Kidney Stone Detection using Deep Learning Methodologies

MSc Research Project
MSc in Data Analytics

Alexander Albert Irudayaraj
Student ID: 20172176

School of Computing
National College of Ireland

Supervisor: Dr. Cristina Hava Muntean

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Alexander Albert Irudayaraj
Student ID:	20172176
Programme:	Msc in Data Analytics — **Year:** 2021-2022
Module:	Msc Research Project
Supervisor:	Dr. Cristina Hava Muntean
Submission Due Date:	31-01-2022
Project Title:	Kidney Stone Detection using Deep Learning Methodologies
Word Count:	10256 words — **Page Count**: 23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………

Date: 31-01-2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Kidney Stone Detection using Deep Learning Methodologies

Alexander Albert Irudayaraj
20172176

Abstract

Nephrolithiasis is a condition in which undesired sediments are deposited in the kidneys, interfering with the normal functioning of the urinary system and, in some cases, blocking urine flow, causing excruciating agony. The ability to detect kidney stones from medical imaging is thus critical for providing effective and timely medication. Deep Learning methods can be used to accurately perform object detection from images. The application of deep learning methods in kidney stone detection will help avert the percentage of errors that currently arise due to human error. In this paper four specific deep learning methods have been employed to detect whether kidney stones are present or not in CT scan images. The four algorithms used are VGG16, ResNet50V2, MobileNetV2, InceptionNetV3. A dataset containing 1799 CT scan images of kidneys was used for building these models to perform kidney stone detection. The classification performance of all four models were assessed using accuracy, precision, and recall metrics. InceptionNet neural network produced the best classification results in terms of accuracy, precision and recall. It produced an accuracy of 0.862, precision of 0.866 and recall of 0.8331. These measures are higher than the corresponding values for other three models by 11%, 10.5%, 2.9% respectively, hence this research confirms that among the four algorithms under consideration, InceptionNet is to be employed for automatic detection of kidney stones.

1 Introduction

The human body contains a complex combination of multiple chemical substances which are responsible and vital for performing different functions such as digestion, blood circulation. The important responsibility of regulating these chemical substances and their concentration in the human body is carried out by a pair of organs called Kidneys. Kidneys are bean-shaped organs located beneath the rib cage on either side. Kidneys also perform excretion of unwanted components in the form of urine, control the production of red cells, have a direct impact on the blood pressure. Each kidney possesses large number of filtering units called the nephrons, which retain only the required chemicals while unwanted substances are removed. The normal operations of the kidneys can be impacted by external factors such as diabetes, high blood pressure. Kidney stones are miniscule lumps of unwanted deposits of minerals, and these are one of the most influential factors for the malfunctioning of the urinary system. The presence of small lumps of sediment in the kidneys, known as kidney stones, is the most common cause of kidney malfunction. Sometimes, excess minerals and salts can stay back in the urinary tract and overtime get built up to hard deposits in the urinary tract and these are called kidney stones. When these stones are small, the functioning of the kidneys may be unaffected; however, when these particles develop to a significant size, they can cause urine blockage, which can result in malfunctioning of the kidneys This causes a great deal of discomfort in the back, thus early detection of such kidney stones is critical in reducing pain in people who are suffering from this uncommon occurrence.

Currently, a variety of imaging modalities are used to detect unwanted sediments in the kidneys. X-ray images, Computed Tomography images and Ultrasound images are some of the examples of imaging techniques being used for kidney stone detection. The CT scan approach is the most widely used of all three procedures (Bierig and Jones 2009) . The radiologists oversee the task of interpreting the scans and determining whether stones are present in the kidney. The practice of clinical radiology has been determined to have a 4% mistake rate (Sabih et al. 2011). Although 4% may seem insignificant, when we consider the total number of scans performed worldwide, it adds up to a significant number of inaccuracies. Due to the increased degree of computational resources accessible today, deep learning approaches are now widely used to fulfil the task of accurate object detection. As a result, if deep learning algorithms are used to detect kidney stones, the number of errors can be greatly reduced, as deep learning approaches have been shown to deliver reliable findings (Ozturk et al. 2020).

The goal of our study is to explore the suitability of deep learning algorithms to accurately detect kidney stones, to eliminate the need for radiologists in the diagnosis of kidney stones. Computed Tomography images of the KUB (Kidney Ureter and Bladder) type will be used for this categorization task. This research employs four deep learning algorithms to try and solve this problem of kidney stone detection from CT scan images: MobileNetV2, VGG16, InceptionNetV3, ResNet50V2. The proposed algorithms for our study have never been employed for kidney stone diagnosis, and our study attempts to fill the gap in the research by attempting to find a solution for detection of kidney stones using these methodologies. Using these fairly new deep learning algorithms for Nephrolithiasis, our research brings value to the domain of kidney stone diagnosis.

The main tasks that are performed as part of the research are Data gathering, Preparation of Data constituting of pre-processing of images, image augmentation, model building using deep neural networks, and evaluation of model performance. The experiments are carried out on a dataset of 1799 CT scan images and due to the limited number of source images, the results of the findings can be biased. Thus, Image augmentation process is carried out to increase the number of source images.

The Research question for the research is as follows:

‘Which of the following deep learning algorithms - MobileNetV2, VGG16, InceptionNetV3, ResNet50V2 provide the best results in terms of accuracy, precision and recall for the identification of kidney stones from Computed Tomography images?’

The roadmap of the additional sections of this report are as follows: Section 1.2 lists the objectives of this research. Section 2 provides critical review of the previous research performed in the domain of kidney stone detection using data analytical methods. Section 3 provides a detailed description of the methodology applied to achieve the research objectives. Section 4 gives an overview of the design specification used for the research and Section 5 lists the techniques performed to perform the implementation of the research. Section 6 lists the different experiments performed and insights into the values obtained for each in terms of evaluation parameters. It also provides a discussion on the results that were obtained during the evaluation process. Finally, Section 7 describes the conclusion of the research and the steps to be taken in future to improve the findings of the research. The objectives of the research are described in the Table 1.

Objective	Description	Metrics
1	Critical Review of Kidney Stone Detection using different Machine Learning and Deep Learning techniques	
2	Perform Exploratory Data Analysis to gain powerful insights on detection of kidney stones from CT scan images	
3	Build Deep Learning model using the MobileNetV2 convolutional neural network and evaluate the performance for kidney stone detection	Accuracy, Precision, Recall
4	Build Deep Learning model using the VGG16 convolutional neural network and evaluate the performance for kidney stone detection	Accuracy, Precision, Recall
5	Build Deep Learning model using the InceptionNetV3 convolutional neural network and evaluate the performance for kidney stone detection	Accuracy, Precision, Recall
6	Build Deep Learning model using the ResNet50V2 convolutional neural network and evaluate the performance for kidney stone detection	Accuracy, Precision, Recall
7	Compare the performance of all four models [MobileNet, VGGNet, InceptionNetV3, ResNet50V2]	Accuracy, Precision, Recall

Table 1: Objectives of the Research

2 Related Work

Over the last decade, deep learning and machine learning approaches have been employed extensively to accomplish accurate Object Detection and Image Classification. These approaches can be used in medicine to diagnose various flaws or diseases in a short amount of time. Kidney stones are undesirable deposits in the urinary system which obstruct the flow of urine. This section provides a critical overview of the techniques and approaches used in previous research to detect kidney stones.

2.1 Detection of Kidney Stones using Deep Learning Techniques

Deep learning is a subset of Machine Learning which follow a multi-layered strategy that aids in performing better classification functions like identifying anomalies and irregularities in medical images, patient clustering based on similar characteristics and level of severity or find the interrelation between outcomes and disease related symptoms. Deep learning, unlike other types of machine learning, has the extra benefit of requiring far less human intervention in decision-making. A neural network's layers work in a similar way to a human brain, learning from a large amount of data. This section summarizes a study of deep learning approaches that were previously utilized for object recognition in kidneys.

