Artificial Intelligence in Nephrology: Detecting Chronic Kidney Disease using Neural Network

by

Md. Farhan Rakib Hridoy
18201070
Ahnaf Asif
18201075
Mashruf Mahmud
18201084
Rashad Rahman
18301004
Md Sahin Siraj
18201104

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
May 2024

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name Signature:

Md. Farhan Rakib Hridoy
18201070

Ahnaf Asif
18201075

Mashruf Mahmud
18201084

Rashad Rahman
18301004

Md Sahin Siraj
18201104

Approval

The thesis titled "Artificial Intelligence in Nephrology: Detecting Chronic Kidney Disease using Neural Network" submitted by

1. Md. Farhan Rakib Hridoy (18201070)
2. Ahnaf Asif (18201075)
3. Mashruf Mahmud (18201084)
4. Rashad Rahman (18301004)
5. Md Sahin Siraj (18201104)

Of Spring, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May, 2024.

Examining Committee:

Supervisor:
(member)

Dewan Ziaul Karim
Lecturer
Department Of Computer Science And Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department Of Computer Science And Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Chronic kidney disease (CKD) is a significant global health concern, impacting more than 800 million people globally. Prompt identification and precise categorization are crucial for optimal therapy. The primary objective of this study is to create a sophisticated machine learning algorithm that can effectively identify and categorise Chronic Kidney Disease (CKD). We use a convolutional neural network (CNN) to examine medical imaging data, namely CT scan pictures. The full dataset was partitioned into training, validation, and testing subsets, and the performance of several pre-trained models, including VGG16, ResNet50, and EfficientNetB0, was assessed. The CNN model suggested obtained exceptional outcomes, showcasing substantial promise in differentiating between normal and diseased kidney states and precisely categorising CKD phases. The model attained a training accuracy of 97.05% and a validation accuracy of 91.79%. The findings emphasise the capability of our technology to aid healthcare practitioners in making prompt and precise choices about the diagnosis and treatment of CKD.

Keywords:

Chronic kidney disease

Early detection

Accurate classification

Medical imaging data

Machine learning algorithms

Convolutional neural networks

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.

Secondly, to our supervisor Dewan Ziaul Karim sir, we would like to express our sincere gratitude for his steadfast assistance, priceless counsel, and insightful academic observations. His advice have been invaluable every time we've needed it.

And finally, We would want to sincerely thank our parents for their unwavering love, support, and prayers. Our academic success and our near-graduation have been largely attributed to their unshakable faith in us.

Table of Contents

Declaration i

Approval ii

Abstract iii

Acknowledgment iv

Table of Contents v

List of Figures vii

List of Tables viii

Nomenclature ix

1 Introduction 1
- 1.1 Problem Statement . . . 2
- 1.2 Research Objectives . . . 3
- 1.3 Thesis Organization . . . 3

2 Literature Review 5
- 2.1 Machine Learning (ML) . . . 5
  - 2.1.1 Machine Learning Models . . . 5
- 2.2 Related Works . . . 6

3 Dataset 8
- 3.1 Data Description . . . 9
- 3.2 Data Splitting and Formation . . . 10

4 Methodology 14
- 4.1 Working Plan . . . 14
- 4.2 Proposed Model Explanation . . . 15
  - 4.2.1 Proposed Model Architecture . . . 15
  - 4.2.2 Input Data . . . 18
  - 4.2.3 Processing . . . 18
  - 4.2.4 Validation . . . 18

5 Results and Discussion 19
- 5.1 Hardware Specification . . . 19
- 5.2 Pre-trained Models Performance . . . 19

5.2.1 VGG16 Model Result 19
5.2.2 VGG19 Model Result 20
5.2.3 ResNet50 Model Result 20
5.2.4 MobileNetV2 Model Result 21
5.2.5 DenseNet121 Model Result 21
5.2.6 EfficientNetB0 Model Result 22
5.2.7 InceptionV3 Model Result 22
5.2.8 ResNeXt Model Result 23
5.2.9 Xception Model Result 23
5.3 Proposed Model Result 24
5.4 Comparison of Models 26

6 Conclusion **29**
6.1 Conclusion 29
6.2 Future Work 29

Bibliography **30**

List of Figures

3.1 Normal 8
3.2 Tumor 8
3.3 Cyst 9
3.4 Stone 9
3.5 Items of Classes 10
3.6 Data Splitting 11
3.7 Training Set 12
3.8 Validation Set 12
3.9 Testing Set 13

4.1 Data Pre-processing 14
4.2 Proposed Model Architecture 16
4.3 Tumor 18
4.4 Stone 18
4.5 Cyst 18
4.6 Normal 18

5.1 VGG16 Accuracy 20
5.2 VGG16 Loss 20
5.3 VGG19 Accuracy 20
5.4 VGG19 Loss 20
5.5 ResNet50 Accuracy 21
5.6 ResNet50 Loss 21
5.7 MobileNetV2 Accuracy 21
5.8 MobileNetV2 Loss 21
5.9 DenseNet121 Accuracy 22
5.10 DenseNet121 Loss 22
5.11 EfficientNetB0 Accuracy 22
5.12 EfficientNetB0 Loss 22
5.13 InceptionV3 Accuracy 23
5.14 InceptionV3 Loss 23
5.15 ResNeXt Accuracy 23
5.16 ResNeXt Loss 23
5.17 Xception Accuracy 24
5.18 Xception Loss 24
5.19 Proposed Model Accuracy 24
5.20 Proposed Model Loss 24
5.21 Confusion Matrix 25
5.22 Training Accuracy 26
5.23 Validation Accuracy 27

List of Tables

4.1 Proposed CNN Model Structure 17
4.2 Training Parameters 17

5.1 Hardware Specification 19
5.2 Classification Report 25
5.3 Training and Validation Accuracy Comparison 27
5.4 Testing Accuracy of Various Models 28

Nomenclature

The list includes many symbols and abbreviations that will be used later in this document.

CKD - Chronic Kidney Disease
MRI - Magnetic Resonance Imaging
CT - Computed Tomography
CNN - Convolutional Neural Network
AI - Artificial Intelligence
ID3 - Iterative Dichotomiser 3
SVM - Support Vector Machine
GFR - Glomerular Filtration Rate
ANN - Artificial Neural Networks
KNN - K-nearest Neighbors
LR - Logistic Regression
PLS DA - Partial Least Squares Discriminant Analysis
ReLU - Rectified Linear Unit
Adam - Adaptive Moment Estimation
GPU - Graphics Processing Unit
VGG - Visual Geometry Group
ResNet - Residual Network
ResNeXt - Residual Next

Chapter 1

Introduction

Machine learning is an exclusive field within the science of artificial intelligence (AI) which is focused on creating algorithms and statistical models. These models allow computer systems to enhance their performance on a particular job by acquiring knowledge from data. The fundamental principles of machine learning, including neural networks and decision trees, were established in the early 1950s.In further time period it evolves and expands with the development of algorithms like ID3, concepts like reinforcement learning. Furthermore, Support Vector Machines (SVMs) and ensemble methods such as Random Forests gained prominence during this period, and there were significant advancements of deep learning which helped in kidney segmentation [1].

This thesis embarks on a journey to address this critical healthcare challenge through the development and application of cutting-edge algorithms. The main objective is to develop a reliable system that can identify and categorize CKD with accuracy. Achieving high diagnostic accuracy is essential since it has a direct impact on when treatment can be started, potentially reducing the effects of this degenerative disorder [2].

The methodology is based on the collection of a wide range of medical imaging data, including modalities like ultrasound, computed tomography, and magnetic resonance imaging.These multi-model datasets serve as a core resource for developing and testing the suggested algorithms, laying the groundwork for sophisticated image-based CKD diagnosis.

This research increases its attention to strengthen preprocessing methods and improve the caliber of medical imaging data in the goal of increased precision. CNNs are used to rigorously separate between healthy kidney states and pathological abnormalities using deep learning, ensuring a high level of accuracy in CKD identification [3].

Furthermore, the research endeavors to transcend binary outcomes by training the model to classify CKD stages, affording invaluable insights into the progression of the disease [4]. This precise understanding gives medical professionals the tools they need to assess the severity of a patient's ailment precisely, which informs the development of personalized treatment regimens.

The results of this research unveil a promising system ready to revolutionize CKD diagnosis and monitoring. This method has the ability to enhance the quality of patient care and reduce the medical expenses associated with treating CKD by providing healthcare practitioners with prompt and accurate information on the state of CKD [5]. A software programme designed for people with chronic kidney disease may provide them with wellness guidelines for their daily lives, which can be quite beneficial. The proposed methodology is further refined as this work draws to a close, opening the door for even more successful CKD detection and classification strategies in the future [3].

1.1 Problem Statement

The prognosis for individual CKD patients is currently lacking in sufficient information. Furthermore, there is an urgent need for expedited prediction of the severity of chronic kidney disease (CKD) utilising easily accessible age group and blood biochemical characteristics throughout the follow-up period [6]. Early detection of CKD is a major challenge. There are four explored techniques, even though the SVM classifier gives the highest accuracy and sensitivity, it is very tough to predict CKD [4].

As CKD is dependable on many factors such as glomerular filtration rate(GFR) and patient's previous medical report must be considered before data collection. These changeable factors can be an issue for predicting the condition of the [7]. By using more and more dataset instances the critical importance of early intervention in improving outcomes can be done.

In addition, diverse medical imaging data sources, including MRI, CT scans, and ultrasound, need to be integrated, processed, and effectively utilized for robust CKD diagnosis, considering the inherent variability in image characteristics. For that particular reason, handling diverse medical imaging data was a huge challenge for collecting dataset information [8]. The lack of information on the dataset used for developing and evaluating the predictive algorithm delays the comprehension of the research's applicability and dependability. It is crucial to know specifics about the dataset's size, origin, and representativeness.

As per data [3], the information provided does not specify the algorithms used for medical image processing in CKD evaluation. Details on the image analysis techniques are essential for understanding the methodology. The current method for assessing CKD severity involves 24-hour urinary protein assessment, which is inconvenient during follow-up and often inconvenient in outpatient clinics [6].

How CNN architectures can be optimized and tailored specifically for CKD diagnosis, ensuring that they accurately differentiate between normal and pathological kidney conditions. As stated in [9], Artificial Neural Networks (ANN) can perform well for the practical challenges and considerations for integrating machine learning-based CKD diagnosis into clinical settings. These challenges can be addressed effectively by real world clinical implementation.

Furthermore, for any diverse patient populations and clinical environments few algorithms and models struggle for maintaining high accuracy and robustness.As per data[10], between all six machine learning algorithms, the random forest model achieved an accuracy rate of 99.75% which is one of the highest. If we can ensure a more multivariate dataset, we can gain more accuracy in the near future.

The ongoing research to refine and improve the proposed methodology, ensuring that CKD detection and classification techniques remain at the forefront of medical advancement[11].

Therefore, the problem remains until getting more individual data without any noise into the images, modeling the scattering properties of CKD patients and developing user-friendly tools for primary care or community-level screening.

