Comparative Analysis of MobileNetV2, EfficientNetB0, and U-Net Models for Kidney Disease Classification from CT Scans

Submitted By

Kongkiat Patwary

Batch: 35(A)

ID: 211-35-720

Department of Software Engineering.

Daffodil International University

Supervised By

Md. Hafizul Imran

Assistant Professor,

Department of Software Engineering,

Daffodil International University

Thesis submitted in fulfillment of the requirements for the award of the degree of Bachelor of Science

Fall – 2024

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Kongkiat Patwary
Date of Birth	:	07/07/2000
Title	:	Comparative Analysis of MobileNetV2, EfficientNetB0, and U-Net Models for Kidney Disease Classification from CT Scans
Academic Session	:	Fall 2024

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☒	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the *thesis for the purpose of research only.*
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

(Student's Signature)

Kongkiat Patwary
Student Name
Student ID: 211-35-720
Date: 15-09-2025

(Supervisor's Signature)

Md Hafizul Imran
Name of Supervisor
Date: 15-09-2025

APPROVAL

APPROVAL

This thesis titled on "**Comparative Analysis of MobileNetV2, EfficientNetB0, and U-Net Models for Kidney Disease Classification from CT Scans**", submitted by **Kongkiat Patwary (ID: 211-35-720)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- **Chairman**

Dr. S M Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 1**

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Khalid Been Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Dr. Md. Sazzadur Rahman
Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

I, Kongkiat Patwary, declared that this thesis was conducted by me under the supervision of Md. Hafizul Imran (Assistant Professor), Department of Software Engineering at Faculty of Science and Information Technology, Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Kongkiat Patwary

Kongkiat Patwary
ID: 211-35-720
Department of Software Engineering Daffodil International University

Certified by

Md. Hafizul Imran
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

Comparative Analysis of MobileNetV2, EfficientNetB0, and U-Net Models for Kidney Disease Classification from CT Scans

Kongkiat Patwary

Thesis submitted in fulfillment of the requirements for the award of the degree of Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

SEPTEMBER 2025

ACKNOWLEDGEMENT

First and foremost, I am grateful to Almighty Allah, who has granted me the strength, wisdom, and perseverance to complete this research. Throughout my academic journey, I have been fortunate to receive the unconditional love, support, and encouragement of my parents. Their belief in me has been a constant source of motivation and inspiration.

I would like to extend my heartfelt gratitude to my supervisor, Assistant Professor Md. Hafizul Imran, for his invaluable advice, support, and guidance throughout the research. His knowledge and insight have significantly influenced this work. I also deeply appreciate the departmental head, Dr. Imran Mahmud, for his support, guidance, and constructive feedback, which played an essential role in the successful completion of my project.

Lastly, I would like to thank all my friends, colleagues, and everyone who has supported and encouraged me during this process. Your help has been greatly appreciated.

List of Abbreviations

AI – Artificial Intelligence
CNN – Convolutional Neural Network
CT – Computed Tomography
CKD – Chronic Kidney Disease
DL – Deep Learning
ESRD – End-Stage Renal Disease
F1-score – F1 Measure (Harmonic Mean of Precision and Recall)
GPU – Graphics Processing Unit
KNN – K-Nearest Neighbors
LR – Logistic Regression
ML – Machine Learning
MRI – Magnetic Resonance Imaging
ReLU – Rectified Linear Unit
SVM – Support Vector Machine
UNet – U-Net Neural Network Architecture
ViT – Vision Transformer
WHO – World Health Organization

ABSTRACT

Kidney disease includes affected cysts, tumors, and stones which is a significant worldwide health problem, where delayed or inaccurate diagnosis can negatively impact results. Computed tomography (CT) affords high anatomical context and is standardly utilized to assess renal parenchyma and identify abnormal areas. This study explores deep learning based automated classification of kidney abnormalities in CT. A publicly accessible multi-class renal CT dataset was developed and preprocessed through resizing, intensity normalization, and augmentation. Three convolutional architectures which are EfficientNetB0, MobileNetV2, and U-Net adapted to image-level prediction those were trained and compared in repeated cross validation using accuracy, precision, recall, and F1-score as primary metrics. EfficientNetB0 delivered the most consistent and highest accuracy performance in evaluations, with some evidence of clinical adequacy. MobileNetV2 achieved competitive, stable performance at lower computational expense, with a preference for deployment in resource-constrained settings. The U-Net baseline, although computationally efficient, demonstrated higher variability for direct classification. In total, the findings confirm the potential of lightweight CNNs for CT-based kidney disease diagnosis and the importance of multi-test evaluation for stability. Future directions include extension to larger, more heterogeneous cohorts, exploration of transformer-based multimodal fusion with clinical text, incorporation of explainable AI for transparent decision-making, and validation in actual radiology workflows.

TABLE OF CONTENT

APPROVAL ________ v

DECLARATION ________ vi

ACKNOWLEDGEMENT ________ vii

List of Abbreviations ________ viii

ABSTRACT ________ ix

CHAPTER 1 ________ 1

INTRODUCTION ________ 1

1.1 Understanding the problem ________ 1

1.2 Motivation of this work ________ 1

1.3 Problem Statement ________ 2

1.4 Research Objectives ________ 2

Chapter 2 ________ 4

Literature Review ________ 4

2.1 Machine Learning (ML) ________ 4

2.1.1 Machine Learning Models ________ 4

2.2 Related Works ________ 5

Chapter3 ________ 9

Methodology ________ 9

3.1 Dataset Description ________ 9

3.2 Data Preprocessing ________ 11

3.4 Training Details ________ 12

3.5 Models ________ 13

3.6 Model Architecture View ________ 17

3.7 Evaluation Metrics ________ 18

Chapter 4 ________ 19

Result and Discussion ________ 19

4.1 Introduction ________ 19

4.2 Data splitting details ________ 19

4.3 Result ________ 20

4.4 Result Comparison : ________ 26

4.4.1 Result analysis: ________ 27

4.5 Comparison of Model ________ 28

4.5.1 Introduction: ________ 28

4.5.2 Advantages and Disadvantages: ________ 28

4.5.3 Comparison Graph: ________ 29

4.5.4 Comparison of Model Performance Metrics: ____________________ 30

Chapter 5 ____________________ 38

Findings & Future Scope ____________________ 38

5.1 Findings ____________________ 38

5.2 Future Work and Scope ____________________ 38

Chapter 6 ____________________ 40

Conclusion ____________________ 40

References ____________________ 42

List of Figure

Figure 1: Ratio of image data distribution among classes 9

Figure 2: Image data distribution among classes 10

Figure 3: Normal Kidney of different groups. 10

Figure 4: Kidney Cyst for various groups. 11

Figure 5: Kidney Tumor of different groups. 11

Figure 6: Kidney Stone of different groups. 11

Figure 7: Overview of process 17

Figure 8: Output of MobileNetV2 20

Figure 9: Confusion matrix of MobileNetV2 21

Figure 10: Output of EfficientNetB0 22

Figure 11: Confusion matrix of EfficientNetB0 23

Figure 12: Output of U-Net V2 24

Figure 13: Confusion matrix of U-Net 25

Figure 14: Comparison graph of accuracy and loss 29

Figure 15:MOBILENETV2 (TIME VS ACCURACY) 30

Figure 16:MOBILENETV2 (TIME VS PRECISION) 31

Figure 17:MOBILENETV2 (TIME VS RECALL) 31

Figure 18:MOBILENETV2 (TIME VS F1-SCORE) 32

Figure 19: EFFICIENTNETB0 (TIME VS ACCURACY) 33

Figure 20:EFFICIENTNETB0 (TIME VS PRECISION) 33

Figure 21: EFFICIENTNETB0 (TIME VS RECALL) 34

Figure 22: EFFICIENTNETB0 (TIME VS F1-SCORE) 34

Figure 23:U-NET (TIME VS ACCURACY) 35

Figure 24:U-NET (TIME VS PRECISION) 36

Figure 25:U-NET (TIME VS RECALL) 36

Figure 26:U-NET (TIME VS F1-SCORE) 37

LIST OF TABLE

Table 1: MobileNetV2 Algorithm 14

Table 2: EfficientNetB0 Algorithm 15

Table 3: U-Net Classification Algorithm 16

Table 4: Data splitting details 19

Table 5: Model Results 26

CHAPTER 1
INTRODUCTION

1.1 Understanding the problem

Kidney diseases, including cysts, tumors, and stones, are among the most common life-threatening diseases worldwide [1] [2]. They affect millions of people every year and they are increasing public health threat [3] [4]. Worldwide, there are over 10 % patients of Chronic Kidney Disease (CKD) [3]. Statistic showed that in 2040, CKD may ranked as the sixth most common cause of death [5]. According to the Bangladesh Kidney Foundation the prevalence of CKD in Bangladesh is about 18–20 million people which is around 10–12% of the total population [6]. If the diseases are not detected in their initial stages, they will evolve to end-stage renal disease (ESRD), which requires costly and life-altering interventions such as dialysis or kidney transplantation [7]. CKD causes a substantial burden of morbidity, mortality, and economic cost worldwide, with current estimates that more than 850 million people have some form of kidney disease [8]. The effects are even more severe in low- and middle-income nations as diagnosis facilities are incompetent, people are unaware and struggle to get to specialized health care [9]. The World Health Organization (WHO) has prioritized kidney disease as a burgeoning non-communicable health threat, which has served as a push for inexpensive screening and scale-friendly diagnostic options [2]. Although Computed Tomography (CT) is a critical tool in the detection of kidney impairments attributable to its high-resolution cross-sectional imaging capabilities manual reporting of CT scans remains time-consuming and strongly dependent on radiologist expertise [10]. In settings where there is limited radiologist availability or excessive workload, risk of diagnostic delay or error caused by fatigue, bias, or inter-observer variation is increased [11]. These types of problems offer strong rationale for computerized diagnostic techniques, and particularly those that utilize deep learning and CNNs, to increase efficiency, consistency, and accuracy of medical imaging interpretation [12].

1.2 Motivation of this work

This work was motivated by curiosity about where Artificial Intelligence and healthcare meet. I was curious about the idea of automating diagnoses through deep learning. This thesis was inspired by browsing open-source medical imaging datasets and reading about recent progress in AI for healthcare. The availability of the CT Kidney Dataset on Kaggle, with thousands of CT images of the kidney, each labelled as a normal, cyst, tumor and stone respectively, gave us solid ground for a project that could potentially serve as a prototype for real-world problems. The question was whether deep learning algorithms could accurately classify kidney problems and help radiologists make faster, more definitive diagnoses.

We elected to work on CT images rather than MRI or ultrasound images since CT is used most frequently as the medical imaging modality for kidney diagnosis in both emergency and routine situations. Because of better imaging of stones and soft tissues, shorter scan time and wider availability and affordability of CT than MRI, in most of the hospitals in Bangladesh, we have chosen to do CT. MRI, useful as a soft tissue imaging, proves more costly and can't be used for penetration screenings of kidneys on large scale. Ultrasound, despite being in extensive use, is limited by operator dependency and lower resolution, and is thus not convenient for deep learning based classification work.

1.3 Problem Statement

Kidney diseases including cysts, tumors and stones are there, need an early and precise diagnosis to avoid life-threatening complications. CT is the most capable modality but manual reading of this imaging is slow, has high errors, and is very experienced radiologist-dependent. Current deep learning methods tend to suffer from instability, low efficiency or are not able to easily deploy to real-world scene. In that sense, this study compares MobileNetV2 and EfficientNetB0 and U-Net, in the search for a stable, accurate, and computationally efficient model for the automated classification of kidney disease in CT.

1.4 Research Objectives

In this research, we chose three deep learning models: MobileNetV2, EfficientNetB0 and U-Net. We selected these models as a compromise between classification performance, time and architectural suitability for medical imaging. The MobileNetV2 is a lightweight architecture for mobile and edge devices which fits for the application in the rural hospital or portable equipment without real-time inference calculation. The EfficientNetB0 was chosen because of its novel compound scaling method that can achieve high performance with less number of parameters. U-Net, a segmentation model, was added to check if the encoder-decoder network could perform well on classification-tasks if properly adjusted. Larger models such as ResNet or DenseNet were not selected in this work to keep the focus on models more suitable for preference to faster training, edge deployment scenarios, or overall efficiency.

In this work, we make a comparison between MobileNetV2, EfficientNetB0, and U-Net on the CT Kidney Dataset. The performances of the models are measured with standard classification metrics accuracy, precision, recall, and F1-score to compare which models have the best capability to discover the kidney disorders. By comparing with these approaches, we hope to find a model that is both accurate and practical for deployment in resource-poor clinical contexts. This work adds to the burgeoning area of AI-enabled diagnostic tools, and represents a step towards the use of deep learning in the medical imaging pipeline for the timely diagnosis of kidney disease.

Chapter 2
Literature Review

Nowadays, machine learning has emerged as a strong methodology for detecting and classifying Chronic Kidney Disease (CKD). With the help of big medical imaging data (e.g., CT and MRI images), the state of the art algorithms can provide more accurate and close to the accurate diagnosis results at very early stage. These models help discriminate between stages of CKD and facilitate the development of personalized treatments. Recently, works such as the deep learning based heterogeneous modified artificial neural network introduced by Ma et al. [13], have shown that the diagnostic performance of neural networks and, especially, convolutional architectures, keeps increasing as they are trained with more and more data. These advances show that the applications of machine learning have the potential to transform CKD diagnostic processes, save on shrinking healthcare resources and improve patient care.

2.1 Machine Learning (ML)

Machine learning is a subset of artificial intelligence and refers to the development of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions. Its history can be traced back to the middle of the 20th century and is rooted in simple algorithms and rule-based methods. Its development was, however, pushed faster with the advent of digital computing and ready access to large data. In past few decades, substantial advances in both deep learning and high performance computing, have catapulted machine learning into an increasingly forefront of technology. Nowadays, it is driving transformation in a wide range of domains including medicine, finance, natural language processing, and image analysis. In medicine especially in chronic diseases such as CKD, machine learning is gaining interest for early diagnosis, risk stratification, and personalized treatment [14].

2.1.1 Machine Learning Models

Xiao et al. [15] studied the performance of four data mining models, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR) and Decision Tree classifiers for Chronic Kidney Disease (CKD) prediction by using clinical dataset. They observed that SVM classifier outperformed the performance in the aspect of precision and sensitivity. Beyond the predictive modeling aspect, recent works have focused on the processing of medical imaging for the automatic detection of kidney-related abnormalities, such as cysts, stones, and tumors. For instance, Islam et al. [16] proposed a vision transformer based explainable transfer learning method to diagnose CT radiography which demonstrated effectiveness in solving the diagnostic problems. In addition, the Kidney Disease: Improving Global Outcomes (KDIGO) project stresses the importance of uniform definitions and staging systems of CKD that could guide the integration of state-of-the art machine learning methodologies in the clinical setting [17].