In (Yildirim et al. 2021) the XResNet-50 deep neural network was used to achieve both detection and localisation of kidney stones from Computed Tomography images. The

XResNet-50 algorithm constitutes a cross-residual network to achieve kidney stone detection. Additionally, augmentation techniques such as zooming, rotation are also applied to overcome the overfitting problem that can be caused when a model memorizes the input data. The Adam Optimizer and the Cross Entropy Loss Function were also utilized to fine-tune the XRestNet-50 model's parameters. During categorization, the Grad-CAM approach was also used to determine the areas of concentration. The dataset used for training the model consists of 1800 images of over 500 patients with two specific classes namely Kidney Stone and Normal with each category containing 790 and 1009 images respectively. The performance of the model has been evaluated using different metrics such as Sensitivity and Specificity which were found to be 95 and 97 percent respectively. The model misclassified 4 normal images as images with kidney stones. The main reason for the misclassification can be attributed to the type of scan images that were used for modelling the data. The images used were of coronal CT type and hence a number of other parts were also present in the scan such as abdomen, thorax and pelvis. The performance of this classification method could have been further increased by performing the segmentation on the source image data to increase the focus on the kidney region in the source CT scan.

Convolutional Neural Network is a deep learning method that can be used to perform kidney stone detection effectively. In (Längkvist et al. 2018), the CNN algorithm has been used specifically for identification of ureteral stones. Ureteral stones are the unwanted impediments that are formed in the ureters which are thin muscular tubes that are responsible for transferring the urine from the kidneys to the bladder and they are approximately 10 meters in length. The model has been contrived using 465 CT scan images of both men and women. Instead of employing extracted features, the method learns directly from raw data using a pixel/voxel-based machine learning approach. The sensitivity was determined to be 100 percent, and pre-processing was performed using the Connected Components approach. In the connected components approach consists of two steps namely the binarization and then grouping all pixels together. The main disadvantage of this strategy is that no images without ureteral stones were included in the dataset when training the model. As a result, the performance metric appears to be skewed because the source data lacks important subsets of data that can be available in real-time. The number of images that have been used to train the model is also small as there are just 465 images in use. The performance of the model was measured via the sensitivity parameter, and it was found that the sensitivity was 100%. The major highlight of this research is that it specifically targets ureteral stones and the detection of the same is complicated compared to detection of kidney stones, as the target area is smaller in case of ureteral stones.

The accuracy of classification based on medical images can be severely influenced by the speckle noise present in the image. In (Nithya et al. 2020) the kidney stone identification has been executed using the Artificial Neural Network algorithm post speckle noise removal using the median filter. Additionally, a specialised optimization algorithm called the crow search optimization algorithm has been employed to decrease the complexity and also to improve the accuracy of the classification. The dataset used for building the model consists of 100 ultrasound images. The model performed classification with sensitivity of 100%, specificity of 90%, and accuracy of 93.45%. Post classification the images which were found to contain kidney stones were subjected to segmentation using the multi kernel k-means algorithm. The major drawback of this approach is the lack of large number of images used in training the model. Since only 99 images have been used, the classification accuracy is questionable considering the overfitting factor.

When two or more approaches are combined to perform a specific task with the goal of increasing the performance, then the approach is termed as ensemble method. In (Cui et al. 2021), the CNN and threshold-based algorithms was employed to identify kidney stones and determine their dimensions and impact. The studies used a total of 625 computed tomography scan pictures, and the segmentation was done using a cascaded application of the 3D U-Net technique. Because a two-step segmentation was used, the classifier received a clean image focusing just on the target region, removing the possibility of classification mistakes due to the presence of several organs in the source image. The accuracy of segmentation was found to be 0.97 and the accuracy post classification was found to be 91 percent. In this research kidney images with significant anomalies were excluded and hence we can consider this as the drawback of this research as the model performance may differ when these images are included.

Modified artificial neural networks can also be utilized for performing object detection form medical images. In (Ma et al. 2020), the Heterogeneous modified Artificial Neural Network has been used to perform detection of renal calculi. For the purpose of this research, 640 ultrasound images have been used and these have been subjected to different processes such as speckle noise removal, image restoration and sharpening, contrast enhancement. The feature extraction process is then carried out on these processed images using the Gray Level Co-occurrence matrix followed by segmentation. The accuracy of the classifier was found to be 98% and the performance of this modified neural network was compared with additional techniques such as SVM, Backpropagation network and multi-Layer perceptron. The HMANN was found to produce the best classification results in comparison with other methods. The importance of segmentation process in improving the accuracy of the classification has been highlighted by the findings of this research.

Transfer learning refers to the storage of knowledge and findings gained from tackling one problem and put in practice for solving another problem which is related to the previous one. In (Parakh et al. 2019) a cascaded convolutional neural network pre-trained with ImageNet has been applied for performing the kidney stone detection. The model was trained by using 535 CT scan images of the abdominopelvic region and pre-processing methods such as segmentation, normalization of the Image orientation and Multi window encoding. The accuracy of the classification can be increased significantly by training using pre-available datasets. The depth of the urinary tract is assessed using the first CNN, and the Nephrolithiasis is performed by the inner CNN. The proposed model performed classification with sensitivity of 94% and specificity of 96%. In order to overcome limited number of images available in the source dataset, the pretraining of the model using ImageNet and GrayNet has been performed. This research shows that the cascaded networks can be used for performing object detection in medical images while the major drawback of the approach is that the dataset used for training consist of very limited number of images (less than 600).

Optimization techniques can be used in conjunction with the deep learning models to perform accurate object detection. In (Raju et al. 2019), an optimal probabilistic neural network along with spider monkey optimizer has been employed to perform renal calculi location in ultrasound images. As ultrasound images contain significant amount of speckle noise, the source images have been processed using the median filter for noise removal. These images are then subject to feature extraction to increase the performance of the classifier. The weight for the classifier is determined optimally using the Spider Monkey Optimizer. The accuracy of the OPNN classifier was found to be 93%. The research does not delve into the specifics of the number of images under the source dataset which is a major drawback of this approach as the results may or may not be biased depending on the source dataset size.

A back propagation network (BPNN) is a convolutional neural network where weights are adjusted to reduce the errors caused by the loss function, lowering the error rate, and enhancing the model's reliability. Kidney stone detection has been achieved using a back propagation network in (Akshaya et al. 2020) wherein Magnetic resonance pictures of the kidneys were included in the dataset utilized for building the model. The pre-processing steps that were performed prior to classification include steps such as Principal Component Analysis for feature extraction. The major drawback of this research is that the number of images in the dataset and the evaluation metric for measuring the performance of the classification are not explicitly mentioned.

To improve image quality, the medical images should be pre-processed before being classified. In (Viswanath and Gunasundari 2016)three different pre-processing approaches are applied on a dataset of 500 ultrasonography images. Initially, the source images are repaired to remove speckle noise and imperfections that have degraded the image. The Gabon filter is then used to smoothen the image and extract texture-based features. After that, the images are modified by Histogram Equalization, which improves the image's contrast. The level set segmentation technique is then used to identify the target areas. The energy levels are then retrieved before the Multi-Layer Perceptron and Back Propagation Artificial Neural Network are used to train the energy levels. The classifier's accuracy was found to be 98 percent. This model does not perform any augmentation method on the source images which could have impacted the performance of the model and improved it.

Thus, prior research publications that used deep learning approaches for kidney stone identification were evaluated in this section, and via this examination, the various algorithms available for object detection, as well as the most widely used evaluation metrics, were understood. It was observed that the most common deep learning techniques employed to perform kidney stoned detection were neural networks which followed the transfer learning approach. This is the reason in this research all four chosen deep learning networks employ transfer learning to perform image classification.

2.2 Kidney Stone detection using Conventional techniques

Object detection in medical images can also be performed using conventional machine learning algorithms and this section explores some of the previous research which utilized such methods for kidney stone detection. The use of these methods is less when compared to the number of research employing deep learning for classification, but these methods offer an alternative to the resource consuming but highly effective deep learning methods.

In (Verma et al. 2017) the kidney stone detection has been performed using the k-nearest neighbor and Support Vector Machine classification techniques. In order to perform the reduction of multi-dimensional data space mapping to lower dimensions, Principal Component Analysis, a feature extraction method is put to use prior to applying the classifiers. The dataset contains ultrasound images, but their exact quantity has not been specified. Methods such as entropy-based segmentation and morphological methods are also applied to increase the accuracy of the classifier. The accuracy of the classification was found to be 89% for KNN and 84% for the SVM. The major observation from this paper is that the accuracy of the machine learning methods is less than the deep learning counterparts. The lack of specification of the number of images in the dataset raises the question whether the findings are reliable for this research.

