

# Ability of Machine Learning and Deep Learning Models for Multiclass Classification of Kidney Stone and Lung Cancer from Computed Tomography Images: A Comparative Study

Anima Kujur\* and Zahid Raza

*School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi - 110 067, India.*

*\*Email: anima.kujur55@gmail.com*

## ABSTRACT

Feature extraction is crucial in biomedical image classification because it determines the accuracy of image representations and significantly impacts the effectiveness of classification models. Deep neural network classification architectures have gained significant interest due to their ability to automatically extract important features from input images, resulting in significant progress in diverse image classification tasks in recent years. However, with the rise of deep learning techniques, traditional machine learning approaches have been largely overshadowed. This study aims to close this gap by undertaking a rigorous comparative analysis of three important machine learning models, namely Gaussian Naïve Bayes, Support Vector Machine, and Random Forest Classifier, and three advanced deep learning models, namely VGG16, InceptionV3, and Xception. The comparison is based on their ability to do multiclass classification, using two datasets kidney stone and lung cancer. Each dataset consists of four different target classes. Both machine learning and deep learning frameworks are trained separately on the datasets, with deep learning models utilizing transfer learning techniques. The performance of each model across the varied output classes is assessed using evaluation measures such as precision, recall, and F1 scores. The results of the simulation analysis reveal that both machine learning and deep learning models perform equally well, as indicated by similar F1 scores across all output classes for both datasets. This study represents a major step towards simplifying classification efforts by promoting the use of machine learning models instead of deep learning models for classifying kidney stone and lung cancer datasets. This approach helps reduce the workload and computing requirements for training.

**Keywords:** Deep learning; Machine learning; Biomedical image classification; Computed tomography; Biomedical image processing; Feature extraction

## 1. INTRODUCTION

Accurate classification of biomedical images, particularly Computed Tomography (CT) scans, is vital for early detection and precise diagnosis of medical conditions such as kidney stones and lung cancer. Feature extraction, a fundamental technique in biomedical image classification<sup>1,2</sup>, involves identifying and highlighting crucial image elements necessary for distinguishing between different disease classes. The effectiveness of any classification model heavily relies on the proficiency with which these features are extracted, directly impacting the accuracy and reliability of the classification process. Deep Neural Networks (DNNs)<sup>3,4</sup> have emerged as formidable contenders in the realm of biological image classification due to their capability to autonomously learn intricate patterns from input images. These models have demonstrated outstanding performance across various image classification tasks, often surpassing traditional Machine Learning (ML)<sup>5</sup> models. However, traditional ML techniques remain pertinent in

biomedical image classification, offering results that are interpretable and computationally efficient, albeit facing challenges with complex datasets.

Deep learning models, particularly DNNs, excel at detecting complex patterns and correlations within extensive and high-dimensional datasets. Their ability to learn directly from raw input data enables them to perform intricate tasks without necessitating domain-specific knowledge. Despite their computational demands, DL models exhibit exceptional generalisation ability, adaptability to changing conditions, and capacity to learn from new data. In contrast, ML techniques, while less computationally intensive, may require additional support when handling large and intricate datasets.

Feature extraction is pivotal in effectively representing image information, thereby playing a crucial role in image classification. Researchers have proposed various comparative approaches and techniques for feature extraction, aiming to enhance classification accuracy by identifying the most suitable method. Moreover, studies have explored the efficacy of deep learning methods in the diagnosis

of kidney stones<sup>6-9</sup> and lung cancer<sup>10-13</sup>, demonstrating significant advancements in automated kidney stone and lung cancer classification. Additionally, researchers have investigated methods to improve image quality and enhance diagnosis using ML and DL techniques. Within the field of medical research, experts are always working to improve diagnostic models specifically designed for the automated classification of kidney stones<sup>14,15</sup> and lung cancer<sup>16,17</sup>. The consistent and determined effort highlights the pressing necessity to improve medical diagnostics<sup>18,19</sup> specifically in crucial fields such as oncology. Kanavati<sup>20</sup> performed a crucial effort in this field, specifically studying the complex distinctions between the main histological kinds of lung cancer. By carefully developing and training a deep learning model utilizing H&E-stained Whole Slide Images of tiny trans bronchial lung biopsy specimens, scientists achieved a significant improvement in classification accuracy, paving the way for more accurate and efficient diagnosis. Wang and Dong<sup>21</sup> introduced an innovative method for identifying lung cancer by utilising the capabilities of CT imaging. Their application of transfer learning, along with a complex neural network structure, not only demonstrates technological progress but also emphasizes the integration of varied datasets and approaches in the pursuit of diagnostic excellence.