1.2 Research Objectives

These research objectives reflect the overarching goals of detecting the stage of CKD, which involve the development of machine learning-enhanced CKD diagnosis and classification techniques to address the pressing need for accurate and early detection of this widespread medical condition. The objectives are given below:

1. To create and apply use a deep learning model that can recognize and categorize various phases of CKD in medical photos.
2. To create image processing techniques that improve the clarity of medical images and the precision of CKD detection.
3. Comparing the system's effectiveness to more established diagnostic methods in terms of sensitivity, specificity, and accuracy.
4. Employ these algorithms to assist in the prompt detection of CKD, recognising the pivotal significance of early intervention in enhancing patient outcomes.
5. Encourage further research to explore and refine the proposed methodology, ensuring continuous advancements in CKD detection and categorization methods for the benefit of patients and healthcare professionals.

1.3 Thesis Organization

This thesis is organized into several chapters, each focusing on a different aspect of the research on detecting Chronic Kidney Disease (CKD) using neural networks:

- Chapter 1: **Introduction**
  This chapter presents a comprehensive summary of the study, including the problem statement, research objectives, and the general arrangement of the thesis.

- Chapter 2: **Literature Review**
  This chapter reviews relevant literature on machine learning and its application in medical diagnostics, specifically focusing on CKD. It discusses various machine learning models and related works that have contributed to this field.

- Chapter 3: **Dataset**
  This chapter describes the dataset used for the research, including data collection methods, data description, and how the data was split and formatted for analysis.

- Chapter 4: **Methodology**
  This chapter outlines the research methodology, detailing the working plan, the proposed model architecture, input data, processing methods, and validation techniques used in developing the CKD detection model.

- Chapter 5: **Results and Discussion**
  This chapter presents the results of the research, including hardware specifications, performance of pre-trained models, results of the proposed model, and a comparison of different models. It also discusses the implications of the findings.

- Chapter 6: **Conclusion**
  The concluding chapter provides a concise overview of the principal discoveries made throughout the research, discusses the limitations, and provides suggestions for future work.

  By structuring the thesis in this manner, it ensures a logical flow of information, making it easier to understand the research process and the results achieved.

Chapter 2

Literature Review

The potential for detecting and categorizing Chronic Kidney Disease (CKD) using machine learning is enormous. Large volumes of medical imaging data, like those from MRI and CT scans, may be processed by machine learning algorithms with amazing precision, enabling early and accurate CKD diagnosis. They enable differentiation between various CKD stages, leading to individualized treatment regimens and actions. Additionally, as machine learning models learn from more data, such as convolutional neural networks (CNNs), their diagnostic accuracy can constantly increase [12]. This has the potential to revolutionize the identification and categorization of CKD, lower healthcare costs, and enhance patient outcomes.

2.1 Machine Learning (ML)

Machine learning is a specialised area within the subject of artificial intelligence that concentrates on creating algorithms as well as statistical models that allow computing devices to gain information and generate assumptions or judgements without the need for programming expertise. It has seen substantial evolution throughout the years, originating from the mid-20th century. Early innovations relied on simple algorithms and rule-based systems. However, the development of digital processing and the accessibility of enormous datasets gave the subject a boost. Machine learning has advanced significantly over the past few decades thanks to developments in deep learning, neural networks, and high-performance computing. Machine learning is currently at the cutting edge of technology, having an impact on a variety of industries like healthcare, finance, natural language processing, and picture recognition, among others. With ongoing research and invention, it is also continuing to evolve quickly [3].

2.1.1 Machine Learning Models

This article [4] used clinical data to investigate four distinct machine learning approaches for the purpose of predicting CKD. The four strategies that have been examined are KNN, SVM, LR, and decision tree classifiers. The results of the study demonstrated that the SVM classifier exhibited the best levels of both precision and sensitivity. The study focuses on the automated identification of kidney cysts, stones, and tumours using CT-radiography [6]. This is achieved via the utilisation of vision transformers and explainable transfer learning models. The authors sug-

gest using sophisticated machine learning approaches to tackle this medical imaging difficulty. [11].

2.2 Related Works

The research article [10] used KNN imputation to replace the missing data. The data collection contains a total of 400 samples. Out of the total 400 samples, 250 belong to the ckd group and 150 belong to the notckd category. It is crucial to acknowledge that the data set includes a substantial amount of missing values. When dealing with incomplete data sets, six machine learning methods, including logistic regression, random forest, support vector machine, k-nearest neighbour, naive Bayes classifier, and feed forward neural network, are used to generate models after successfully completing the missing data. Random forest demonstrated superior performance compared to other machine learning models, with a diagnostic accuracy rate of 99.75%.

The objective of the Project, as stated in the research article [5], is to develop a digital tool for CKD patients that may provide them with self-care guidelines for their daily lives. The flash solution for CKD healthcare instruction was developed with Adobe Flash CS5.5.An Adobe Dreamweaver website has been established to promote this multimedia application, providing clear directions for self-care for patients with CKD. A novel user-friendly interface to guide CKD individual care has been created as part of the project, resulting in the creation of a more effective information channel for patients.

The research paper [13] explores various systems, molecules, and reactions that contribute to the development of pathological fibrosis in chronic kidney disease. Specifically, it examines the role of pain, the renin-angiotensin system, parathyroid hormone, fibroblast growth factor 23, Klotho, microRNAs, and the vitamin D hormonal system in this process. All of them are crucial constituents of the regulatory and fundamental pathways that facilitate fibrosis, a condition that has a significant detrimental impact on the kidney and heart in chronic kidney disease.

Over time, Chronic Kidney Disease has undergone development and the diagnostic criteria and categories of CKD have been modified. Chronic kidney disease is influenced by several variables such as age, gender, dietary habits, geographical location, and past medical history. Currently, the globally accepted approach to detect chronic kidney disease relies on the measurement of glomerular filtration rate. A glomerular filtration rate below 60 mL/min per 1.73m2 indicates reduced kidney function, whereas a GFR below 15 mL/min per 1.73 m2 signifies renal failure or end stage kidney disease [7].

A comparative research was conducted to evaluate nine prediction models for forecasting the course and severity of chronic kidney disease based on non-urinary clinical and demographic characteristics. Linear models, such as logistic regression, had the greatest predictive capability. Common blood tests, including as albumin, serum creatinine, triglycerides, low-density lipoprotein, and estimated glomerular filtration rate values, were shown to be valuable indicators. The citation for the

source is "Islam et al., 2022". A unique method using ultrasonic imaging was developed to evaluate proteinuria in different phases of chronic kidney disease, specifically targeting the speckle effect seen in the pictures. The system utilises the Nakagami distribution and Local Binary Pattern to represent the scattering characteristics, with a particular focus on the age distribution as a significant aspect. The system has a high level of sensitivity and specificity [14].

A research study applied artificial intelligence and deep learning algorithms to identify chronic kidney disease by analysing retinal pictures in persons residing in community settings. The programme, which used retinal pictures as input data, was verified on several datasets. The hybrid deep learning algorithm exhibited the most superior performance in accurately assessing chronic kidney disease, so indicating the viability of using retinal imaging as a non-invasive method for screening [15].

There is a research that provides a clinical update sourced from Kidney Disease: Improving Global Outcomes [9]. It specifically examines the correlation between cardiovascular disease and chronic kidney disease. The article offers a comprehensive examination of the present knowledge on cardiovascular problems in patients with CKD. This includes an analysis of risk factors, techniques for prevention, and recommendations for therapeutic care [2].

The study [1] focuses on identifying and diagnosing chronic kidney disease via the utilisation of a deep learning-derived heterogeneous refined artificial neural network. The paper introduces a method that utilises advanced deep learning methods and an enhanced artificial neural network to enhance the precision of detecting and diagnosing Chronic Kidney Disease.

The study suggests using fuzzy classifiers for diagnosing individuals with chronic kidney disease. The research use two fuzzy classifiers to construct a diagnosis model for Chronic Kidney Disease[13]. Fuzzy classifiers are a machine learning method specifically designed to handle situations when there is ambiguity and imprecise input. A total of 386 dataset samples were utilised for the investigation. Among the 400 samples, 251 were classified as belonging to the chronic kidney disease group, while the remaining 149 samples were classified as belonging to the non-chronic kidney illness category. The research examines the use of these classifiers to enhance the precision of CKD diagnosis. The diagnostic model produced in the PLS-DA classifier obtains an accuracy of $95.5 \pm 0.6\%$ and a specificity of $89.5 \pm 1.3\%$, which are the lowest among the tested models. However, it achieves the highest sensitivity of 100%.
The above discussion indicates that this issue is multifaceted and has a substantial body of literature associated with it. The evaluated research together emphasise the considerable potential of machine learning in the identification of Kidney disease.Furthermore, the enhanced availability of vast datasets and the rise in computational capabilities might facilitate the creation of a system that aids healthcare practitioners in making clinical choices by providing them with prompt and precise details regarding patients' CKD condition.

Chapter 3

Dataset

The initial dataset consists of 12,446 distinct data points, with 3,709 representing cysts, 5,077 representing normal cases, 1,377 representing stones, and 2,283 representing tumours.The dataset used in this study is sourced from the CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone dataset [6]. To address the issue of dataset inequalities, we use class trimming to ensure an equal amount of photographs in each class.We augmented several classes by expanding them, resulting in an overall of 12000 photos utilised. Out of the total of 12,000 photos, 8,400 were allocated for training, 2,400 were designated for validation, and 1,200 were set out specifically for testing. All the images were in black and white, with dimensions of 224 X 224 pixels. Below are examples from several classes:

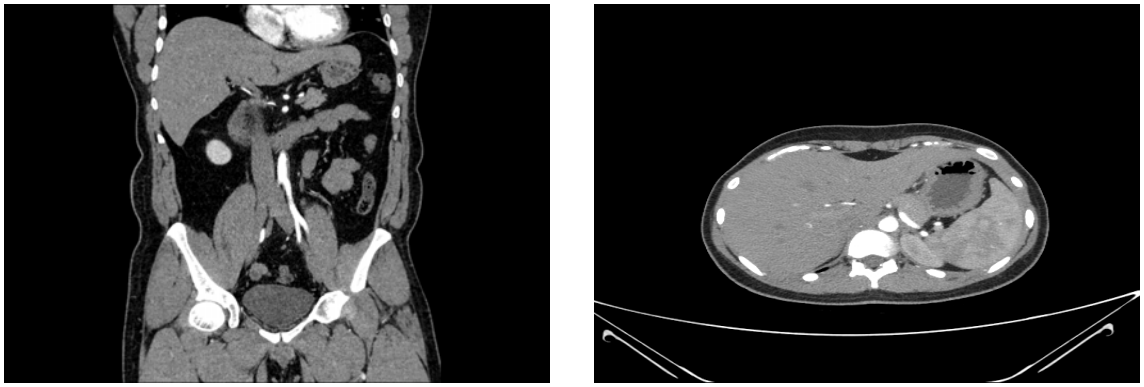

Figure 3.1: Normal

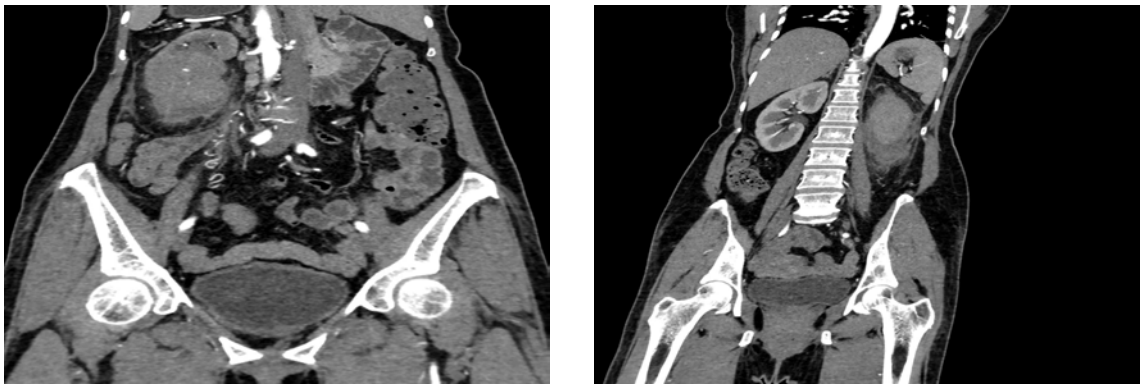

Figure 3.2: Tumor

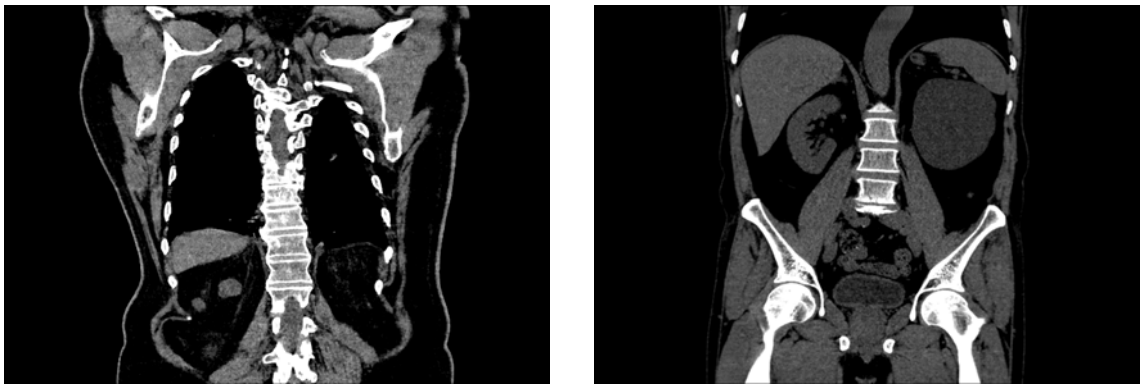

Figure 3.3: Cyst

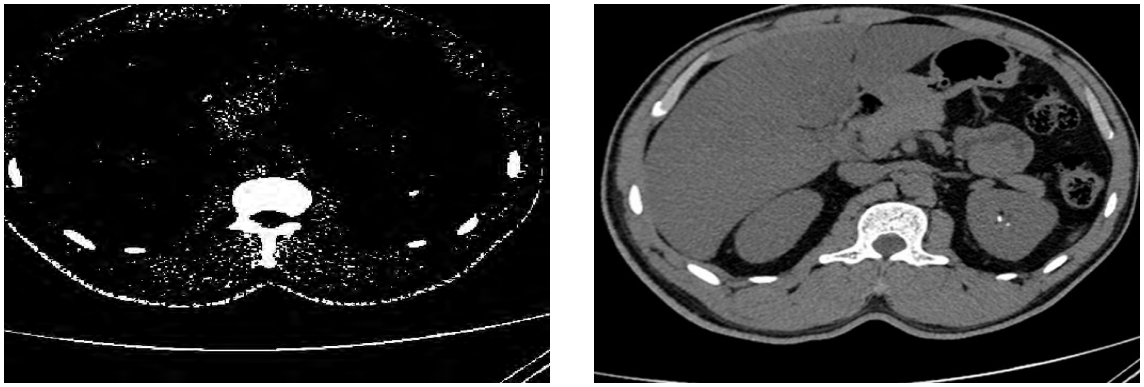

Figure 3.4: Stone

3.1 Data Description

- Resizing Images: The images were resized to a pixel resolution of 224 × 224 as CNN models often need inputs of a set size.
- Dataset Rescaling: The study emphasizes the importance of preprocessing a dataset before training a neural network model. This involves rescaling pixel values, which is done by dividing each pixel value by 255. This standardizes the input data and facilitates faster convergence during training. The normalized dataset is then used for further augmentation and training of the CNN model. This normalization process scales the pixel values from the original range of [0, 255] to the normalized range of [0, 1].

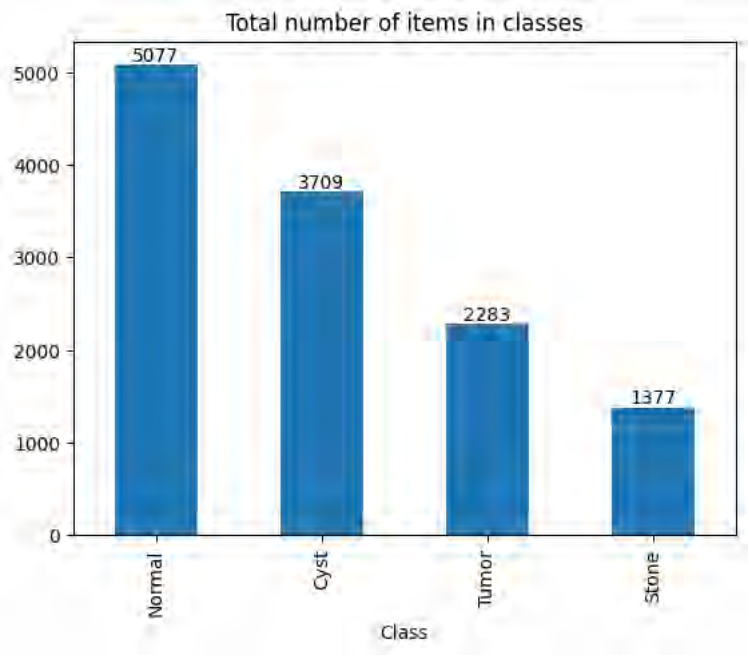

Figure 3.5: Items of Classes

- Data Augmentation: Typically, CNN models exhibit improved performance when provided with a larger number of pictures. Therefore, we used several data augmentation techniques to increase the size of our training data. Various techniques, like shearing, rotating, shifting, flipping, etc., were used to introduce diversity into the dataset and enhance the model's resilience.

3.2 Data Splitting and Formation

Firstly we manually divide our working dataset into 70%, 20%, 10% for training, validation and testing respectively. In addition, we try to distribute the data for every class evenly. Secondly, we trained our custom CNN model using training data and validated our custom model using validation Data.

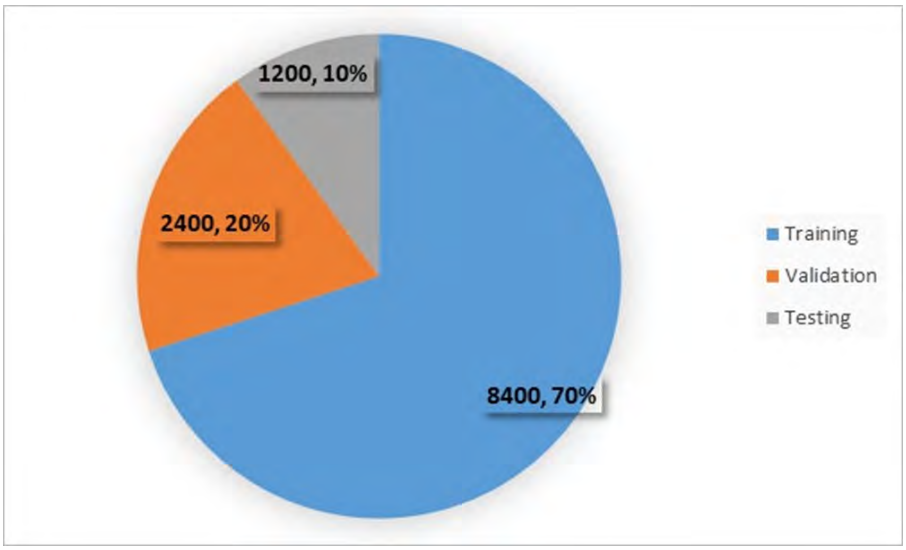

Figure 3.6: Data Splitting

We trimmed or enhanced the number of photos in each class to achieve equal representation in order to solve the imbalance in the dataset. We reduced the number of photographs in the "Cyst" class from 3,709 to 3,000. The 1,377 and 2,283 photographs in the "Stone" and "Tumor" classes, respectively, were increased to 3,000 images each. Comparably, we lowered the 5,077 photographs in the "Normal" class to 3,000 images in order to make it consistent with the other classes.Solving image dataset imbalance by trim and expand between classes. For the dataset split, Each class contributed 2,100 images to the training set. 600 images from each class were allocated to the validation set. The remaining 300 images from each class were reserved for the testing set.