2.2 Related Works

This research article [18] considered the problem of missing data on Chronic Kidney Disease (CKD) dataset with 400 samples (250 samples as CKD patients and 150 as non-CKD). Incomplete entries were managed by K-Nearest Neighbors (KNN) imputation to produce a completed dataset for further analysis. After imputation, the six algorithms, including Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbor, Naive Bayes, and Feedforward Neural Networks, were used to construct the predictive models of CKD diagnosis. Of these, the Random forest model had the highest statistical performance with a diagnostic accuracy of 99.75%. Such a finding demonstrates that ensemble estimators are able to handle missing values in clinical datasets in a reliable way and give highly reliable predictions that can be useful to help in the early screening and management of CKD.

The study by Chen et al. [19] intended to create digital platforms to help CKD patients with self-care on a daily basis. To this end, an interactive multimedia-based solution was developed with Adobe Flash CS5. 5, hosted by a website designed with Adobe Dreamweaver in order to offer patients a reference for CKD management. One of the focus points of this project was to develop an interface, that would serve as a support for unambiguous and coherent communication of self-care. Association digital technology and tangible health information made a valid reliable information channel for CKD patients, thus contributing for better patient empowerment and disease control.

Panizo et al. [20] elucidated the complex molecular and physiological mechanism responsible for pathological fibrosis in Chronic Kidney Disease (CKD). The paper emphasizes the implication of various agents such as pain, the renin-angiotensin system, parathyroid hormone, fibroblast growth factor 23, Klotho, microRNAs, and the vitamin D endocrine system, and is furthermore focused on the relevant signaling pathways participating in renal fibrogenesis. Fibrosis, on the other hand, has deleterious effects on renal and cardiovascular functions in CKD patients as well. By explaining the mechanisms, the study offers insights into candidate targets for therapy and emphasizes the value of understanding the hosts perspective of molecular pathophysiology in the management of CKD.

CKD has a long history of development and the diagnostic criteria and classification of CKD have been regularly updated. Several factors contribute to the occurrence and progression of CKD such as age, sex, diet, geographic location, and history of disease. Currently, the worldwide standard for identification of CKD is the measure of glomerular filtration rate (GFR). A GFR lower than 60 mL/min/1.73 m^2 is evidence of decreased kidney function, and one lower than 15 mL/min/1.73 m^2 is an indication of renal failure or end-stage kidney disease. These indicators are important performance standards for clinical evaluation, risk stratification and early intervention management in patient care with CKD [21].

A detailed investigation by Akter et al. [22] evaluated nine predictive models for determination of CKD progression and its severity, based on non-urinary clinical and demographic data. Of these models, linear techniques- and logistic regression in particular-had the highest predictive performance. Levels of albumin, serum creatinine, Trig, LDL, and estimated glomerular filtration rate in blood tests were considered as indicators for monitoring development of the illness. In addition to statistical model, they developed a new ultrasonic imaging method for proteinuria diagnosis at different CKD stages using speckle effect of ultrasonic pictures. This methodology is based on the Nakagami distribution and Local Binary Patterns in order to efficiently encode tissue scattering feature memory, where age distribution is a crucial factor in the analysis process. The high sensitivity and specificity of the system underscored its potential use as a non-invasive, dependable tool for early CKD detection and risk prediction. Together, these observations emphasize the increasing role of sophisticated computational modeling and experimental imaging approaches in early diagnosis, patient monitoring and personalized treatment planning in CKD therapy.

Nickolas et al. [23] used AI and deep learning algorithms to determine CKD from retinal images of people in a community setting. The resulting system, using retinal fundus images as input was tested on different datasets to guarantee the fairness and generalization. Of the analyzed models, a hybrid deep learning method achieved the best accuracy for CKD detection, indicating the possible role of retinal imaging as a non-invasive and more reliable screening method. This methodology highlights a developing possibility of using AI-based image analysis for early disease detection, providing an adjunct to traditional clinical assessments, and augmenting the ability for early intervention pertaining to community based populations.

In a recent publication a clinical update based on Kidney Disease: Improving Global Outcomes (KDIGO) guidelines [24] focusing on the linkage between Cardiovascular Disease (CVD) and Chronic Kidney Disease (CKD) has taken place. This article provides an extended summary of current understanding of cardiovascular complications in CKD patients with focus on major risk factors, preventative measures and the proposed therapeutical approaches. Through incorporation of this new information the paper emphasizes the necessity of addressing CVD in CKD patient populations in order to mitigate morbidity and improve patient results [25].

Hsieh et al. [26] elaborated on the detection and diagnosis of CKD, using deep learning based heterogeneous improved artificial neural network. The study suggested a novelty method which combined the complex deep learning algorithms with more advanced neural network model architecture to enhance the accuracy and robustness in CKD diagnose and stage criterion. With the use of imaging data, the method highlights the potential for AI-driven models to offer accurate diagnostic assistance, supporting early clinical action and more efficient disease control.

The use of deep learning for medical image analysis has seen tremendous growth over the last decade as first experimental results matured into successful applications. Esteva et al. (2017) were among the first to propose the use of convolutional neural networks (CNNs) for skin cancer diagnosis, reporting dermatologist-equivalent performance when using their model trained on more than 129,000 dermoscopic images. This study was one of the earliest to demonstrate that deep learning model performances were on par with clinical professionals in image-based diagnostic tasks. Yet, they focused on external lesions of the skin but did not consider internal diagnostic imaging such as CT or MRI [27].

In another study Rajpurkar et al. (2017) proposed a deep learning model, CheXNet, which is based on the DenseNet-121, for pneumonia detection in chest X-ray images. Their network yielded a substantial improvement over practicing radiologists in several diagnostic tasks, and validated the capability of deep CNNs learn to understand and interpret 2D radiological images. Nevertheless, the study provided with a positive performance only explored the field of 2D chest X-ray images without considering the more sophisticated 3D CT image data, which is, for example, widely used for kidney disease diagnosis [28].

And Dosovitskiy et al. (2021) proposed Vision Transformers (ViT) where convolutions are replaced by self-attentions to capture long range dependencies in image data. Though ViTs have demonstrated state-of-the-art performance in large-scale classification datasets, their use on medical imaging is still restricted by data-hungry need and long training duration. Their high complexity renders them less applicable for real-time clinical systems without massive annotated databases [29].

More recently, Yuan et al. (2023) presented a model from a cross-breed of Swin Transformers and U-Net for kidney CT image segmentation. By introducing global attention and local feature preservation, this model obtained superior boundary accuracy for both tumors and cysts. While the hybrid architecture achieved superior performance in segmentation, the inference time and computation overhead were problematic for a real-time classification than in-resource environment [30].

Although these studies indicate the increasing efficacy of deep learning in the diagnosis of kidney diseases, there are still several limitations. Still, light-weighted crowd counting models capable of high classification accuracy need to be of interest for deploying in portable and/or embedded devices. There is a scarcity of studies directly comparing several of these model architectures especially both classification and segmentation networks, such as U-Net, MobileNet V2 and EfficientNetB0 that are trained and tested on the same dataset. There is not much work on application of U-Net for classification despite its popularity in the field of segmentation. Attention-based models, like Vision Transformers, are less investigated in kidney CT image analysis because of the complexity of training. Most prior work also ignores practical issues such as class imbalance, model interpretability, and compatibility with multi-modal imaging data. Last and probably most important, there is a notable absence of clinical validation since most papers avoid testing the models, on a real hospital environment, with actual patient cases, using only open datasets.

Chapter3

Methodology

3.1 Dataset Description

This experiment applies the CT Kidney Dataset (Nazmul0087, Kaggle) [31] containing 12,446 CT scans data in four class whose labels are Normal, Cyst, Tumor, and Stone. They are all grayscale and have different resolution, so we resize every image to 150x150 pixels for better compatibility with a model.

The data is arranged in folders by class labels which makes it easy to load and preprocess the data. The pictures display various kidney normal and abnormal structures, which proves to be a good prior for training the classifiers. Nevertheless, the data source is without metadata such as demographic or sex and personal data, and no segmentation mask, restricting its application in personalized or localization related applications.

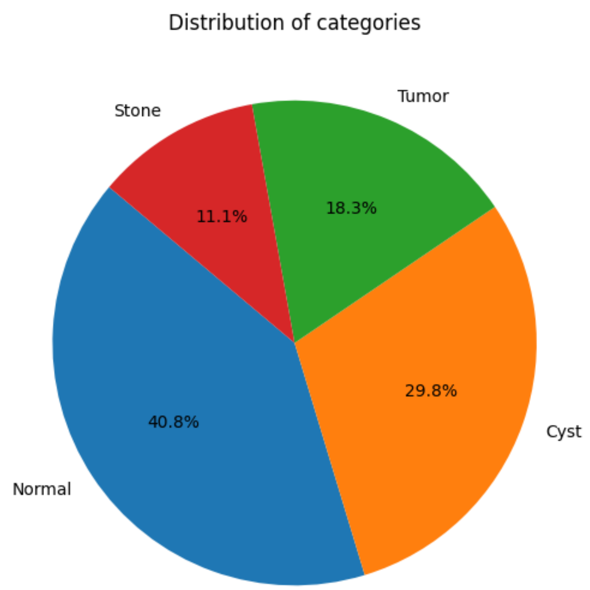

Figure 1: Ratio of image data distribution among classes

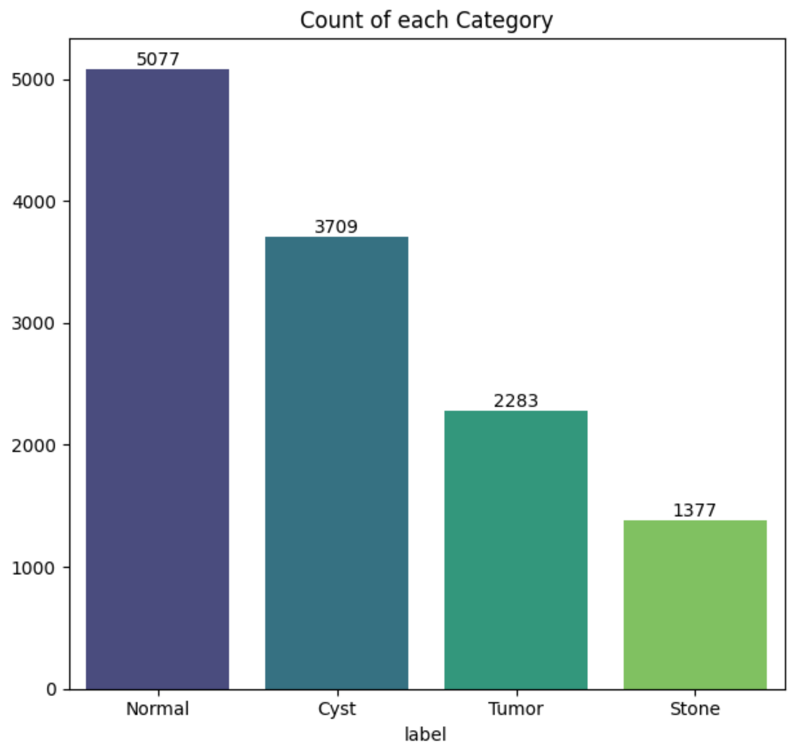

Figure 2: Image data distribution among classes

A summary of the dataset used for image classification is shown in the table, along with the number of images that belong to the four different classes of Normal, Cyst, Tumor, and Stone. There are four categories in all in the dataset, and each one has a different quantity of folders and images:

Normal: There are 5077 images in 47 groups in the dataset that represent healthy kidney.

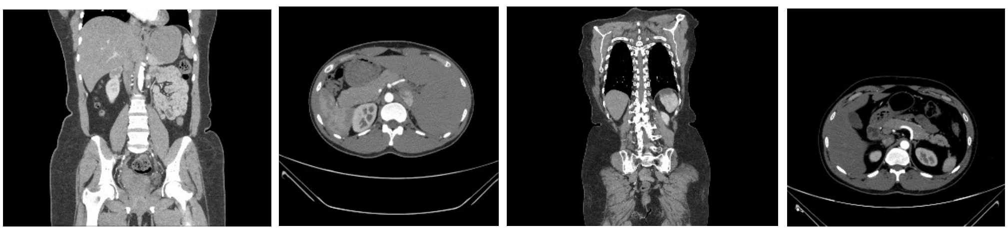

Figure 3: Normal Kidney of different groups.

Cyst: This category includes 3709 photos and 80 groups which record abnormalities related to cysts.

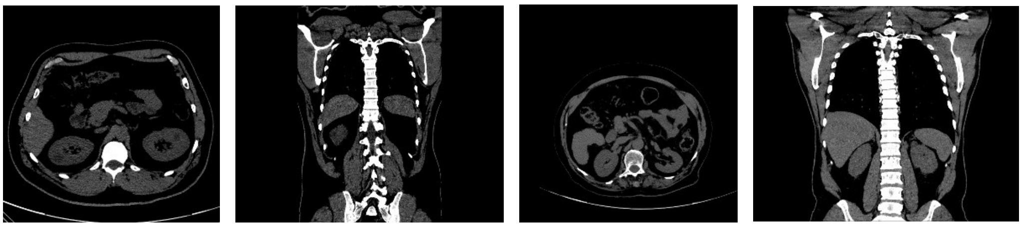

Figure 4: Kidney Cyst for various groups.

Tumor: The study's focus is tumor cases, represented by the 24 groups containing 2283 images.

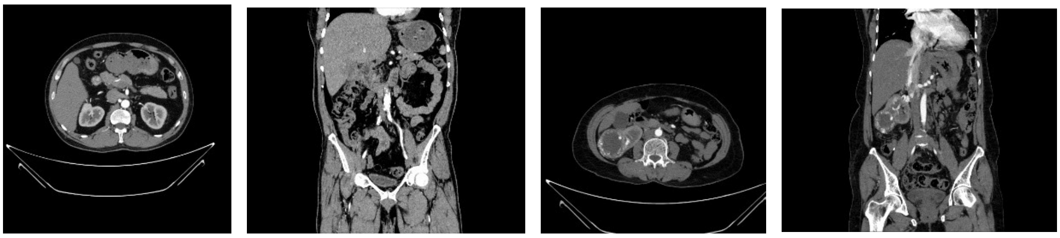

Figure 5: Kidney Tumor of different groups.

Stone: This class contains 1377 images in 62 groups related to stone formations in the medical field.

Figure 6: Kidney Stone of different groups.

However, despite all the shortcomings of the dataset, including size, class symmetry and image quality, it is considered a convincing dataset for testing the performance of deep learning classifiers for kidney abnormal class prediction.

3.2 Data Preprocessing

For maximum performance and to be compatible with the deep learning models, several preprocessing steps were performed on the raw CT scan images. These steps aimed at normalizing the input data, accelerating the training process, and optimizing the generalization capacity of the model to unknown samples.

Image Resizing: Finally, all images were rescaled to 150×150 pixels through simple interpolation, so as the dimensions would be the same for the selected model configurations (MobileNetV2, EfficientNetB0 and UNet).

Channel Standardization: CUDA_BGR2RGB (rgb). Note that images are automatically converted to 3-channel format to fit the input requirements of pre-trained models that assume RGB inputs.

Normalization: Each image pixel value was normalized between 0 and 1 by dividing all pixels by 255 to improve training stability and convergence.

Data Augmentation: The data set was augmented as follows to gain diversity and to minimize overfitting:

- Random horizontal and vertical flips
- Rotations with fixed angle-deviation limits
- Zooming in/out
- Width and height shifts

Train-Test Split: The dataset was split into training and testing sets at a ratio of 80:20. Besides, 20% of the training data was set aside as validation during training.

