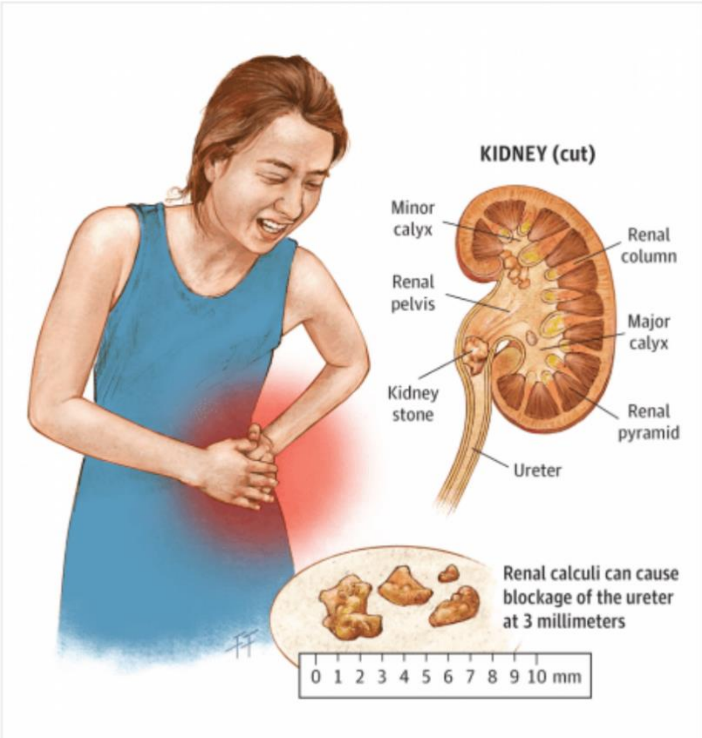
Proof of Concept of Automated Kidney Stone Detection Using Machine Learning And A Dataset of Augmented CT Images

A proposed machine learning model engineered to detect urolithiasis by automatic interpretation of synthetic CT abdominal images.



Executive Summary:

Kidney stone disease is a common health disorder in the Western world with a reported lifetime prevalence of 10–12% in men and 5–6% in women. Symptoms of the disease include severe flank pain, painful urination, and vomiting. Complications can include kidney injury, infections within the kidneys, and septicaemia. The rapid diagnosis and treatment of kidney stones reduces pain and the incidents of complications.

Diagnosis is made via Computer Tomography (CT) imaging of the kidney, ureters, and bladder (KUB). A radiologist interprets the CT scan volume to determine whether a stone or 'calculus' is present and whether it is pathological. This interpretation could possibly be automated by machine learning.

In this project we seek to engineer a machine learning model which can reliably detect urolithiasis in CT images. Currently, a dataset of CT KUB volumes is being labelled at Austin Health for this purpose. However, the use of this dataset is restricted by medical privacy permissions, and it is not available yet, nor will it be allowed to be stored on the cloud. Neither is there a public dataset of labelled CT KUB volumes available online. Therefore, in order to develop the model, we will augment a publicly available dataset of CT volumes to add features that resemble kidney stones to them. We will then train the model on that dataset and report its performance.

We hope to establish that the automated interpretation of CT KUB volumes in order to detect kidney stones is feasible, and to estimate how large the dataset must be to achieve acceptable performance across relevant metrics.

Background

Kidney stone disease is a common reason for presentation of patients to the Emergency Department. The passing of a kidney stone down the urinary tract can be incredibly painful. Some say that it is more painful than childbirth. Yet they are also very common. Patients who have this condition can be treated with agents that cause the ureter to relax, which helps the stone pass more quickly and with less trauma through the urinary tract. We can also use Lithotripsy, also known as extracorporeal shockwave lithotripsy (ESWL), which uses shockwaves to break up the kidney stone into small pieces. In severe cases can perform an emergency surgery, and cut the stone out.