On the other hand, Marentakis and Karaiskos<sup>22</sup> undertook an extensive investigation, examining various methods for classifying tumors using CT images. Their comprehensive analysis not only enhances our comprehension of the subtle interplay between various approaches but also underscores the intrinsic difficulty of medical imaging jobs. Moreover, the ground breaking research conducted by Adriana and Dinh-Hoan<sup>23</sup> in the automation of kidney stone classification represents a major achievement in the discipline, introducing a novel age of precision and effectiveness through the utilisation of supervised learning methodologies. However, in the middle of all this innovation, there remains a significant gap - there is no comparative study that explains the ability of ML and DL models to extract features for classifying kidney stones and lung cancer. This unexplored domain invites investigation, offering the possibility for profound understanding and potentially significant progress in the field of medical diagnostics.

This study aims to compare ML and DL models in classifying CT images for kidney stone and lung cancer detection. The evaluated ML models include Support Vector Machines (SVM<sup>24</sup>), Random Forest Classifier (RFC<sup>25</sup>), and Gaussian Naïve Bayes (GNB<sup>26</sup>), while DL models comprise VGG16<sup>27</sup>, InceptionV3<sup>28</sup>, and Xception<sup>29</sup>. By leveraging transfer learning, DL models utilize pre-trained weights to enhance feature extraction. Performance evaluation is conducted using precision, recall, and F1 score metrics across multiple output classes. Our objective is to provide insights into choosing suitable ML or DL models for multi-class classification using CT scans, focusing on kidney stone and lung cancer detection. Additionally,

we aim to elucidate the feature extraction capabilities of these models in the context of biological image classification, identifying their comparative strengths and limitations. The findings of this study have the potential to improve patient outcomes and healthcare effectiveness by developing more accurate and reliable classification models for early disease detection using CT scans and potentially expanding to other medical conditions.

The subsequent sections of this paper are organised as follows: Section 2 outlines the proposed methodology and provides an overview of the simulation study. Section 3 delves into the empirical findings and evaluates performance, while Section 4 discusses the outcomes of the experiment. Finally, Section 5 concludes by suggesting directions for future research.

## 2. MATERIAL AND METHODS

Fig. 1 represents the flow diagram of the proposed framework for kidney and lung cancer classification. It has five phases. The first phase is the data collection. The second phase is data pre-processing. The third phase corresponds to the model training where three ML and three DL algorithms are implemented. The result evaluation is done in the fourth phase. Model comparison over the selected parameters is done in the fifth phase.

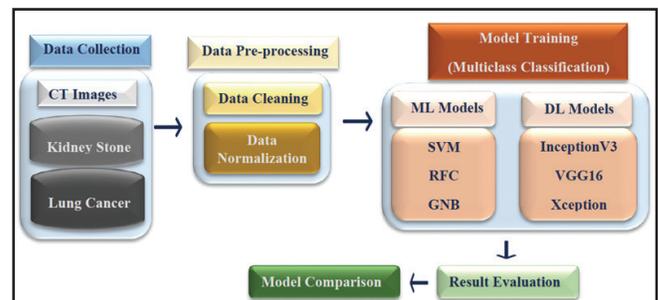


Figure 1. The flow diagram of the proposed classification framework.

### 2.1 Data Collection

This study utilised two distinct datasets to perform multiclass classification tasks. One dataset focused on categorizing kidney stones, while the other centered on classifying lung cancer. The kidney dataset comprised 12,446 CT images, categorised into four groups: cyst (3,709 images), normal (5,077 images), stone (1,377 images), and malignant (2,283 images). Conversely, the lung cancer dataset contained 1,000 CT images, distributed among four categories: adeno carcinoma (338 images), large cell carcinoma (260 images), squamous carcinoma (187 images), and normal cell (215 images). Both datasets were collected from Kaggle, a publicly available data repository. Fig. 2 visually represents the class distribution of the lung cancer (left) and kidney stone (right) datasets. Within the lung cancer dataset, adeno carcinoma emerges as the most prevalent subtype, constituting nearly 40% of all lung cancer cases. Large cell carcinoma, characterised by atypical features compared to other types, and squamous cell carcinoma, often associated with a history of smoking and primarily affecting central

lung regions, are also depicted. Figure 3 shows sample CT scans from the lung cancer dataset, providing visual insights into various lung cancer pathologies. Moving to the kidney dataset, four distinct classes were identified: kidney tumor, cyst, stone and normal, each reflecting specific clinical conditions impacting kidney health. Kidney tumors denote abnormal growths within kidney tissue, which can be either benign or malignant. Conversely, kidney cysts represent fluid-filled sacs, typically harmless and asymptomatic. Kidney stones, composed of calcium oxalate, and other compounds, can form in the kidneys, leading to various health issues. Fig. 4 exhibits sample CT scans from the kidney stone dataset, illustrating different disease manifestations related to kidney health.

