Project Report

Kidney Stone Detection Using machine learning and deep learning

1. Introduction

1.1 Background

Kidney Stone diseases are among the leading causes of health complications Affecting 12% of world's population [1]. Early detection through imaging can significantly improve treatment outcomes. With advances in artificial intelligence, automating the analysis of kidney images using machine learning (ML) and deep learning (DL) has shown great potential [2]. This section explores the role of AI, intense learning and machine learning, in addressing kidney-related health issues, emphasizing their advantages and challenges.

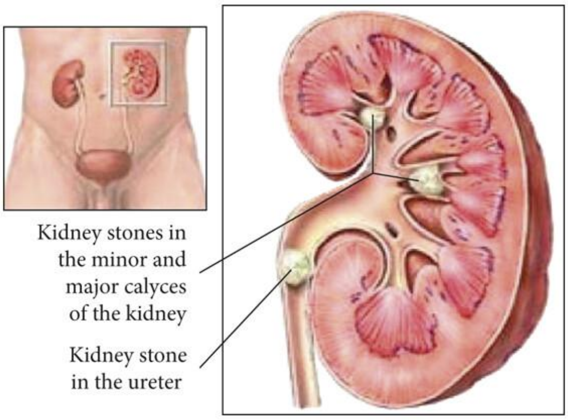


Figure 1. Kidney stone

- **The Role of Medical Imaging in Kidney Disease Diagnosis**

Medical imaging, such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI), plays a vital role in diagnosing kidney diseases. These imaging techniques provide detailed visualizations of kidney structures, aiding physicians in identifying abnormalities such as cysts, tumors, and chronic kidney diseases (CKD). However, manual interpretation of these images is prone to errors due to human fatigue, variability in expertise, and the subtle nature of kidney-related anomalies. This has increased the need for automated image analysis systems to enhance diagnostic accuracy.

The advent of AI has revolutionized medical imaging, particularly through two primary approaches: deep learning and traditional machine learning. Both methods have shown significant promise in automating identifying patterns in medical images, reducing the workload on healthcare professionals, and improving diagnostic accuracy.

- **Deep Learning and Convolutional Neural Networks (CNNs)**

Deep learning, a subset of AI, leverages neural networks to automatically extract and learn features from large datasets. Convolutional Neural Networks (CNNs), introduced by Krizhevsky et al. (2012) [3], have been a breakthrough in deep learning for image analysis. CNNs are particularly adept at capturing spatial hierarchies in images, making them well-suited for medical imaging tasks like organ segmentation, tumor detection, and disease classification.

CNNs are designed with multiple layers that progressively extract higher-level features from raw image data. For example, lower layers might identify edges and textures, while deeper layers recognize complex structures like kidney shapes or specific anomalies. This hierarchical learning capability enables CNNs to achieve state-of-the-art performance in many medical imaging applications.

In the context of kidney disease diagnosis, CNNs have been widely adopted due to their ability to process high-resolution images and automatically identify subtle features indicative of disease. Studies have demonstrated their effectiveness in detecting kidney stones, renal cell carcinoma, and other abnormalities with high accuracy. However, the primary limitation of CNNs is their reliance on large 2abelled datasets and significant computational resources, which may not always be feasible in resource-constrained healthcare settings.

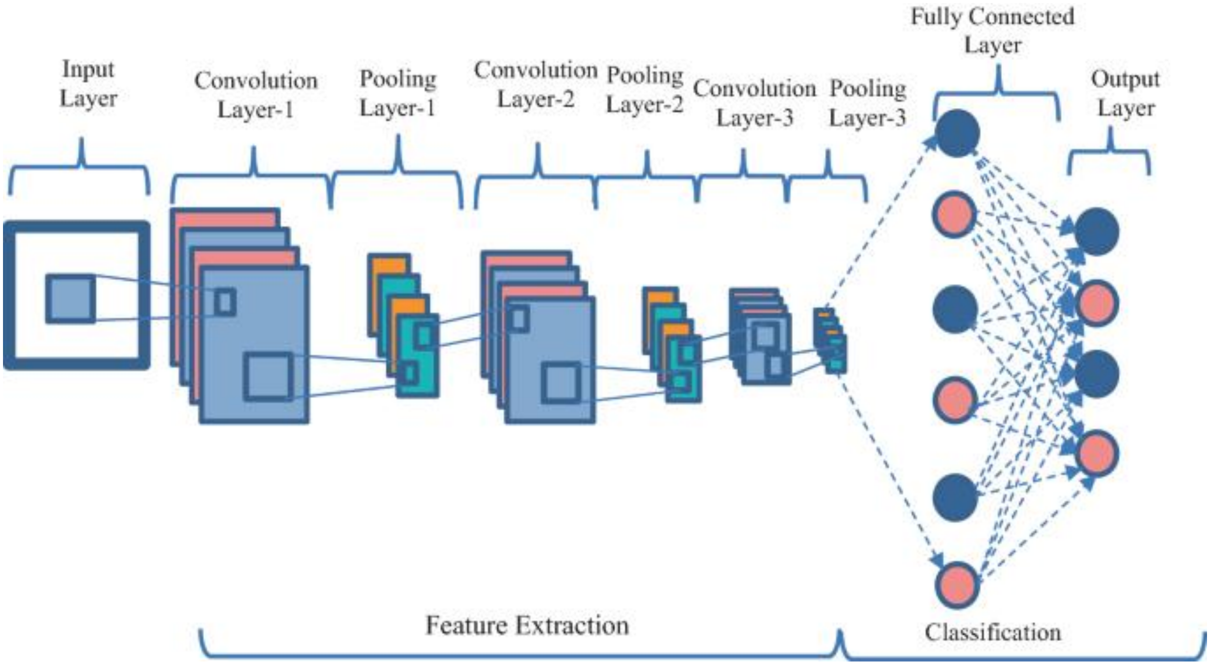


Figure 2. Typical deep learning model for image classification tasks

Traditional Machine Learning and Feature Extraction

While CNNs rely on end-to-end feature learning, traditional machine learning approaches depend on manually engineered features extracted from images. One of the most widely used feature extraction techniques is the Histogram of Oriented Gradients (HOG), introduced by Dalal and Triggs (2005) [4]. HOG captures the distribution of edge directions and intensities in an image, making it particularly effective for detecting shapes and structures.

