ANALYSIS OF TRANSFORMER-BASED APPROACHES FOR CLASSIFYING RENAL CANCER FROM IMAGE DATA

by

S. M. Mushfiq Reza
20101254
Amreen Rahman
20301479
Ankita Roy
23141059
Abu Bakar Hasnath
20301037
Abdullah Bin Faruk
20301303

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

Department of Computer Science and Engineering
Brac University
January 2023

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:

S. M. Mushfiq Reza
20101254

Amreen Rahman
20301479

Ankita Roy
23141059

Abu Bakar Hasnath
20301037

Abdullah Bin Faruk
20301303

Approval

The thesis/project titled "ANALYSIS OF TRANSFORMER-BASED APPROACHES FOR CLASSIFYING RENAL CANCER FROM IMAGE DATA" submitted by

1. S. M. Mushfiq Reza(20101254)
2. Amreen Rahman(20301479)
3. Ankita Roy(23141059)
4. Abu Bakar Hasnath(20301037)
5. Abdullah Bin Faruk(20301303)

Of Spring, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 14, 2023.

Examining Committee:

Supervisor:
(Member)

Md. Tanzim Reza
Lecturer
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Dr. Farig Sadeque
Assistant 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

This research highlights the pressing need to revise the current diagnostic framework for kidney cancer due to the projected increase in its global prevalence. Renal Cell Carcinoma (RCC), the most prevalent form of kidney cancer, comprises a substantial majority, accounting for approximately 80-85% of all renal tumors. Recognizing the escalating trends of this disease, proactive measures are warranted to overcome upcoming challenges in accurate diagnosis and management. By addressing the evolving landscape of kidney cancer, healthcare professionals and researchers can better equip themselves to confront the imminent rise in cases and improve patient outcomes worldwide. Just like there are no specific causes, it is also hard to detect renal cancer because there are no such symptoms or clear signs to diagnose. As dangerous as it is, it can be cured if detected and treated early. Therefore, to determine this type of cancer early in this modern time, deep learning and machine learning are highly used. As part of this research, we additionally graded RCC using a fresh dataset with five distinct grades. From the Department of Pathology at Kasturba Medical College (KMC), Mangalore, India, we received 722 Hematoxylin & Eosin (H &E) stained slides of various patients with related grades. These histopathology dataset images are stored to detect renal cancer in the kidney by using different models and frameworks of deep learning and machine learning. Our approach incorporates the topologies of Convolutional Neural Networks (CNN) and Multi-Layer Perceptrons (MLP), as well as data augmentation methods and precise hyperparameter tuning (learning rate, batch size, dropout rate, and regularization strength, etc). Transformer-based deep learning methods are the latest trend in cancer classification from medical images. Currently, there are various types of transformer-based approaches; namely Vision Transformer, Swin Transformer, Multi-Axis Transformer, and so on. Our approach is a unified framework that allows for the diagnosis without manual intervention while simultaneously identifying tumors and classifying subtypes. A model that leverages this state-of-the-art classification approach should surely help doctors in automated renal cancer detection.

Keywords: Deep Learning; Malignant Cell; Renal Cell Carcinoma; Kidney Cancer; Transformers; Vision Transformer

Dedication (Optional)

A dedication is the expression of friendly connection or thanks by the author towards another person. It can occupy one or multiple lines depending on its importance. You can remove this page if you want.

Acknowledgment

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

Secondly, to our Supervisor Mr. Md. Tanzim Reza sir for his kind support and advice in our work. He helped us whenever we needed help.

Thirdly, the whole judging panel and our Co-Supervisor Mr. Dr. Farig Sadeque sir gave helped us a lot in our later works.

And finally to our parents without their throughout sup-port it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ANN Artifical Nueral Network