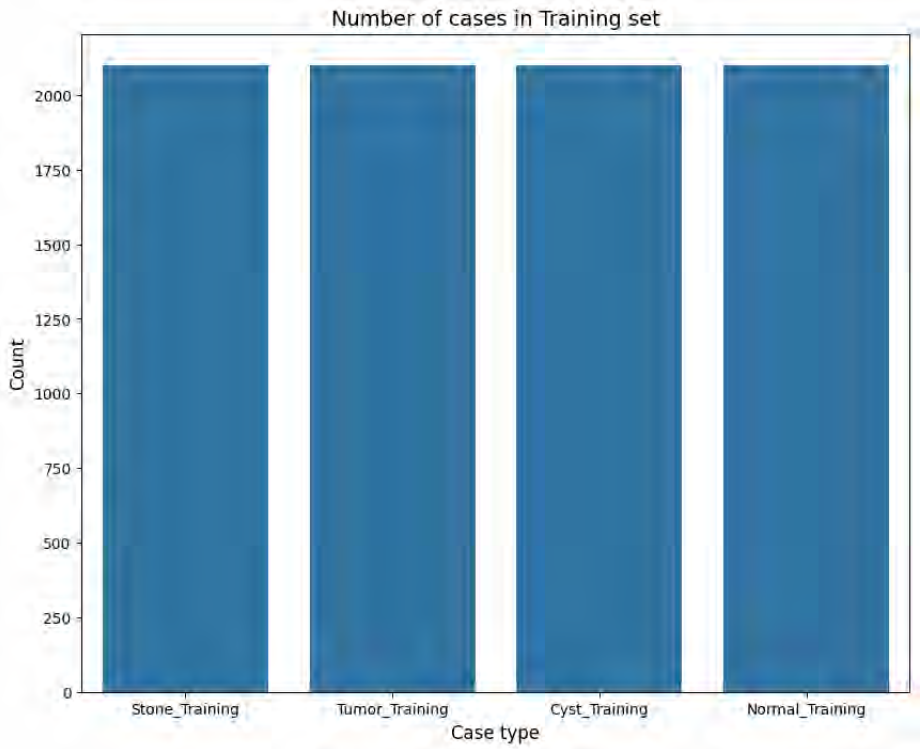

Figure 3.7: Training Set

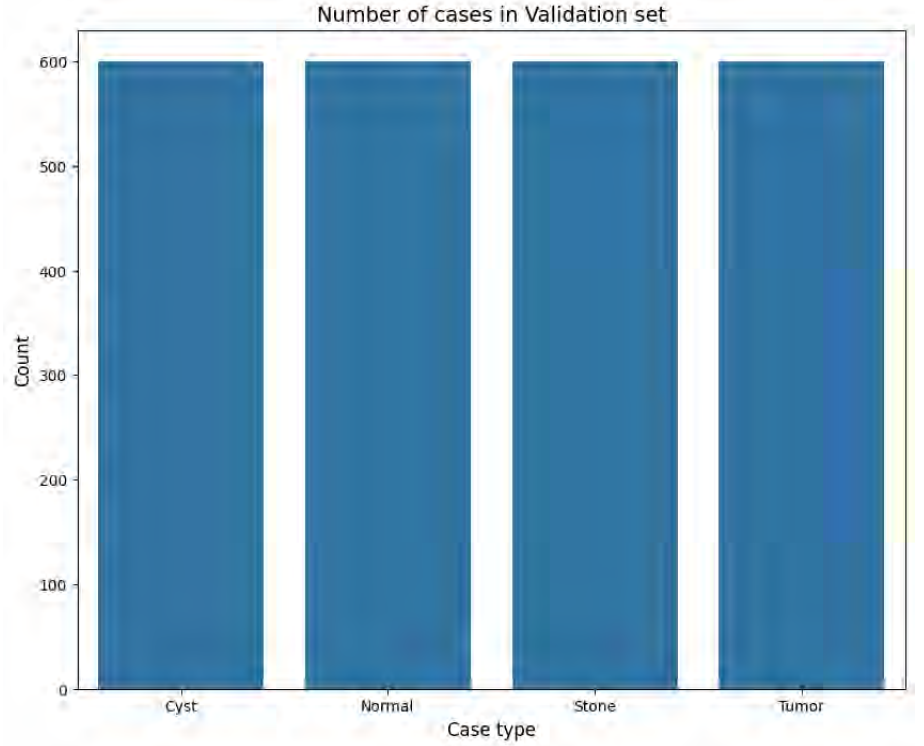

Figure 3.8: Validation Set

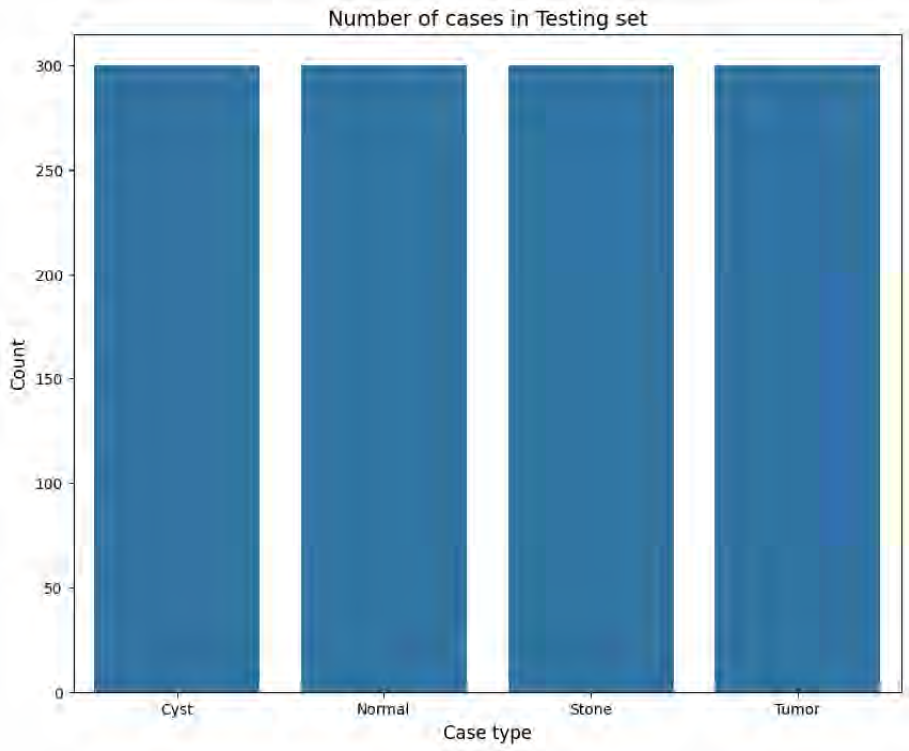

Figure 3.9: Testing Set

Chapter 4

Methodology

4.1 Working Plan

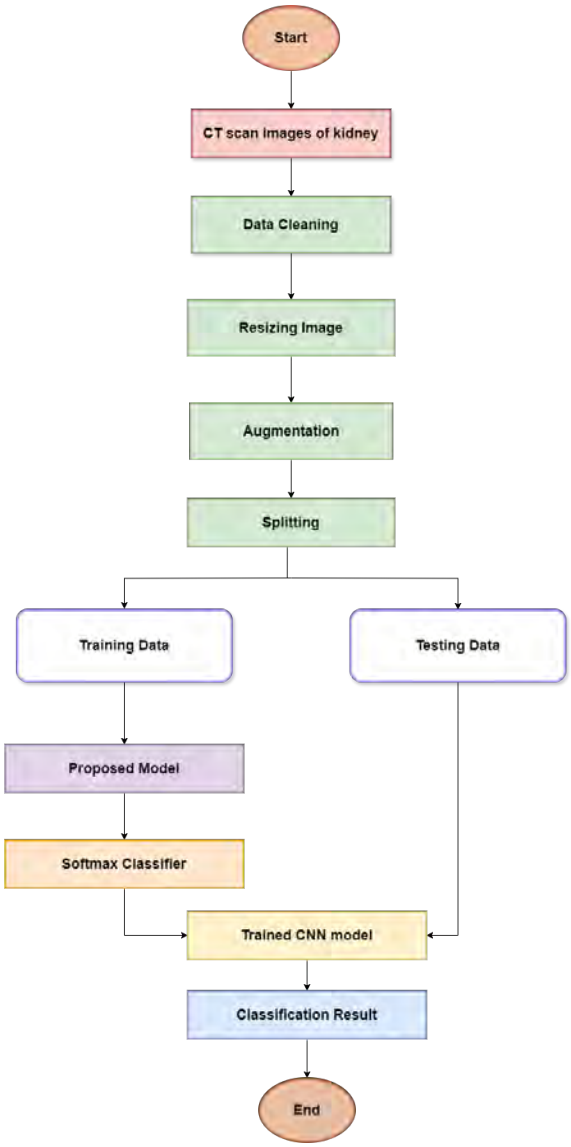

Figure 4.1: Data Pre-processing

After augmenting the dataset, with the train set we will train our model with that. For our model, we developed a custom architecture tailored specifically to our needs. For effective training, we created the dataset format to work with our custom model. We used our testing set to verify the model's performance and make sure it correctly predicts the disease class after training.

4.2 Proposed Model Explanation

A Multi-layered CNN architecture is intended to learn from input images automatically and adaptably. By using deep learning, our algorithm is very good at identifying kidney illness from CT scans. Convolutional, pooling, and fully linked layers were employed. Convolutional layers create feature maps that capture different elements like edges, textures, and forms by applying a collection of filters to the input image. The spatial dimensions of the feature maps are decreased by pooling layers, usually by max-pooling, which aids in lowering the computational effort and managing overfitting. To determine the final classification, these features are interpreted by the fully linked layers at the end of the network. By introducing non-linearity, activation functions such as ReLU allow the network to learn intricate patterns [16]. Our suggested mode can be trained on labeled CT scan images to detect and categorize abnormalities in the context of kidney disease detection with high accuracy, helping with early diagnosis and therapy planning.

$$\mathbf{Y = max(0,X)}$$

4.2.1 Proposed Model Architecture

In this research, we propose a CNN model to detect stone, tumor, cyst, and cancer in the kidney using CT scan images.

- Input Layer: The input dataset consists of images with (227,227) pixels.
- Convolutional Layers: There are 7 convolutional layers with filters 96,256,256,384,384,384,256 respectively. Filter size added in decreasing order. Relu is used as activation. After each convolution layer, the batch normalization method is used to speed up the learning process of the model.
- Pooling Layers: After the first convolutional layer a max-pooling operation with strides (2x2) is added. Another max-pooling layer with (2x2) strides at the end.
- Fully Connected Layers: 2 consecutive dense layers added with 4096, 4096 units respectively.
- Output Layer: The output layer consists of 4 nodes for 4 types of kidney disease and softmax classifier was used as an activation function to produce a prediction table. [17]

$$\frac{\mathbf{e^{x_i}}}{\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{e^{x_i}}}$$