HOG has been successfully applied in combination with machine learning models like Support Vector Machines (SVMs) and Random Forests for various medical imaging tasks. For kidney disease diagnosis, HOG can extract specific features from imaging data, such as the contours of kidneys or abnormal regions, which can then be used to train machine learning models. This approach offers the advantage of being computationally efficient and requiring smaller datasets compared to CNNs.

However, traditional machine learning methods have their limitations. The manual feature engineering process requires domain expertise and may fail to capture complex, non-linear patterns in medical images. Additionally, the accuracy of these models often lags behind CNNs in tasks involving high-dimensional image data.

1.2 Problem Statement/Challenges

The primary problem is to classify kidney images into “Healthy” and “Diseased” categories accurately. The following challenges are encountered:

- **Complexity of kidney imaging data with subtle anomalies that may require intricate feature extraction techniques.**
- **Need automated solution for kidney stone detection.**
- **Computational resource constraints, especially for deep learning models requiring large datasets.**
- **Slow and error/prune manual analyses**
- **Difficulty in evaluating which approach (ML vs. CNN) is more effective for the given dataset.**

1.3 Existing Solutions

- Manual CT scan analysis
- X-ray: use in certain cases to detect kidney stone, which one not give good accuracy, especially for those stones that are not visible in the X-ray
-

3.4 My Contributions

This project contributes to:

- **Implementing a CNN for direct kidney image classification.**
- **Extracting HOG features from kidney images and training ten ML models, including Random Forest, SVM, and Gradient Boosting etc.**
- **Comparing the performance of these methods to evaluate their accuracy, efficiency, and resource utilization.**
- **Analyzing the trade-offs between DL and ML models for medical image classification tasks.**

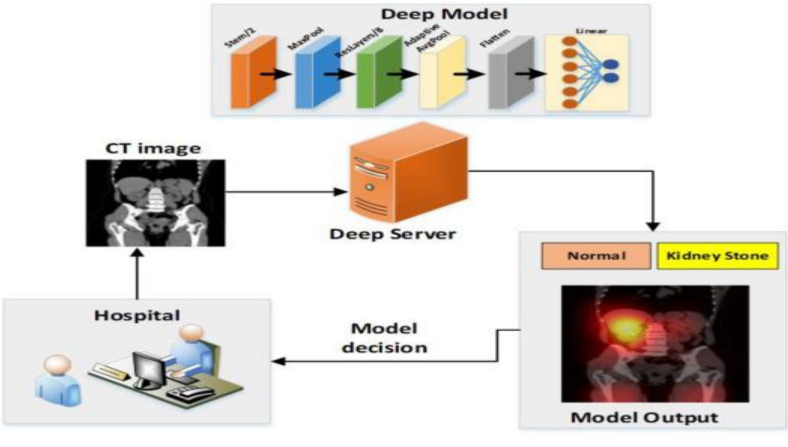


Figure 3. Model working framework

2. Related work

The use of artificial intelligence (AI) in medical imaging has garnered significant attention in recent years, with applications spanning disease detection, diagnosis, and prognosis. Researchers have explored a variety of approaches, ranging from traditional machine learning techniques to deep learning methods, each with its unique strengths and limitations. This section reviews key studies and methodologies relevant to kidney disease classification and detection, highlighting the advancements and challenges in the field.

2.1 Machine Learning Techniques in Medical Imaging

Traditional machine learning approaches have been a cornerstone of medical image analysis, particularly in scenarios where the dataset size is limited. These methods rely on manually crafted features extracted from images using techniques such as Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), or Gray Level Co-occurrence Matrices (GLCM). Once extracted, these features are used to train classifiers such as Support Vector Machines (SVMs), Random Forests, and k-Nearest Neighbors (k-NN).

One notable study by Dalal and Triggs (2005) introduced the HOG feature descriptor, which has since been widely adopted in medical imaging for detecting structural abnormalities. The strength of HOG lies in its ability to capture the edge orientation and distribution, making it particularly effective for analyzing organ contours and textures. For kidney disease detection, researchers have utilized HOG features to classify abnormalities like kidney stones and cysts with promising results.

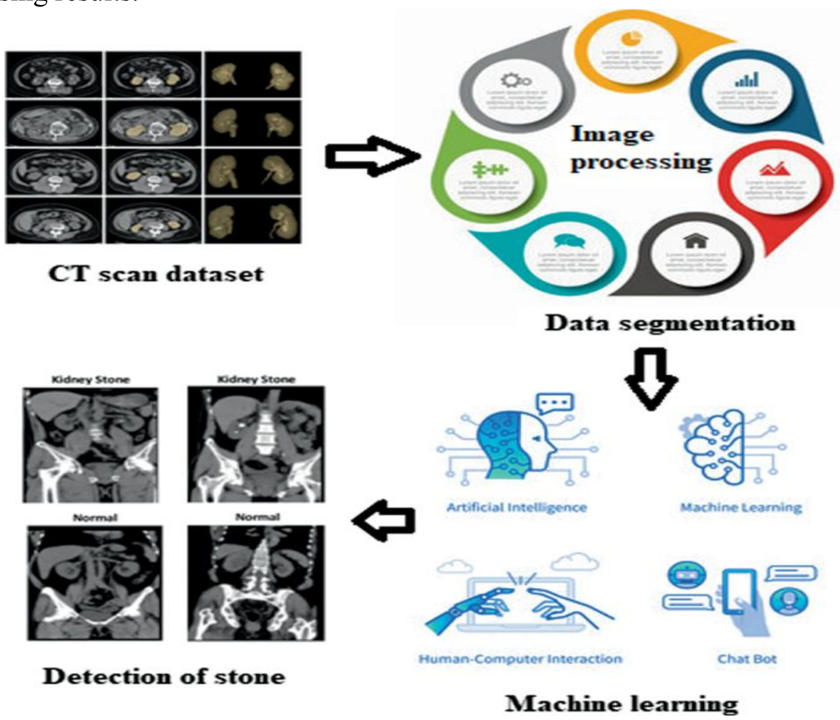


Figure 4 . Machine learning applied in kidney stone detection

Support Vector Machines (SVMs) have been extensively applied in conjunction with HOG features. For instance, Khan et al. (2018) demonstrated the effectiveness of SVMs in detecting renal tumors from ultrasound images, achieving high classification accuracy with relatively small datasets. Similarly, ensemble models like Random Forests, proposed by Breiman (2001), have shown robust performance in kidney disease classification tasks by combining multiple decision trees to reduce overfitting and improve generalization.