Caching and Prefetching: The data pipeline from TensorFlow was leveraged for batch caching, shuffling, and prefetching to speed up the training and make it more efficient.

These pre-processing techniques helped in making dataset ready for better training model and will enhance consistency and variability. Consequently, the models were able to learn better meaningful patterns while avoiding overfitting and computational complexity.

3.4 Training Details

After data preprocessing and model selection, all models were trained under the same settings for fair comparison among all architectures. All models were trained based on a supervised real-image learning model (except for transferred base layers from MobileNetV2 and EfficientNetB0). Training was done with TensorFlow and Keras on Kaggle notebook with GPU support.

The concrete training settings with specific options are:

Environment Used: All models were developed in Kaggle's cloud platform with the use of a Tesla P100 GPU (16 GB VRAM) and RAM (13 GB), allowing Swift training and prediction of the models.

Input Image Size: We resized all input images to 150×150 with 3 channel (RGB) to fit the input image size of the classifiers.

Batch Size: All experiments used a batch size of 32 to compromise between consuming memory and training time.

Number of Epochs: All models were trained to 40 epochs, with the use of early stopping applied according to the validation loss to avoid overfitting.

Loss Function: The Categorical Cross-Entropy Loss was applied, which is applicable to multi-class classification tasks.

Optimizer: We chose the Adam optimizer for all models with a learning rate of 0.001, which had adaptive parameter updates and reasonably fast convergence.

Learning Rate Scheduler: A learning rate decay callback (ReduceLROnPlateau) was used to decrease the learning rate if validation performance did not improve.

Callbacks Used:

- EarlyStopping: to stop the training once the validation loss started increasing.
- ModelCheckpoint: to save the best model w.r.t validation accuracy.
- ReducLROnPlateau: for learning rate annealing on plateaus.

Validation Split: 20% of the training set was used as a validation set to observe the performance during training.

Data Augmentation: Image augmentation performed at training time using TensorFlow's ImageDataGenerator.

As a result, using the same training parameters, models could be evaluated under the same conditions and performance metrics could be more easily compared (i.e., accuracy, precision, recall and F1-score).

3.5 Models

Three deep learning models were utilized for the classification of kidney abnormalities from the CT Kidney Dataset. **MobileNetV2**, **EfficientNetB0**, and **U-Net** were the models selected for this comparison.

MobileNetV2: MobileNetV2 was chosen because of its lightweight architecture, using depthwise separable convolutions to reduce computational complexity without compromising high accuracy. The base model was pre-trained with ImageNet weights and fine-tuned on the CT Kidney Dataset.

Algorithm 1 MobileNetV2-based Classification Model

1. Input: Training dataset train_ds, testing dataset test_ds
2. Output: Trained MobileNetV2 classification model
3. LOAD PRE-TRAINED MODEL:
4. *base_model ← MobileNetV 2(weights = 'imagenet', include_top = False, input_shape = (150, 150, 3))*
5. FREEZE BASE MODEL LAYERS:
6. base_model.trainable ← F alse
7. ADD CUSTOM CLASSIFICATION HEAD:
8. *x ← base_model.output*
9. *x ← GlobalAverageP ooling2D(x)*
10. *x ← Dense(128, activation = 'relu')(x)*
11. *x ← Dropout(0.5)(x)*
12. *outputs ← Dense(4, activation = 'softmax')(x)*
13. CREATE FINAL MODEL:
14. *model ← Model(inputs = base_model.input, outputs = outputs)*
15. COMPILE MODEL:
16. *Optimizer ← Adam(lr = 0.0001)*
17. *Loss ← Sparse Categorical Crossentropy*
18. *Metrics ← Accuracy*
19. TRAIN MODEL:
20. *history_mobilenet ← model.f it(train_ds, epochs = 40, validation_data = test_ds)*
21. EVALUATE AND VISUALIZE:
22. Evaluate trained model on test dataset
23. Plot accuracy, loss curves, and confusion matrix

Table 1: MobileNetV2 Algorithm

EfficientNetB0: EfficientNetB0 was the second model picked, which was a highly efficient and performing model. Like MobileNetV2, EfficientNetB0 also had pre-trained weights for ImageNet loaded along with fine-tuning on data. The same custom layers were added to EfficientNetB0 as were added to MobileNetV2.

Algorithm 2 EfficientNetB0 -based Classification Model

1. Input: Training dataset train_ds, testing dataset test_ds
2. Output: Trained EfficientNetB0 classification model
3. LOAD PRE-TRAINED MODEL:
4. *base_model_eff icientnet ← Eff icientNetB0(weights = 'imagenet', include_top =*
    a. *False, input_shape = (150, 150, 3))*
5. FREEZE BASE MODEL LAYERS:
6. *base_model_efficientnet.trainable ← False*
7. ADD CUSTOM CLASSIFICATION HEAD:
8. *x ← base_model_efficientnet.output*
9. *x ← GlobalAverageP ooling2D(x)*
10. *x ← Dense(128, activation = 'relu')(x)*
11. *x ← Dropout(0.5)(x)*
12. *outputs ← Dense(4, activation = 'softmax')(x)*
13. CREATE FINAL MODEL:
14. *efficientnet_model ← Model(inputs = base_model_eff icientnet.input, outputs = outputs)*
15. COMPILE MODEL:
16. *Optimizer ← Adam(lr = 0.0001)*
17. *Loss ← Sparse Categorical Crossentropy*
18. *Metrics ← Accuracy*
19. TRAIN MODEL:
20. history_eff icientnet ← efficientnet_model.f it(train_ds, epochs = 40, validation_data = test_ds)
21. EVALUATE AND VISUALIZE:
22. Evaluate trained model on test dataset
23. Plot accuracy, loss curves, and confusion matrix

Table 2: EfficientNetB0 Algorithm

U-Net: Image segmentation specialist U-Net was employed to segment and classify abnormalities in the kidneys in CT scans. U-Net architecture consists of encoder-decoder with skip connections and pixellevel classification (segmentation) capability. U-Net was optimized for kidney abnormalities classification, and segmentation map was used for highlighting regions of interest in images.

Algorithm 2 U-NET -based Classification Model

1. **Define Model Architecture (Build_UNet function):**
2. Input Layer: *$(150,150,3)$*
3. Encoder Block 1: *Conv2D(32, 3×3, ReLU, "same") → BatchNormalization → MaxPooling2D(2×2)*
4. Encoder Block 2: *Conv2D(64, 3×3, ReLU, "same") → BatchNormalization → MaxPooling2D(2×2)*
5. Bridge: *Conv2D(128, 3×3, ReLU, "same") → BatchNormalization*
6. Classification Head:
   a) GlobalAveragePooling2D
   b) Dropout(0.5)
   c) Dense(128, ReLU)
   d) Dense(4, Softmax)
7. Return Compiled U-Net Model $\mathcal{M}$ with:
   • *Optimizer = Adam(lr=0.0001)*
   • *Loss = Sparse Categorical Crossentropy*
   • *Metric = Accuracy*
8. **Create Model:** *$unet_model \leftarrow Build_UNet()$*
9. **Compile Model:** as defined above
10. **Train Model:**
    $history_unet \leftarrow unet_model.fit(train_ds,; epochs=40,; validation_data=test_ds)$
11. **Evaluate and Visualize:**
12. Evaluate trained U-Net model on test dataset
13. Plot training/validation accuracy and loss curves
14. Generate confusion matrix

Table 3: U-Net Classification Algorithm

3.6 Model Architecture View

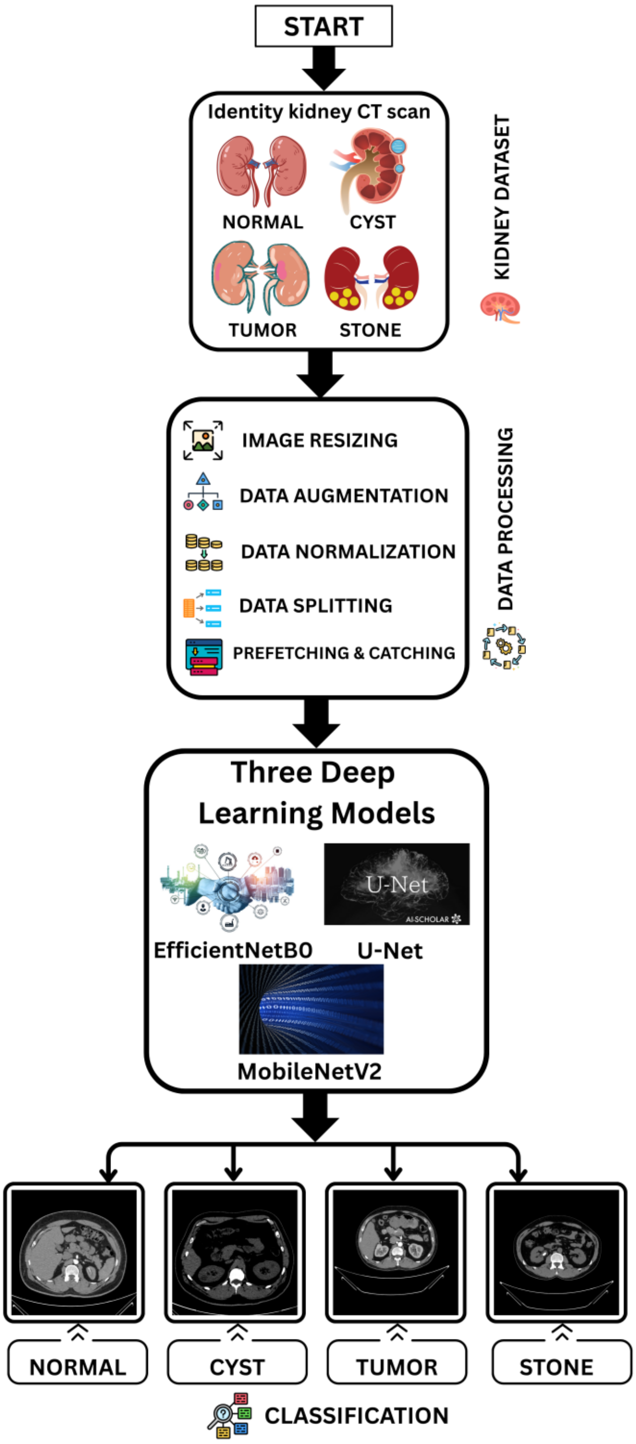

Figure 7: Overview of process

3.7 Evaluation Metrics

To verify the performance of each model, the following metrics were calculated for assessment:

Accuracy: Accuracy is the count of correct predictions among all predictions. This metric provides a general idea of how well the model is performing for all classes.

$$\boldsymbol{Accuracy = \frac{TP+TN}{TP+TN+FP+FN}}$$

Precision: Precision is utilized to find the number of true positive predictions as a ratio of total positive predictions predicted by the model. It is important when false positives are expensive.

$$\boldsymbol{Precision = \frac{TP}{TP+FP}}$$

Recall: Recall determines the proportion of the actual positive predictions out of all the positive instances. Recall is effective where the cost of false negatives is significant.

$$\boldsymbol{Recall = \frac{TP}{TP+FN}}$$

F1-Score: F1-score is the harmonic mean of precision and recall. It is very useful when class distribution is not balanced.

$$\boldsymbol{F1\ Score = \frac{2 \text{ x } Precision \text{ x } Recall}{Precision + Recall}}$$

These evaluation metrics were used in the analysis and comparison of different deep learning models for the detection and classification of kidney abnormalities from CT scan images.

Chapter 4

Result and Discussion

4.1 Introduction

In this section, we present and discuss the experimental results of three considered models (MobileNetV2, EfficientNetB0, U-Net). Each model was trained and tested on the CT Kidney dataset with identical preprocessing and training configurations. For robustness and consistency, the model performance of each model is reported based on 10 independent tests by showing the lowest, highest, and average performance. The performance metrics were accuracy, precision, recall, F1-score, and training time. Besides the average performance, this approach offers also insight into stability and variability over iterative experiments.

4.2 Data splitting details

Data Split	Percentage	Number of sample
Training data	80%	9957
Test data	20%	2489

Table 4: Data splitting details

4.3 Result

Output of MobileNetV2:

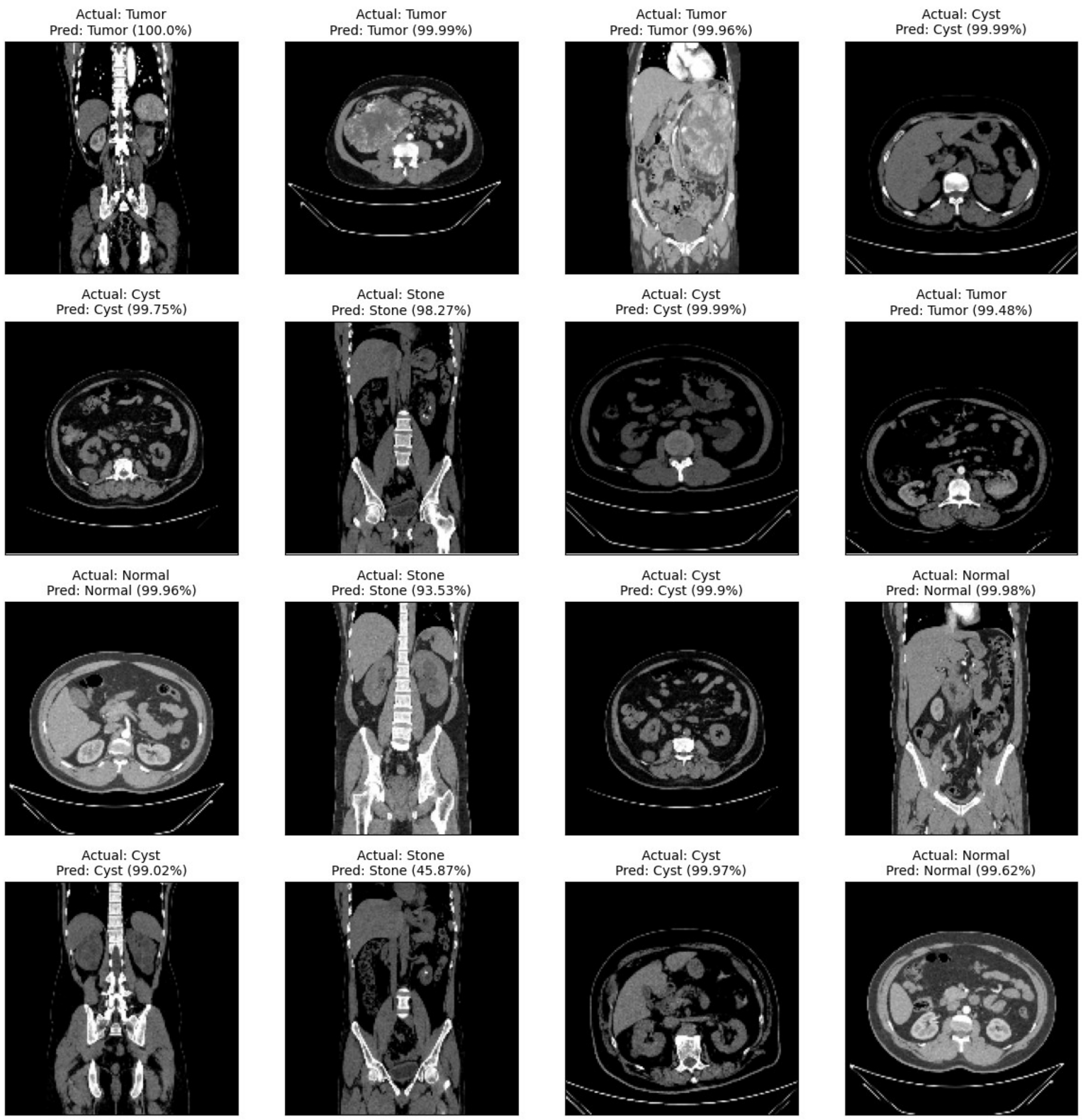

Figure 8: Output of MobileNetV2

This demonstrates an array of CT scan images recorded under 4 structures: Tumor, Cyst, Stone and Normal. For each image there are actual and predicted diagnoses. The model shows high precision, with most predictions being over 99%, suggesting that the performance of the model in the classification of medical images is good.