### 2.2 Data Pre-Processing

Data cleaning is the process of detecting and rectifying discrepancies in a dataset to ensure its integrity and dependability for analysis. The images which are not labelled are discarded. Data normalisation aims to scale the values of features in a dataset to a comparable range, hence enhancing the performance of machine learning algorithms. The dimensionality of each image varies, in this stage, all the images are resized into one dimension for both the datasets. The input pixels are between range (0-255) and all the image pixels are normalised to the range (0-1).

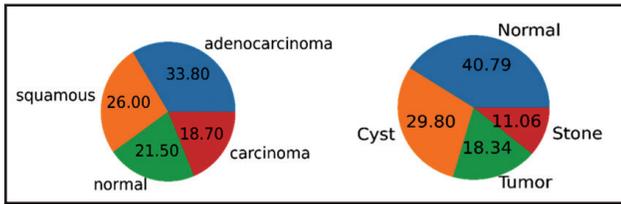


Figure 2. Class distribution for lung cancer dataset(left) and kidney stone dataset(right).

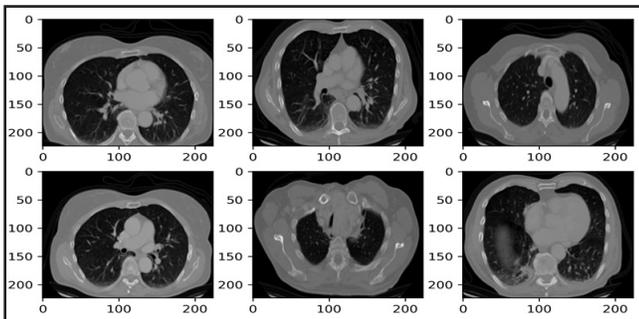


Figure 3. The sample CT images for lung cancer dataset.

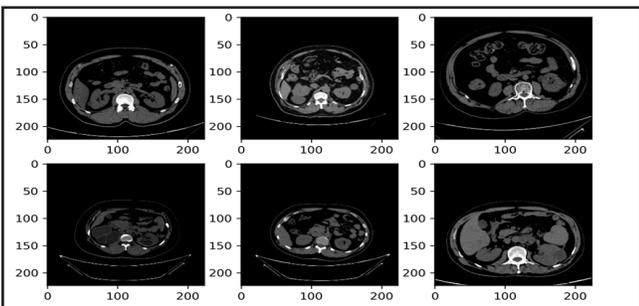


Figure 4. The sample CT images for kidney stone dataset.