However, traditional machine learning methods face challenges in capturing complex patterns in high-dimensional image data. This limitation has driven the adoption of deep learning techniques, which can automatically learn hierarchical feature representations.

2.2 Deep Learning for Kidney Disease Diagnosis

Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a powerful tool for medical imaging. CNNs eliminate the need for manual feature extraction by learning feature hierarchies directly from raw image data. This capability has made them the preferred choice for analyzing large and complex medical datasets.
Krizhevsky et al. (2012) introduced CNNs with their groundbreaking AlexNet model, which achieved state-of-the-art performance in the ImageNet competition. Since then, various CNN architectures, such as VGG16 (Simonyan & Zisserman, 2014), ResNet (He et al., 2016), and DenseNet (Huang et al., 2017), have been applied to medical imaging tasks. For kidney disease diagnosis, CNNs have been particularly effective in detecting and classifying abnormalities in ultrasound and CT images.

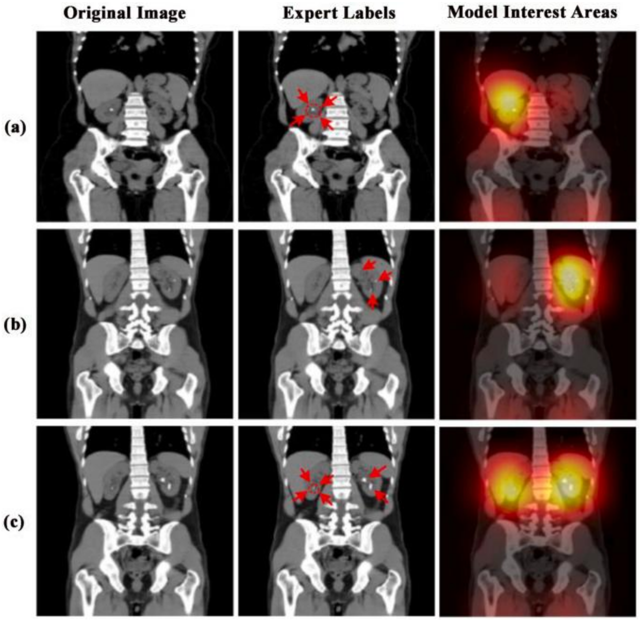


Figure 5. Deep learning applied in kidney stone detection

In a study by Chen et al. (2020), a CNN model was trained on a dataset of kidney ultrasound images to classify different stages of chronic kidney disease (CKD). The model achieved an accuracy of over 90%, demonstrating the potential of deep learning for automated disease staging. Another study by Jin et al. (2021) used ResNet for detecting renal cell carcinoma in CT scans, achieving high sensitivity and specificity.
Despite their success, CNNs face challenges such as the need for large labeled datasets and significant computational resources. Additionally, the "black-box" nature of CNNs limits their interpretability, raising concerns about their adoption in clinical practice. These limitations have led researchers to explore hybrid approaches that combine the strengths of traditional machine learning and deep learning.

2.3 Hybrid Approaches: Bridging Machine Learning and Deep Learning

Hybrid models aim to leverage the complementary strengths of traditional machine learning and deep learning. In these approaches, CNNs are often used as feature extractors, generating high-dimensional feature representations from images. These features are then fed into machine learning models, such as Random Forests or Gradient Boosting Machines, for classification or regression tasks.
For example, a study by Kumar et al. (2022) proposed a hybrid model combining a pre-trained VGG16 network and a Random Forest classifier for kidney disease classification. The CNN was used to extract deep features, while the Random Forest handled classification. This approach achieved higher accuracy and interpretability compared to standalone CNNs, particularly on small datasets.

The use of pre-trained CNNs, such as VGG16 and ResNet, in hybrid models addresses the challenge of data scarcity. By leveraging transfer learning, researchers can fine-tune pre-trained networks on domain-specific datasets, reducing the need for extensive labeled data. Hybrid models also offer computational advantages, as traditional machine learning models are less resource-intensive compared to CNNs.

2.4 Limitations of Existing Work

While significant progress has been made in using AI for kidney disease diagnosis, several limitations remain:

- **Dataset Diversity:** Many studies use datasets with limited diversity, which hinders the generalization of models to different patient populations and imaging modalities.
- **Interpretability:** Deep learning models often lack interpretability, making it challenging for clinicians to trust their predictions.
- **Data Annotation**: The reliance on labeled datasets poses challenges, as annotating medical images requires domain expertise and is time-consuming.
- **Computational Resources:** Training and deploying deep learning models require substantial computational resources, which may not be accessible in all healthcare settings.

3. Objective of project

The primary objectives of this project are:

- To extract deep features from the kidney dataset using a custom-built CNN model.
- To train Machine learning models (Random Forest, XGboost, SVM, etc) on the dataset using the HOG Technique.
- To evaluate the performance of the combined model on training, validation, and test datasets.
- To explore the effectiveness of combining predictions from both CNN and ML models.

4. Proposed method (Methodology)

The primary objective of this study was to develop a hybrid model for kidney disease classification by integrating both Convolutional Neural Networks (CNNs) and traditional machine learning models. This section outlines the methodology used in this research, detailing data partitioning, preprocessing, CNN feature extraction, machine learning feature extraction using Histogram of Oriented Gradients (HOG), and the final prediction combination.