Confusion Matrix of MobileNetV2:

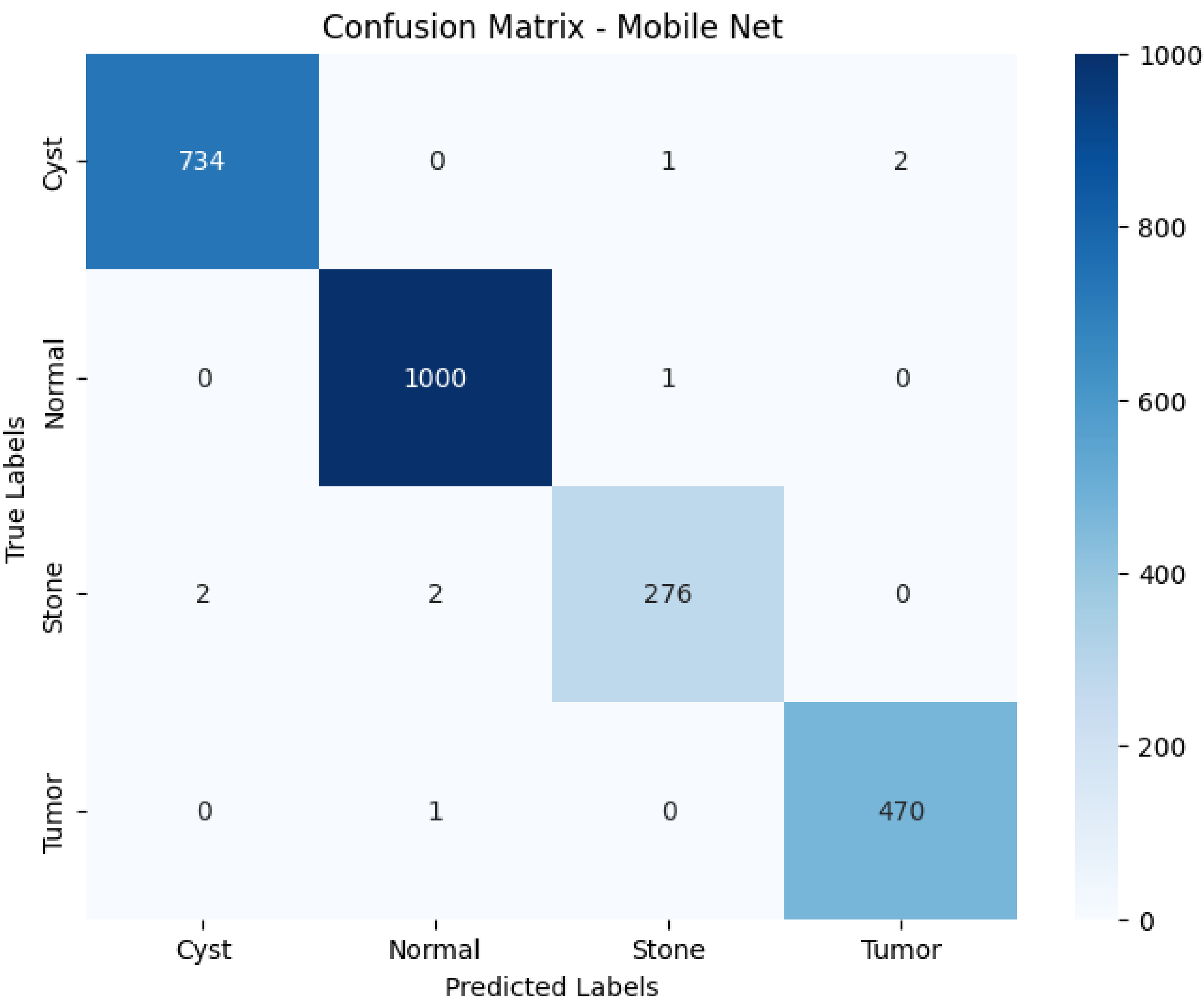

Figure 9: Confusion matrix of MobileNetV2

This confusion matrix shows how well our MobileNetV2 model is doing across four classifications: Cyst, Normal, Stone, and Tumor. The true labels are shown on the rows of the matrix, while the predicted labels are shown on the columns.

Output of EfficientNetB0:

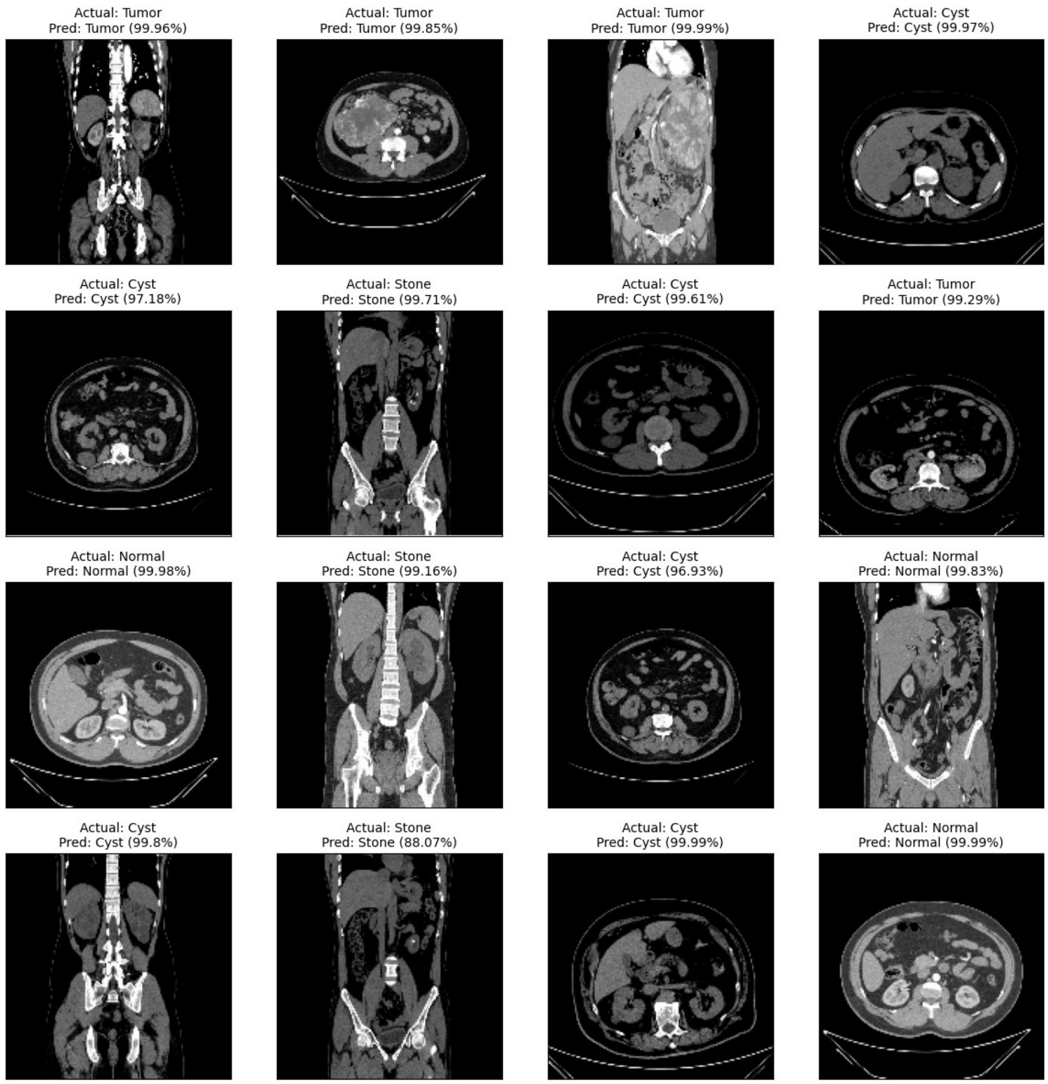

Figure 10: Output of EfficientNetB0

The CT scan images in this figure are the real diagnosis and model prediction, respectively, and the accuracy of each classification is shown in percentage format. The model demonstrates high degree of accuracy, and most predictions achieve rates of accuracy higher than 99%, demonstrating the ability of the model to distinguish between various medical cases.

Confusion matrix of EfficientNetB0:

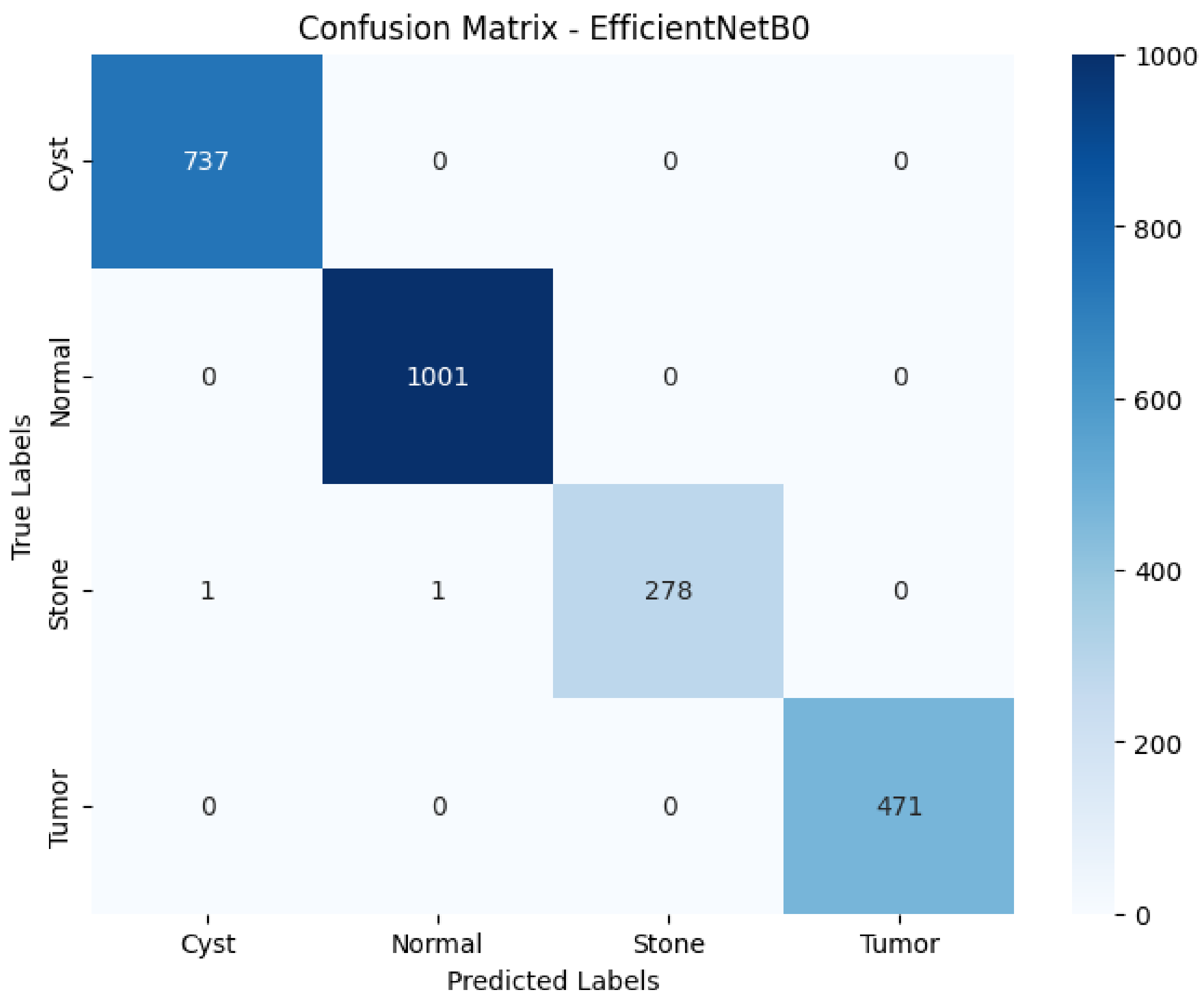

Figure 11: Confusion matrix of EfficientNetB0

This is a confusion matrix for EfficientNetB0 model. It shows you how good the model is at Cyst, Normal, Stone or Tumor. The matrix appears to have high accuracy, Slight misclassifications between Cyst and Normal have been identified, but the model performs consistently well across all types.

