Kidney Stone Detection Using CNN’s

Submitted by:

Anirudh Allaparthi(hm1093)

Ashwini Awathe(ho4124)

Dinesh Arasavalli(hm3731)

Harsha Vardhan Aitha(hq4797)

Sravanthi Lankalapalli(hp8500)

Department of Industrial Engineering

Wayne State University, MI

Submitted to Dr. Ratna Babu Chinnam

IE-7860

ABSTRACT

We provide a comparative study of Convolutional Neural Networks (CNNs) for kidney stone diagnosis from CT scan pictures in the ongoing effort to advance medical diagnostics through machine learning. Our study uses a robust dataset of standardized 128x128 pixel images to distinguish between two different methods: one using a CNN without transfer learning and the other using a transfer learning framework. In order to improve image quality and enable more effective feature extraction, preprocessing techniques including median blurring and histogram equalization were used, which closely resembled the conditions of medical imaging in the real world.

Our results showed that the bespoke CNN model, built and trained from scratch, outperformed its transfer learning model, obtaining an accuracy of 92% compared to 90.50%, which defies prevalent trends in deep learning. The conventional CNN approach demonstrated a higher true positive rate, as evidenced by the confusion matrix and classification metrics, indicating its robustness in reliably detecting kidney stones. With this study, we contribute to a more prudent use of transfer learning in the field of medical image analysis and add to the complex discourse on the effectiveness of deep learning models in this area.

Kidney stones have become quite common in recent years. If they are not detected early, they can cause problems and require surgery to remove the stone. According to earlier studies, volumetric measures of kidney stones are more reliable and reproducible than linear measurements. Stone detection may be aided and the work involved in manual

stone detection reduced by Deep Learning-based algorithms that employ non-contrast abdominal computed tomography (CT) scans. A dataset of CT scans is used to identify the stone, which includes CT scans with and without manually indicated kidney stones. After performing image processing techniques on the raw CT scan images, Convolutional neural network is implemented to find the optimal parameter values for the deep learning model.

Introduction

Kidney stones are a widespread health issue that affect millions of people worldwide. The increase in cases indicates a growing health issue that frequently remains undetected until it produces excruciating agony. The development of kidney stones can result in a range of symptoms, from severe stomach pain to none, prompting several treatment techniques. In order to prevent more serious issues and to protect the well-being of people afflicted, early and accurate detection is crucial. Large stones may necessitate complex medical procedures.

Medical diagnostics have made significant strides due to developments in Deep Learning (DL), specially in identifying health concerns by analyzing images. The use of DL in urology has tremendous potential for differentiating kidney stones from other anatomically similar sections on CT imaging. The objective of this study is to accurately detect kidney stones by analyzing the subtle differences in CT images by utilizing the sophisticated capabilities of DL algorithms. By doing this, we aim to improve the field of urinary health diagnostics and reduce diagnosis errors.

Objectives:

- To acquire and preprocess a comprehensive dataset of CT scan images for kidney stone detection.
- Implement advanced image processing techniques to enhance image quality and facilitate accurate detection.
- Developed two models: one built with custom configurations, and the other utilizing a pretrained architecture.
- Evaluate the performance of the model using various metrics and validate its effectiveness in a clinical setting with and without transfer learning and evaluate the effective method.

METHODOLOGY

Data gathering

The kind of problem we are attempting to address determines how the dataset is gathered. Considering the primary focus of this project is image classification, we need to acquire the required resources from open-source websites like Kaggle, Github, etc. The link to the dataset is given below:

https://github.com/yildirimozal/Kidney_stone_detection/tree/main/Dataset

Data Characteristics

The dataset consists of approximately 1700 images categorized into two folders named "Kidney_Stone" and "Normal". These images are of variable sizes and are unlabeled; however, their folder names serve as class labels for classification purposes.