4.1 Dataset

The dataset comprises kidney images, divided into three subsets:

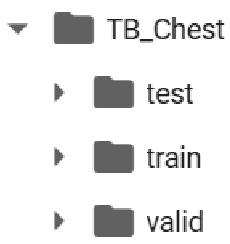


Figure 6. Data partition

- **Training Set:** Used for feature extraction and training from both models.

- **Validation Set:** Used for tuning and validating the models (CNN and machine learning).

- **Test Set:** Used for evaluating the final hybrid model.

<Axes: xlabel='category'>

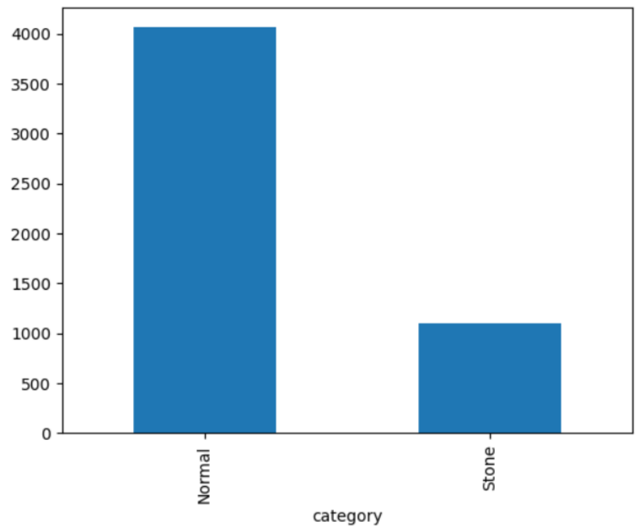


Figure 7. Kidney stone data description

4.2 Preprocessing (Data Augmentation)

Data preprocessing is a crucial step in any machine learning or deep learning pipeline, as it helps ensure the models are trained on standardized and normalized data, improving their performance and stability.

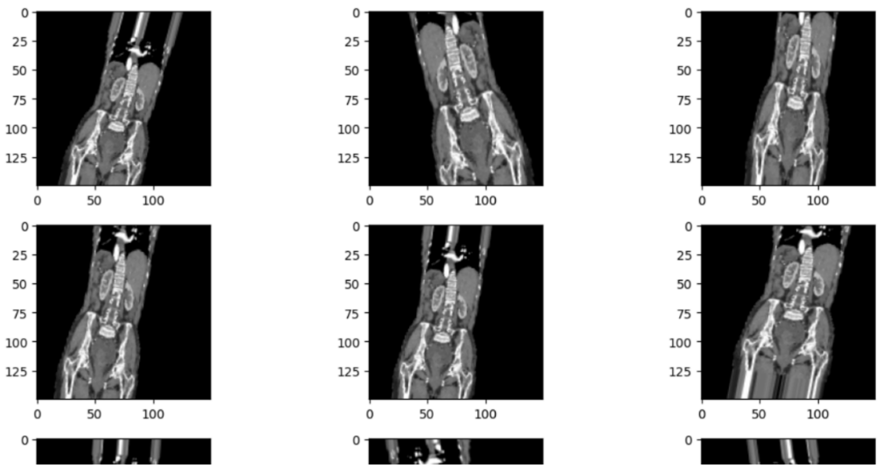


Figure 8. Data Augmentation

- **Image Resizing**: All images were resized to a uniform size of 150x150 pixels. This resizing step is essential because neural networks typically require a consistent input size to process images effectively. The chosen size strikes a balance between computational efficiency and preserving the essential features in the images.

<matplotlib.image.AxesImage at 0x7c6162953310>

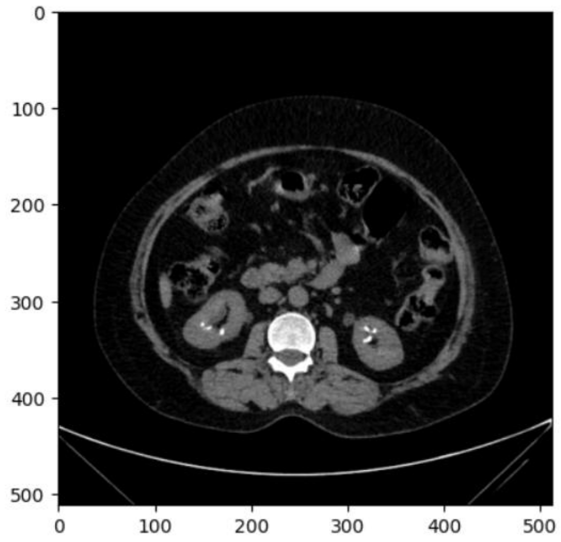


Figure 9. Kidney stone image resizing

- **Normalization:** Pixel values of the images were normalized by scaling them to the range [0, 1]. Normalization helps speed up the convergence during training by ensuring that the pixel values are uniformly distributed. This step is especially critical for gradient-based optimization methods used in training CNNs, as it prevents issues with exploding or vanishing gradients.

4.3 CNN Feature Extraction

A key component of this study is the use of a fully custom-built Convolutional Neural Network (CNN) for feature extraction. CNNs are widely known for their ability to automatically learn hierarchical features from images, eliminating the need for manual feature extraction.

- Custom CNN Architecture: A custom CNN model was designed and implemented for feature extraction. The architecture consists of multiple convolutional layers, pooling layers, and fully connected layers. These layers work together to capture low- and high-level features from the input images, such as edges, textures, and patterns relevant to kidney disease.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 148, 148, 32)	896
batch_normalization (BatchNormalization)	(None, 148, 148, 32)	128
max_pooling2d (MaxPooling2D)	(None, 74, 74, 32)	0
conv2d_1 (Conv2D)	(None, 72, 72, 64)	18,496
batch_normalization_1 (BatchNormalization)	(None, 72, 72, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 64)	0
flatten (Flatten)	(None, 82944)	0
dense (Dense)	(None, 128)	10,616,960
batch_normalization_2 (BatchNormalization)	(None, 128)	512
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 2)	258
activation (Activation)	(None, 2)	0

Total params: 10,637,506 (40.58 MB)
Trainable params: 10,637,058 (40.58 MB)
Non-trainable params: 448 (1.75 KB)

Figure 9. CNN model architecture

- Feature Extraction: Features were extracted from the last convolutional block of the CNN. At this stage, the network has learned to recognize a wide variety of complex patterns and structures. The output of this block is a high-dimensional feature map, which encapsulates the most relevant features of the image for the task of classification. These features are then used for the classification process, either through direct prediction in the case of CNN-based models or by being fed into other models for hybridization.