ASR Age-Standardised Rates

AUC Area Under Curve

CNN Convolutional Nueral Network

$CTScan$ Computed Tomography Scan

DNN Deep Nueral Network

DSC Dice Similarity Coefficient

DT Decision Tree

FPN Feature Pyramid Network

$GELU$ Gaussian Error Linear Unit

HE Hematoxylin Eosin

$HPRC$ Hereditary Papillary Renal Cancer

KNN K-Nearest Neighbors

KUB Kidney, Ureter & Bladder X-Ray

MLP MultiLayer Perceptron

$MRIScan$ Magnetic Resonance Imaging Scan

RCC Renal Cell Cancer

ROC Receiver Operating Characteristic

$SMOTE$ Synthetic Minority Oversampling Technique

SVM Support Vector Mechine

VHL von Hippel-Lindau

ViT Vision Transformer

Chapter 1

Introduction

1.1 Background of Renal Cancer

The kidney, a vital organ of the human body helps to filter out waste products from blood from the body to make urine. The primary kidney filtration units are called nephrons. The portion of the nephron where electrolytes are balanced and water is absorbed is termed the renal tubule. The term "renal" describes the kidneys. The malignant cell or abnormal growth of cells that forms in these tubules is commonly called renal cancer or kidney cancer [22]. The malignant cells form due to mutation of DNA in between cells. However, the cells develop and spread rapidly so that cells can break free and affect other cells of different parts of the body. Most of the cells of the bones and lungs are mostly affected. There is no specific reason why these malignant cells are formed in renal tubules, but doctors researched some of the causes of renal cancer. If someone has a bad habit of chain smoking, then he has a lot of risk for renal cancer. It depends on how much one smokes. If someone is a non-smoker, the risk is reduced a lot. Patients with obesity or high blood pressure(HBP) are at the highest possibility of having renal cancer. If any patient is already suffering from kidney failure and going through regular dialysis has a possibility of renal cancer.

1.2 Types of Renal Cancer

Genetically this disease can affect anyone. Most of the time, renal cancer does not pass from father to child. But if anyone has a history of renal cancer in their family, renal cancer can affect multiple generations of that family which is also known as hereditary syndromes. It is so rare that only 5%-8% of the cases are hereditary syndromes. And there are four types of hereditary syndromes[1].

Von Hippel-Lindau disease - a rare type of inherited disease that causes the growth of tumors, and cysts in certain parts of the body like kidneys. The patients have a huge risk of having renal cancer as there are no signs or symptoms of von Hippel-Lindau disease until young age or adulthood. This happens because of a mutation of the VHL gene, that why it's also called VHL syndrome[1].

Birt-Hogg-Dubé syndrome - a rare type of inherited disease that abnormally changes skin called fibrofolliculomas, it is not cancer but renal cancer, benign kidney tumors,

lung cysts, and a condition called pneumothorax is more likely to develop in people with Birt-Hogg-Dubé syndrome.

Hereditary Leiomyomatosis & Renal Cell Cancer - a rare inherited disease in an autosomal dominant manner and caused by the FH gene mutation. In addition, in this disease, smooth muscle tissues change abnormally which is called leiomyomas[1]. Benign leiomyomas can also form as fibroids in the uterus in females. The patient has an increased risk of renal cancer.

Hereditary Papillary Renal Cancer(HPRC) - is a rare inherited disease that increases the risk of type 1 papillary kidney cancer. Usually, HPRC is of two types: type-1 and type-2. This cancer develops in the cell that lines the small tube of the kidney that use to make urine by filtering out waste from the blood[1]. There can be more than one tumor in each kidney in HPRC and HPRC often affects both kidneys. HPRC is caused because of mutations in the MET gene.

Hereditary renal cancer gets detected at an early stage than sporadic renal cancer. The treatment and conditions are also different from sporadic cancer. Sporadic Renal cancer can be of different types.

Renal Cell Carcinoma - is the most frequent type of renal cancer found in adults. About 85% of the cases are renal cell carcinoma[1]. These cancer cells form in the renal tubule of the kidney which filters the waste from the blood, which is a part of the Nephrons. And there are about 10 lac nephrons in each kidney.

Urothelial carcinoma - this type of cancer is also called transitional cell carcinoma which is a type of cancer found in adults. About 5%-10% of the cases are of urothelial carcinoma. This cancer develops in the renal pelvis, the part where urine is collected before moving to the bladder[1]. Bladder cancer also develops in the cells of that same line, so urothelial carcinoma is also treated as bladder cancer.

Sarcoma - this type of cancer is rare in kidneys. This type of cancer develops in capsules, a thin layer of connective tissue surrounding the kidney. Sarcoma usually comes back after surgery and can spread to different parts of the human body, that's why more treatments and chemotherapies are recommended by doctors.

Wilms tumor - this type of cancer is commonly found in children. About 1% of the cases are Wilms tumors. Wilms tumor is successfully treated by radiation therapy and chemotherapy.

Lymphoma - this type of case is rare and the kidneys are enlarged as a lone tumor mass and associated with enlarged regional lymph nodes called lymphadenopathy.