The major disadvantage of employing Ultrasound in medical imaging is that these images can be easily affected by speckle noises which may directly affect the performance of a model, when these images are fed as the input. In (Selvarani and Rajendran 2019), an adaptive mean median filter approach has been used to perform speckle noise removal. The kidney stone detection is performed using Support Vector Machine learning algorithm on 250 ultrasound images. To eliminate the undesired parts of the image techniques such as erosion and dilation are followed. Further segmentation has been carried out using the k-means segmentation algorithm. The classification is performed using SVM aided by Particle Swarm optimization which determines the optimal hyperplane. The accuracy of the classification was found to be 98%. The major advantage of this method is that it addresses non-Gaussian speckle noises that influenced the effectiveness of SVM classifiers.

One of the most often used techniques for classification is the Support Vector Machine. The SVM classifier was employed to detect the presence of kidney stones in the urinary tract in (Soni and Rai 2020). The model was trained using 156 Computed Tomography scan samples, and it attained a 98 percent accuracy rate. The source images were improved using a process known as Histogram Equalization, which improves the image's contrast. Embossing was applied to the enhanced images for easier object identification and edge recognition. Embossing creates a layer similar to a mask on the source image, making it easier to identify stones. During the embossing step, a directional differential filter with both vertical and horizontal kernels was used. The SVM classifier is used after embossing to ensure correct categorization. The main disadvantage of this strategy is that the source data only contains 190 photos, which is not sufficient for proper training of the model, and the resulting model would have overfit the data. If augmentation approaches to enhance the volume of data to be used for training had been employed, an efficient and able model to perform classification on fresh data would have resulted.

Thus, a review of studies that used traditional machine learning methods to detect kidney stones was provided in the preceding section. The variety of strategies for detecting kidney stones was highlighted by these techniques. From this review it clear that although machine learning techniques can be employed to detect kidney stones from medical images, the time taken to perform such methods are greater than deep learning techniques. This conclusion enabled the choice of deep learning methods to achieve kidney stone detection.

2.3 Pre-Processing methods for Kidney Stone Detection

Image pre-processing constitutes a series of processes that are applied to photographs in order to improve their visual quality and correct any flaws or interruptions that may be present. Due to the complicated machinery and several types of rays used to generate medical images, there will be a certain degree of speckle noise in generated images. This section summarizes an overview of research papers that have emphasized the significance of image pre-processing techniques.

Multiple phases of image pre-processing can be undertaken to improve the quality of the images. A three-stage preprocessing methodology has been proposed by (Thein et al. 2018) that can be used before the segmentation procedure. The Sensitivity parameter was used to evaluate the performance of each preprocessing procedure. The final output images were found to have a sensitivity of 95%. The CT scan images from 30 patients were used for performing the different pre-processing techniques. Unwanted region removing has been performed using the three thresholding algorithms on the basis of intensity, location and size. Limited number of source data for building the model is a drawback of this method. This research sheds light on the different techniques that can be employed for unwanted region removal.

Feature extraction was successfully achieved using Intensity Histogram Analysis for extraction of 19 unique GLCM features in (Hafizah et al. 2012).Ultrasound images were used as the source images for model building. The input US pictures were first cropped to determine the region of interest, and kidney contour recognition was also done. The Ultrasound pictures utilized in this experiment were divided into four groups based on the type of renal dysfunction. The feature extraction method clearly highlights the various features in the source images and the results from this step can be used to effectively perform the determination of presence of kidney stones from medical images.

Thus, in this section, by analyzing earlier research in the Image Pre-processing area, the importance of the various pre-processing procedures to be carried out prior to classification was brought into sharp focus. The major pre-processing techniques used are feature extraction, image resizing and image normalization. Image resizing and image normalization are thus employed in this research to enhance the quality of images and also to aid in the modelling process.

2.4 Conclusion

The Related Work section examined the various approaches for determination of presence of kidney stones through deep learning methods that have already been employed, as well as the advantages and disadvantages of each study methodology. Based on the review the deep learning techniques were chosen for modelling the kidney stone data. Additionally, the basic pre-processing techniques that should be performed when performing image classification using deep learning were also identified and the overall flow of the research was framed post this critical analysis and comparison of different techniques in the previous research. It was observed that in all the aforementioned research the combination of the four deep learning networks- ResNet, VGG16, MobileNet, InceptionNet were never employed to perform kidney stone detection and an effective comparison of these methods' classification performance for detection of kidney stones has not been provided. This research addresses this need and provides a comprehensive comparison for these networks and their suitability for detecting kidney stones from CT scan images.

3 Research Methodology

This research has been conducted to determine the suitability and performance of the four deep learning networks in determining the presence of kidney stones from CT scan images. This research follows the CRISP-DM data mining approach to examine the suitability of these 4 different neural networks in classification of kidney stones. The CRISP-DM approach is a cross-industry standard process for data mining and is an iterative methodology, which contains a sequence of steps such as Business understanding, Data understanding, Data preparation, Modelling, Evaluation, Deployment (Kalgotra and Sharda 2016). CRISP-DM approach is the most widely used methodology to manage data mining projects due to its cost-effectiveness, and its uniformity towards the planning and management of the project. The different stages of the CRISP-DM approach in the research are as follows:

3.1 Business Understanding

The Business understanding phase constitutes the identification and description of the problem in a precise manner along with techniques to measure the fulfilment of the target at hand. In the medical industry, radiologists are responsible for identifying the presence of particles in the kidney from medical images. In this research the automation of this detection process is performed by the construction of deep learning models on CT scan images dataset. The research thus eliminates the human error in the diagnosis of kidney stones by modelling using deep learning techniques. Four specific deep learning methods which employ transfer learning have been utilised to build the models and these models have been assessed using different evaluation metrics such as Precision, Recall, Accuracy. Early and error-free detection of kidney stones from medical images will have a positive impact in the provision of appropriate and timely treatment to the patients under consideration and will also help in mitigating the pain caused by the phenomenon.

3.2 Data Understanding

The Data understanding phase aims at performing critical inspection of the data to be used for the modelling process and this phase enables the identification of any potential problems that may arise in the following phase that may have ramifications in the results of the methodology. This phase involves the accessing of data and performing exploratory analysis to assess the quality of the data at hand. The dataset that has been used for the modelling process in this research have been obtained from Github[1]. The dataset consists of 1799 CT scan images of 500 patients and these images belong to two classes and these are – Kidney_Stone, Normal. Kidney_Stone class refers to images wherein stones are present in the kidneys.

The images in the dataset all belong to the .png format and the resolutions of images are not equal. The distribution of the images basis their resolution is shown in Figure 1. From Figure 1, it is evident that the most common resolution for the images is 1010*1286 with 140 images present in this resolution. Only 21 images are of the resolution 1110*1320 making it the least common resolution for the source images.

---

[1] https://github.com/yildirimozal/Kidney_stone_detection

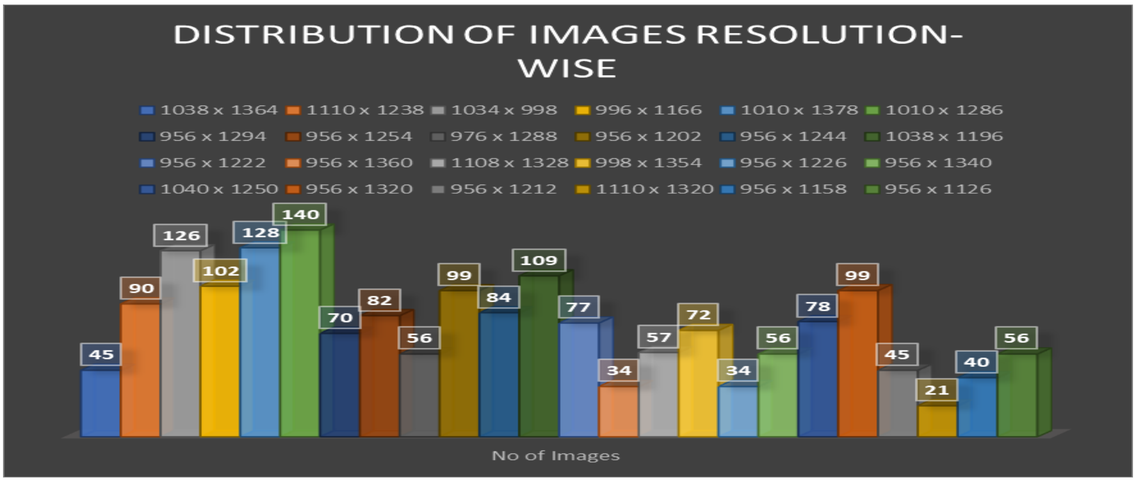

Figure 1: Distribution of images in terms of Resolution

3.3 Data Preparation

The Data Preparation phase encompasses all activities to build the final dataset from the initial dataset to achieve meaningful modelling. In our research the different images are initially converted to images of the same resolution. Then these images are rescaled/normalized to adjust the pixel values in the image to be in the desired format as what is expected for building the deep neural network. The images are then augmented using different parameters such as flipping, rotation to increase the credibility and predictive power of the models to be constructed. The different data preparation operations are discussed in detail below:

3.3.1 Adjustment of Image Resolution

The source images to a deep learning network should all be in the same resolution, as these images will be processed in batches and hence, they should be of the same resolution. In the dataset the images are available in 25 different resolutions, as shown in Figure 1. The MobileNet model require the images to be in the 160*160 resolution while the other three models require the images to be in the 240*240 resolution. The adjustment of resolution for the source images is performed using the in-built functions of the keras API.

The parameter input_shape specifies the resolution of the input images being fed as input to the deep learning model and in this research the resolution 240*240 has been used as the common value for all models. The reason MobileNet model has been fed with input images of resolution 160*160 is because MobileNet model expects the resolution to be in 160*160 by default. This is the reason for providing resolution as 160*160 for MobileNet while other models were provided with 240*240.

3.3.2 Image Rescaling/ Normalization

Images are comprised of components called as pixels and each pixel provides valid information about the image as they are samples of the image under consideration. Pixels are represented using matrices and depending upon the image colour the values in the matrix and its structure varies. Typical pixel values range between 0 to 255. The process of scaling

/normalization involves the conversion of such pixel values into a standard format in the range between 0-1 (Shafiei et al. 2020). The major advantage of such an approach is that it reduces the time taken to train the deep learning models. The normalization is achieved by dividing each pixel value using the value 255 and this is performed for all channels. The ImageDataGenerator function in keras API has been used to achieve the normalization of the pixel values.

3.3.2 Image Augmentation

Image augmentation is the technique by which the size of the dataset can be increased in an artificial manner through addition of modified versions of existing images in the dataset (Farhadi and Foruzan 2019). The training of deep learning networks on increased amounts of data will result in the construction of robust models and will also address the problem of overfitting or model generalization. Using ImageDataGenerator function in the Keras API, different augmentations techniques such as zooming, rotation, flipping have been performed to the existing images in the dataset. The augmentation happens during runtime and hence newly generated data will not be saved locally to the source folder which contains the original dataset. The different parameters used to perform augmentation and the specific values provided are shown in the Table 2. The effect of augmentation is shown in Figure 2.

The zoom_range specifies the extent to which the random zooming takes place for an image. The rotation_range parameter signifies the rotation in degress for an image. The horizontal and vertical flip parameters signify the extent to which flipping of an image is performed.

Augmentation Parameter	Value
zoom_range	0.4
rotation_range	10
horizontal_flip	True
vertical_flip	True

Table 2: Validation parameters

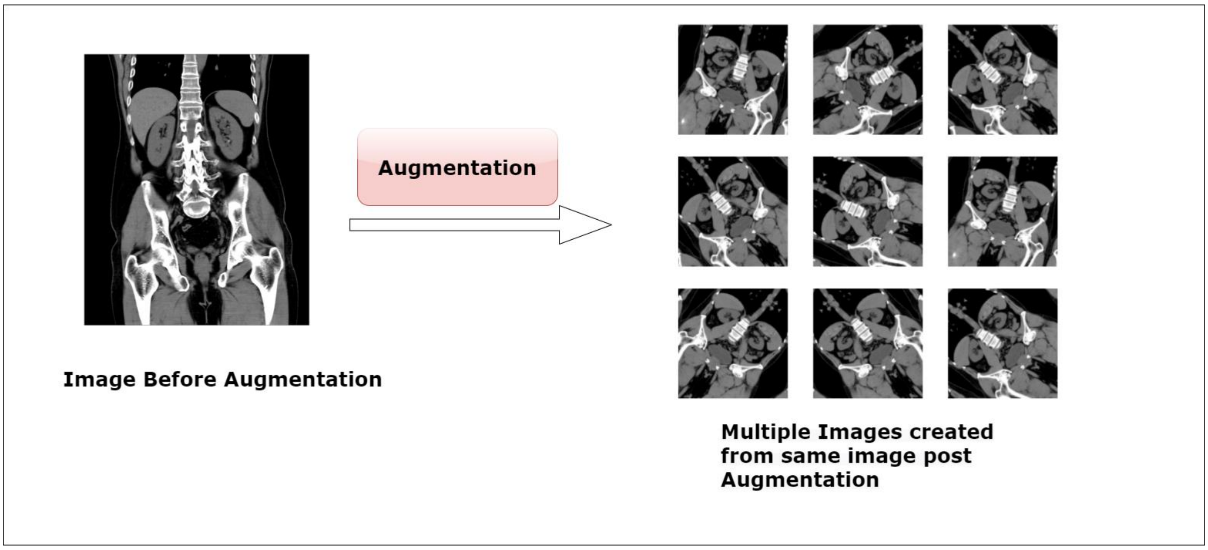

Figure 2: Representation of Image Augmentation

3.4 Modelling

This section describes the different models to be used for achieving the aim of the research which is the classification of whether kidney stone are present or not from the CT scan images. The modelling phase of the CRISP-DM approach involves the three specific tasks such as the selection of models, designing of tests and the building of models. The different models the research are as follows:

3.4.1 MobileNet

MobileNet is a deep learning convolutional neural network which has been created specifically for use in mobile applications (Chen and Su 2018). The major difference in the MobileNet network in comparison with other neural networks is that it employs lesser number of parameters compared to other networks with same depth. MobileNet utilizes special type of convolutions called the depth wise separable convolutions which are made up of two specific operations -the depthwise convolution and the pointwise convolution. MoblieNet has been designed to increase the accuracy in an effective manner while also being considerate towards the limited number of resources available in an embedded or mobile application. A single convolution is performed on every channel by the depthwise convolution filter and the output of the depthwise filters are combined linearly by the point convolution filter. This factorization results in drastic decrease in the size of the model and the computational cost. The network consists of two hyper parameters namely -Width Multiplier and Resolution Multiplier. The main function of the width multiplier is to reduce the size of the network in a uniform manner. The Resolution multiplier reduces the resolution of the images that are employed as part of the dataset. Excluding the depthwise and pointwise convolution layers, there are 28 layers in the baseline MobileNet neural network. The architecture of MobileNet is shown in the Figure 3.

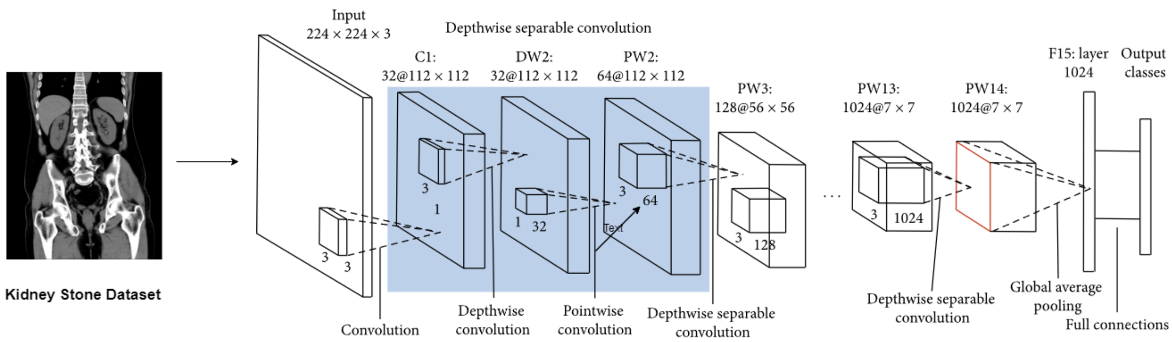

Figure 3: Architecture of MobileNet

3.4.2 VGGNet

VGGNet is a convolutional neural network and is expanded as ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’. The VGGNet produced an accuracy of 92.7% when trained with the ImageNet dataset (Saiharsha et al. 2020). The ImageNet dataset contains over 14 million images and these images belong to thousand different classes. The high performance is achieved by the modification of the kernel-sized filters with number of kernel size filters smaller in size. The input images to the VGGNet dataset should be in the form of 224*224 resolution. The major drawback of the VGGNet is the time taken for model training. Due to the number of connected nodes and the depth of the layers in the VGG, it consumes over 533 MB of space. There are two variations of the VGG network available, and these are VGG16 and VGG19 denoting the number of convolutional layers present in the model. The convolutional filters that are employed in VGG16 are comparatively small. There are 13 convolutional and 3 fully connected layers present in the VGG16 network. VGGNet employs the ReLU activation function which stands for rectified linear unit. This function will propagate the input if positive, otherwise zero will be produced as the ouput. The architecture of the VGG16 network is shown in Figure 4.