Data Preprocessing

As part of preprocessing, all images are standardized to a uniform size of 128 x 128 pixels to maintain consistent input dimensions for the machine learning model, which helps prevent biases that could arise from varying image sizes. The dataset is thoroughly shuffled to ensure an unbiased training process. Subsequently, images are converted into Numpy arrays with corresponding class labels. For the development and evaluation of the model, the data is split into two subsets: 90% is used for training, and the remaining 10% is set aside for testing the model's performance.

Image before processing:

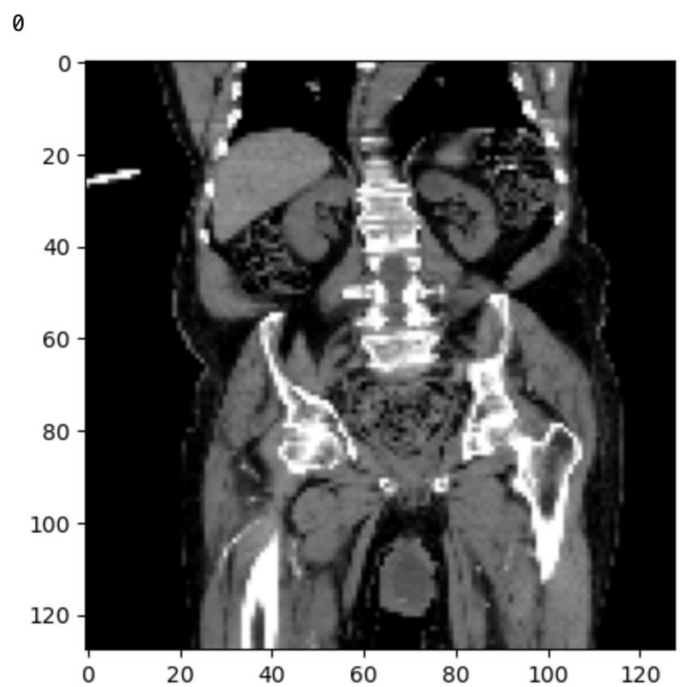


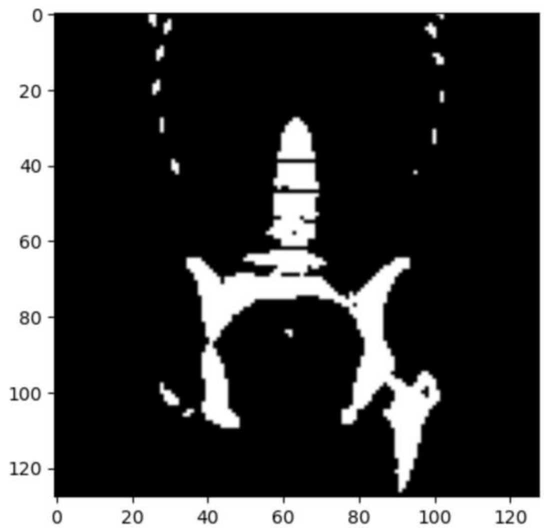
Image Enhancement

It is basically one of the critical tasks because CT Scan images may have noise.To enhance the image quality and improve the model's learning capability, we implemented several preprocessing steps on the training images. These steps included

Median Blurring is a technique we applied extensively to reduce image noise that could obscure kidney stone features. The values of each pixel inside a specified kernel or window size are substituted with the median value of the surrounding pixels. With this technique, the 'salt-and-pepper' noise that is frequently present in CT scans.
Histogram Equalization plays a crucial role in enhancing the contrast of CT scan images by redistributing the intensity values across the available range, particularly improving visibility in low-contrast areas. It adjusts the overall lighting of the images thereby creating a more uniform dataset.

This preprocessing not only aids in better feature extraction by the neural network but also mimics conditions that healthcare professionals might face, dealing with varying quality of medical images.

Below shows the image of CT scan after converting to grey scale highlighting the stone area.



Model Building

1. Building a CNN model

The construction of a convolutional neural network (CNN) is done using Keras, a high-level neural networks API. This model is designed for binary classification tasks, such as determining kidney stones in CT scan images.