### 2.3 The Model Training

The model training phase of this study involves the investigation and use of different ML and DL algorithms. More precisely, we are considering three supervised ML algorithms: SVM, RFC, and GNB. The selection of these algorithms is based on their effectiveness in classification tasks, and they are then utilised on the dataset to evaluate their performance. In addition, deep learning models, particularly CNNs, are used because of their exceptional capacity to acquire intricate patterns from input images. This study employs three widely-used CNN architectures, specifically InceptionV3, VGG16, and Xception. The CNN models utilised in this work have been pre-trained on extensive ImageNet dataset. And then further implemented on the dataset used in this research using transfer learning method. CNNs are composed of several layers that perform convolution and pooling operations. These operations allow the network to extract important features from input images at various levels of abstraction. The convolutional layers utilize filters to analyze the input images, identifying patterns such as edges, textures, and forms. Afterwards, the pooling layers gather the input from the convolutional layers, decreasing the spatial dimensions of the features while maintaining their fundamental qualities. CNNs utilize a hierarchical feature extraction technique, enabling them to effectively capture complex patterns and representations contained in the input data. This makes them very suitable for image classification. The model training phase consists of implementing and evaluating a variety of ML and DL algorithms. Each algorithm is selected based on its specific strengths and capacities to handle classification task being performed. The work seeks to determine the best efficient algorithm for appropriately classifying the dataset using a comprehensive approach. Following are the pseudo codes for each algorithm implemented in this work.

```

Pseudo Code for InceptionV3
1. base_model ← InceptionV3(include_top=False, weights='imagenet', input_shape=IMG_SIZE)
2. x ← base_model.output
3. x ← GlobalAveragePooling2D()(x)
4. x ← Dense(256, activation='relu')(x)
5. predictions ← Dense(4, activation='softmax')(x)
6. model ← Model(inputs=base_model.input, outputs=predictions)
7. for layer in base_model.layers:
   layer.trainable = False
    
```

```

Pseudo Code for VGG16
1. base_model ← VGG16(include_top=False, weights='imagenet', input_shape=IMG_SIZE)
2. x ← base_model.output
3. x ← GlobalAveragePooling2D()(x)
4. x ← Dense(256, activation='relu')(x)
5. predictions ← Dense(4, activation='softmax')(x)
6. model ← Model(inputs=base_model.input, outputs=predictions)
7. for layer in base_model.layers:
   layer.trainable = False
    
```

```

Pseudo Code for Xception
1. base_model ← Xception(include_top=False, weights= 'imagenet',
input_shape= IMG_SIZE)
2. x ← base_model.output
3. x ← GlobalAveragePooling2D()(x)
4. x ← Dense(256, activation='relu')(x)
5. predictions ← Dense(4, activation='softmax')(x)
6. model←Model(inputs=base_model.input, outputs=predictions)
7. for layer in base_model.layers:
    layer.trainable = False
    
```

```

Pseudo Code for SVM
1. model ← SVC()
2. model.fit(train_x,train_y)
3. predictions ←model.predict(val_x)

Pseudo Code for GNB
1. model ← GaussianNB()
2. model.fit(train_x,train_y)
3. predictions ← model.predict(val_x)

Pseudo Code for RFC
1. model ←RandomForestClassifier()
2. model.fit(train_x,train_y)
3. predictions ←model.predict(val_x)
    
```

The remaining phases of the proposed framework corresponds to the results evaluation and model comparison and are discussed in Section 3.

### 3. RESULTS

The experimental results from the simulation study are analysed in this section. It has two subsections. The section 3.1 contains the result analysis for kidney stone classification and section 3.2 contains the result analysis for lung cancer classification. All the implemented models are evaluated using three metrics: precision, recall and F1 score. A model’s precision measures how well it predicts the True Positives (TP) out of all the occurrences predicted to be positives for a given class. It is determined as the proportion of TP to all other true and False Positives (FP). Recall measures a model’s ability to pick out TP from all the other positive examples in that class. It is calculated as the proportion of TP to all true positives and False Negatives (FN). The harmonic mean of accuracy and recall, or the F1 score, is a measurement of the precision against recall trade-off. A higher F1 score denotes greater performance in terms of both accuracy and recall. It gives a single value that combines both precision and recall.

#### 3.1 Result Analysis Of Kidney Stone Classification

Table 1-3 displays the Precision, Recall, and F1 scores for various models in a classification problem involving kidney stones. The table consists of the models GNB, RFC, SVM, VGG16, InceptionV3, and Xception, along with their respective scores for the classes cyst, normal, stone, and tumor. Table 1 displays the precision scores of the RFC, SVM, VGG16, InceptionV3, and Xception models, all of which have achieved a perfect score of

1.00 for all classes. This indicates that these models have accurately predicted all positive cases without any false positives. The GNB model has an average precision score of 0.86, while the RFC, SVM, VGG16, InceptionV3, and Xception models have an average precision score of 1.00.

**Table 1. Precision score for kidney stone classification**

Models	Cyst	Normal	Stone	Tumor	Average F1 score
GNB	0.97	0.83	0.95	0.69	0.86
RFC	1.00	1.00	1.00	1.00	1.00
SVM	1.00	1.00	1.00	1.00	1.00
VGG16	1.00	0.99	1.00	1.00	1.00
InceptionV3	1.00	0.99	1.00	1.00	1.00
Xception	1.00	1.00	1.00	0.99	1.00

Table 2 demonstrates that the RFC, SVM, VGG16, InceptionV3, and Xception models exhibit recall scores of 1.00 for the majority of classes, signifying their ability to accurately identify all positive examples without any false negatives. Nevertheless, the VGG16 and InceptionV3 models have a recall score of 0.99 for the tumor class, suggesting that they failed to detect certain positive cases for that particular class.