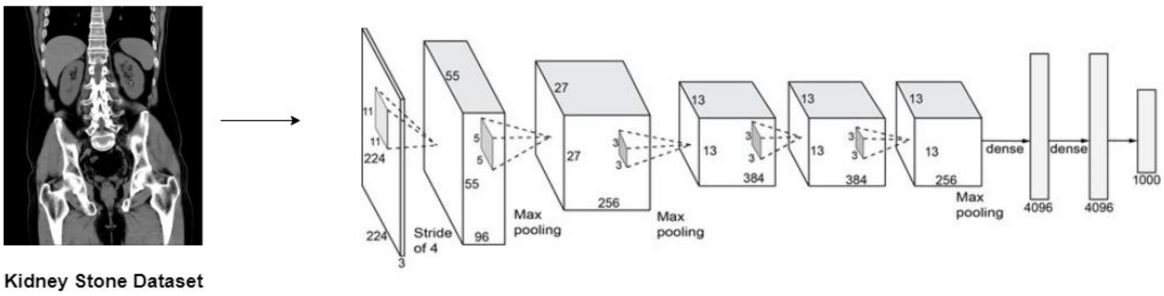

Figure 4: Architecture of VGGNet

3.4.3 InceptionNetV3

InceptionNet is an image recognition model that produces an accuracy of over 78.1% when trained with the ImageNet dataset (Shrivastava et al. 2021). The building blocks of the model consists of both symmetric and asymmetric components such as dropout layers, convolutions, concats, average pooling layers and also fully connected layers. The model extensively uses the Batch Normalization layers for activation of the inputs. These batch normalization layers contribute to the standardization of the inputs to the layers of the model, thus stabilizing the process of learning and also leads to the reduction in the number of epochs. An input with three filter sizes is used as the starting block for InceptionNet wherein Maximum Pooling is also performed. The main functionality of a max pooling layer is to calculate the largest value within a feature map. The convolutions of the neural network are made more efficient through the use of smart factorization methods. The Simplified representation of the InceptionNet V3 model is shown in the Figure 5.

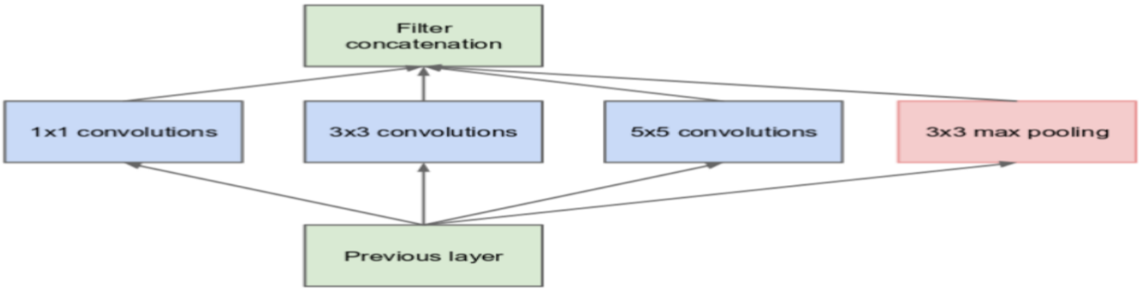

Figure 5: Simplified Representation of InceptionNet

3.4.4 ResNet50 V2

ResNet50 is a convolutional neural network which constitutes of 50 layers and is a residual neural that consists of the stacking of numerous residual blocks on top of each other to comprise a network (Budhiman et al. 2019). The residual networks were created to overcome the problem of vanishing gradients. As the layers in a convolutional neural network increase, the ability of the network to detect and model features of the image increases. The drawback of increasing the number of layers is that after a certain point of time, saturation of the accuracy levels may be triggered resulting in degradation. This degradation is called the problem of vanishing gradient. Using residual blocks, the accuracy is improved by the concept of skip connections. The skip connections provide an alternate path for the passing of the gradients. They also provide the facility for the model to learn an identity function. A three-layer block specifically created to overcome the bottleneck resulting in the Resnet 50 architecture. The number of filters employed in ResNet are lower when compared with VGGNet.

4 Design Specification

This section identifies and presents the methodologies, architecture, and framework that underpin the implementation, as well as the accompanying requirements. The project design process flow is illustrated in Figure 6. The images to be used for modelling the data are initially extracted from the github repository and these images are subjected to different pre-processing operations to enhance the quality of the images. There are two phases in the research execution. Phase 1 involves the use of source images without any cropping for model building. In Phase 2, an alternative flow is followed wherein the original images of the dataset are subjected to cropping to focus on the target area of kidneys. In this research the pre-processing activities such as image normalization, image size resizing is performed. The pre-processed images are then augmented to overcome the overfitting problem of the modelling process. The classification is then performed using the four deep learning networks- MobileNet, ResNet, VGGNet, InceptionNet by training these networks on the dataset under consideration. The classification performance of each of these networks are assessed on the basis of different evaluation metrics such as Accuracy, Precision and Recall.

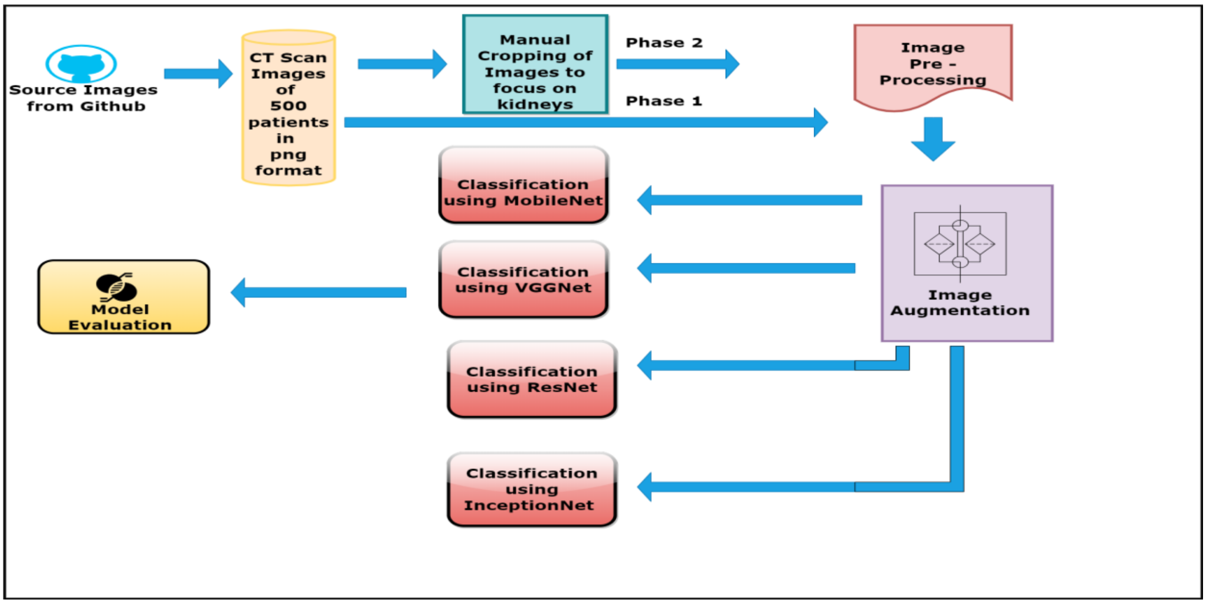

Figure 6: Process flow of the Research

5 Implementation

This section describes the different data transformations, the modifications made to existing models along with the tools and languages used to produce the outputs.