Output of U-Net:

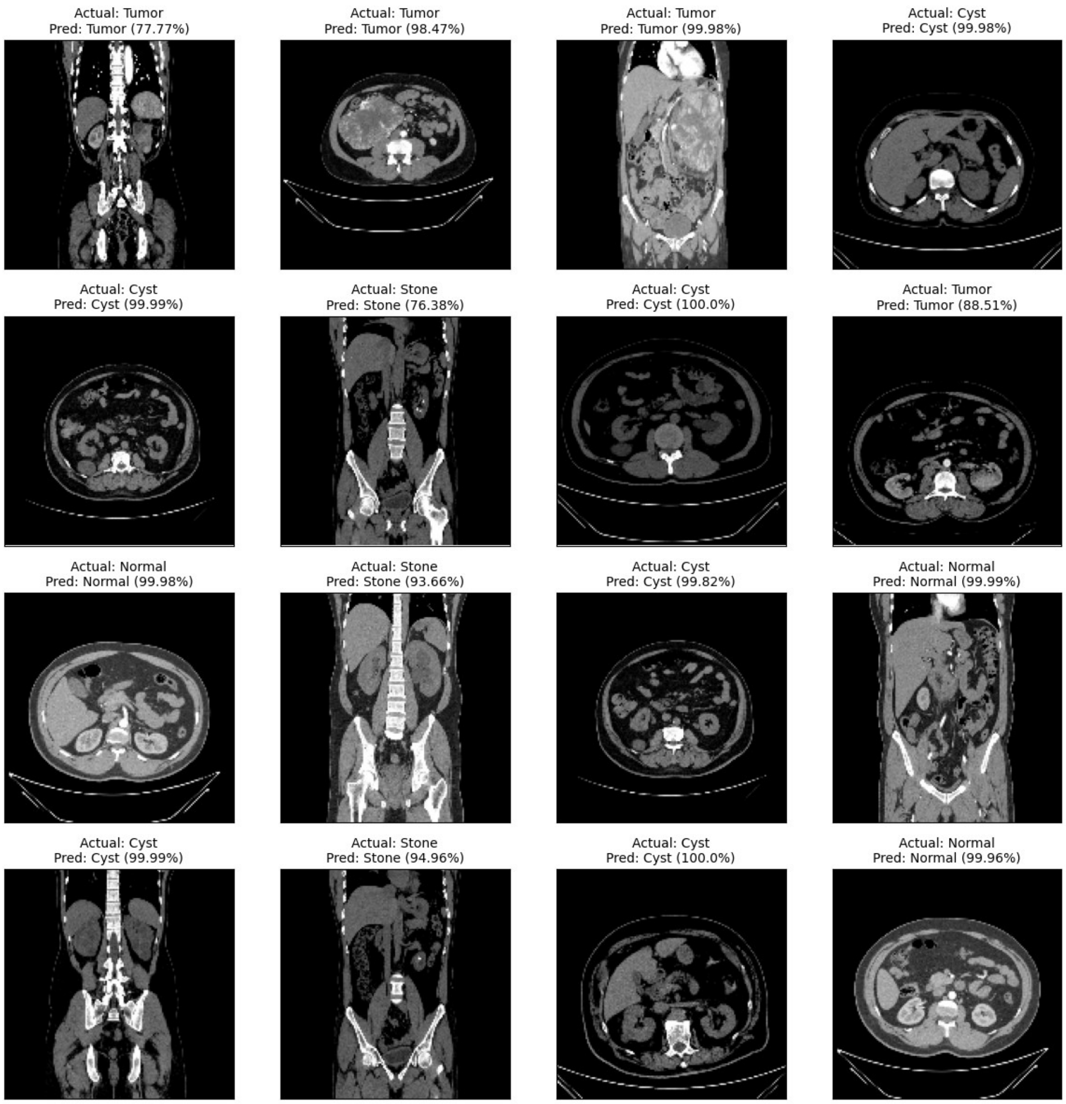

Figure 12: Output of U-Net V2

This figure contains a serial CT scan image set with true diagnoses and model predictions with accuracy also in percentages for each class. The model predictions, although reasonably accurate, have some degree of variation, with most Cyst and Normal cases with high percentage correctness, but miss-classified a few Tumor and Stone cases. However, such occasional misclassifications do not significantly affect overall performance, since the average value of the majority of predictions is above 90% correctness.

Confusion matrix of U-Net:

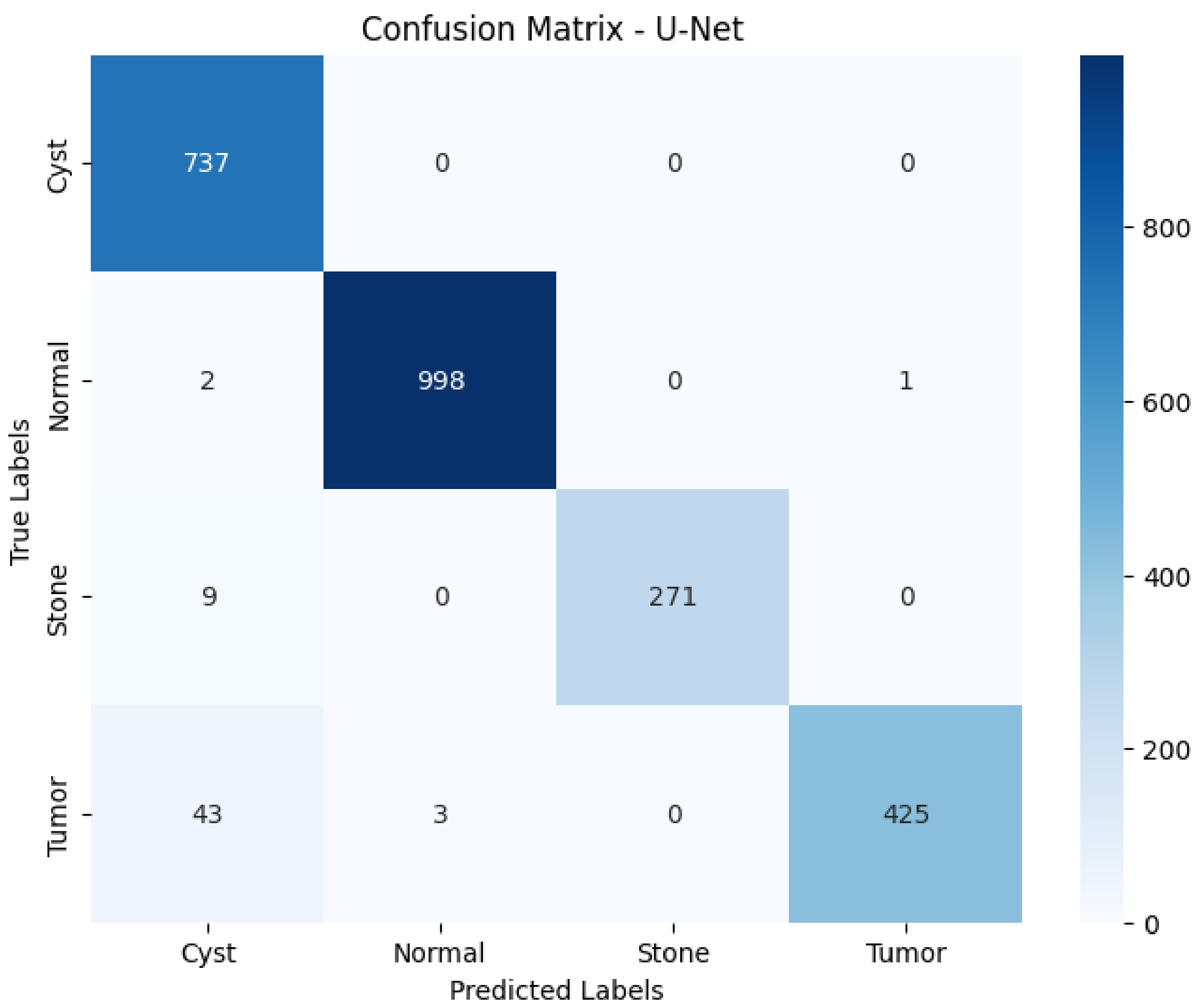

Figure 13: Confusion matrix of U-Net

This image is the confusion matrix of U-Net models, with the real identifies (Cyst, Normal, Stone, Tumor) and predicted identities. The diagonal represent the correct predictions. However, misclassifications are observed in case of Cyst and Normal especially, yet the model performs well overall with high accuracy.

4.4 Result Comparison **:**

Model Name	**Test No**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Training Time (sec)**
MobileNetV2	1	98.48%	99.48%	99.05%	99.26%	200.95
MobileNetV2	2	99.40%	99.36%	98.87%	99.11%	229.32
MobileNetV2	3	99.60%	99.58%	99.23%	99.40%	231.83
MobileNetV2	4	99.44%	99.51%	99.91%	99.20%	211.18
MobileNetV2	5	99.64%	99.61%	99.36%	99.48%	199.6
MobileNetV2	6	99.52%	99.48%	99.14%	99.31%	218.66
MobileNetV2	7	99.60%	99.63%	99.23%	99.42%	193.42
MobileNetV2	8	99.44%	99.38%	98.93%	99.14%	193.36
MobileNetV2	9	99.60%	99.58%	99.26%	99.42%	193.09
MobileNetV2	10	99.44%	99.47%	98.96%	99.21%	191.02
MobileNetV2 Average		99.42%	99.51%	99.19%	99.30%	206.243
EfficientNetB0	1	100%	100%	100%	100%	292.89
EfficientNetB0	2	99.96%	99.97%	99.91%	99.94%	302.15
EfficientNetB0	3	100%	100%	100%	100%	282.79
EfficientNetB0	4	100%	100%	100%	100%	288.23
EfficientNetB0	5	99.92%	99.88%	99.88%	99.88%	284.68
EfficientNetB0	6	100%	100%	100%	100%	279.36
EfficientNetB0	7	99.84%	99.85%	99.71%	99.78%	278.22
EfficientNetB0	8	99.92%	99.93%	99.82%	99.88%	277.76
EfficientNetB0	9	99.92%	99.94%	99.82%	99.88%	277.08
EfficientNetB0	10	99.92%	99.88%	99.88%	99.88%	278.04
EfficientNetB0 Average		99.95%	99.95%	99.90%	99.92%	284.12
UNet	1	97.87%	98.41%	96.18%	97.21%	195.18
UNet	2	99.52%	99.37%	99.36%	99.36%	201.54
UNet	3	99.48%	99.65%	99.94%	99.28%	194.02
UNet	4	99.52%	99.53%	99.34%	99.43%	195.47
UNet	5	98.35%	98.43%	97.61%	97.98%	192.15
UNet	6	98.79%	98.50%	98.90%	98.67%	192.05
UNet	7	94.90%	93.39%	95.59%	94.16%	189.02
UNet	8	98.63%	97.69%	98.63%	98.14%	188.83
UNet	9	99.80%	99.67%	99.73%	99.70%	190.85
UNet	10	96.91%	95.72%	96.99%	96.28%	190.96
UNet Average		98.38%	98.04%	98.23%	98.02%	193.01

Table 5: Model Results

4.4.1 Result analysis:

For stability of evaluation, every model was trained and tested 10 times consistently. The performance results measure over the five metrics (accuracy, precision, recall, F1-score, and training time) were logged for each test, and the minimum, maximum, and average values were computed to have a better sense of the measuring the variability of the performance of each model.

MobileNetV2: Lowest across test values are Accuracy 98.48%, Precision 99.36%, Recall 98.87%, F1-score 99.11%, Training Time 199.60 sec. **Highest** across test values are Accuracy 99.64%, Precision 99.61%, Recall 99.91%, F1-score 99.48%, Training Time 231.83 sec **An average** performance over 10 tests are Accuracy 99.42%, Precision 99.51%, Recall 99.19%, F1-score 99.29%, Training Time 206.24s.

These results show that MobileNetV2 is very stable and further fairly consistent with only minor oscillation between different tests. It was found to strike a better balance between performance and training time over EfficientNetB0 and is appropriate in resource constrained settings.

EfficientNetB0: Lowest among tests are Accuracy 99.84%, Precision 99.85%, Recall 99.71%, F1-score 99.78%, Training Time 277.08 sec. **Highest** among all tests are Accuracy 100%, Precision 100%, Recall 100%, F1-Score 100%, Training Time 302.15 sec. **An average** performance over 10 tests are Accuracy 99.95%, Precision 99.95%, Recall 99.90%, F1-score 99.92%, Time to Train 284.12 sec.

Performance of EfficientNetB0 was highest in all metrics for all tests. The difference of the lowest and highest tests were negligible - that is really stable. However, the longest training time for this model was needed, indicating the compromise between computational cost and better performance.

U-Net: Lowest results across tests are Accuracy 94.90%, Precision 93.39%, Recall 95.59%, F1score 94.16%, Training Time 188.83 sec. **Highest** among all tests are Accuracy 99.80%, Precision 99.67%, Recall 99.94%, F1-Score 99.70%, Training Time 201.54 sec. **An average** performance over 10 tests are Accuracy 98.38%, Precision 98.04%, Recall 98.23%, F1-score 98.02%, Training Time 193.01 sec.

The U-Net had the largest variability of the three models. Though it performed a decent highest accuracy at 99.80%, its lowest test levelled at 94.90%. This instability as well as lower means in comparison to CNN-based models, illustrates U-Net is not appropriate for straight classification tasks albeit with training time efficiency.

Discussion: From the relative results, it is clear that we observe the most accurate, most precise, recalling and F1-scoring results on each test of EfficientNetB0. This very capability to perfectly classify is what one would want for diagnostic applications at the cost, however, of long training times. Even with slightly lower accuracy, MobileNetV2 is still competitive. Its higher training speed and better stability over multiple tests will be of great interest for real-time or limited-resource applications, such as edge device deployment in rural clinics. On the contrary, U-Net showed the worst average and largest variation in performance. Although it sometimes achieved results similar to MobileNetV2 in the optimal tests, its lower minimums indicate less reliability. This supports our observation that we should not directly apply U-Net to the classification problem, but to its segmentation and localization.

Lastly, the utilization of averaging across 10 independent test tests adds confidence in these findings. Through the reporting of minimum, maximum, and mean values, the study gives an insight into how stable and how practical it is to use each model.

4.5 Comparison of Model

4.5.1 Introduction:

Based on the outcomes, EfficientNetB0 has better performance than both MobileNetV2 and U-Net in precision, recall, and accuracy. While MobileNetV2 is, however, quite close in terms of high-performance with a relatively lower computation load, U-Net, as very good as it is for segmentation tasks, had comparatively low performance for classification accuracy and recall and hence not as sufficient for this specific task of kidney abnormality classification.

4.5.2 Advantages and Disadvantages:

MobileNetV2 is most suitable to deploy in low-resource environments but ought to be highly efficient nonetheless. EfficientNetB0 is the highest performing model in terms of both accuracy and recall and therefore is most suitable in scenarios where there is more computationally available and model efficiency is of utmost concern. U-Net, being primarily for segmentation, provides a good insight for detection and localization but falls short of the overall classification accuracy level compared to the rest of the models.

4.5.3 Comparison Graph:

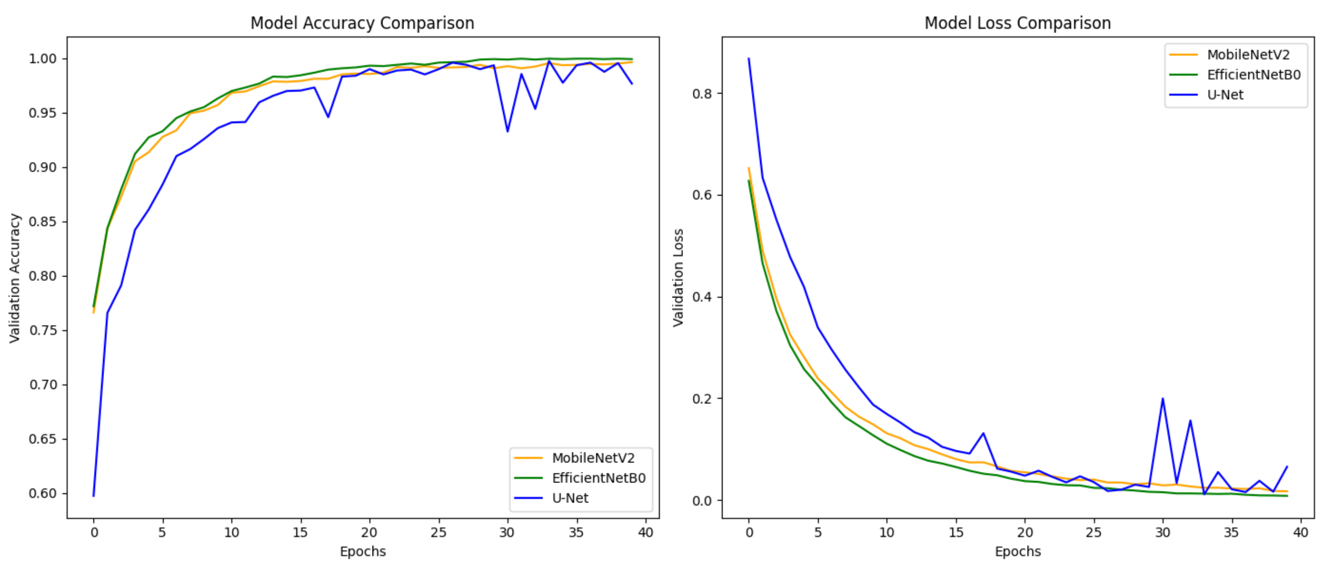

Figure 14: Comparison graph of accuracy and loss

Description of the Graph: The validation accuracy and validation loss of the three models-MobileNetV2, EfficientNetB0, and U-Net-are graphed over 40 epochs. The conclusions can be arrived at from the graph: Validation Accuracy Comparison **(Left Plot):** MobileNetV2 (orange line) and EfficientNetB0 (green line) both indicate the same trend in validation accuracy, with both models each reaching a high level of accuracy. Both MobileNetV2 and EfficientNetB0 show gradual improvement in accuracy across the epochs, with the best at around 0.90 before plateauing. U-Net (blue line) shows gradual but slower improvement in accuracy. The accuracy plateaus at a lower level than for the CNN models, indicating somewhat poorer performance on the task of classification.