**Table 2. Recall score for kidney stone classification**

Models	Cyst	Normal	Stone	Tumor	Average F1 score
GNB	0.76	0.93	0.65	0.87	0.84
RFC	1.00	1.00	1.00	1.00	1.00
SVM	1.00	1.00	1.00	1.00	1.00
VGG16	1.00	1.00	1.00	0.99	0.99
InceptionV3	1.00	1.00	1.00	0.98	0.99
Xception	1.00	0.99	1.00	0.99	1.00

Table 3 shows the F1 scores for the classes cyst, normal, stone, and tumor. In the GNB model, the F1 scores for these classes are 0.85, 0.88, 0.77, and 0.77, respectively. The RFC, SVM, VGG16, InceptionV3, and Xception models have perfect F1 scores of 1.00 for the majority of classes, indicating their effective ability to maintain a balance between recall and precision for certain classes. Nevertheless, the VGG16 and InceptionV3 models exhibit F1 scores of 0.99 for the tumor class, suggesting a somewhat diminished performance for that class when considering the balance between precision and recall. The GNB model has an average F1 score of 0.84, while the RFC, SVM, VGG16, InceptionV3, and Xception models have an average F1 score of 1.00.

**Table 3. F1 score for kidney stone classification**

Models	Cyst	Normal	Stone	Tumor	Average F1 score
GNB	0.85	0.88	0.77	0.77	0.84
RFC	1.00	1.00	1.00	1.00	1.00
SVM	1.00	1.00	1.00	1.00	1.00
VGG16	1.00	1.00	0.99	1.00	1.00
InceptionV3	1.00	0.99	1.00	0.99	0.99
Xception	1.00	1.00	1.00	0.99	1.00

**3.2 Result Analysis Of Lung Cancer Classification**

Table 4-6 displays the Precision, Recall, and F1 scores for various models used in the classification of lung cancer. The tables in this section consist of models GNB, RFC, SVM, VGG16, InceptionV3, and Xception, along with their respective scores for the classes adenocarcinoma, carcinoma, normal, and squamous, similar to the previous section.

The precision scores for the classes adeno carcinoma, carcinoma, normal, and squamous are displayed in Table 4. The GNB model yields precision scores of 0.58, 0.34, 0.79, and 0.50 for the adeno carcinoma, carcinoma, normal, and squamous classes, respectively. The RFC, SVM, VGG16, InceptionV3, and Xception models exhibit differing precision scores across various classes. The precision scores for different models vary between 0.56 and 0.89 on average. Higher precision scores indicate greater accuracy in predicting relevant outcomes for the respective groups.

According to Table 5, the GNB model demonstrates recall scores of 0.20, 0.76, 0.93, and 0.40 for the classification of adeno carcinoma, carcinoma, normal, and squamous, respectively. The RFC model achieved recall scores of 0.78, 0.61, 0.98, and 0.81 for the classes of adenocarcinoma, carcinoma, normal, and squamous, respectively. For several models, the average recall scores vary between 0.51 and 0.88.

According to Table 6, the GNB model achieves F1 scores of 0.29, 0.47, 0.85, and 0.45 for the adeno carcinoma, carcinoma, normal, and squamous classes,

respectively. The RFC model achieves F1 scores of 0.78, 0.72, 0.95, and 0.72 for the classification of adeno carcinoma, carcinoma, normal, and squamous, respectively. The F1 scores for several models in this situation vary between 0.48 and 0.88 on average. Higher F1 scores signify a superior balance between precision and recall, resulting in enhanced overall model performance for the relevant classes in terms of both FP and FN.

**4. DISCUSSION**

Tables 3 and 6 present the F1 scores attained by different models in classifying kidney stones and lung cancer, respectively. The F1 score, a composite measure of precision and recall, is used to evaluate the accuracy of a model. When it comes to both classification challenges, models like RFC, SVM, VGG16, InceptionV3, and Xception continuously demonstrate high performance, routinely achieving F1 scores above 0.76. In contrast, the GNB model has comparatively lower F1 scores in both situations, which suggests a decrease in accuracy. VGG16 and Xception demonstrate the highest F1 scores in lung cancer classification, indicating their greater accuracy in discriminating between various forms of lung cancer. However, choosing the best model requires careful evaluation of other criteria such as computational complexity, interpretability, and practical application. This necessitates comprehensive study and testing for each unique use case.