5.1 Environment Setup

This section describes the different tools that were used to carry out the experiments in the research. The dataset to be used for the modelling process was obtained from Github. Python was used as the programming language to model the deep neural networks. The construction and modelling of Deep Learning networks require considerable computing capacity and resources. Hence Google Collaboratory, a web based Integrated Development Environment was used for framing the code base in Python. Google collab enables connection to high performance Graphical Processing Units through the cloud enabling the smooth modelling of even complex neural networks. For this research, a NVIDIA GPU with 27.3 GB of available RAM was utilised for model building. The keras is a high-level neural network library that has been used in the research to enable seamless processing and prototyping. Along with keras, the Tensorflow framework has also been used. To access the dataset through Google Collaboratory, the dataset was uploaded to the Google drive.

5.2 Data Handling

As mentioned in Section 5.1, the kidney stone dataset consisting of 1799 images was uploaded to Google drive. The source dataset consists of two folder Train and Test indicating the images to be used for the Training and Testing phases. With keras APIs such as ImageDataGenerator and flow_from_directory the validation dataset is generated from the Training images. The ImageDataGenerator also enables the rescaling parameter to be specified through which the Image Normalization is performed as mentioned in Section 3.3.2. The data augmentation process is performed using the keras.sequential method. As specified in Section 4, in Phase 2 of the process flow the images are cropped to focus on the target area of the kidneys in CT scan images. For the purpose of cropping the images to obtain only images of the target region, the in-built photo editing application called Windows Photos was used.

5.3 Transfer Learning – Modifications to Pretrained Models

Transfer learning is the process by which the information/knowledge obtained while resolving a particular problem can be applied to solve another problem on similar lines (Hastuti et al. 2021). The four neural networks that are used for building the models in this research, all employ the transfer learning approach. All the 4 networks under consideration- MobileNet, VGGNet, ResNet, InceptionNet are pre-trained using the ImageNet dataset. Instead of building each layer in these networks from scratch, the base model for all these networks were obtained through the keras.applications package. To retain the features extracted during previous rounds, the layers are the model are frozen. Then to train the model on the kidney stone dataset, new layers are added to the existing network thus enabling the conversion of old features into new predictions. The set of new layers that were added to the 4 pre-trained models are shown in the Table 3. The newly created models are then trained on the new dataset to observe the validation and test accuracy.

Layer Type	Corresponding Parameter
Dropout layer	0.2
Batch Normalization layer	NA
Dense layer	1024
Dropout layer	0.2
Dense layer	1024
Batch Normalization layer	NA
Dropout layer	0.8
Dense layer	2

Table 3: Additional layers added to existing Base Deep learning models

6 Evaluation

This section provides an overview of the different metrics that can be used to assess the classification performance of the neural networks used for the modelling process. The three metrics for performing the evaluation are Precision, Recall and Accuracy (Vakili et al. 2020). The Precision is a metric which describes the ability of a model to classify a particular sample as positive. Precision is calculated as the ratio of number of correct classifications of samples that are positive to the actual number of samples that have been determined as positive. In short, precision provides a reflection of the reliability of the model in classifying a particular sample as positive. Recall is a metric which deals only with the classification of positive samples and the negative samples are not taken into consideration at all. Recall is defined as number of samples which are positive and correctly specified as positive, divided by the final number of positive samples. A high value of recall indicates that a greater number of positive samples have been detected. Accuracy is a metric which is used to describe the classification performance across all classes. Accuracy is obtained by dividing the quantity of correct predictions by the total number of predictions. For each of the below experiments, the precision, recall and accuracy have been calculated using the keras in-built metrics attribute.

In this research the number of epochs is decided by an iterative process. Initially, when each model was run, the number of epochs is by default set to 10 [since a random number is to be assigned to this attribute to build the model]. On observation of the training accuracy of each epoch it is found that validation accuracy increases for each epoch which is the expected behavior. This increase in accuracy will cease at some point and we can call this the saturation point. It is discerned that in case of models built using the cropped dataset, the saturation point occurs post the 6th epoch and for models built using the original image dataset, this saturation occurs after the 10th epoch. This is the reason for choosing the epochs as 6 and 10 for models built using cropped and uncropped dataset respectively.

6.1 Experiment 1: Building a VGGNet Model with unmodified dataset and cropped dataset

In this experiment the VGG16 neural network model is accessed through the keras API and the base layers of the model are made non trainable. As described in section 5.3 an additional set of layers are added to the base VGG16 and the model is compiled to produce results based on Accuracy, Precision, Recall. The number of epochs is set at 10 and 6 for the original and cropped dataset respectively. The results from Experiment 1 are illustrated in Table 4. It can be observed that as the epoch increases, the accuracy, precision and recall values all increase both in terms of unmodified as well as cropped datasets. It can be seen that for the VGG model built using the cropped dataset, the accuracy has increased by 26 %, the precision has increased by 20% and recall has increased by 20% as that of the results of the model built using unmodified dataset. The Accuracy of the VGGModel built using the unmodified dataset increases sturdily with increase of epoch up to epoch 9, after which deterioration is observed in the accuracy values.

Epoch	Val Accuracy	Val Precision	Val Recall
1	0.5068	0.5141	0.5
2	0.5724	0.57854	0.52068
3	0.5724	0.5909	0.53793
4	0.5896	0.5729	0.55517
5	0.56206	0.5796	0.58965
6	0.5793	0.5862	0.5862
7	0.589	0.5993	0.5931
8	0.6103	0.6035	0.5931
9	0.6	0.5992	0.5827
10	0.5931	0.5938	0.6

Epoch	Val Acc	Val Precision	Val Rec
1	0.66206	0.5397	0.914
2	0.6586	0.608	0.834
3	0.6758	0.6434	0.797
4	0.6827	0.6676	0.776
5	0.6999	0.6956	0.772
6	0.7482	0.7178	0.807

Table 4: Validation Results of VGG16 Network for unmodified and cropped dataset

6.2 Experiment 2: Building a ResNet Model with unmodified dataset and cropped dataset

In this experiment the ResNet50 model is imported through the keras library, and the base model is modified to include additional layers. Two versions of the model are built for the unmodified and cropped dataset respectively. The results of the model building process are shown in the Table 5. It can be observed that, for the unmodified dataset, the number of epochs taken before the saturation of training accuracy is more than the number of epochs required for the model built on the cropped dataset. The accuracy, precision and recall values for the ResNet model built using the unmodified dataset, all range within 0.57 to a maximum value of 0.60. As a result of the cropping, the accuracy increases by 23.2%, precision increases by 34% and recall by 21.4% and this rapid increase in performance based on all three metrics indicates that the cropping of images in the original dataset was an effective operation to increase model performance.

Epoch	Val Accuracy	Val Precision	Val Recall
1	0.5689	0.5689	0.5689
2	0.562	0.5597	0.5172
3	0.5517	0.568	0.4896
4	0.5689	0.5631	0.5379
5	0.5793	0.57894	0.5689
6	0.5793	0.58389	0.6
7	0.5793	0.57284	0.5965
8	0.5758	0.58108	0.5931
9	0.5931	0.58666	0.6068
10	0.5655	0.54934	0.5758

Epoch	Val Accu	Val Prec	Val Recall
1	0.68	0.7167	0.6461
2	0.735	0.75	0.6738
3	0.7723	0.79729	0.72615
4	0.779	0.813	0.743
5	0.793	0.845	0.765
6	0.797	0.892	0.789

Table 5: Validation Results of ResNet50 Network for unmodified and cropped dataset

6.3 Experiment 3: Building an InceptionNet Model with unmodified dataset and cropped dataset

The InceptionNet model is obtained through keras API and additional layers are added to this base model. The version of InceptionNet used for this experiment is InceptionNet V3. The results from the two models built on the basis of unmodified and cropped kidney stone dataset are shown in the Table 6. The InceptionNet model built using the cropped dataset produces an accuracy of 0.86 which is the highest among all models built in this research. The precision and recall values of 0.866 and 0.831 for the same dataset is also the highest for any model in this research. The increase in accuracy, precision, recall when model built using original dataset is compared with model built using modified dataset is as follows: 29.5%, 32.90% and 19.3%.