The model starts with a Sequential base, indicating that the layers are stacked linearly. It begins with a **Convolution2D** layer with 64 filters and a kernel size of 3, using the ReLU activation function, optimized for processing 128x128 grayscale images (as indicated by the input shape). This is followed by a MaxPooling2D layer with a pool size of 3x3, which reduces the spatial dimensions of the output from the convolutional layer, helping decrease the computational load and control overfitting.

Another convolutional layer follows, this time with 112 filters, again using the ReLU activation function and followed by another max pooling layer. After the convolutional and pooling layers, the model employs a **Flatten** layer to convert the multi-dimensional input into a one-dimensional array for processing by the dense layers. The network includes a dense layer with 80 units and a dropout layer set at 0.5 to prevent overfitting by randomly omitting a fraction of the neurons during training.

The final layer is a dense layer with a single unit and a sigmoid activation function, which outputs a probability indicating the presence or absence of a kidney stone. The model is compiled with the RMSprop optimizer (with a learning rate of 0.001) and binary crossentropy loss function, appropriate for binary classification tasks. It tracks accuracy as a metric to evaluate the model's performance during training. This architecture is well-suited for extracting and learning complex features from medical images, ultimately aiding in effective medical diagnostics.

Model Summary:

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 64)	640
max_pooling2d (MaxPooling2D)	(None, 42, 42, 64)	0
conv2d_1 (Conv2D)	(None, 40, 40, 112)	64,624
max_pooling2d_1 (MaxPooling2D)	(None, 13, 13, 112)	0
flatten (Flatten)	(None, 18928)	0
dense (Dense)	(None, 80)	1,514,320
dropout (Dropout)	(None, 80)	0
dense_1 (Dense)	(None, 1)	81

Total params: 1,579,665 (6.03 MB)
Trainable params: 1,579,665 (6.03 MB)
Non-trainable params: 0 (0.00 B)

Results:

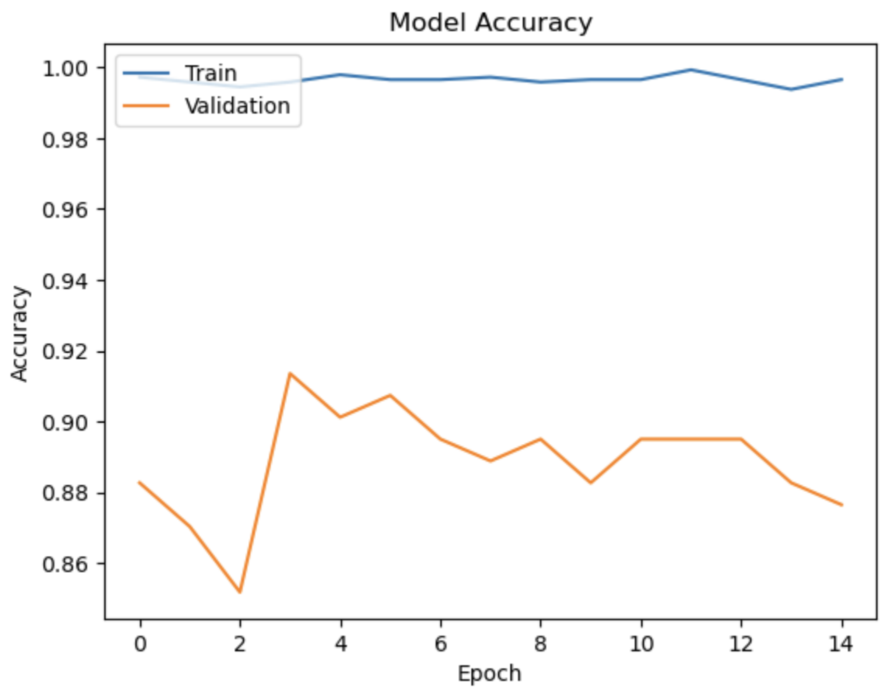
```
[ ] from sklearn.metrics import confusion_matrix
    cc=confusion_matrix(ytest,(ypred>0.5)*1)
    cc
```

```
array([[94,  4],
       [10, 71]], dtype=int64)
```

```
from sklearn.metrics import accuracy_score
accuracy_score(ytest,(ypred>0.5)*1)*100
```

```
92.17877094972067
```