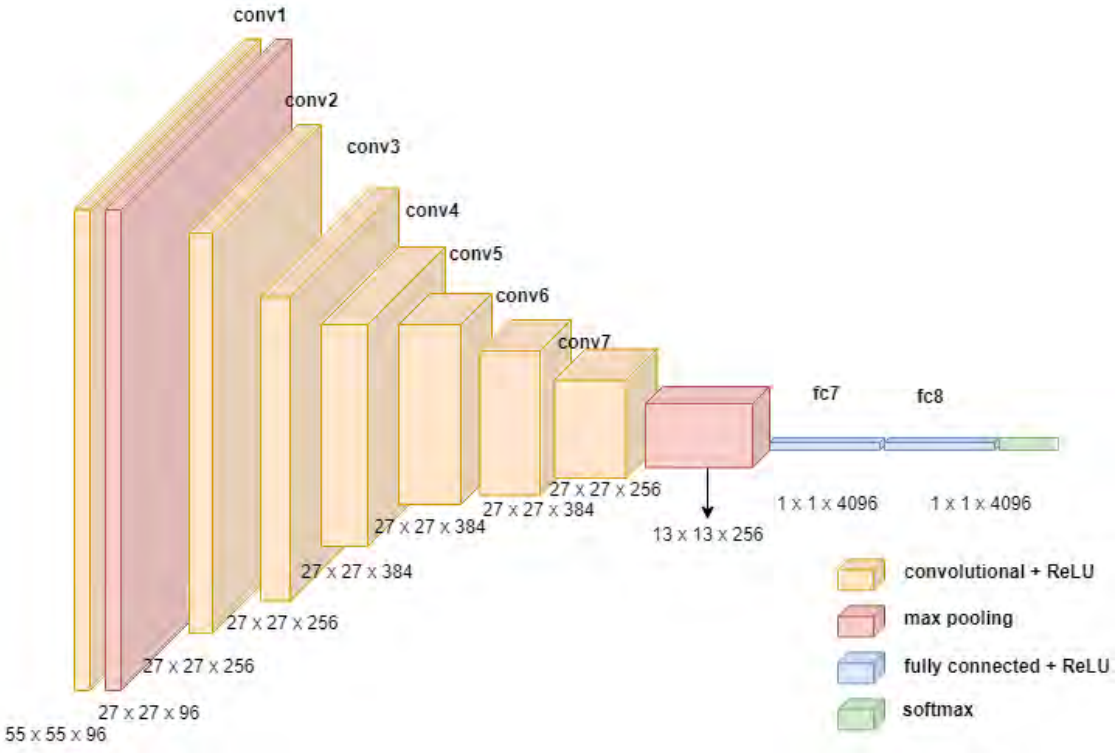

Figure 4.2: Proposed Model Architecture

We proposed a multi-layered CNN model where a total of 7 convolutional layers were added and each conv layer had batch normalization. The model begins with a convolutional layer (conv2d-0), then reduces spatial dimensions and improves feature representation by batch normalization and max-pooling. More complex features are extracted by subsequent convolutional layers (conv2d-1 to conv2d-6) with batch normalization. To more effectively capture important patterns, the feature maps are further downsampled using max-pooling2d-1. Dense layers (dense-0 to dense-2) with dropout regularization allow the model to learn complex patterns and generate predictions after the feature maps have been flattened. The classification probabilities for the dataset's classes are produced by the last dense layer.
Firstly, we prepare the dataset with an appropriate optimizer and loss function specific to our classification objective before building and training this model on it. After that, by giving the training data and their matching labels, we fit the model to the dataset. Under the optimizer's direction, the model iteratively modifies its parameters throughout training in order to minimize the specified loss function. The model's capacity to generalize is further improved by data augmentation. After training, performance parameters including accuracy, precision, recall, and F1 score measured on a different validation set. To maximize performance even more, model architecture and hyperparameter adjustments were made.

Layers	Shape of Output
conv2d-0	(None, 55, 55, 96)
batch-normalization-0	(None, 27, 27, 96)
max-pooling2d	(None, 27, 27, 96)
conv2d-1	(None, 27, 27, 256)
batch-normalization-1	(None, 27, 27, 256)
conv2d-2	(None, 27, 27, 256)
batch-normalization-2	(None, 27, 27, 256)
conv2d-3	(None, 27, 27, 384)
batch-normalization-3	(None, 27, 27, 384)
conv2d-4	(None, 27, 27, 384)
batch-normalization-4	(None, 27, 27, 384)
conv2d-5	(None, 27, 27, 384)
batch-normalization-5	(None, 27, 27, 384)
conv2d-6	(None, 27, 27, 256)
batch-normalization-6	(None, 27, 27, 256)
max-pooling2d-1	(None, 13, 13, 256)
flatten-0	(None, 43264)
dense-0	(None, 4096)
dropout-0	(None, 4096)
dense-1	(None, 4096)
dense-2	(None, 4)
Total params: 206237572 Trainable params: 206233540 Non-trainable params: 4032	

Table 4.1: Proposed CNN Model Structure

For efficient and faster training, Adam was used as the optimizer, in combination with the categorical crossentropy loss function. Besides that, an L2 regularizer was used to reduce overfitting with a weight decay of 0.0001. For machine learning library, we used TensorFlow and for neural network library, we used Keras.

Name of Parameter	Value
Used Optimizer	Adam
Regularizer	l2(0.0001)
Activation	relu
Epochs	30
Steps per Epoch	263
Loss Function	Categorical Crossentropy
Metrics	Accuracy

Table 4.2: Training Parameters

4.2.2 Input Data

Our working dataset consists of CT scan images of kidneys each are 227x227 pixels. The whole dataset is split into three categories, which are training, testing, and validation comprising 70%, 20%, and 10% of the whole dataset respectively.

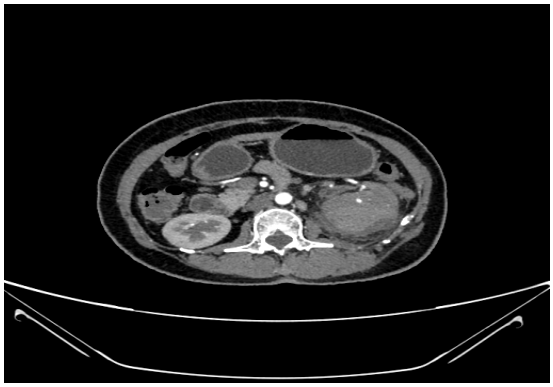

Figure 4.3: Tumor

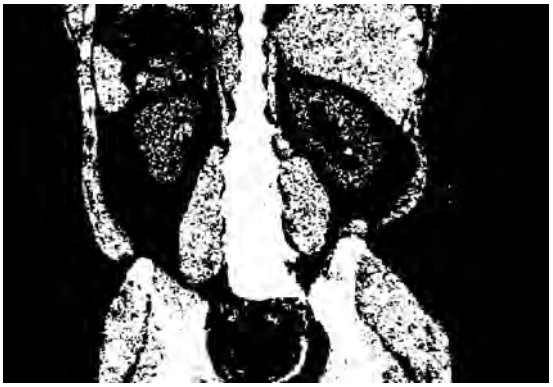

Figure 4.4: Stone

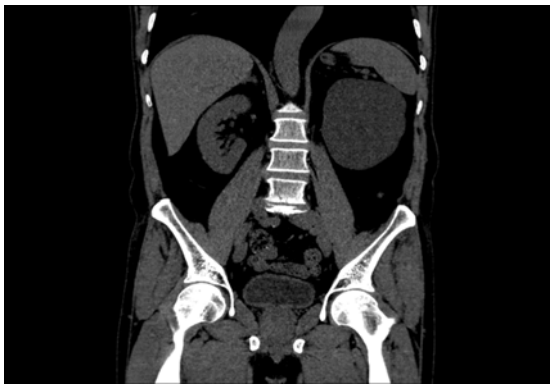

Figure 4.5: Cyst

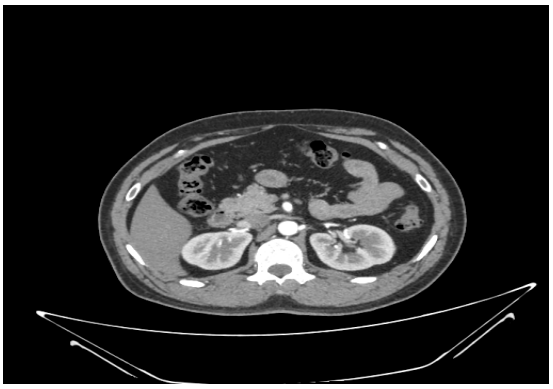

Figure 4.6: Normal

4.2.3 Processing

Our proposed model trained with 30 epochs and training process is done in multiple batches. Model validation accuracy picked with accuracy value of 91%. The model ultimately obtained a training accuracy of 97% and a validation accuracy of 89%.

4.2.4 Validation

Every iteration or epoch following training on the provided dataset, it predicts from validation set frames and compares with ground truth to provide prediction accuracy. The analysis of validation curvature and slope is important in this process. By monitoring the validation loss curve, smoothing out fluctuations was needed. Regularization methods like dropout or early stopping into practice to reduce overfitting and improve the model's generalization performance.

Chapter 5

Results and Discussion

5.1 Hardware Specification

Python 3.x

Hardware	GPU RAM	System RAM
T4 GPU	16 GB	12.7 GB
L4 GPU	22.5 GB	62.8 GB
A100 GPU	40 GB	90 GB

Table 5.1: Hardware Specification

A100 and L4 gpu used for pre-trained model since pre-trained models needs more resource than proposed model. T4 GPU used for proposed model instead of A100 GPU.

5.2 Pre-trained Models Performance

5.2.1 VGG16 Model Result

With a last training accuracy of 99.92%, the VGG16 model demonstrated remarkable performance on the training set. But the validation accuracy was much lower at 53.84% and the validation loss was rather high at 2.3587, indicating that the model had difficulty generalizing to new validation data. In spite of this, the test accuracy, which gauges performance on an entirely different dataset, was 89.42%, suggesting that the model still has some reasonable generalization capacity but still needs work. The last epoch's learning rate of 1.0000e-06 indicated that the model was in the training phase of fine-tuning.

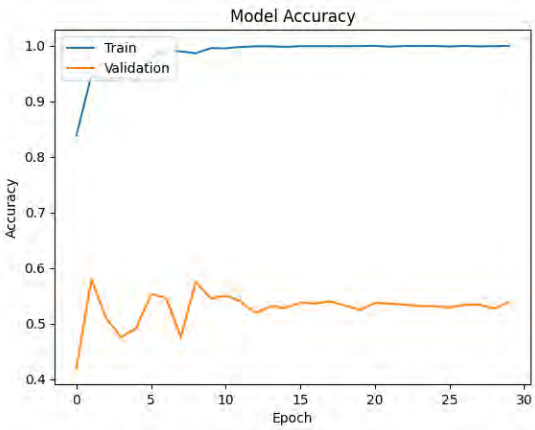

Figure 5.1: VGG16 Accuracy

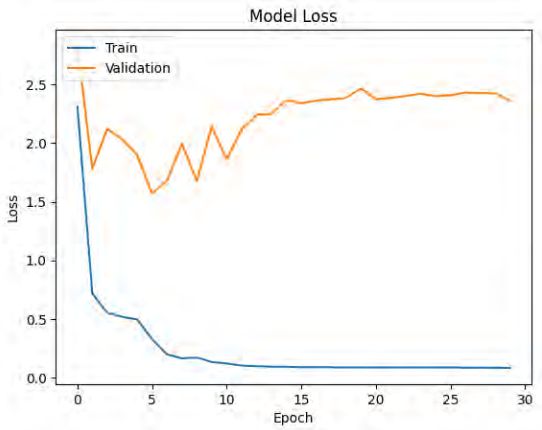

Figure 5.2: VGG16 Loss

5.2.2 VGG19 Model Result

We trained our dataset on the very popular architecture, VGG19 which achieved a high training accuracy of 99.47%, indicating strong performance on the training data. However, with a validation loss of 0.8780 and a validation accuracy of 78.82%, it shows better generalization compared to VGG16 model. The test accuracy was 79.50%, aligning closely with the validation performance, which suggests that the model generalizes well.