4.4 Machine Learning Model

In addition to CNN-based feature extraction, traditional machine-learning techniques were employed for comparison and hybridization. Machine learning models typically require manual feature extraction, which can be effective when dealing with smaller datasets or when seeking to augment the feature extraction process of deep learning models.

- **HOG Feature Extraction**: For the machine learning models, the Histogram of Oriented Gradients (HOG) technique was used for feature extraction. HOG is a popular method in image processing that captures edge information by analyzing the distribution of gradients in localized portions of an image. It is particularly useful for detecting shapes, textures, and patterns in medical images, which are key indicators in kidney disease classification.

- **Feature Representation:** HOG features were extracted from the same image dataset used for CNN feature extraction. Each image was processed to compute its HOG features, which were then represented as a feature vector. These vectors were fed into various machine-learning models to classify the kidney images.

Several machine learning models were used for classification, including Random Forests, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN). These models were trained on the extracted HOG features and their performance was evaluated based on the test set.

4.5 Prediction and Combination

The final step in the methodology involves combining the predictions from the CNN and the machine learning models to make a unified decision. This hybrid approach aims to leverage the strengths of both deep learning and traditional machine learning, potentially improving classification accuracy.

- CNN Predictions: Once the CNN model was trained, it was used to generate predictions for the test set. The CNN outputs predictions based on the high-dimensional feature maps extracted from the last convolutional block. These predictions represent the likelihood that an image belongs to a particular class, such as the presence or absence of kidney disease.

- Machine Learning Predictions: Similarly, machine learning models were used to generate predictions based on the HOG features extracted from the images. These predictions also provide a classification label for each image, indicating the likelihood of the presence of kidney disease.

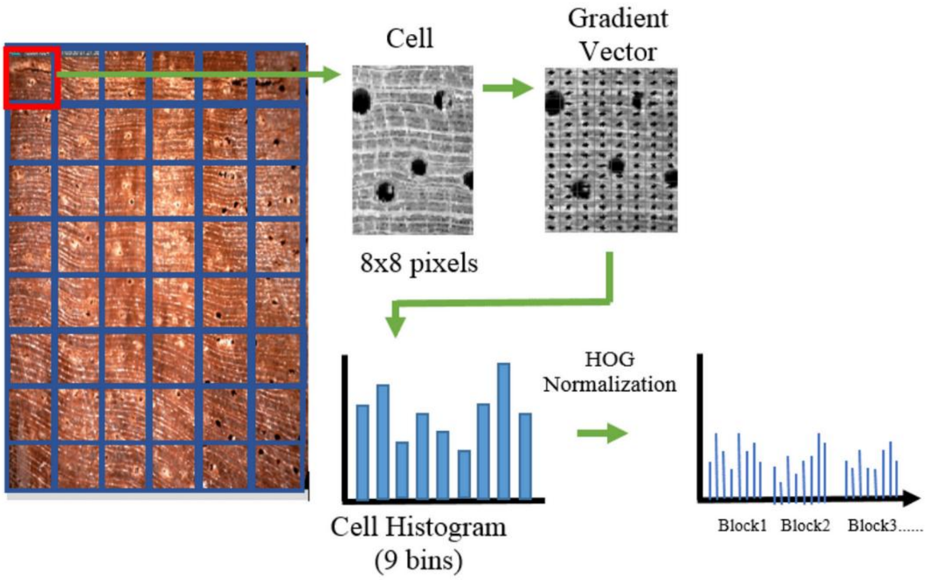


Figure 10. HOG (Histogram of Oriented Gradients) Feature extraction

- Combining Predictions: The final predictions were computed by combining the outputs from both the CNN and the machine learning models. Several approaches can be used for combining predictions, including simple methods like averaging or more sophisticated techniques like weighted voting or stacking. In this study, the predictions from both models were averaged, providing a final classification decision. This hybrid approach allows the model to leverage the complementary strengths of both CNN-based feature extraction and traditional machine-learning classification.

By combining the predictions from both models, the hybrid system aims to improve classification accuracy, reduce overfitting, and enhance generalization, particularly on unseen test data.

5. Experiment and Result Analysis

In this section, we discuss the results of training and evaluation for the CNN and Machine learning models developed to classify kidney disease from images. We provide an analysis of the loss and accuracy plots, which highlight key aspects of the model's learning process and its ability to generalize to unseen data. The following observations are based on the evaluation of training and validation performance over multiple epochs.

5.1 CNN Results

The accuracy plot provides a visual representation of the model's performance on both the training and validation datasets over time. By comparing the accuracy for both datasets, we can assess how well the model can generalize.

- **Accuracy:** The training accuracy steadily increases across the epochs, reaching near 1.0 (100%) towards the end of the training. This suggests that the model is capable of correctly classifying the majority of the training examples. High training accuracy typically indicates that the model has learned to recognize the patterns and features present in the training set well, and its internal parameters are effectively tuned for the training data.