Epoch	Val Accuracy	Val Precision	Val Recall
1	0.5931	0.5848	0.6655
2	0.61379	0.619	0.6517
3	0.6379	0.6293	0.6793
4	0.6379	0.6289	0.6896
5	0.6482	0.638	0.6931
6	0.6517	0.6424	0.6999
7	0.6482	0.6426	0.7068
8	0.6689	0.6477	0.7103
9	0.6689	0.6466	0.7068
10	0.6655	0.6516	0.6965

Epoch	Val Accuracy	Val Prec	Val Recall
1	0.8793	0.8888	0.827
2	0.8586	0.8534	0.803
3	0.862	0.8561	0.8206
4	0.862	0.857	0.831
5	0.8517	0.8576 5	0.831
6	0.862	0.866	0.831

Table 6: Validation Results of InceptionNetV3 Network for unmodified and cropped dataset

6.4 Experiment 4: Building a MobileNet Model with unmodified dataset and cropped dataset

The MobileNet model is a lightweight model and the MobileNetV2 is used as the base model in this experiment. The Table shows the different results for the models built for the original and cropped dataset in terms of Validation Accuracy, Validation Precision and Validation Recall. As seen from the Table 7, the number of epochs for the above two versions of the model are set as 10 and 6 respectively.

Epoch	Val Accuracy	Val Precision	Val Recall
1	0.5758	0.56	0.7241
2	0.5862	0.5694	0.7206
3	0.6206	0.5945	0.7482
4	0.6137	0.6005	0.7
5	0.6034	0.5797	0.6517
6	0.6034	0.5867	0.64137
7	0.5931	0.5876	0.6241
8	0.5827	0.5838	0.62413
9	0.5896	0.5836	0.6137
10	0.5931	0.5874	0.613

Epoch	Val Accuracy	Val Precision	Val Recall
1	0.8551	0.8161	0.8724
2	0.8034	0.8027	0.8
3	0.7896	0.8055	0.8
4	0.7827	0.7903	0.7931
5	0.762	0.7705	0.7758
6	0.762	0.7751	0.7965

Table 7: Validation Results of MobileNetV3 Network for unmodified and cropped dataset

Through the impact of the cropping process the accuracy, precision, recall have all increased by the following percentages: 28.4% increase,31.95% increase and 20.3% increase respectively.

6.5 Discussion

The aim of this research is to examine the suitability of the four deep neural networks- VGGNet, MobileNet, ResNet, InceptionNet to classify whether the kidney stones are present from CT scan images. The Accuracy of each model provides a good overview of the model performance across all classes. It takes into account the total number of predictions as well as the number of correct predictions. A high value of the Accuracy metric indicates that the model produces accurate classification across all the classes under consideration. From Figure 7, it can be seen that the accuracy of the InceptionNet model is greater than all other models built using the unmodified dataset without cropping by an average percentage value of 20%. The accuracy of the InceptionNet model without cropping is measured as 0.66.

The accuracy of ResNet , MobileNet, VGGNet for the modified dataset is measured as 0.56, 0.59 ,0.59 respectively. Hence on average the InceptionNet performs the best in terms of accuracy when compared with other models. The low value of the accuracies can be attributed to the complexity of the dataset wherein each image contains a CT scan image of the human body. In addition to the data about kidneys, these scan images also contain information of other organs such as the lungs, rib cage, spinal cord, thorax. Hence to narrow down the target area for the models, manual cropping of the source images is performed as shown in Figure 8.

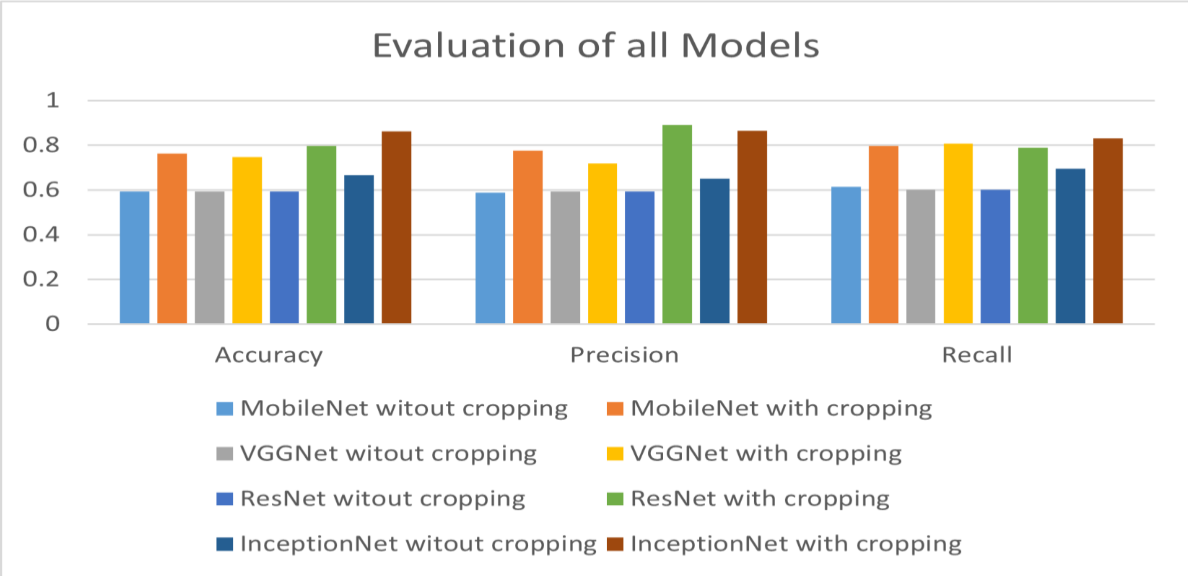

Figure 7: Comparison of all models based on evaluation metrics

The models that were built using the cropped images as the source dataset performed better classifications as is evident from the Figure 7. It is observed that the InceptionNet model performed the best for the cropped images dataset providing accuracy value of 0.86 which is the highest accuracy value for any model for both unmodified and cropped dataset

Precision is a metric which is calculated as the ratio of the True Positives and all the Positives. It provides a measure of the number of samples that the model correctly predicted as having kidney stones. Hence in this research a high value of Precision is desired. From the Figure 7, it can be observed that InceptionNet model has the highest precision value of 0.866, for models built on both unmodified and cropped dataset.

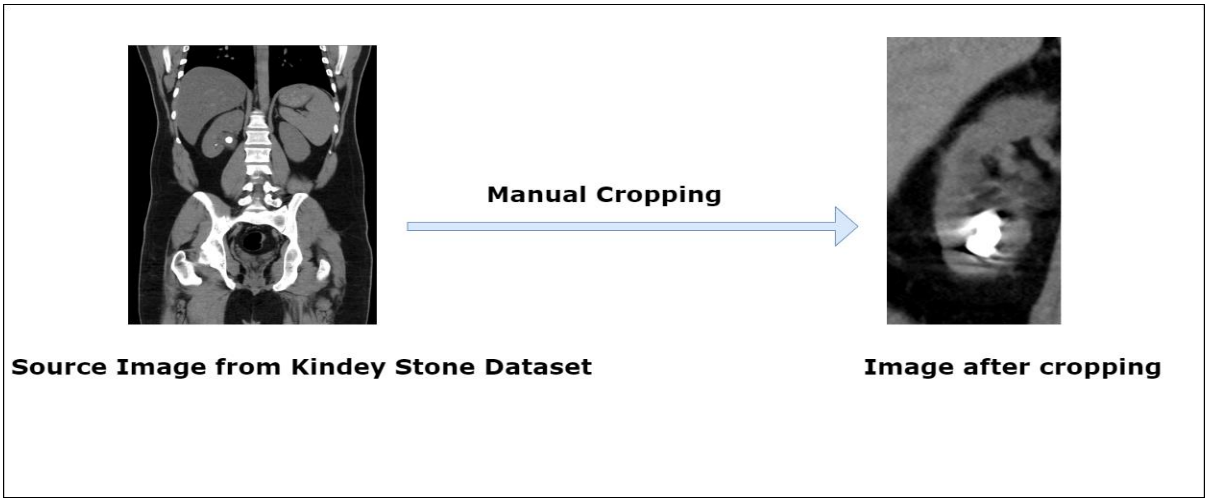

Figure 8: Representation of Manual Cropping

The Recall metric measures the ability of a model to correctly detect the positive samples and the high values of recall indicate that more positive samples were successfully detected.