Visualizations:



2. CNN with Transfer Learning

For the transfer learning approach, we leveraged existing pretrained models like ResNet50 and Xception. These models are highly sophisticated and have been trained on vast datasets like ImageNet, providing a robust set of feature detectors that can be beneficial for medical image analysis.

The VGG16 model was employed as the foundational architecture, with the top fully connected layers excluded to allow for customization specific to our task. The base layers were frozen to retain the learned features from ImageNet, preventing them from being updated during training. Custom layers were added on top of the base model to cater to our binary classification needs, including a Flatten layer, a Dense layer with ReLU activation for learning non-linear relationships, a Dropout layer to reduce overfitting, and a final Dense layer with sigmoid activation for binary output.

Model Compilation and Training

The model was compiled with the Adam optimizer and binary crossentropy loss function, reflecting our focus on binary classification. It was trained over 10 epochs with a batch size of 32, using both training and validation datasets to monitor performance and avoid overfitting:

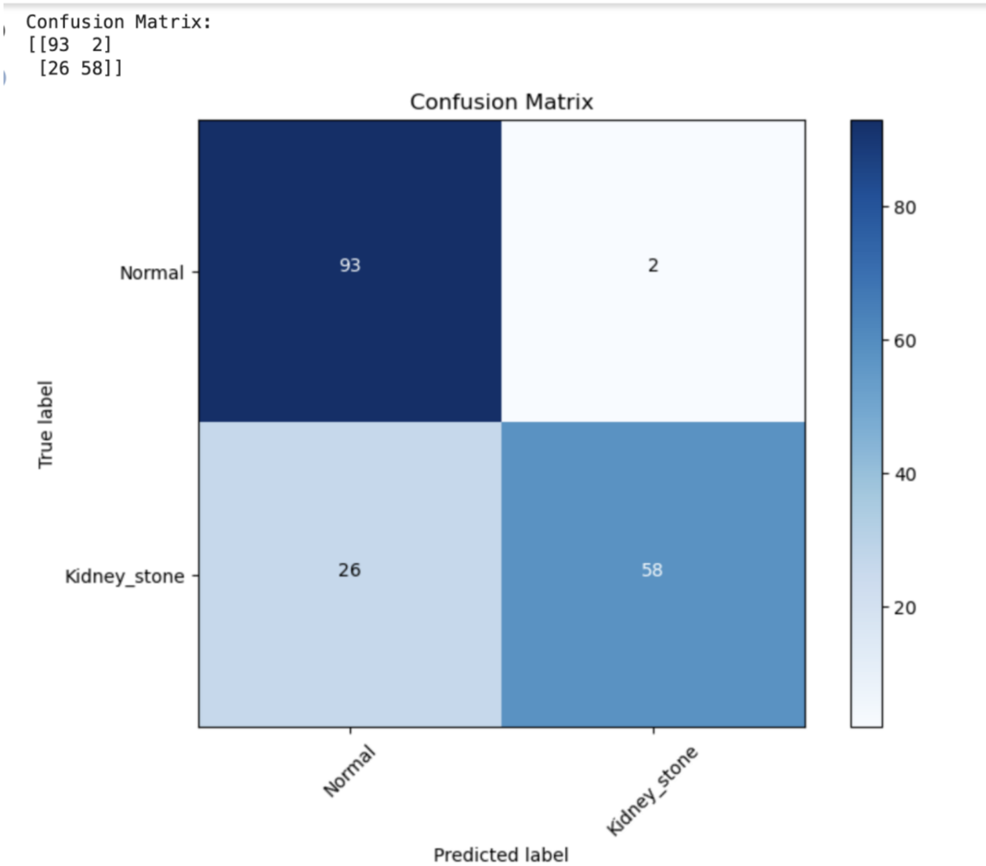
Performance Evaluation

The evaluation phase involved several key metrics:

- **Accuracy**: Measures the overall correctness of the model on the test data.
- **Confusion Matrix**: Provides a detailed breakdown of the model's predictions, offering insights into the true positives, false positives, true negatives, and false negatives.
- **Classification Report**: Includes precision, recall, and F1-score for a more nuanced understanding of model performance.