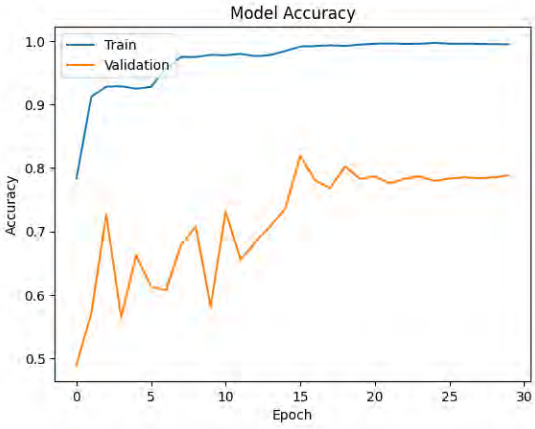

Figure 5.3: VGG19 Accuracy

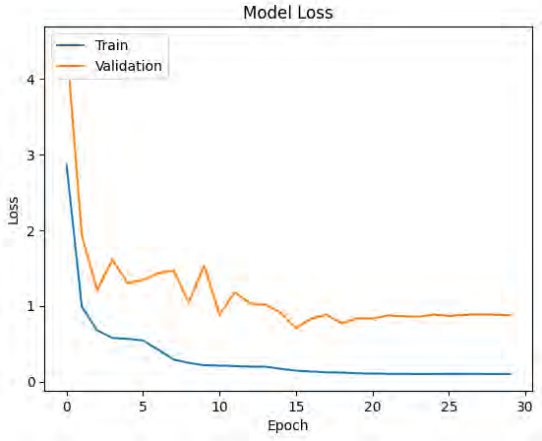

Figure 5.4: VGG19 Loss

5.2.3 ResNet50 Model Result

With a training accuracy of 99.06%, the ResNet50 model proved to be adept at learning on the training set. It does, however, show moderate generalization to the validation data. Despite being less than the training accuracy, the test accuracy of 76.42% nevertheless demonstrates respectable performance. The model appeared to be at the fine-tuning stage based on the learning rate of 1.0000e-06 in the final epoch.

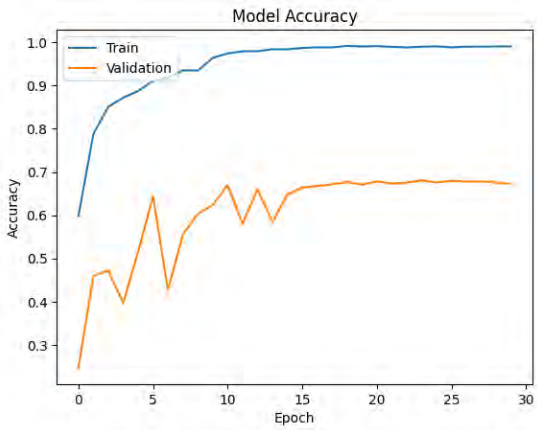

Figure 5.5: ResNet50 Accuracy

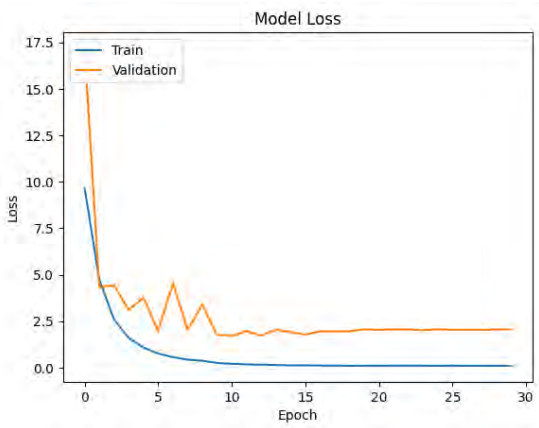

Figure 5.6: ResNet50 Loss

5.2.4 MobileNetV2 Model Result

The perfect training accuracy of 100% shows how successfully the MobileNetV2 model learnt the training data. With a high validation accuracy of 94.92% and a validation loss of 0.3281 on the validation set, the model also demonstrated remarkable performance. Though, the test accuracy dropped to 75.33%, suggesting that there is an overfitting issue preventing the model from properly generalizing to new, unseen data in testing set. The learning rate for the last epoch which represents the fine-tuning stage was 4.0000e-06.

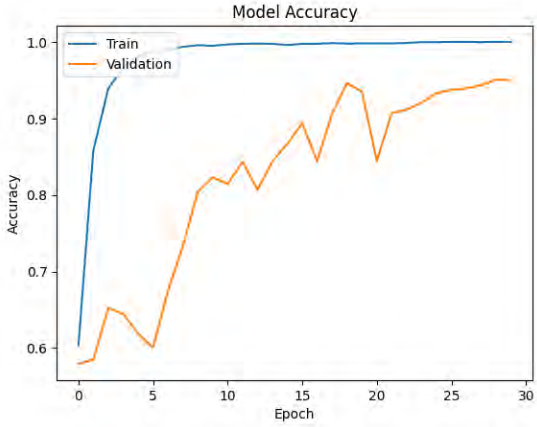

Figure 5.7: MobileNetV2 Accuracy

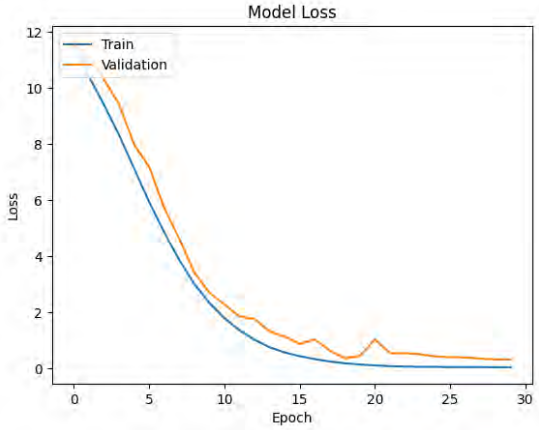

Figure 5.8: MobileNetV2 Loss

5.2.5 DenseNet121 Model Result

With an astounding training accuracy of 99.93%, the DenseNet121 model demonstrated exceptional learning on the training set. The model showed good generalization to the validation data, with a validation accuracy of 85.50% and a validation loss of 0.8616. The test accuracy acquired 88.75% which is roughly matches the validation accuracy. The model was in the fine-tuning stage, as shown by the final epoch's learning rate of 2.0000e-05

From the above output and accuracy and loss graph of DenseNet121 model we can say that the model training accuracy is 99% with 80% validation accuracy.

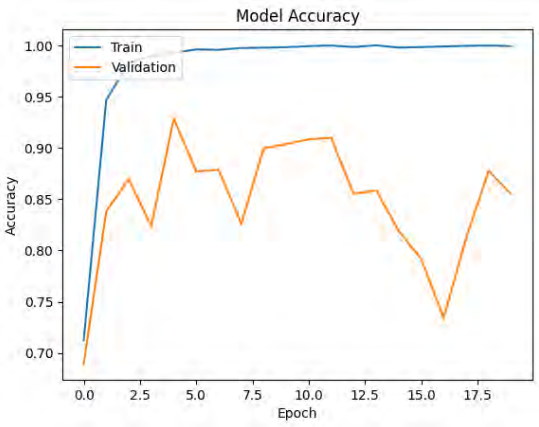

Figure 5.9: DenseNet121 Accuracy

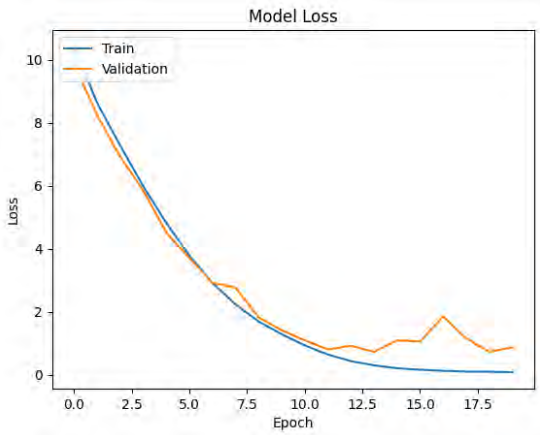

Figure 5.10: DenseNet121 Loss

5.2.6 EfficientNetB0 Model Result

The EfficientNetB0 model demonstrated reasonable learning on the training set. Nevertheless, the model demonstrated challenges in generalizing to the validation data, with a high validation loss of 4.3232 and a validation accuracy of 68.32%. The model's difficulty with generalization is still evident in the test accuracy of 75.67%, which is just marginally better than the validation accuracy. The last epoch's learning rate was 1.0000e-04, suggesting that the fine-tuning process was still in progress.

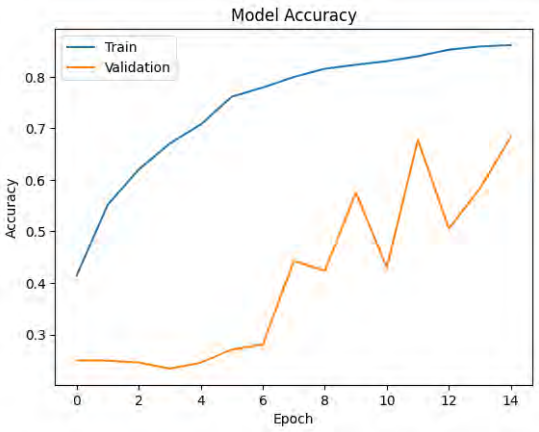

Figure 5.11: EfficientNetB0 Accuracy

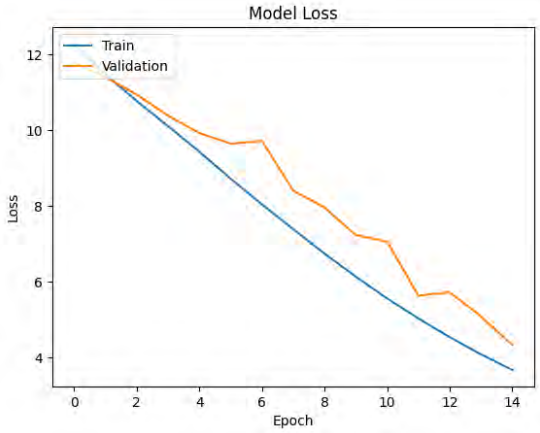

Figure 5.12: EfficientNetB0 Loss

5.2.7 InceptionV3 Model Result

Another very popular model was chosen, InceptionV3 gained exceptional training accuracy of 99.96%. Strong generalization of the model to the validation set was demonstrated, with a validation accuracy of 84.55%. With a test accuracy of 83.17%, which was nearly in line with the validation results. The model was in its fine-tuning stage, as evidenced by the learning rate of 1.0000e-06 in the final epoch. From the above output and accuracy and loss graph of InceptionV3 model, it was exceptionally effective on our dataset.

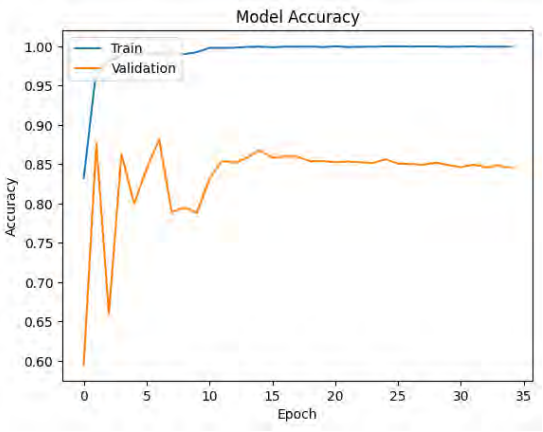

Figure 5.13: InceptionV3 Accuracy

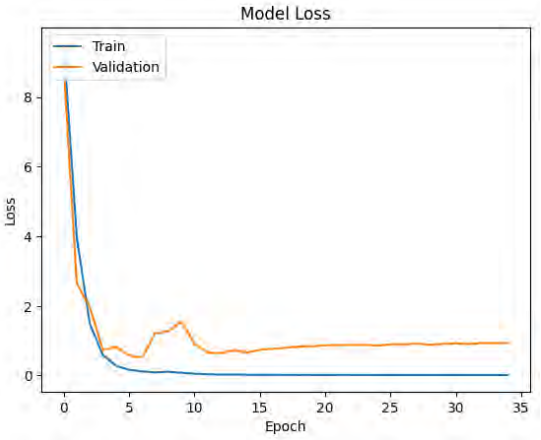

Figure 5.14: InceptionV3 Loss

5.2.8 ResNeXt Model Result

With an astounding accuracy of 99.93% and a very low training loss of 0.0824, the ResNeXt model learning curve was better on the training set. The model showed high generalization to the validation set with a validation accuracy of 85.50% and a validation loss of 0.8616. With a test accuracy of 88.75%, it appears that the generalization to untested data was successful. The model was in the fine-tuning stage, as shown by the final epoch's learning rate of 2.0000e-05.