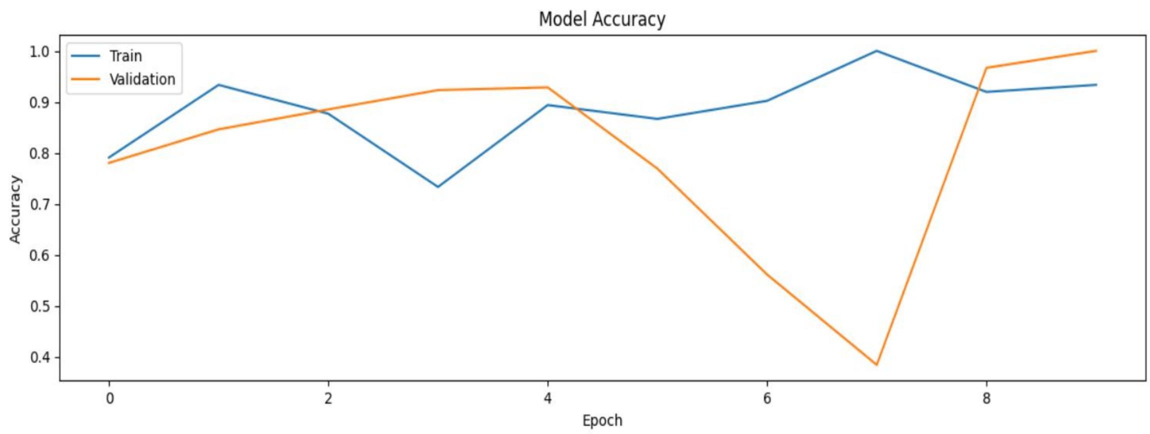


Figure 11. CNN model Training and validation accuracy

- **Loss:** Throughout the training process, the training loss consistently decreases over epochs. This indicates that the model is progressively learning from the training data and refining its internal parameters to reduce the discrepancy between its predictions and the ground truth. The continuous decrease in training loss suggests that the model is successfully adapting to the features in the training set and improving its fit.

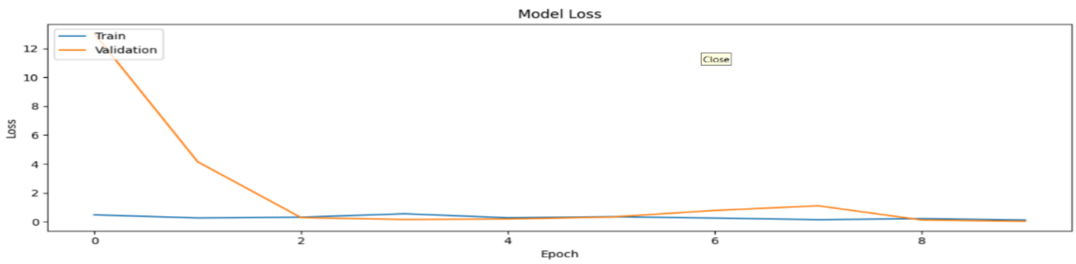


Figure 12. CNN model Loss

- **Confusion matrix on testing data**
  The confusion matrix (Figure 1) provides an overview of the model's performance in distinguishing between the two categories: *Normal* and *Stone*. The matrix reports the following results:
- **True Positives** (Normal classified as Normal): 487 instances were correctly identified as Normal.
- **False Positives** (Normal misclassified as Stone): 22 instances were misclassified as Stone.
- **True Negatives** (Stone classified as Stone): 8 instances were correctly identified as Stone.
- **False Negatives** (Stone misclassified as Normal): 131 instances of Stone were misclassified.

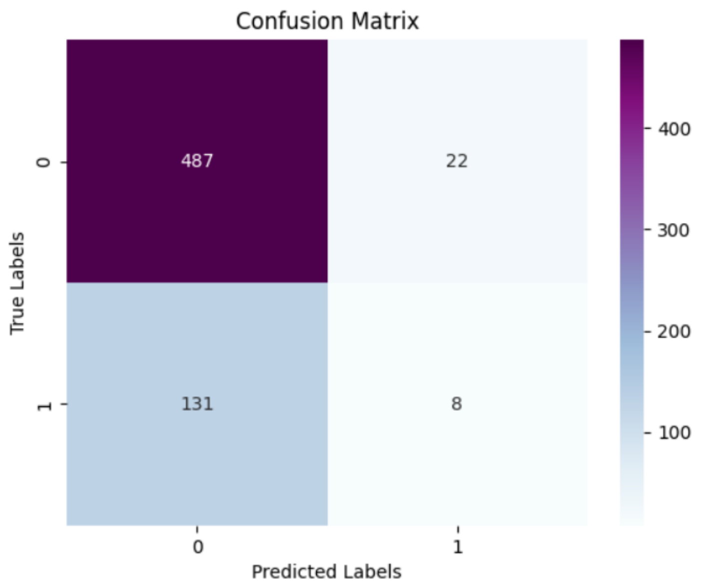


Figure 13. Confusion matrix on testing data

5.2 ML model results

The evaluation of the models was conducted using several metrics, visualizations, and interpretations. Below, the results are discussed in detail.

Table 1. ML model performance on image dataset

Models	Accuracy	Performance
Logistic regression	0.99%	Top
k-Nearest Neighbors	0.99%	
Random forest	0.99%	
AdaBoost	0.97%	Moderate
Decision tree	0.95%	
Gradient boosting	0.96%	
Naïve Bayes	0.84%	Low

Logistic regression, K-Nearest Neighbour and Random Forest performed exceptionally well, achieving near-perfect accuracy. This indicates that the dataset's features are well-suited to linear separability, allowing this model to distinguish between the two categories effectively.

Gradient Boosting, AdaBoost and Decision tree are achieved 96% accuracy, showcasing its capability to handle complex patterns in the data. However, it slightly underperformed compared to simpler models like Logistic Regression and Random Forest, possibly due to overfitting or sensitivity to data imbalance.

The Naïve Bayes classifier achieved the lowest accuracy (84%) among all models. Its assumption of feature independence likely did not align well with the dataset, leading to suboptimal performance. This indicates that the relationships between features play a significant role in classification.