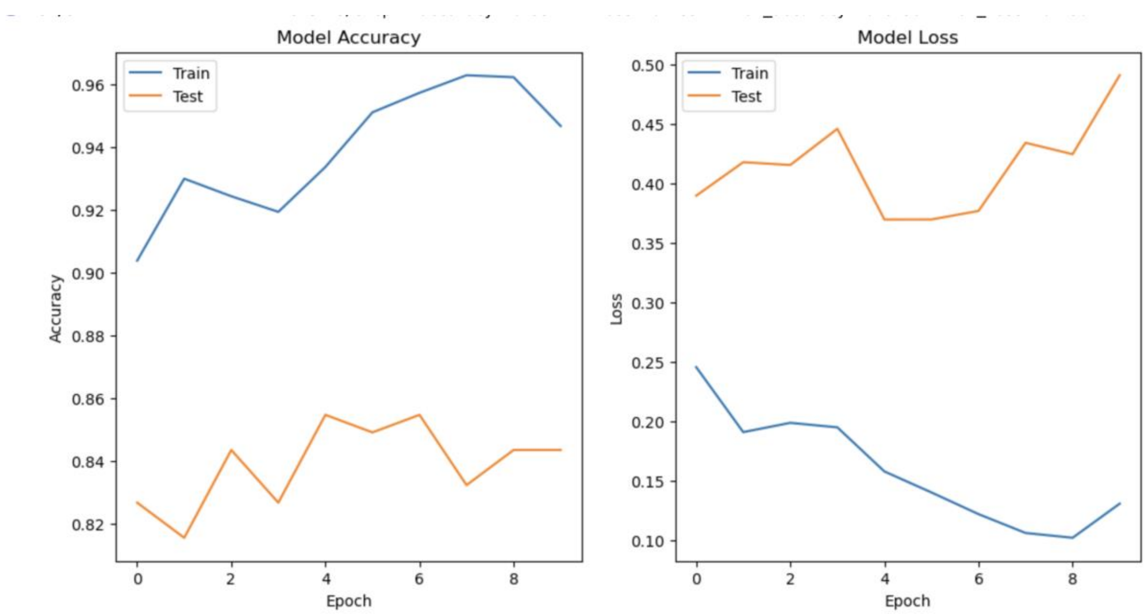
	precision	recall	f1-score	support
Normal	0.78	0.98	0.87	95
Kidney_stone	0.97	0.69	0.81	84
accuracy			0.84	179
macro avg	0.87	0.83	0.84	179
weighted avg	0.87	0.84	0.84	179

The accuracy of the model presented is 84%.



As per the confusion matrix, the model correctly predicted 'Normal' 93 times and 'Kidney_stone' 58 times, while it incorrectly predicted 26 'Normal' cases as 'Kidney_stone' and 2 'Kidney_stone' cases as 'Normal'.

Model Accuracy and Loss Graph:



To visually assess the training process and outcomes, we plotted accuracy and loss metrics over the training epochs. These plots are vital for diagnosing issues like overfitting or underfitting and for understanding the learning dynamics of the models.

The training accuracy started high and increased over epochs, while the test accuracy varied but generally increased, indicating some variation between the training and test datasets.

The model loss graph shows a decreasing trend in training loss over epochs, implying that the model is learning effectively, while the test loss exhibits volatility, peaking before rising again.

Comparative analysis

The model without transfer learning achieved an accuracy of approximately 92%. In contrast, the model with transfer learning showed an overall accuracy of approximately 84%.