Validation Loss Comparison **(Right Plot):** All the models' validation loss declines significantly at an initial stage of training, with MobileNetV2 and EfficientNetB0 exhibiting a similar trend of decline. The U-Net model does experience some oscillations and does not decline as uniformly as the CNN models. A smaller validation loss indicates that the model's predictions are converging to the actual labels. Both MobileNetV2 and EfficientNetB0 display steady low loss in later training, demonstrating superior performance at classification and generalization. MobileNetV2 and EfficientNetB0 are clearly the top performers, both demonstrating high accuracy as well as small loss values, which indicate superior ability to generalize to the test set. U-Net, as much as it is good at segmentation tasks, has a more volatile performance both in loss and accuracy and hence is not quite as good at this classification task as the CNN models. From this graph, there is clear visual insight as to how every model performs with time, MobileNetV2 and EfficientNetB0 both having better performance regarding validation accuracy and loss.

4.5.4 Comparison of Model Performance Metrics:

A details analysis of the Model performance metrics is given below:

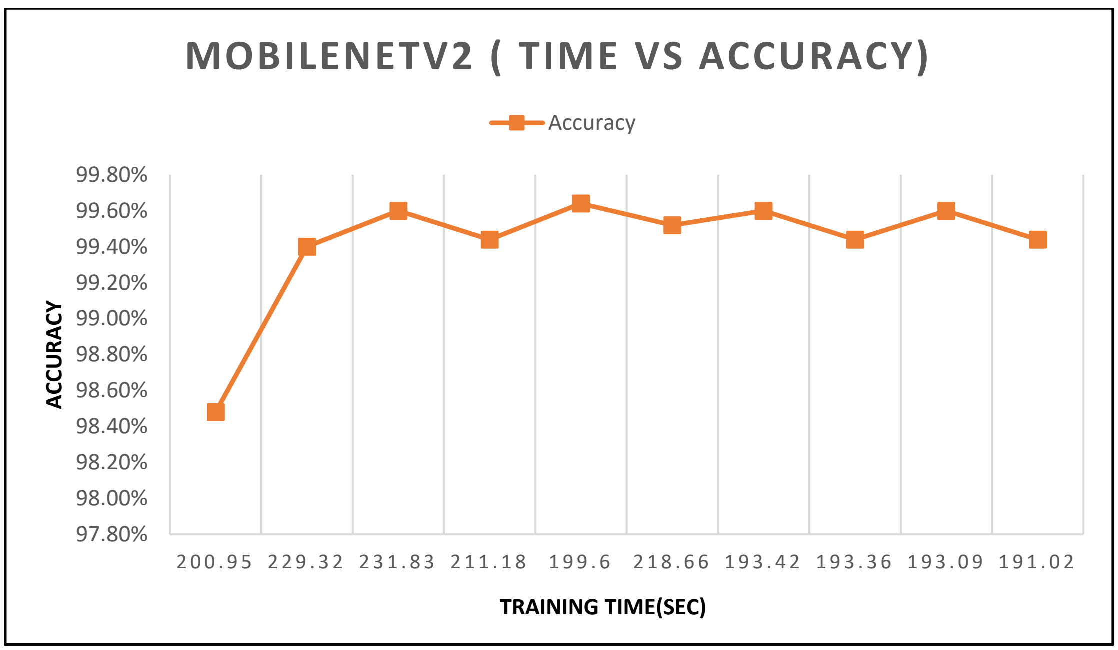

Figure 15:MOBILENETV2 (TIME VS ACCURACY)

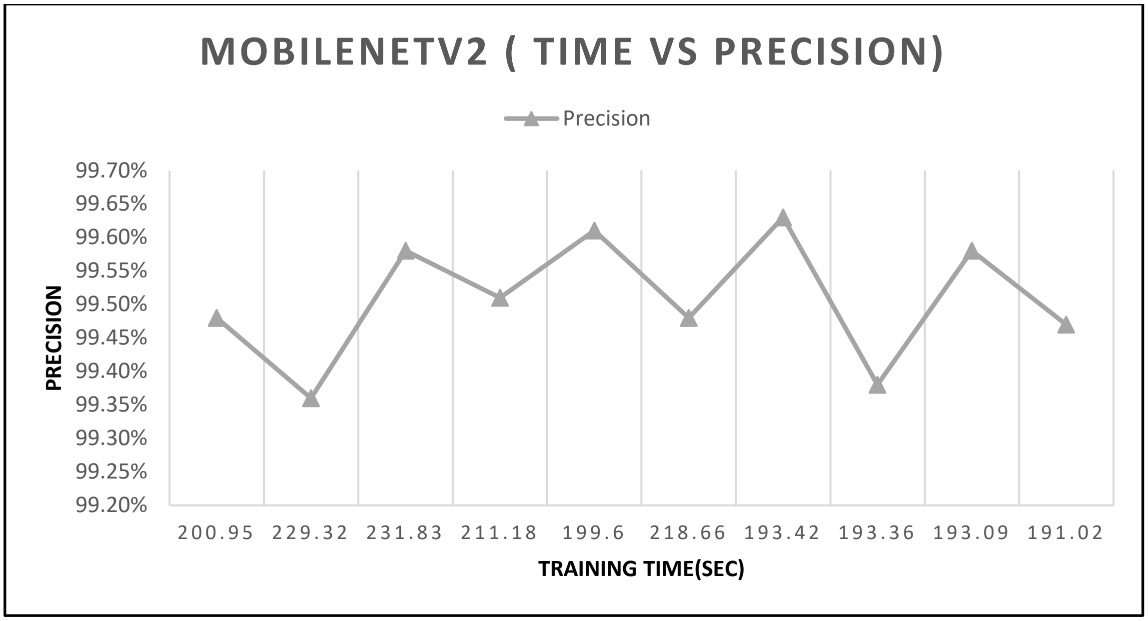

Figure 16: MOBILENETV2 (TIME VS PRECISION)

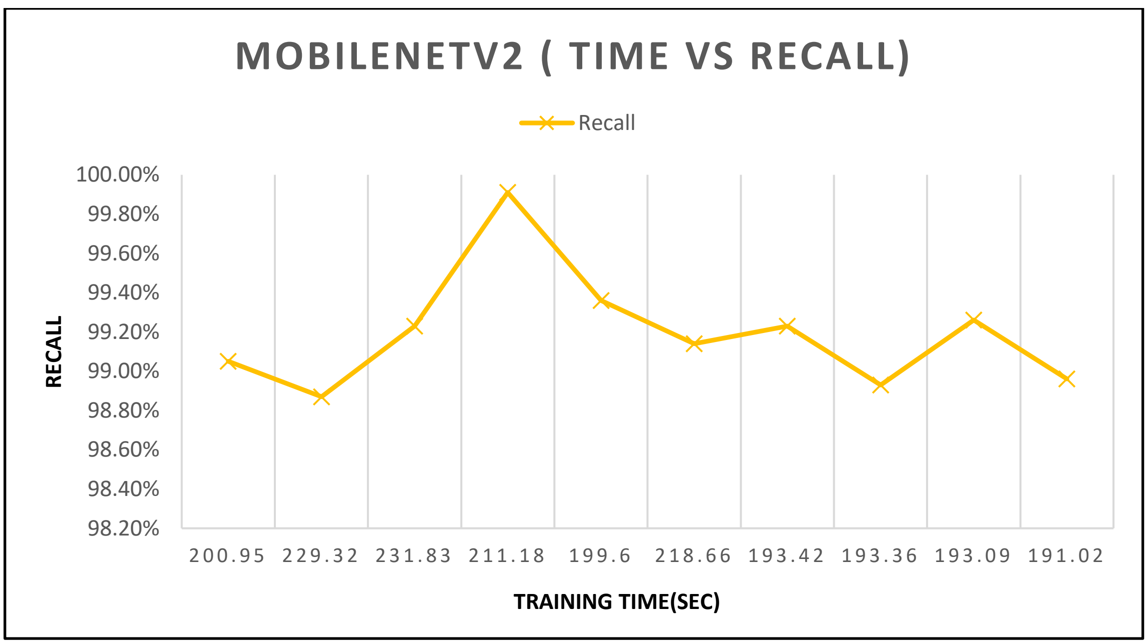

Figure 17:MOBILENETV2 (TIME VS RECALL)

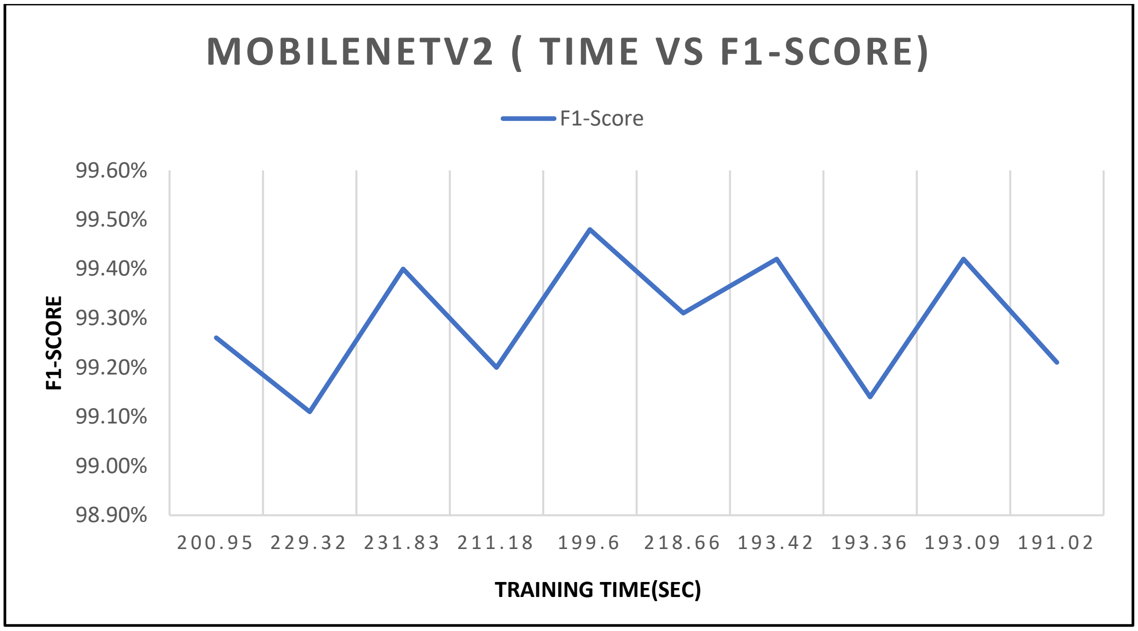

Figure 18:MOBILENETV2 (TIME VS F1-SCORE)

Analysis of Model Performance Metrics of MobileNetV2: In figure 15, the accuracy range around 98% to 99.8%. The graph shows a steady improvement and stabilization. It proves that it's a lightweight yet efficient model. In figure 13b, MobileNetV2 is excellent in avoiding false positives. Figure 17 suggests that MobileNetV2 is dependable in capturing true positives. Moreover, figure 16 and 17 confirms that this model balances true precision and recall well.

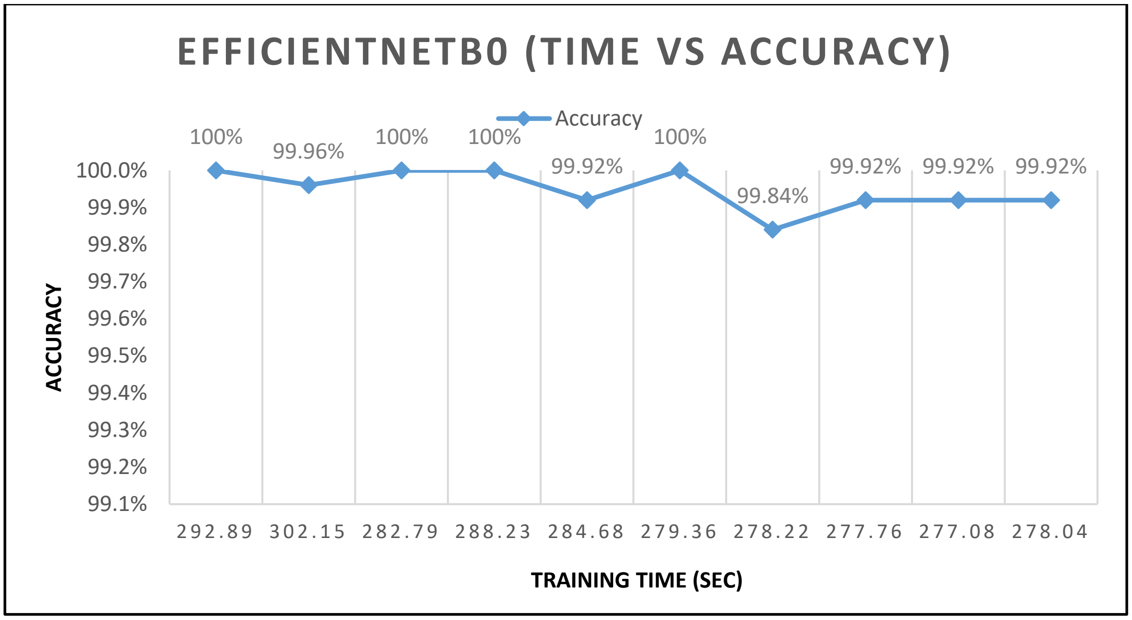

Figure 19: EFFICIENTNETB0 (TIME VS ACCURACY)