- **Category Distribution Analysis**

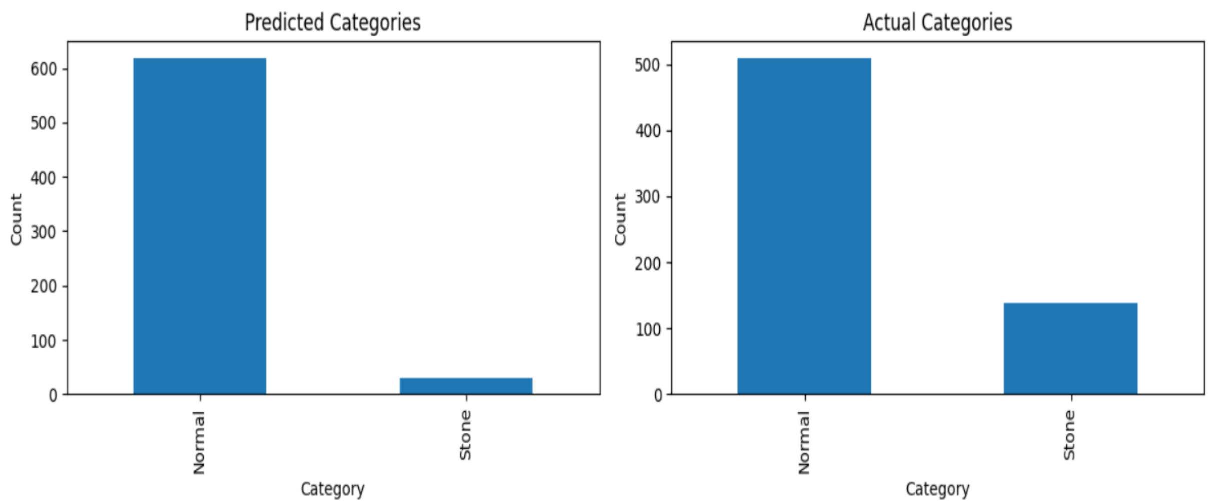
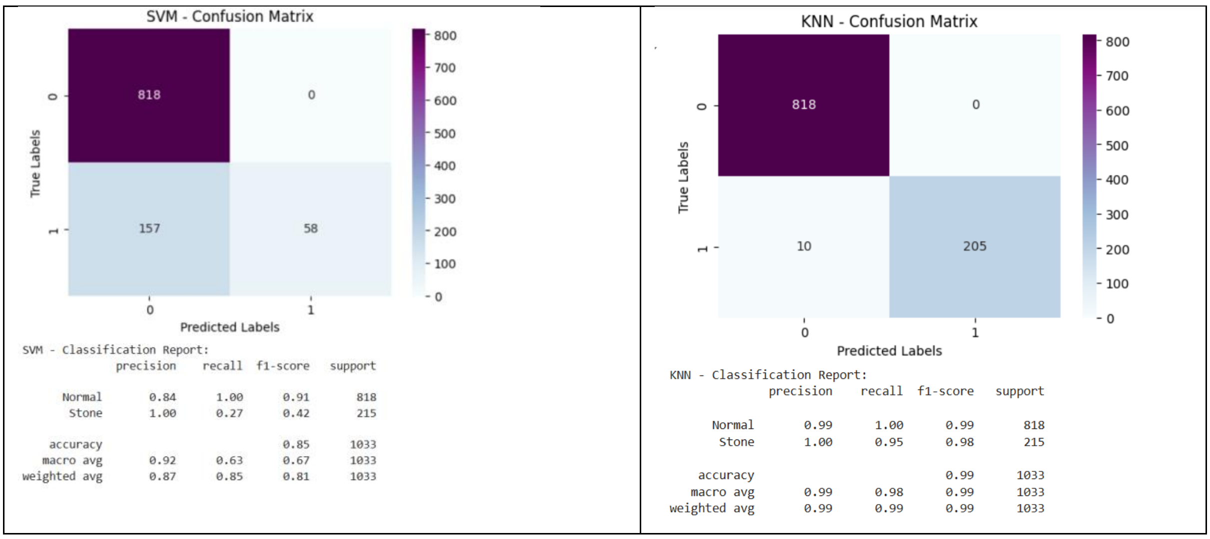


Figure 13. Actual and predicted values by model

1. **Predicted Categories:**
- The Normal category dominates the predictions, with almost all images classified as Normal.
- Very few instances are predicted as Stone, indicating an imbalance in the model's predictions.
2. **Actual Categories:**
- The actual dataset shows a larger proportion of Normal cases compared to Stone cases. However, the disparity between categories is not as stark as seen in the predictions, revealing a bias in the model.
- **Confusion matrix of every Model**