Transfer learning usually provides a robust starting point due to pre-trained weights on large datasets, often leading to better generalization; however, in this case, the model without transfer learning performed exceptionally well. This could be due to various reasons such as the specificity of the data, overfitting of the transfer learning model to the training data, or even the architecture and hyperparameters being better tuned in the model without transfer learning. The confusion matrix also supports this observation, where the non-transfer model misclassified fewer 'Normal' cases as 'Kidney stones' but struggled slightly more with the opposite.

The comparative analysis reveals that while transfer learning is a powerful tool, it doesn't always guarantee superior results, and there are scenarios where a well-constructed and trained model from scratch can outperform despite the theoretical advantages of transfer learning.

Explainable AI

In our image classification project, we did not incorporate Explainable AI (XAI) methods primarily due to the inherent complexity of interpreting image data and the challenges posed by the multi-layered structure of deep learning models. Image data is high-dimensional and contains complex spatial relationships, making it difficult to pinpoint exactly what features the model is using to make predictions. Additionally, convolutional neural networks, which are extensively used in image classification, involve multiple layers of abstraction. These layers progressively extract and process features from basic textures and edges in earlier layers to more complex shapes and object parts in deeper layers. This layered complexity complicates the direct interpretation of how each layer influences the final decision, presenting a significant barrier to implementing straightforward XAI methods.

Moreover, the general applicability and scalability of existing XAI techniques also influenced our decision to forego their use in this project. Techniques that are effective for simpler, more transparent models or structured tabular data often do not translate well to the complexities of image classification models without significant adaptations. Methods such as LIME or SHAP, while useful, would require considerable modification to adequately address the unique spatial and feature dependencies present in image data. Additionally, the computational overhead associated with these XAI methods can be prohibitively high, particularly when dealing with the large models and extensive datasets typical in image classification tasks. This combination of required adaptations and high computational demands made the integration of XAI methods impractical within the scope and resources of our project.

Conclusion and Future Work

The project showed CNNs' capabilities in medical imaging, specifically regarding kidney stone identification. The distinction between a transfer learning strategy and a custom CNN architecture shed light on the potential applications of cutting-edge deep learning methods in the medical field. The results suggest potential paths for further research, including the exploration of other architectures and the integration of additional data sources for model training. In the future, these models will be used in real-world medical settings and clinically validated with the goal of improving patient outcomes and diagnostic accuracy at healthcare facilities across the globe.

References

1. Chen Z, Prosperi M, Bird VY. Prevalence of kidney stones in the USA: the National Health and Nutrition Evaluation Survey. J Clin Urol 2018 Nov 26 [Epub ahead of print] https://doi.org/10.1177/2051415818813820.

2. Foster G, Stocks C, Borofsky MS. Emergency Department Visits and Hospital Admissions for Kidney Stone Disease, 2009: Statistical Brief #139. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville, Md: Agency for Healthcare Research and Quality, 2012. https://www.ncbi.nlm.nih.gov/pubmed/23016164.

3. American College of Radiology. ACR appropriateness criteria. Acute onset flank pain-suspicion of stone disease (urolithiasis). https://acsearch.acr.org/docs/69362/Narrative/.

4. Westphalen AC, Hsia RY, Maselli JH, Wang R, Gonzales R. Radiological imaging of patients with suspected urinary tract stones: national trends, diagnoses, and predictors. Acad Emerg Med 2011;18(7):699–707. Crossref, Medline, Google Scholar

5. Fwu CW, Eggers PW, Kimmel PL, Kusek JW, Kirkali Z. Emergency department visits, use of imaging, and drugs for urolithiasis have increased in the United States. Kidney Int 2013;83(3):479–486.

6. Wang DC, Parry CR, Feldman M, Tomlinson G, Sarrazin J, Glanc P. Acute abdomen in the emergency department: is CT a time-limiting factor? AJR Am J Roentgenol 2015;205(6):1222–1229.