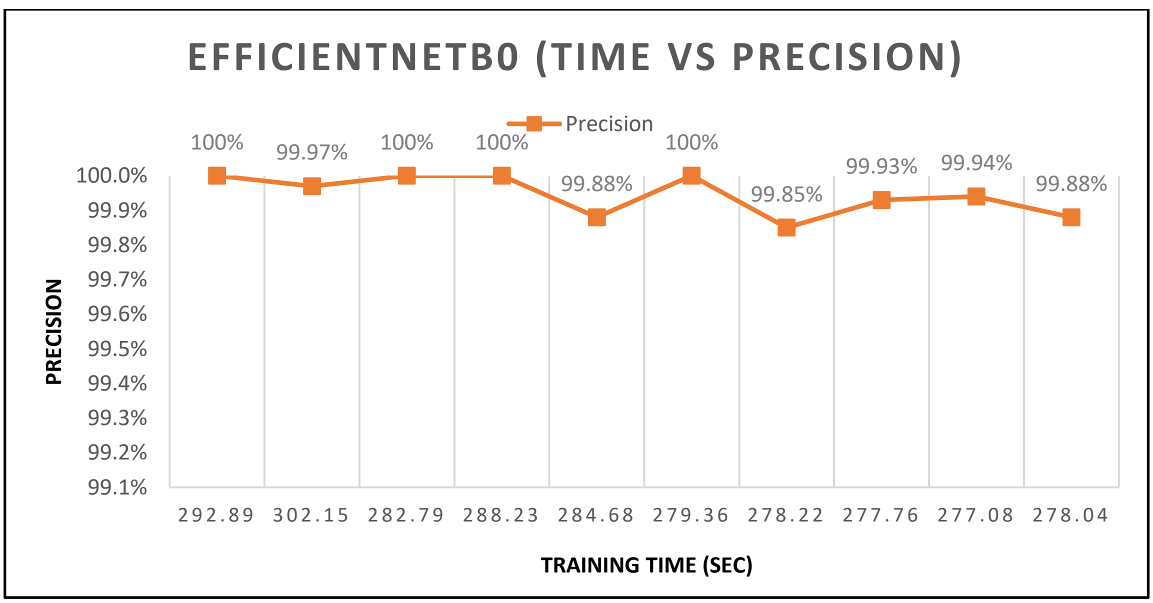

Figure 20:EFFICIENTNETB0 (TIME VS PRECISION)

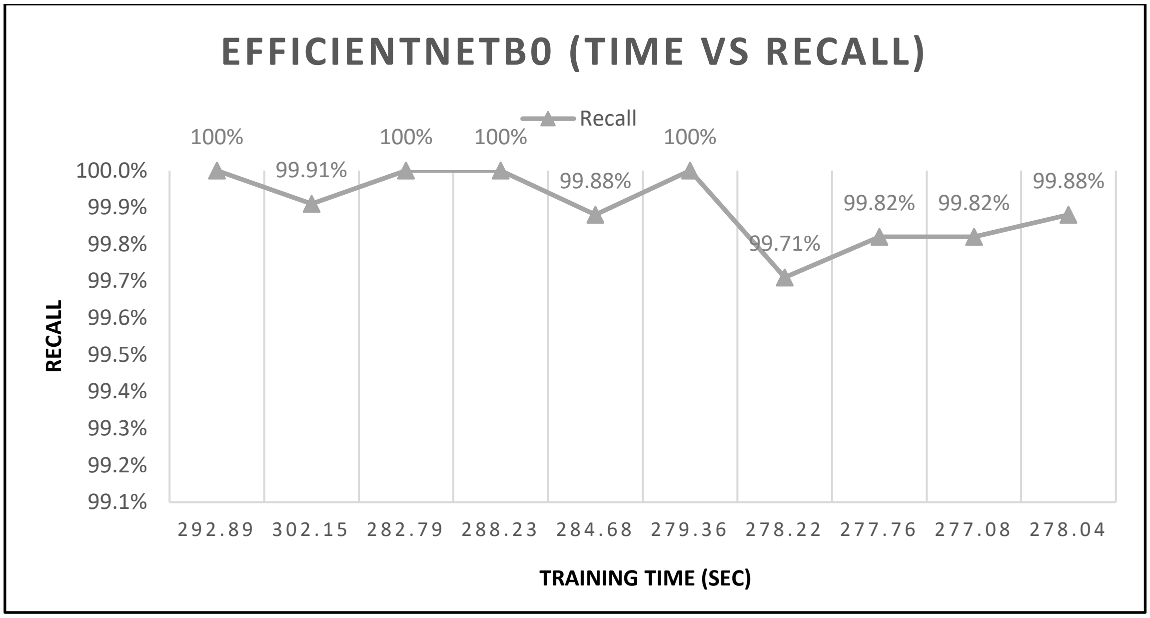

Figure 21: EFFICIENTNETB0 (TIME VS RECALL)

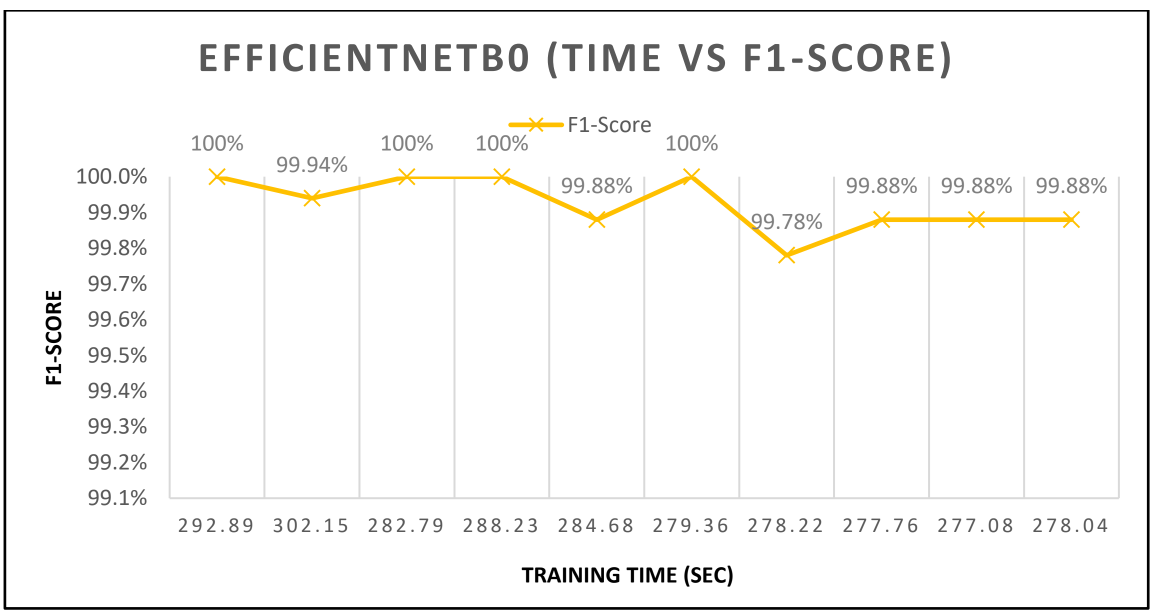

Figure 22: EFFICIENTNETB0 (TIME VS F1-SCORE)

Analysis of Model Performance Metrics of EfficientNetB0: Figure 19 tells that the accuracy is very high. There are no sharp drops. This concludes that the model is reliable and consistent. From figure 20, it says this model is very good at correctly identifying positive cases with minimal false positives. Again, the line is nearly flat, showing consistency over time. Figure 21 indicates the model rarely misses true positives. The graph is stable. To conclude, since both the precision and recall are strong, figure 22 is smooth.

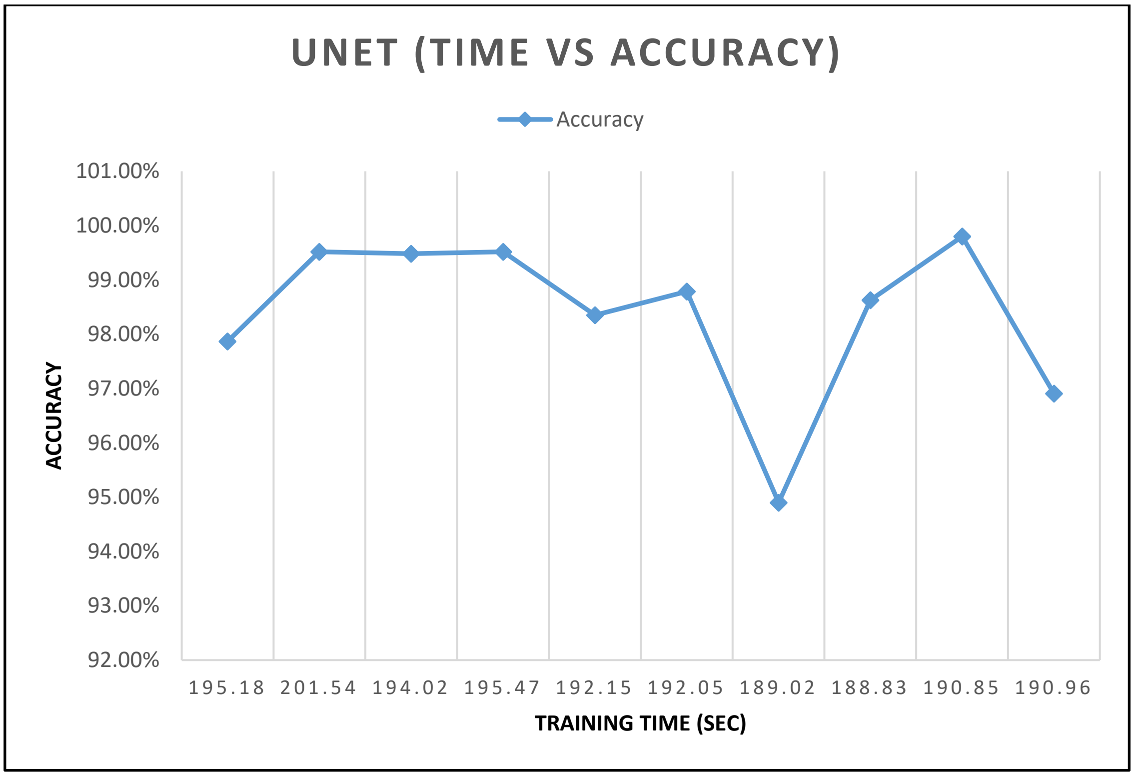

Figure 23:U-NET (TIME VS ACCURACY)

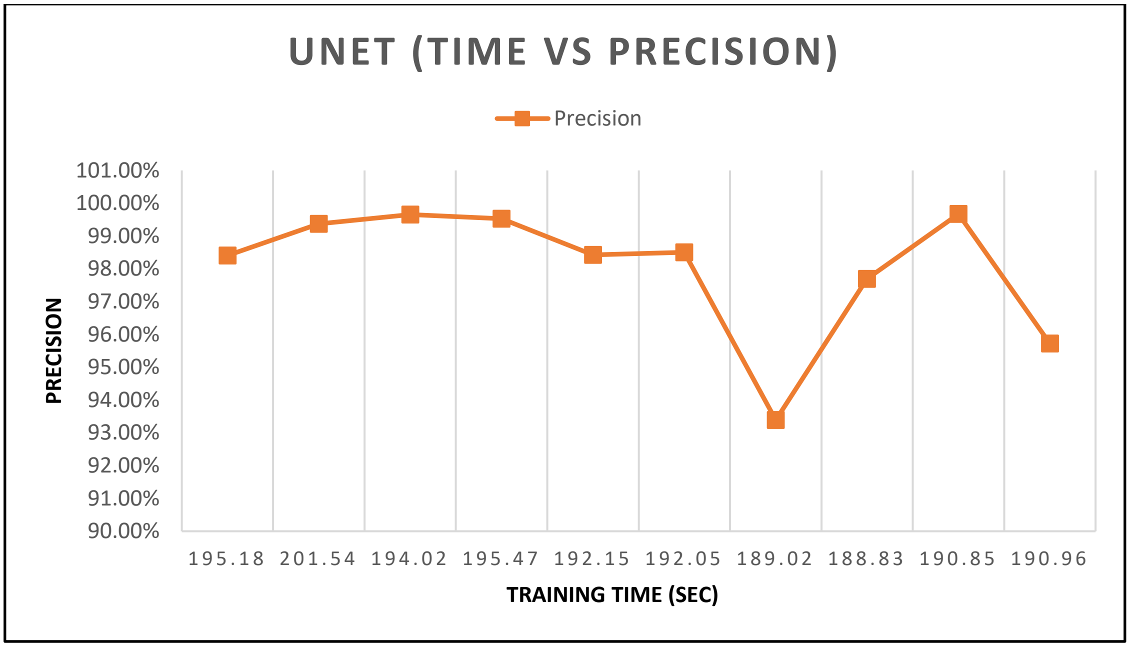

Figure 24:U-NET (TIME VS PRECISION)

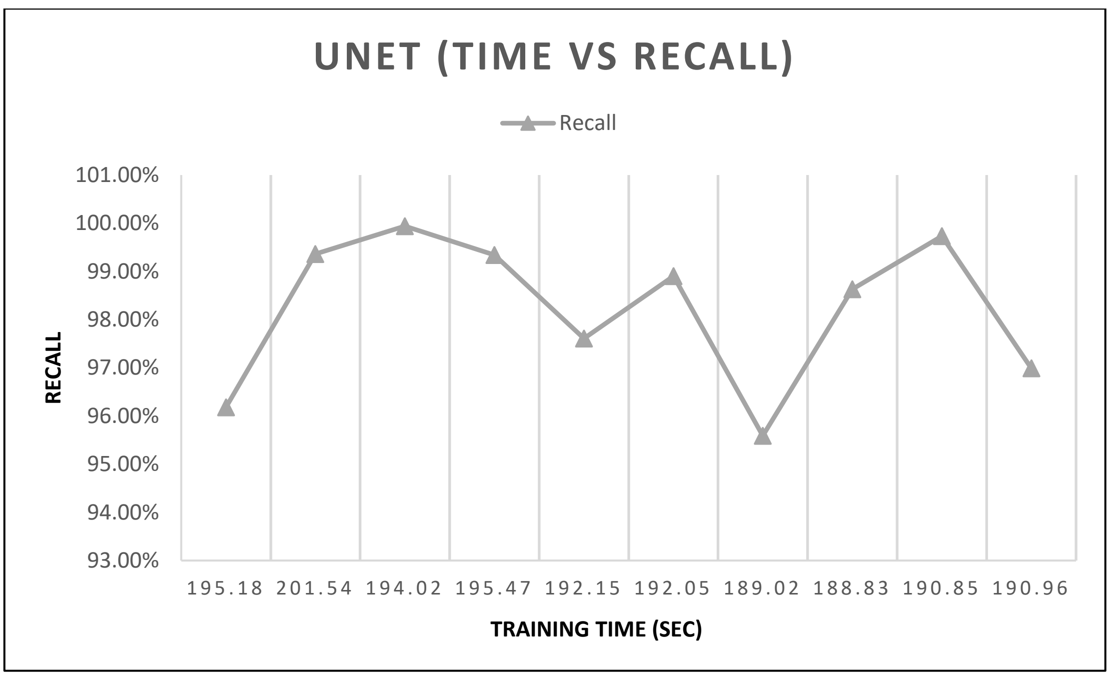

Figure 25:U-NET (TIME VS RECALL)