Figure 5.15: ResNeXt Accuracy

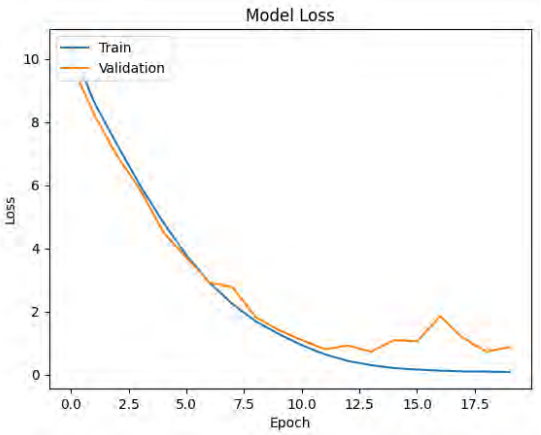

Figure 5.16: ResNeXt Loss

5.2.9 Xception Model Result

The Xception model demonstrated exceptionally successful learning on the training set. However, the model displayed a discernible decline in performance on the validation set, indicating some overfitting. The test accuracy of 75.83% showed a moderate degree of generalization in prediction. The final epoch's learning rate was 1.0000e-04, indicating continued fine-tuning attempts.

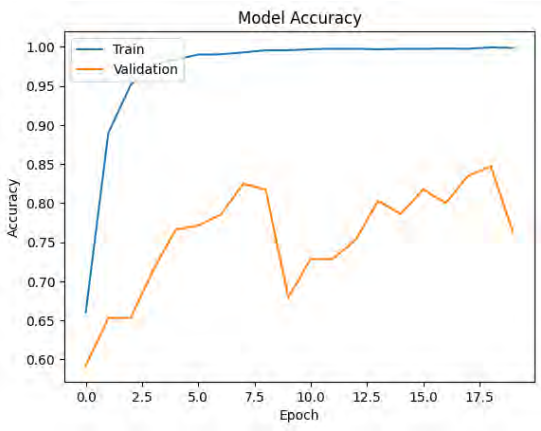

Figure 5.17: Xception Accuracy

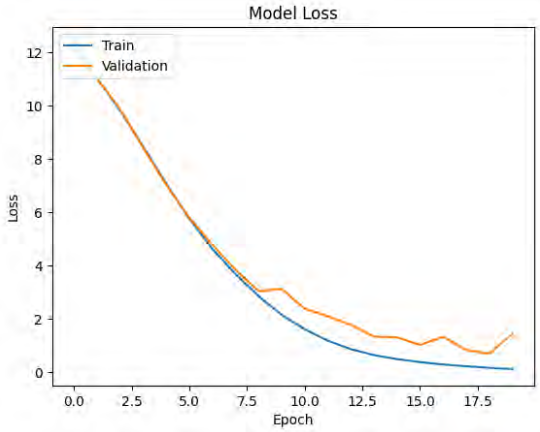

Figure 5.18: Xception Loss

5.3 Proposed Model Result

From below 5.19 and 5.20 graph, we can say that there was a huge rise in training accuracy and the significant drop in training loss indicates that our proposed CNN model's training procedure across 30 epochs displays excellent learning and convergence. Along with a significant improvement in validation loss from 18.5614 to 0.4114, the accuracy of validation increased from 25.25% to 89.58% and maximum validation accuracy gained 91.79%. The validation loss and accuracy show periodic spikes and fluctuate that may indicate overfitting, but generally the trend suggests that the model generalizes well.

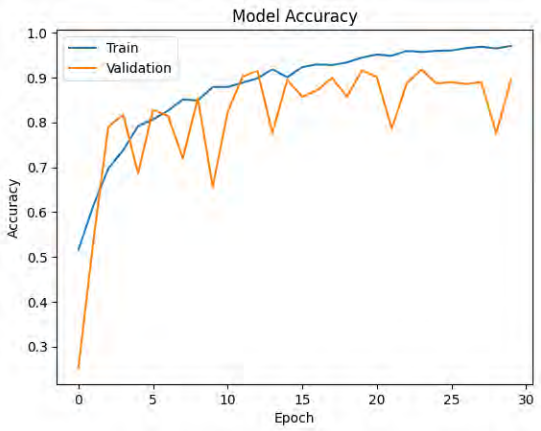

Figure 5.19: Proposed Model Accuracy

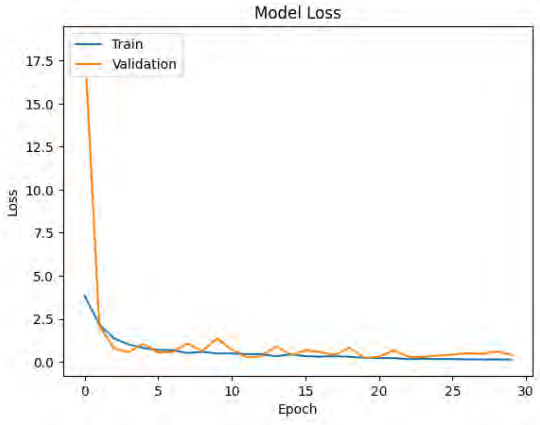

Figure 5.20: Proposed Model Loss

Confusion Matrix

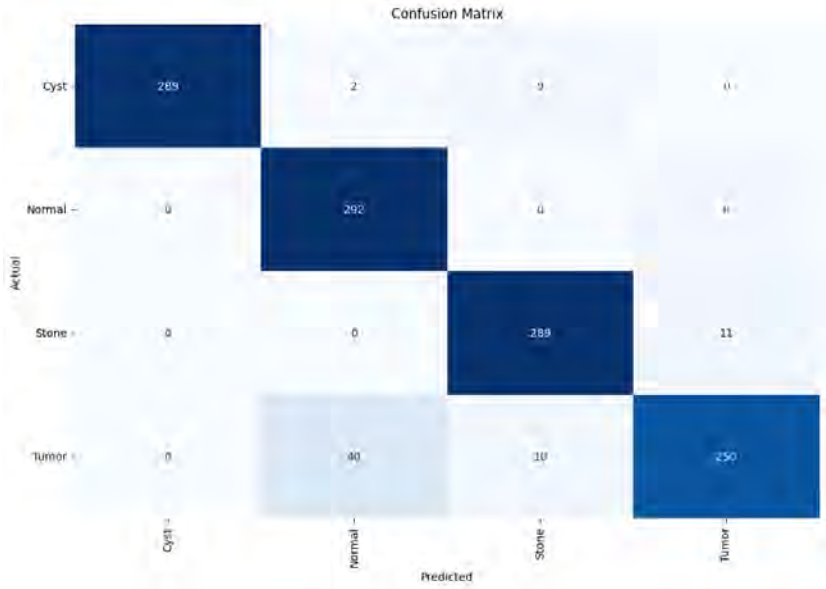

Figure 5.21: Confusion Matrix

From the 5.21, the confusion matrix shows that Cyst provides 100% accuracy to detect actual Cyst. But Normal class is having some error around 40 images as detecting Tumor. Stone class predicted 289 images as actual stone but 10 images as Tumor and 9 images as Cyst. Tumor predicted 250 images as actual Tumor and 11 images as Stone, 8 images as Normal.

Classification Report

	Precision	Recall	F1-score	Support
Cyst	1.0000	0.9633	0.9813	300
Normal	0.8743	0.9733	0.9211	300
Stone	0.9383	0.9633	0.9507	300
Tumor	0.9294	0.8333	0.8787	300
Accuracy		0.9333		1200
Macro Avg	0.9355	0.9333	0.9330	1200
Weighted Avg	0.9355	0.9333	0.9330	1200

Table 5.2: Classification Report

From 5.2, the classification report shows that 1200 tests were performed, and 80 errors were found in the confusion matrix, yielding an accuracy of 93.33%.

