant renal masses." *European radiology,* Vol. 31pp. 1011-21, (2021).

41- Johannes Uhlig *et al.*, "Radiomic features and machine learning for the discrimination of renal tumor histological subtypes: a pragmatic study using clinical-routine computed tomography." *Cancers,* Vol. 12 (No. 10), p. 3010, (2020).

42- M. Cellina *et al.*, "Artificial Intelligence in Emergency Radiology: Where Are We Going?" (in eng), *Diagnostics (Basel),* Vol. 12 (No. 12), Dec 19 (2022).

43- Firouz Amani, Masoud Amanzadeh, Mahnaz Hamedan, and Paniz Amani, "Diagnostic accuracy of deep learning in prediction of osteoporosis: a systematic review and meta-analysis." *BMC Musculoskeletal Disorders,* Vol. 25 (No. 1), p. 991, 2024/12/04 (2024).

44- Ali Tarighatnia, Masoud Amanzadeh, Mahnaz Hamedan, Alireza Mohammadnia, and Nader D. Nader, "Deep learning-based evaluation of panoramic radiographs for osteoporosis screening: a systematic review and meta-analysis." *BMC Medical Imaging,* Vol. 25 (No. 1), p. 86, 2025/03/12 (2025).

45- Quan-shi Zhang and Song-Chun Zhu, "Visual interpretability for deep learning: a survey." *Frontiers of Information Technology & Electronic Engineering,* Vol. 19 (No. 1), pp. 27-39, (2018).

46- Matthew D Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks." in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, (2014): *Springer*, pp. 818-33.