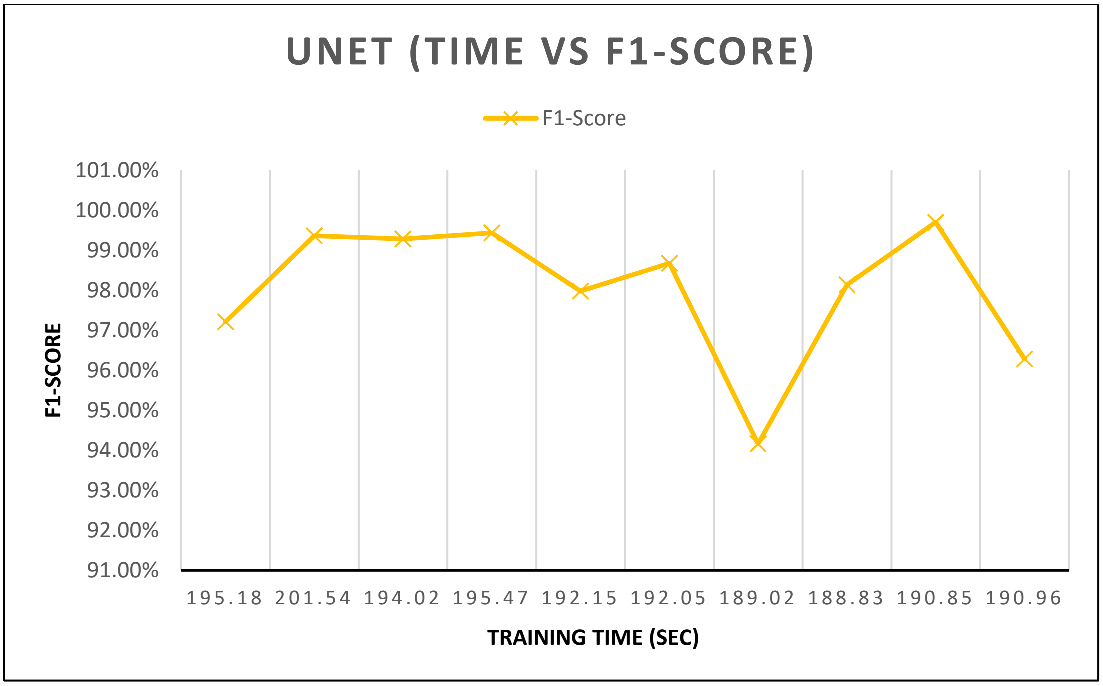

Figure 26:U-NET (TIME VS F1-SCORE)

Analysis of Model Performance Metrics of U-Net: Figure 23 shows slightly more variation than EfficientNetB0. The curve hits some fluctuations, suggesting training is a little less steady. Figure 24, precision is also high but not perfectly flat. U-Net sometimes struggles with false positives. Figure 25 suggest that the recall sits between 95% and 99%. This depicts that U-Net occasionally misses positive cases. The curve shows ups and downs and sensitive to training time. In conclusion, figure 26, shows that the F1-Score is strong enough, but a bit more fluctuation than EfficientNetB0 model.

Chapter 5
Findings & Future Scope

5.1 Findings

The following facts could be concluded from the findings of the present study:

- EfficientNetB0 resulted best performance, with all accuracy, precision, recall and F1-score reaching to nearly 100%, and was regarded the most informative model for diagnostic use.
- MobileNeV2 performances similarly well with high reproducibility and much shorter training time, hence making it an ideal candidate for a real-time and resource-limited clinical environment.
- The U-Net model was more variable across tests, with accuracy between 94.90% and 99.80%, which proves its lack of efficacy in classification, although is still strong in segmentation.
- Results were averaged out for 10 test tests, to give reliable and robust indications of the stability of the models making allowances for the effect of one-off exceptional results.
- The comparison analysis also illustrates the complementary property between accuracy and computation cost, especially for realistic medical deployment in diversified environments.

5.2 Future Work and Scope

Even though showing the good ability of both MobileNetV2 and EfficientNetB0, our case still has plenty of rooms for further research in the abnormal kidney classification. Adding more diverse and clinically collected CT images to the dataset will enhance generalization and reduce bias of the model. Clinical validation on real-world should be highly preferred to test for reliability with respect to diagnostic work-flows on actual patient data. Follower Research may further integrate multimodal images (e.g., CT, MRI and ultrasound) to construct more comprehensive diagnosis systems. Model optimization methods, such as testing and quantization, also might make more accurate models, such as the EfficientNetB0, more applicable to the real-time deployment on mobile or edge devices. Investigation into novel transformer-based architectures could potentially lead to better performance, and introduce explainable AI methods such as Grad-CAM for higher clinical trust in predictions. The treatment of class imbalance and rare abnormalities through synthetic data generation or dedicated training strategies will be crucial. Furthermore, incorporating such models into clinical decision support systems may aid the radiologists by decreasing burnout and avoiding diagnostic errors. If these models can be shown to track progression of the disease and help in prognosis, then longitudinal studies may be be especially valuable. Finally, ethical and legal concerns of patient privacy and fairness and regulatory compliance need to be tackled to guarantee a responsible and safe implementation of AI in healthcare.

Chapter 6
Conclusion

The objective of this study was to explore the applications of deep learning models on the differentiation of kidney abnormalities from CT scans, among which three architecture models (i.e., MobileNetV2, EfficientNetB0, and U-Net) were used. The models were tested throughout ten independent test tests with significant experiments and the evaluation metrics include accuracy, precision, recall, F1-score, and training time. The experiment results validated the great potential of CNNs for medical image classification and showed that EfficientNetB0 achieved the best performance. Nearly perfect performance close to 100% in all evaluation measures with very small standard deviations across tests indicates that it may be a highly trusted tool in diagnostic practice where accuracy is paramount. MobileNetV2 also achieved impressive accuracy, and not only accuracy never fell below 98.48% in this case, but it also achieved the property of fast learning, which is particularly important for real-time processing and resource-limited applications in rural clinics, mobile diagnostics units, or other embedded healthcare devices. Though U-Net was computationally efficient, it did not work well in consistently predicting like that of ours, indicating that it tends to be more a model for segmentation and localization related tasks but not a direct classifier.

Apart from the numerical results, there are several important implications for AI in healthcare based on this work. First, it shows that lightweight deep learning models can reach world-class level of diagnostic accuracy without designing very deep and complicated network architecture, which makes them deployable in health systems that have limited computing resources. The second implication of this study is that it is critical to assess the models over different test tests to maintain reliable and stable model, something that is often neglected in single-test research. Third, it create a comparison between classification-focused CNNs and segmentation-based network, allowing to understand their strengths and weaknesses while working on CT image classification.

In general, this study adds to the increasing evidence that AI-based diagnostic aids can support radiologists to make appropriate decisions for faster, more consistent, and accurate diagnoses. The efficientNetB0 may be used in hospitals or research institutions with high resources because of its excellent diagnostic accuracy (and the MobileNetV2 may be used to screen patients in general practice, or in field to enable real-time diagnosis. The results of this work not only prove the applicability of deep learning on kidney abnormality detection but also suggest opportunities for future development, among which larger dataset, cross-modal integration, and clinically verified application. In summary, this study demonstrates the transformative potential of deep learning in the medical imaging realm, and with appropriate optimization and deployment strategies, AI can help bridge fundamental gaps in the delivery of healthcare and facilitate timely, accurate and wider-spread diagnosis of kidney diseases across the globe.

References

[1] S. Jacobson, "Chronic kidney disease public health problem?," *Lakartidningen,* vol. 110, no. 21, pp. 1018-1020, 2013.

[2] W. H. Organization, "Kidney disease," WHO, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/kidney-disease.

[3] V. J. e. al., " Chronic kidney disease: global dimension and perspectives," *The Lancet,* vol. 382, no. 9888, pp. 260-272, 2013.

[4] G. B. D. C. Collaboration, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis," *The Lancet,* vol. 395, no. 10225, p. 709–733, 2020.

[5] K. J. F. e. al., "Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality…," *The Lancet,* vol. 392, no. 10159, p. 2052–2090, 2018.

[6] P. D. S Das, "Chronic Kidney Disease Prevalence among Health Care Providers in Bangladesh," *Mymensingh medical journal,* 2010.

[7] R. L. e. al., "Sustainable development goals relevant to kidney health: an update," *Kidney International,* vol. 96, no. 1, pp. 37-39, 2019.

[8] K. C. L. R. R. M. J. V. Z. C. Jager KJ, "A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases," *Kidney International,* vol. 96, no. 5, p. 1048–1050, 2019.

[9] R. M. C. L. B. N. S. P. M. D. Alvarenga L, "A Silent Dialogue Shaping Chronic Kidney Disease.," *J Xenobiot,* vol. 15, no. 3, p. 73, 2025.

[10] H. Y. e. al., "Kidney abnormality segmentation from CT using Swin-UNet: A transformer-based hybrid approach," *IEEE Journal of Biomedical and Health Informatics,* 2023.

[11] T. L. e. al., "A survey on deep learning in medical image analysis," *Medical Image Analysis,* vol. 42, pp. 60-88, 2017.

[12] G. Z. e. al., "Computer-aided kidney disease diagnosis using deep learning," *Computers in Biology and Medicine,* vol. 131, 2021.

[13] T. S. L. L. H. J. F. Ma, "Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network," *Future Generation Computer Systems,* vol. 111, pp. 17-26, 2020.

[14] E. V. N. R. L. M. P. M. A. C. Webster, "Chronic kidney disease," *The Lancet,* vol. 389, no. 10075, p. 1238–1252, 2017.

[15] R. D. X. X. H. G. X. F. T. S. S. Z. Z. Y. J. Xiao, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression," *Journal of Translational Medicine,* vol. 17, no. 1, pp. 1-13, 2019.

[16] M. H. M. K. H. M. G. R. A. M. Z. U. A. S. M. N. Islam, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography," *Scientific Reports,* vol. 112, no. 1, p. 11440, 2022.

[17] K.-U. E. Y. T. A. L. J. C. J. R. D. D. Z. T. H. H. N. L. G. E. A. S. Levey, "Definition and classification of chronic kidney disease: a position statement from kidney disease: Improving Global Outcomes (KDIGO)," *Kidney International,* vol. 67, no. 6, pp. 2089-2100, 2005.

[18] U. H. M. A. F. I. D. S. Z. A. F. Rosmani, "I-KS: Composition of chronic kidney disease (CKD) online informational self-care tool," in *2015 International Conference on Computer, Communications, and Control Technology (I4CT)*, 2015.

[19] Z. Z. R. Z. Y. X. P. B. H. Z. Chen, "Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers," *Chemometrics and Intelligent Laboratory Systems,* vol. 153, pp. 140-145, 2016.

[20] L. M.-A. C. A.-M. P. C. B. M.-C. J. L. F.-M. M. N.-D. N. C.-L. J. B. C.-A. S. Panizo, "Fibrosis in chronic kidney disease: pathogenesis and consequences," *International Journal of Molecular Sciences,* vol. 22, no. 1, p. 408, 2021.

[21] H. F. S. N. R. K. R. K. A. R. E. J. J. A. S. A. S. O. O. N. A. Almansour, "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," *Computers in Biology and Medicine,* vol. 109, pp. 101-111, 2019.

[22] A. H. M. A. I. M. S. H. W. A. F. P. R. S. M. A. S. Akter, "Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease," *IEEE Access,* vol. 9, p. 165184–165206, 2021.

[23] E. M. S. E. D. K. K. N. M. K.-K. C. A. Z. D. J. M. X. S. L. S. B. S. C. e. a. T. L. Nickolas, "Rapid cortical bone loss in patients with chronic kidney disease," *Journal of Bone and Mineral Research,* vol. 28, no. 8, p. 1811–1820, 2013.

[24] R. W. A. A. K. B. D. M. C. J. D. R. G. H. K.-U. E. B. L. K. P. A. M. R. S. P. e. a. C. A. Herzog, "Cardiovascular disease in chronic kidney disease: A clinical update from Kidney Disease: Improving Global Outcomes (KDIGO)," *Kidney International,* vol. 80, no. 6, p. 572–586, 2011.

[25] D. X. D. S. T. S. N. R. B. H. H. C. L. Y.-C. T. C. Y. C. E. S. T. e. a. C. Sabanayagam, "A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations," *The Lancet Digital Health,* vol. 2, no. 6, p. e295–e302, 2020.

[26] C.-H. L. Y.-C. C. W.-S. L. H.-F. C. J.-W. Hsieh, "Stage classification in chronic kidney disease by ultrasound image," in *Proceedings of the 29th International Conference on Image and Vision Computing, New Zealand*, 2014.

[27] B. K. R. N. J. K. S. S. H. B. S. T. A Esteva, "Dermatologist-level classification of skin cancer with deep neural networks.," *Nature,* vol. 542, pp. 115-118, 2017.

[28] P. Rajpurkar, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv preprint ,* 2017.

[29] L. B. A. K. D. W. X. Z. T. U. .. &. N. H. A. Dosovitskiy, "An image is worth 16×16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[30] L. C. &. H. L. H. Yuan, "Kidney abnormality segmentation from CT using Swin-UNet: A transformer-based hybrid approach," *EEE Journal of Biomedical and Health Informatics,* 2023.

[31] H. M. H. M. A. M. R. G. U. M. S. A. Islam MN, "Kaggle," 2021. [Online]. Available: https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone.

[32] Y. Z. a. X. Z. Y. Wang, "A deep learning approach for kidney tumor detection using ResNet-50 on CT scans," *Journal of Medical Imaging and Health Informatics,* vol. 13, no. 1, pp. 112-1210, 2023.

[33] S. L. &. H. L. Y. Zhang, "DenseNet for kidney stone segmentation and classification from CT images," *International Journal of Medical Informatics,* vol. 161, p. 104229, 2022.

[34] W. X. &. W. Z. J. Liu, "Using VGG-16 for kidney disease classification from CT scan images," *Journal of Healthcare Engineering,* vol. 2022, 2022.

[35] A. K. &. R. G. S. Kumar, "MobileNetV2 for kidney tumor classification from CT scan images," *Journal of Medical Imaging,* vol. 38, no. 5, pp. 240-249, 2023.

[36] M. C. &. D. Z. L. Zhang, "EfficientNetB0 for kidney stone detection in CT images," *BMC Medical Imaging,* vol. 23, no. 1, p. 17, 2023.

[37] J. X. &. F. Y. Z. Chen, "U-Net for segmentation and detection of kidney tumors from CT scan images," *Medical Image Analysis,* vol. 78, 2022.

211-35-720

ORIGINALITY REPORT

20%	16%	17%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Dublin Business School Student Paper	1%
2	doctorpenguin.com Internet Source	1%
3	assets-eu.researchsquare.com Internet Source	1%
4	arxiv.org Internet Source	1%
5	www.techscience.com Internet Source	1%
6	www.hindawi.com Internet Source	1%
7	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
8	www.nature.com Internet Source	<1%
9	mdpi-res.com Internet Source	<1%
10	pmc.ncbi.nlm.nih.gov Internet Source	<1%
11	www.researchsquare.com Internet Source	<1%
12	Krisna Nuresa Qodri, Muhammad Rausan Fikri, Luthfi Ardi. "Analytical Prediction for	<1%