5.4 Comparison of Models

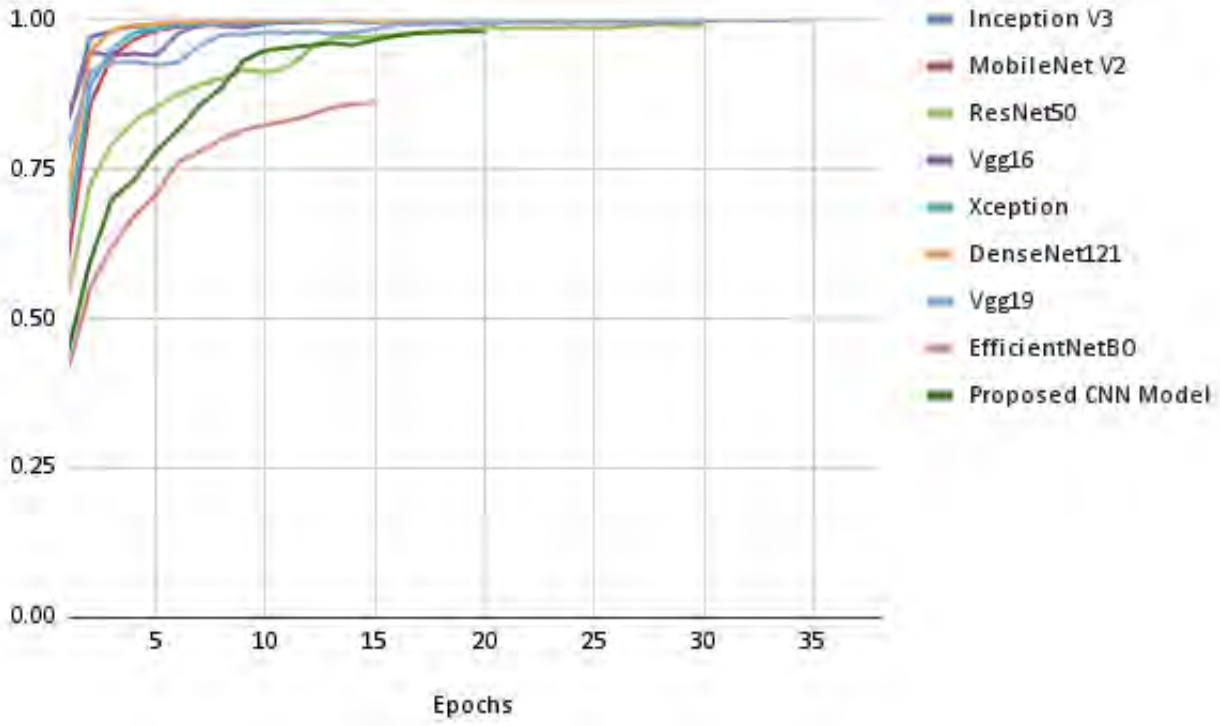

Figure 5.22: Training Accuracy

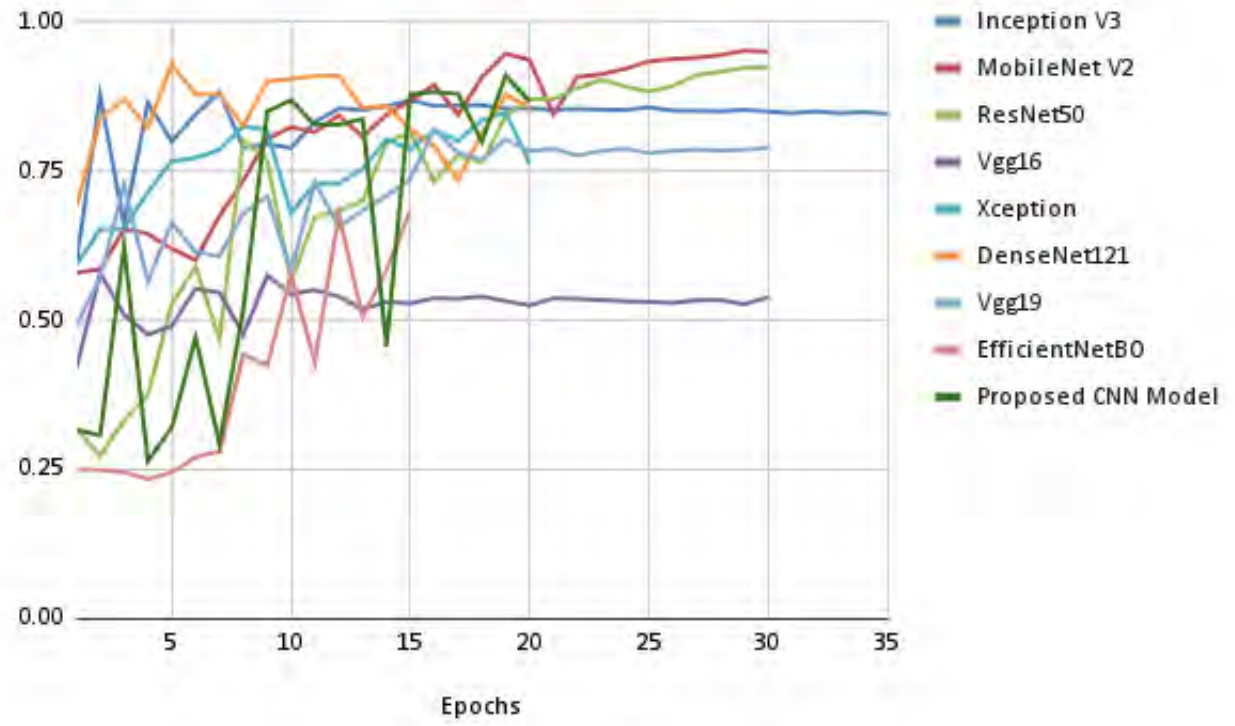

Figure 5.23: Validation Accuracy

With a validation accuracy of 91% and a training accuracy of 97%, our suggested model showed excellent performance and generalization. In contrast, pre-trained models with high training accuracies (99.9% and 86.2%, respectively) and significantly lower validation accuracies (53.8% and 68.3%, respectively) showed considerable overfitting. Other models like ResNet50, DenseNet121, InceptionV3, and ResNeXt also performed well but their training accuracy is very low compared to our proposed model. Moreover, these pre-trained models use more computes and more resources. Our model is more lightweight compared to all of these pretrained models. We developed our own model and trained it from scratch and gained equivalent result or better than those models.

Model	Training Accuracy	Validation Accuracy
Proposed Model	97.05%	91.79%
VGG16	99.9%	53.8%
VGG19	99.4%	78.8%
Resnet50	99%	87.2%
Mobilenetv2	100%	94.9%
DenseNet12	99.9%	85.5%
EfficientNetB0	86.2%	68.3%
InceptionV3	99.9%	84.5%
ResNeXt	99.9%	85.5%
Xception	99.8%	76.2%

Table 5.3: Training and Validation Accuracy Comparison

The proposed model performs better in this model comparison than pre-trained models, with testing accuracy of 93.33%. At 89.41% and 88.74%, respectively, VGG16 and DenseNet12 attain comparable accuracies, but other well-known designs like VGG19, ResNet50 and EfficientNetB0 fall short. Notably, Mobilenetv2 performs noticeably worse in testing and 296 errors were found in the confusion matrix. In the meanwhile, the accuracies of InceptionV3, ResNeXt, and Xception range from 75.66% to 83.16%. This shows that the suggested model performs better at discriminating than a number of popular architectures in a range of tasks, suggesting that it might be used more extensively in real-world scenarios.

Model	Testing Accuracy
Proposed Model	93.33%
VGG16	89.41%
VGG19	79.50%
Resnet50	76.41%
Mobilenetv2	75.33%
DenseNet12	88.74%
EfficientNetB0	75.66%
InceptionV3	83.16%
ResNeXt	88.74%
Xception	75.83%

Table 5.4: Testing Accuracy of Various Models

Chapter 6

Conclusion

6.1 Conclusion

Machine learning is currently at the cutting edge of technology, having an impact on a variety of industries like healthcare, finance, natural language processing, and picture recognition, among others. With ongoing research and invention, it is also continuing to evolve quickly. This work has succeeded in creating algorithms that can diagnose chronic kidney diseases with exceptional accuracy by utilizing machine learning, deep learning, and advanced image processing designed specifically for CKD. The use of convolutional neural networks has made it possible to distinguish precisely between healthy and unhealthy kidney states as well as to categorize CKD stages to inform targeted therapies. The findings provided here have a great deal of potential for helping healthcare professionals by providing fast and precise insights regarding CKD status. Future study and improvement of this methodology are crucial in order to create the way for even more accurate CKD detection and classification methods, ultimately leading to better patient care and results for the management of kidney disease.

6.2 Future Work

For further improvements of our model, we will consider to increase dataset size with real data collected from hospitals. In future we will adjust CNN architecture parameters more to minimize the error and maximize the accuracy of our model. We extend our model to work with humans to detect kidney disease. Our approach will be implemented in an online application. User-friendliness is a top concern for an application so that patients may obtain initial CT scan report data. In order to aid with scanning through the complexity of diverse CT scan images, we will concentrate on image processing in order to acquire additional information from the built-in high resolution camera. It will thus provide real-time analysis to the user and increase demand for our products. We will also invest more time to improve the model's design, data quality, and computing resource use.

Bibliography

[1] J.-W. Hsieh, C.-H. Lee, Y.-C. Chen, W.-S. Lee, H.-F. Chiang, Stage classification in chronic kidney disease by ultrasound image, in: Proceedings of the 29th international conference on image and vision computing New Zealand, 2014, pp. 271–276.

[2] C. Sabanayagam, D. Xu, D. S. Ting, S. Nusinovici, R. Banu, H. Hamzah, C. Lim, Y.-C. Tham, C. Y. Cheung, E. S. Tai, et al., A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations, The Lancet Digital Health 2 (6) (2020) e295–e302.

[3] A. C. Webster, E. V. Nagler, R. L. Morton, P. Masson, Chronic kidney disease, The lancet 389 (10075) (2017) 1238–1252.

[4] J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun, S. Zhu, Z. Ye, Comparison and development of machine learning tools in the prediction of chronic kidney disease progression, Journal of translational medicine 17 (1) (2019) 1–13.

[5] Z. Chen, Z. Zhang, R. Zhu, Y. Xiang, P. B. Harrington, Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers, Chemometrics and Intelligent Laboratory Systems 153 (2016) 140–145.

[6] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin, A. Soylu, Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from ct-radiography, Scientific Reports 12 (1) (2022) 11440.

[7] N. A. Almansour, H. F. Syed, N. R. Khayat, R. K. Altheeb, R. E. Juri, J. Alhiyafi, S. Alrashed, S. O. Olatunji, Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study, Computers in biology and medicine 109 (2019) 101–111.

[8] S. Yin, Q. Peng, H. Li, Z. Zhang, X. You, K. Fischer, S. L. Furth, G. E. Tasian, Y. Fan, Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks, Medical image analysis 60 (2020) 101602.

[9] C. A. Herzog, R. W. Asinger, A. K. Berger, D. M. Charytan, J. Díez, R. G. Hart, K.-U. Eckardt, B. L. Kasiske, P. A. McCullough, R. S. Passman, et al., Cardiovascular disease in chronic kidney disease. a clinical update from kidney disease: Improving global outcomes (kdigo), Kidney international 80 (6) (2011) 572–586.

[10] A. F. Rosmani, U. H. Mazlan, A. F. Ibrahim, D. S. Zakaria, I-ks: Composition of chronic kidney disease (ckd) online informational self-care tool, in: 2015 International Conference on Computer, Communications, and Control Technology (I4CT), IEEE, 2015, pp. 379–383.

[11] A. S. Levey, K.-U. Eckardt, Y. Tsukamoto, A. Levin, J. Coresh, J. Rossert, D. D. Zeeuw, T. H. Hostetter, N. Lameire, G. Eknoyan, Definition and classification of chronic kidney disease: a position statement from kidney disease: Improving global outcomes (kdigo), Kidney international 67 (6) (2005) 2089–2100.

[12] F. Ma, T. Sun, L. Liu, H. Jing, Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network, Future Generation Computer Systems 111 (2020) 17–26.

[13] S. Panizo, L. Martínez-Arias, C. Alonso-Montes, P. Cannata, B. Martín-Carro, J. L. Fernández-Martín, M. Naves-Díaz, N. Carrillo-López, J. B. Cannata-Andía, Fibrosis in chronic kidney disease: pathogenesis and consequences, International Journal of Molecular Sciences 22 (1) (2021) 408.

[14] S. Akter, A. Habib, M. A. Islam, M. S. Hossen, W. A. Fahim, P. R. Sarkar, M. Ahmed, Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease, IEEE Access 9 (2021) 165184–165206.

[15] T. L. Nickolas, E. M. Stein, E. Dworakowski, K. K. Nishiyama, M. Komandah-Kosseh, C. A. Zhang, D. J. McMahon, X. S. Liu, S. Boutroy, S. Cremers, et al., Rapid cortical bone loss in patients with chronic kidney disease, Journal of Bone and Mineral Research 28 (8) (2013) 1811–1820.

[16] L. Nie, L. Zhang, Y. Yang, M. Wang, R. Hong, T.-S. Chua, Beyond doctors: Future health prediction from multimedia and multimodal observations, in: Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 591–600.

[17] P. Kumar, G. Bhattacharyya, S. Dattatreya, H. Malhotra, Tackling the cancer tsunami (2009).