Similar to the results observed in terms of Precision and Accuracy, the recall value is highest for the InceptionNet model.

As InceptionNet model provides the best results in terms of Accuracy, Precision and Recall it can be affirmed that InceptionNet model is the best suited deep learning neural network among ResNet, MobileNet, VGGNet, InceptionNet. From the Figure 7, it is evident that the classification performance increased considerably for all four models post the cropping process by an average of 15% for all metrics. The cropping process which was carried out manually in all four experiments consumed considerable amount of time and in future experiments this cropping process is to be done through code for automated execution. This is one of the limitations of this research. The Dataset used in this research contains only 1800 images and the classification performance can be further improved by using a dataset which contains more images.

7 Conclusion and Future Work

The research aims at detection of kidney stones from the CT scan images using deep learning models and evaluating the performance in terms of accuracy, precision, and recall. From the results of the four experiments, it can be concluded that the InceptionNet model provides the best performance in terms of all metrics under consideration. The performance of the models was improved by performing manual cropping to focus on the target areas and this resulted in higher accuracy, higher recall, and improved precision values. Thus, through this research all the proposed objectives mentioned in Section 1.2, were achieved. The findings from this research indicate that InceptionNet model can be used for building an end-to-end system capable of automatic detection of kidney stones when CT scan images are provided as input to the system. This interactive system is the future work that can be performed from the research findings. Using this system, the role of radiologists in detection of kidney stones can be removed and automatic detection can be performed.

References

Akshaya, M., Nithushaa, R., Raja, N.S.M., Padmapriya, S. (2020) ‘Kidney Stone Detection Using Neural Networks’, *2020 International Conference on System, Computation, Automation and Networking, ICSCAN 2020*.

Bierig, S.M., Jones, A. (2009) ‘Accuracy and cost comparison of ultrasound versus alternative imaging modalities, including CT, MR, PET, and angiography’, *Journal of Diagnostic Medical Sonography*, 25(3), 138–144.

Budhiman, A., Suyanto, S., Arifianto, A. (2019) ‘Melanoma Cancer Classification Using ResNet with Data Augmentation’, *2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019*, 17–20.

Chen, H.Y., Su, C.Y. (2018) ‘An Enhanced Hybrid MobileNet’, *2018 9th International Conference on Awareness Science and Technology, iCAST 2018*, 308–312.

Cui, Y., Sun, Z., Ma, S., Liu, W., Wang, X., Zhang, X., Wang, X. (2021) ‘Automatic Detection and Scoring of Kidney Stones on Noncontrast CT Images Using S.T.O.N.E. Nephrolithometry: Combined Deep Learning and Thresholding Methods’, *Molecular Imaging and Biology*, 23(3), 436–445.

Farhadi, M., Foruzan, A.H. (2019) ‘Data Augmentation of CT Images of Liver Tumors to Reconstruct Super-Resolution Slices based on a Multi-Frame Approach’, *ICEE 2019 - 27th Iranian Conference on Electrical Engineering*, 1783–1786.

Hafizah, W.M., Supriyanto, E., Yunus, J. (2012) ‘Feature extraction of kidney ultrasound images based on intensity histogram and gray level co-occurrence matrix’, *Proceedings - 6th Asia International Conference on Mathematical Modelling and Computer Simulation, AMS 2012*, 115–120.

Hastuti, E.T., Bustamam, A., Anki, P., Amalia, R., Salma, A. (2021) ‘Performance of True Transfer Learning using CNN DenseNet121 for COVID-19 Detection from Chest X-Ray Images’, *InHeNce 2021 - 2021 IEEE International Conference on Health, Instrumentation and Measurement, and Natural Sciences*.

Kalgotra, P., Sharda, R. (2016) ‘Progression analysis of signals: Extending CRISP-DM to stream analytics’, *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 2880–2885.

Längkvist, M., Jendeberg, J., Thunberg, P., Loutfi, A., Lidén, M. (2018) ‘Computer aided detection of ureteral stones in thin slice computed tomography volumes using Convolutional Neural Networks’, *Computers in Biology and Medicine*, 97(April), 153–160.

Ma, F., Sun, T., Liu, L., Jing, H. (2020) ‘Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network’, *Future Generation Computer Systems*, 111, 17–26, available: https://doi.org/10.1016/j.future.2020.04.036.

Nithya, A., Appathurai, A., Venkatadri, N., Ramji, D.R., Anna Palagan, C. (2020) ‘Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images’, *Measurement: Journal of the International Measurement Confederation*, 149, 106952, available: https://doi.org/10.1016/j.measurement.2019.106952.

Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Rajendra Acharya, U. (2020) ‘Automated detection of COVID-19 cases using deep neural networks with X-ray images’, *Computers in Biology and Medicine*, 121(April), 103792, available: https://doi.org/10.1016/j.compbiomed.2020.103792.

Parakh, A., Lee, H., Lee, J.H., Eisner, B.H., Sahani, D. V., Do, S. (2019) ‘Urinary Stone Detection on CT Images Using Deep Convolutional Neural Networks: Evaluation of Model Performance and Generalization’, *Radiology: Artificial Intelligence*, 1(4), e180066.

Raju, P., Malleswara Rao, V., Prabhakara Rao, B. (2019) ‘An efficient optimized probabilistic neural network based kidney stone detection and segmentation over ultrasound images’, *International Journal of Recent Technology and Engineering*, 8(3), 7465–7473.

Sabih, D.-, Sabih, A., Sabih, Q., Khan, A.N. (2011) ‘Image perception and interpretation of abnormalities; can we believe our eyes? Can we do something about it?’, *Insights into Imaging*, 2(1), 47–55.

Saiharsha, B., Abel Lesle, A., Diwakar, B., Karthika, R., Ganesan, M. (2020) ‘Evaluating performance of deep learning architectures for image classification’, *Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020*, (Icces), 917–922.

Selvarani, S., Rajendran, P. (2019) ‘Detection of Renal Calculi in Ultrasound Image Using Meta-Heuristic Support Vector Machine’, *Journal of Medical Systems*, 43(9), 1–9.

Shafiei, S., Safarpoor, A., Jamalizadeh, A., Tizhoosh, H.R. (2020) ‘Class-Agnostic Weighted Normalization of Staining in Histopathology Images Using a Spatially Constrained

Mixture Model', *IEEE transactions on medical imaging*, 39(11), 3355–3366.

Shrivastava, P., Singh, A., Agarwal, S., Tekchandani, H., Verma, S. (2021) 'Covid detection in CT and X-Ray images using Ensemble Learning', *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, (Iccmc), 1085–1090.

Soni, A., Rai, A. (2020) 'Kidney Stone Recognition and Extraction using Directional Emboss & SVM from Computed Tomography Images', *MPCIT 2020 - Proceedings: IEEE 3rd International Conference on 'Multimedia Processing, Communication and Information Technology'*, 57–62.

Thein, N., Nugroho, H.A., Adji, T.B., Hamamoto, K. (2018) 'An image preprocessing method for kidney stone segmentation in CT scan images', *2018 International Conference on Computer Engineering, Network and Intelligent Multimedia, CENIM 2018 - Proceeding*, 147–150.

Vakili, M., Ghamsari, M., Rezaei, M. (2020) 'Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification', available: http://arxiv.org/abs/2001.09636.

Verma, J., Nath, M., Tripathi, P., Saini, K.K. (2017) 'Analysis and identification of kidney stone using Kth nearest neighbour (KNN) and support vector machine (SVM) classification techniques', *Pattern Recognition and Image Analysis*, 27(3), 574–580.

Viswanath, K., Gunasundari, R. (2016) 'VLSI implementation and analysis of kidney stone detection from ultrasound image by level set segmentation and MLP-BP ANN classification', *Advances in Intelligent Systems and Computing*, 394, 209–227.

Yildirim, K., Bozdag, P.G., Talo, M., Yildirim, O., Karabatak, M., Acharya, U.R. (2021) 'Deep learning model for automated kidney stone detection using coronal CT images', *Computers in Biology and Medicine*, 135(April), 104569, available: https://doi.org/10.1016/j.compbiomed.2021.104569.