**Table 4. Precision score for lung cancer classification**

Models	Adeno carcinoma	Carcinoma	Normal	Squamous	Average F1 score
GNB	0.58	0.34	0.79	0.50	0.56
RFC	0.78	0.89	0.93	0.64	0.80
SVM	0.72	1.00	0.93	0.62	0.80
VGG16	0.88	1.00	0.91	0.77	0.89
InceptionV3	0.96	0.64	0.70	0.77	0.77
Xception	0.96	0.76	0.91	0.84	0.88

**Table 5. Recall score for lung cancer classification**

Models	Adeno carcinoma	Carcinoma	Normal	Squamous	Average F1 score
GNB	0.20	0.76	0.93	0.40	0.51
RFC	0.78	0.61	0.98	0.81	0.79
SVM	0.84	0.46	0.95	0.74	0.77
VGG16	0.96	0.88	0.84	0.87	0.88
InceptionV3	0.96	0.44	0.82	0.74	0.77
Xception	1.00	0.81	0.82	0.91	0.88

**Table 6. F1 score for lung cancer classification**

Models	Adeno carcinoma	Carcinoma	Normal	Squamous	Average F1 score
GNB	0.29	0.47	0.85	0.45	0.48
RFC	0.78	0.72	0.95	0.72	0.79
SVM	0.78	0.63	0.94	0.67	0.76
VGG16	0.92	0.93	0.88	0.82	0.88
InceptionV3	0.96	0.52	0.76	0.76	0.76
Xception	0.98	0.79	0.86	0.87	0.88

The results of the simulation study highlight the complex and diverse parameters that affect the effectiveness of ML and DL models in tasks involving image classification. The kidney stone dataset, which consists of 12,446 CT images, shows enhanced performance in both ML and DL models. However, the smaller lung cancer dataset, which includes 1000 CT scans, produces relatively inferior results. Factors like as the relevance of the data, the presence of noise, the size of the dataset, and the quality of the features have significant impacts on the performance of the model. Images that are of excellent quality and clearly labeled, and that accurately depict real-life situations, have a tendency to improve the accuracy of models. On the other hand, images that are of poor quality, containing a lot of noise, arti facts, or inconsistencies, might negatively impact the accuracy of classification. The presence of noise in image data might introduce irrelevant information or mask important patterns, which can negatively impact the effectiveness of the model. Likewise, when there is a large amount of variation in the image data, it becomes difficult for the model to apply what it has learned to new and unseen images, resulting in reduced accuracy. The size of the image dataset used for training is crucial, as larger and more diverse datasets result in more accurate and representative features. However, this comes at the cost of requiring extensive resources for data collecting and management. Moreover, the accuracy of the model is greatly influenced by the quality of the features used for training, specifically the image representations. Relevant features that enhance the ability to distinguish and accurately classify, contribute to the model's effectiveness. Conversely, poorly chosen or irrelevant features decrease the model's effectiveness. This research highlights the important relationship between data quality, noise, data variance, dataset size, and feature quality in determining the effectiveness of ML and DL models in image classification tasks. Although larger datasets that are of good quality often result in more accurate models, it is important to consider and address aspects such as noise and feature quality in order to enhance the performance of the model for various classification tasks.

## 5. CONCLUSIONS

This work has produced numerous noteworthy contributions that have been clarified by the collected results: Firstly, it thoroughly evaluates the effectiveness of ML and DL models in classifying images of two different medical illnesses, specifically kidney stones and lung cancer. This analysis offers vital insights into how these models compare in terms of their usefulness in healthcare applications. Additionally, the analysis presents the F1 scores of several models used in the classification tasks. It demonstrates that models like RFC, SVM, VGG16, InceptionV3, and Xception consistently exhibit enhanced performance in both kidney stone and lung cancer classification. Furthermore, the study finds key characteristics that

have a substantial impact on the performance of image classification models. These elements include the quality of the data, the level of noise present in the data, the variation in the data, the size of the dataset, and the effectiveness of feature selection. It is important to note that noise in image data negatively affects the accuracy of models, while large variation makes it difficult to apply the models to new, unseen images. Moreover, the size of the image dataset utilised for model training is a significant factor, since larger datasets enable more accurate feature extraction and improved accuracy. However, this comes with the drawback of requiring extensive resources for data collecting and management. Furthermore, the selection of high-quality features for model training is crucial. Relevant features improve model accuracy, while irrelevant or poorly chosen features reduce the model's ability to distinguish and overall effectiveness.