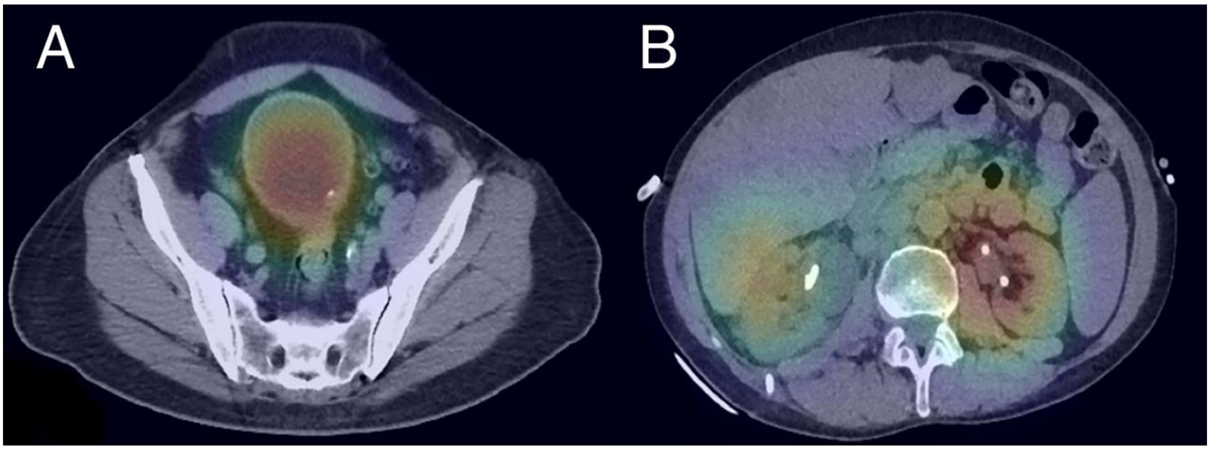
It is good for patients with this condition to be CT-scanned and diagnosed quickly, to avoid the agony of waiting hours for treatment. However, even after the scan is complete, a radiologist still needs to interpret the resultant volume. There are factors in this process which lead to poor outcomes for the patient. Some stones are very small, and can be missed. Sometimes a stone is present, but it is not causing the patient's symptoms, and the risk of a misdiagnosis is therefore raised.

The timely and correct interpretation of CT KUB volumes is necessary to avoid these poor outcomes. Given the successes of machine learning (ML) applications across other domains in radiology, it is likely that the interpretation of CT KUB volumes could be automated accurately by similar methods.

Relevant Major Publications

1. **Urinary Stone Detection on CT Images Using Deep Convolutional Neural Networks: Evaluation of Model Performance and Generalization**

This 2019 study used a dataset of 535 CT KUB volumes to train a cascading Convolutional Neural Network (CNN) model to detect urinary tract stones on unenhanced CT scans with a high accuracy (AUC, 0.954). This study was of high quality, published in RSNA, and the model received full CT KUB volumes as input.



2. **Deep learning model for automated kidney stone detection using coronal CT images**

This 2022 study used a dataset of 433 subject's CT KUB volumes to train a CNN model which showed an accuracy of 96.82% in detecting kidney stones on CT slices. This study had limitations. The model only received a few pre-selected coronal (vertical) slices of the CT KUB volume as input, and therefore does not technically predict kidney stones from CT KUB volumes, but rather gives a result 'per slice' of a preselected subset of slices.

Proposed Methods

Radiologists typically view CT volumes as a number of discrete slices, which they scroll through. This is sufficient in most cases to make a diagnosis. In some more challenging instances, the Radiologist will scroll through the slices in three different ways, once in in each of the three orientations: coronal (front-to-back), sagittal (side-to-side), and horizontal (head-to-toe).

The interpretation of CT volumes will require the model to use a 3D volume as input. The question is whether to use a 3-D convolutional neural network, or to use a model which takes discrete sets of 2-D slices as inputs, such as a 2-D convolutional neural network which iterates through slices.

It is apparent from the most popular Kaggle competitions that use CT volumes as inputs that the latter approach is usually superior. Especially if a pretrained 2-D Convolutional Neural Network is utilised, such as GreyNet, which was implemented in the RSNA publication listed above. Additionally, the preprocessing methods for extracting useful features from 2-D images which were taught in CSCI-E25 can be implemented in the model and their effects on performance analysed.

Given that this is a proof of concept, I will also attempt to train a 3-D convolutional neural network which receives 3-D volumes. This will be possible due to the processing power available with cloud computing.