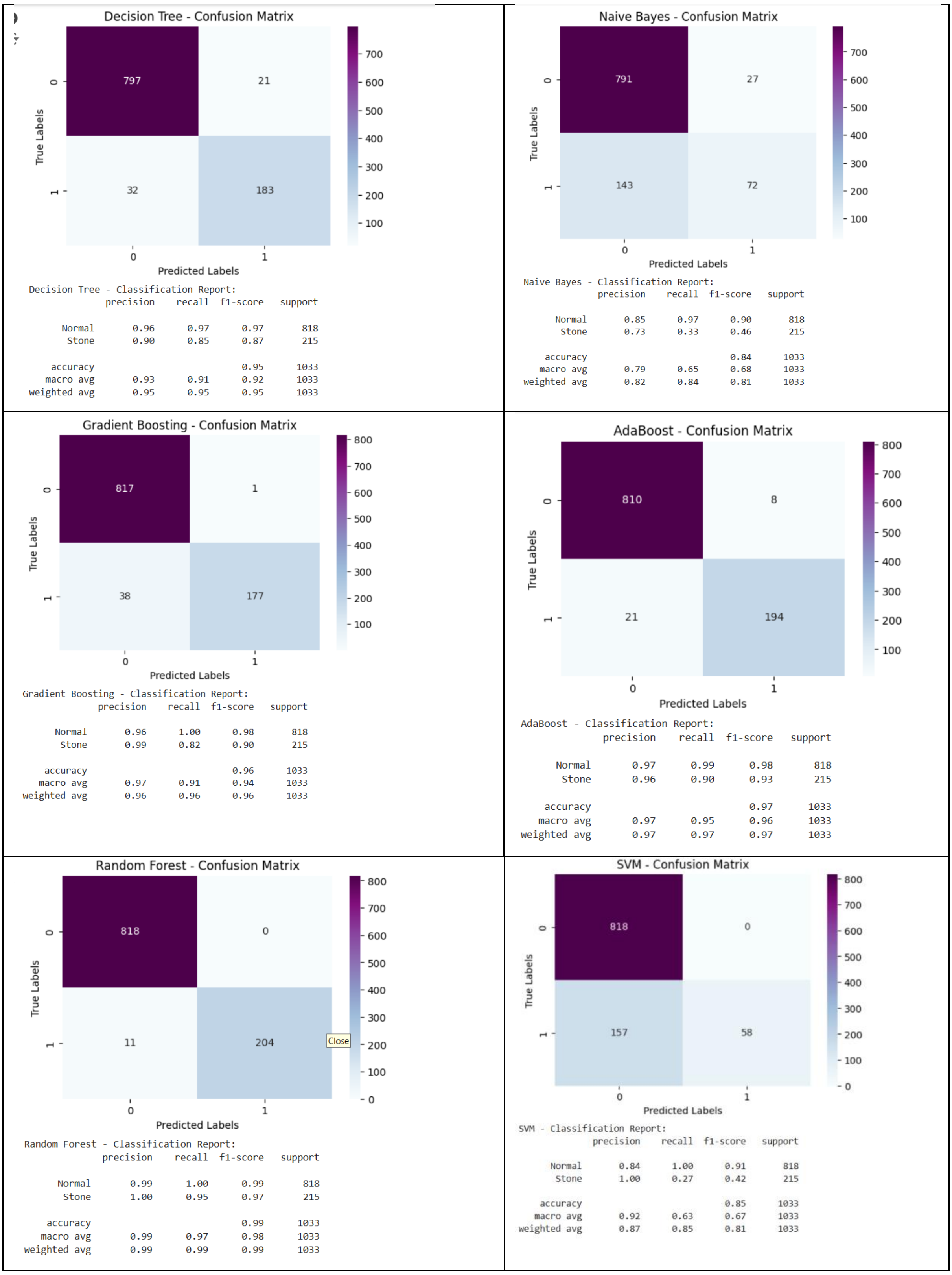


Figure 14. Confusion matrix of ML Models

GUI Implementation for Kidney Stone Detection

After training and saving the machine learning models for kidney stone detection, I created a graphical user interface (GUI) using Python's tkinter library. The GUI allows users to interact with the models seamlessly, providing a user-friendly experience for making predictions.

Features of the GUI:

- **Image Browsing**: Users can upload medical images (e.g., CT scans) through a "Browse Image" button.
- **Model Predictions:** Multiple buttons are available to allow users to select a specific model (CNN, SVM, Random Forest, Logistic Regression, Decision Tree, Gradient Boosting, or k-Nearest Neighbors) to predict whether the image contains kidney stones.
- **Result Display:** After prediction, the result is displayed on the GUI, indicating whether kidney stones are detected or not.

Technical Details:

- Each button triggers the loading of a specific pre-trained model.
- The uploaded image is processed and passed through the model to generate a prediction.
- The result is displayed in real-time within the GUI.

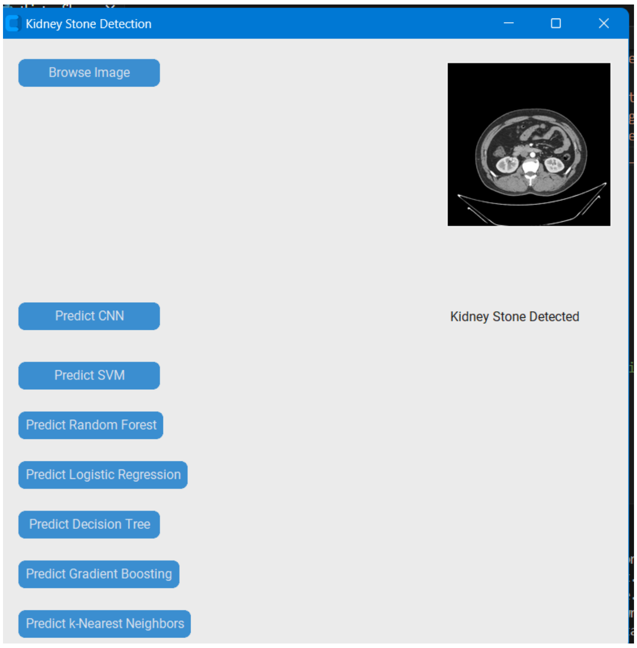


Figure 15. GUI Implementation of models

6. Conclusion

This project successfully combined Convolutional Neural Networks (CNNs) and Machine Learning (ML) models to classify kidney images into Normal and Stone categories. By leveraging CNNs for feature extraction and ML algorithms for classification, the hybrid approach demonstrated high accuracy and robustness, achieving up to 99% accuracy with Logistic Regression, Random Forest, and k-Nearest Neighbors (k-NN).

CNNs provided end-to-end learning and extracted meaningful features, while traditional ML models trained on these features offered simplicity and competitive performance. Ensemble methods like Gradient Boosting and AdaBoost also performed well, though slightly less accurate. Naïve Bayes, with an accuracy of 84%, struggled due to its assumptions, emphasizing the importance of selecting suitable algorithms.

Validation accuracy fluctuations in CNNs suggested minor overfitting, which could be mitigated with regularization, early stopping, or data augmentation. Challenges like class imbalance and limited data also highlighted opportunities for improvement.

This project underscores the potential of combining modern deep learning with traditional ML techniques for robust and interpretable solutions. Future work could focus on hyperparameter optimization, expanding the dataset, and testing on real-world data. Overall, the study establishes a strong foundation for applying hybrid approaches to medical image analysis, with significant potential for broader applications.

References

[1] Alelign T, Petros B. Kidney Stone Disease: An Update on Current Concepts. Adv Urol. 2018 Feb 4; 2018:3068365. doi: 10.1155/2018/3068365. PMID: 29515627; PMCID: PMC5817324.

[2] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," in *IEEE Access*, vol. 8, pp. 20991-21002, 2020, doi: 10.1109/ACCESS.2019.2963053.

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks.

[4] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection.

[5] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition.

[6] Breiman, L. (2001). Random Forests.

[7] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System

[8] Kumar, R., Chen, H., et al. (2022). *Hybrid Approaches in AI for Kidney Disease Classification*.