The study highlights the crucial role of image quality in datasets, emphasizing the importance of clear, accurately labelled, and representative images in improving the accuracy of model performance. It also emphasizes the negative impact of low-quality images on misclassifications and reduced accuracy of both ML and DL models. Furthermore, the results highlight the need of taking these characteristics into account when choosing the most suitable model for particular use cases, while balancing computational complexity, interpretability, and real-world applicability requirements. It emphasizes the significance of conducting further analysis and testing to ascertain the most suitable model for certain situations. Future research should prioritize overcoming the problems outlined in order to improve the accuracy and dependability of image classification models for classifying kidney stones and lung cancer. Furthermore, there is a requirement for the advancement of more efficient feature extraction techniques that can accurately capture significant image attributes to enhance classification results.

Future work will be focused on improving the accuracy and reliability of image classification models for kidney stones and lung cancer. Efforts will be directed towards enhancing image quality, mitigating noise in data, collecting larger and diverse datasets, and developing efficient feature extraction techniques

## REFERENCES

1. Kumar, A.; Kim, J.; Lyndon, D.; Fulham, M.; & Feng, D. An Ensemble of Fine-Tuned convolutional neural networks for medical image classification. *IEEE J Biomed Health Inform.*, 2017, **21**(1), pp. 31-40. doi: 10.1109/JBHI.2016.2635663
2. Hegde, R.B. & Prasad, K. Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. *Biocybern Biomed Eng.*, 2019, **39**(2), pp. 382-392. doi: 10.1016/J.BBE.2019.01.005

3. Lecun, Y.; Bengio, Y. & Hinton, G. Deep learning. *Nature.*, 2015, **521**(7553), pp. 436-444.  
doi: 10.1038/NATURE14539
4. Shen, D.; Wu, G. & Suk, H.I. Deep learning in medical image analysis. 2017, **19**, pp. 221-248.  
doi: 10.1146/annurev-bioeng-071516-044442
5. Nayyar, A.; Gadhavi, L. & Zaman, N. Machine learning in healthcare: Review, opportunities and challenges. *Machine learning and the internet of medical things in healthcare.* 2021, pp. 23-45.  
doi: 10.1016/B978-0-12-821229-5.00011-2
6. Akshaya, M.; Nithushaa, R.; Raja, N.S.M. & Padmapriya, S. Kidney stone detection using neural networks. *International conference on system, computation, automation and networking, ICSCAN*, 2020.  
doi: 10.1109/ICSCAN49426.2020.9262335
7. Ebrahimi, S. & Mariano, V.Y. Image quality improvement in kidney stone detection on computed tomography images. *Journal of Image and Graphics*, 2015, **3**(1).  
doi: 10.18178/joig.3.1.40-46
8. Caglayan, A.; Horsanali, M.O. & Kocadurdu, K. Deep learning model-assisted detection of kidney stones on computed tomography. *International Braz J. Urol.*, 2022, **48**(5), pp. 830-839.  
doi: 10.1590/S1677-5538.IBJU.2022.0132
9. Yildirim, K.; Bozdog, P.G.; Talo, M. & Yildirim, O. Deep learning model for automated kidney stone detection using coronal CT images. *Comput Biol Med.*, 2021, **135**.  
doi: 10.1016/j.compbiomed.2021.104569
10. Kuruvilla, J. & Gunavathi, K. Lung cancer classification using neural networks for CT images. *Comput methods programs biomed.* 2014, **113**(1), pp. 202-209.  
doi:10.1016/j.cmpb.2013.10.011
11. Lakshmanprabu, S.K., Mohanty, S.N. & Shankar, K. Optimal deep learning model for classification of lung cancer on CT images. *Future generation computer systems.* 2019, **92**, pp. 374-382.  
doi: 10.1016/j.future.2018.10.009
12. Alakwaa, W.; Nassef, M. & Badr, A. Lung cancer detection and classification with 3D Convolutional Neural Network (3D-CNN). *International Journal of Advanced Computer Science and Applications*, 2017, **8**(8), pp. 66-73.  
doi: 10.14569/IJACSA.2017.080853
13. Dandil, E.A computer-aided pipeline for automatic lung cancer classification on computed tomography scans. *J. Healthc. Eng.*, 2018.  
doi: 10.1155/2018/9409267
14. Abu-Faraj, M. & Zubi, M. Analysis and implementation of kidney stones detection by applying segmentation techniques on computerized tomography scans. 2020, **43**, pp. 590-602.
15. Verma, J.; Nath, M.; Tripathi, P, & Saini, K.K. .Analysis and identification of kidney stone using Kth nearest neighbour (KNN) and support vector machine (SVM) classification techniques. *Pattern Recognition and Image Analysis*, 2017, **27**(3), pp. 574-580.  
doi: 10.1134/S1054661817030294
16. Hussain, L.; Alsolai, H.; Hassine, S.B.H. & Nour, M.K. Lung cancer prediction using robust machine learning and image enhancement methods on extracted gray-level Co-occurrence matrix features. *Applied Sciences*, 2022, **12**(13).  
doi: 10.3390/app12136517
17. Toda, R.; & Teramoto, A.; Saito K, & Fujita, H. Lung cancer CT image generation from a free-form sketch using style-based pix2pix for data augmentation. *Sci. Rep.*, 2022, **12**(1).  
doi: 10.1038/s41598-022-16861-5
18. Yadav, S.S; & Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data*, 2019, **6**(1).  
doi: 10.1186/s40537-019-0276-2
19. El-Shafai, W.; El-Nabi SA; El-Rabaie, E.S.M., Ali, A.M. Efficient deep-learning-based autoencoder denoising approach for medical image diagnosis. *Computers, materials and continua.* 2022, **70**(3), pp. 6107-6125.  
doi: 10.32604/cmc.2022.020698
20. Kanavati, F.; Toyokawa, G.; Momosaki, S, & Takeoka H. A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images. *Sci. Rep.*, 2021, **11**(1). doi: 10.1038/s41598-021-87644-7
21. Wang, S.; Dong, L.; Wang, X. & Wang, X. Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy. *Open medicine (Poland).* 2020, **15**(1), pp. 190-197.  
doi: 10.1515/med-2020-0028
22. Marentakis, P.; Karaiskos, P.; Kouloulis V, & Kelekis, N. Lung cancer histology classification from CT images based on radiomics and deep learning models.  
doi: 10.1007/s11517-020-02302-w/Published
23. Adriana, M. & Dinh, H. Towards an automated classification method for ureteroscopic kidney stone images using ensemble learning. 42nd Annual international conferences of the IEEE engineering in medicine and biology society. 2020, pp. 1936-1939.  
doi: 10.1109/EMBC44109.2020.9176121.
24. Evgeniou, T. & Pontil, M. Support vector machines: Theory and applications. *Lecture notes in computer science, Springer Verlag*, 2001, **2049**, pp. 249-257.  
doi: 10.1007/3-540-44673-7\_12
25. Paul, A.; Mukherjee, D.P.; Das, P. & Chintha, A.R. Improved random forest for classification. *IEEE, transactions on image processing.* 2018, **27**(8), pp. 4012-4024.  
doi: 10.1109/TIP.2018.2834830
26. Anand, M.V.; Kiranbala, B.; Srividhya, S.R. & Rahman, M.H. Gaussian Naïve Bayes algorithm: A reliable technique involved in the assortment of the segregation in cancer. *Mobile information systems.* 2022.  
doi: 10.1155/2022/2436946