Datasets for this project

Public dataset rather than private

This project will develop the methods required to detect urolithiasis on real CT images. Since I will use a public dataset, the data used within this project will not be constrained by the same medical privacy conditions as the Austin Health dataset of CT images. Therefore, the necessary computations can be run on a cloud computing setup. This will enable us to train quickly, in an environment in which we can utilise state-of-the-art methods, and brings the project within the scope of CSCI E-25's Graduate Independent Project. The necessary trade-off is that I may be required to use augmented data rather than real CT KUB images of kidney stones.

Candidate Datasets

1. **Kits 2019 Kidney cancer dataset. Size 300 https://kits19.grand-challenge.org/data/**

This dataset is the best candidate on preliminary review. It is a dataset which was curated for competition and as such was made accessible. It contains CT images of the kidney, and I believe I can likely augment it to my purposes. Although I would prefer a larger dataset than size 300, I am reassured by the fact that the 2019 RSNA study listed above used a dataset of size 535, which is on the same order of magnitude.

I will need to investigate whether the dataset's CT volumes extend to the full visualisation of the urinary tract. A CT KUB volume extends to the ureter, bladder, and urethra.

2. **ACRIN CT Colonography trial dataset, Size 836 https://wiki.cancerimagingarchive.net/display/Public/CT+COLONOGRAPHY#3539213deca74644fb24f819129e8bb8f2a3658**

This dataset is the next best candidate on preliminary review. It is a subset of a dataset of CT Colonography images related to a trial conducted by ACRIN and published in NEJM in 2008. This subset which was made publicly available in 2021 via The Cancer Imaging Archive (TCIA) Public Access. It is of good size, and images the abdomen, which is a similar region to that imaged by CT KUBs.

The limitation of this dataset is that CT Colonography studies are not typically used to detect kidney stones. If I use this dataset, I will make it clear that my dataset is systematically different to the CT KUB volumes on which kidney stones are actually detected. Any performance metrics will need to be conservatively reported and discussed, as successful results may not generalise CT KUB.

Augmentation and Preprocessing of Data

If we are only able to obtain a dataset of CT abdominal volumes *without* kidney stones, I believe it will be easy to add stone-like features to them (stones are simple obloid-like objects) using basic CV methods (e.g add one sphere of constant density select radius to a select region of the volume) I think it's feasible that the resultant augmented dataset will resemble a real dataset closely enough for practical purposes. Since I don't have a full real dataset I can use as a comparator, I will make subjective comments about the similarities, and why I believe the result of my augmentation is sufficiently similar to a real volume.

In fact, the distinguishability of these augmented volumes from the real slices could itself be tested by training a model to perform the task. There is also the option, which is outside the scope of this project proposal, to design a generative adversarial network (GAN) that generates CT volumes with certain characteristics.

If we also work with universal image types (e.g DICOM), we can run this project in such a way that it will be possible to apply the methods which we develop to a real dataset of CT KUB images. To that end, I will design the project so that the preprocessing of the individual DICOM files is automated, and that the data will not be unnecessarily reworked by a human before it can be used an input to the model.

Final Comments

I believe that this project is valuable for a number of reasons.

First, there's a lot of room for improvement in the task of automated detection of kidney stones by machine learning. The leading publication in RSNA (as far as I'm aware used a dataset of size 300, which is not large. The automated flagging of CTs containing kidney stones and the computer assisted diagnosis of these CTs can both assist the Radiologist in delivering excellent care. I believe that what is being used in the Emergency Department today is extremely far from what is possible today, and we *should* be giving the patients the best possible care.

The methods will not be difficult to implement. Python has several libraries for dealing with DICOM files. FastAI provides a suitable high level library of functions for processing and displaying them, which will make communicating the elements of the project much easier. There are a multitude of implementations of CT interpretation which are published on Kaggle, and already made to run on cloud environments with reasonable hardware requirements.

Finally, the methods developed here will be implemented at a later date on a real dataset of approximately 3000 labelled CT volumes which already exists with at Austin Health. If the methods are transferrable, there will be a publication that I believe will be strong for two reasons. First, it will be strong on the basis of the size of the dataset, which is five times the size of the RSNA publication's dataset. Secondly, in 80 hours of project development here, I can engineer high performing model which makes the best use of the state-of-the-art architecture and pretraining currently available.