27. Kavukcuoglu, K.; Sermanet, P.; Boureau, Y.L.; Gregor, K.; Mathieu, M. & Lecun, Y. Learning convolutional feature hierarchies for visual recognition. 2010, pp.1090-1098.
28. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlen, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. *IEEE Computer Society*, 2016, pp. 2818-2826.  
doi: 10.1109/CVPR.2016.308
29. Chollet, F. Xception: Deep Learning with Depth wise Separable Convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.

### CONTRIBUTORS

**Ms. Anima Kujur**, is presently a PhD scholar in the School of Computer and Systems Sciences at Jawaharlal Nehru University, New Delhi, India. She has received her master's degree in

computer application, from Jawaharlal Nehru University, New Delhi. She obtained her MSc degree in applied mathematics, from University of Hyderabad. Her research interests are in the area of Biomedical image processing, data-driven healthcare, machine learning, deep learning and radiological image data. She has contributed towards conceptualisation, methodology, implementation, result analysis and writing of the manuscript.

**Dr. Zahid Raza**, is presently a Professor in the School of Computer and Systems Sciences at Jawaharlal Nehru University, New Delhi, India. He has a Master's degree in Electronics in which, he was the Gold Medalist. He also has a Master's degree in Computer Science with the first position, and a PhD degree in Computer Science. His research interests are in the area of parallel and distributed systems, grid computing, cloud computing and machine learning. He has proposed various scheduling models for job scheduling for computational grid, cloud, and parallel systems. He has 20 years of experience and has published 30+ research publications. He has contributed towards conceptualisation, methodology, supervision, review and editing of final manuscript.