1.3 Diagnosis & Treatment

As there are no specific signs or symptoms of renal cancer, it is hard to diagnose renal cancer. But doctors initially mentioned some problems which commonly cancer patient faces. Like unexplained weight loss, nausea, malaise, vomiting, loss of

appetite, and high fever. For renal cancer, the patient may develop a palpable mass in the abdomen or lower back. In clinical manifestations, the tumor grows enough or the cancer spreads enough to physically obstruct the urinary flow which will make urine build up in the ureter. Although, Cancer can compress the nearby nerves which may cause pain in the hip bone or the flank. Renal cancer can be diagnosed through lab studies, imaging tests, and renal biopsy. In a lab test, Complete Blood Count (CBC), Kidney Function Tests, and Urinalysis are conducted. In an imaging test, Pelvic or Abdominal CT scans, MRI scans, Renal Ultrasound, KUB (Kidney, Ureter & Bladder X-Ray), and rediomics are conducted by which stage tumors can be classified. The objective behind radiomics is to build models that match image attributes to pathological outcomes and extract numeric features from radiographic pictures. However, some radiomics algorithms have been presented forth in recent years to identify kidney cancers[19]. Once suspicious lesions are identified, a renal biopsy is performed and confirms the diagnosis.

When small localized tumors are diagnosed, surgery takes place to remove the affected part which is known as a partial nephrectomy. When a large place is affected, a patient has to go through a radical nephrectomy in which the whole kidney may be removed or nearby lymph nodes are removed and chemotherapy and radiation therapy to terminate the cancer cells.

1.4 Statistics of Renal Cancer

More than 432,000 fresh kidney cancer cases are anticipated globally in 2020. Only Bangladesh has more than 1500 new cases of kidney cancer. Lithuania has the highest overall rate of kidney cancer in ASR 14.5[17]. If we separate genders then the rate of new cases for males is double the number of new cases for females. And in death cases, more than 179 thousand people lost their lives to kidney cancer in 2022. In Bangladesh, the death rate is 0.89% which is 969. Slovakia has the highest rate of death in ASR 4.7. In 2022, the USA has more than 79 thousand new cases of kidney cancer and a death rate of more than 13 thousand[23]. Globally, kidney cancer caused 138.5 thousand (95% UI: 128.7-142.5) fatalities, 393,000 (95% UI: 371.0-404.6) new incident cases, and 3.3 million (95% UI: 3.1-3.4) DALYs in 2017[22].

Chapter 2

Problem Statement

The present modern era has brought many scientific successes in the field of medicine. Now people do not have to die for lack of treatment and no one has to suffer. Although, in this scientific age, there are still some diseases whose treatment and solutions have yet to be wholly found. Only early detection and treatment of life-threatening illnesses can preserve a person's life, and this is only possible if the cancer is found in its earliest stages. Renal cancer is also called renal adenocarcinoma or hypernephroma. worldwide every year, kidney cancer accounts for 144,000 fatalities and 338,000 new cases. Moreover, proper treatment planning and patient care, and the precise categorization of renal cancer are essential. Renal cancer is a complicated illness with a variety of symptoms.

Adult kidney cancer is most frequently diagnosed as renal cell carcinoma, accounting for 85% of cases. The proximal renal tubules, which make up the kidney's filtration system, are the site of development for this kind of cancer. In the United States, kidney cancer is anticipated to affect 431,288 individuals worldwide in 2020. Besides, since the middle of the 1990s, the mortality rate has been declining. Deaths due to kidney cancer will be down by about 2% per year between 2013 and 2020. Globally, kidney cancer is anticipated to claim 179,368 lives in 2020. Additionally, in the US for 2022 the most recent projections for renal cancer from the American Cancer Society will be 50,290 new instances of kidney cancer in males and 28,710 in women. Total of almost 79,000 new cases. Furthermore, in total, 13,920 persons will pass away from this illness (8,960 males and 4,960 women)[28]. Also, kidney cancer will be discovered in 81,800 individuals in 2023 (52,360 males and 29,440 women). When the cancer is limited to the kidney, about two-thirds of patients receive a diagnosis. The rate of 5-year survival for this category is 93%. The rate of 5-year survival is 71% if renal cancer has disseminated to adjacent tissues or organs and/or the local lymph nodes. It may frequently be treated when it is still contained in the kidney and the surrounding tissue.

It's vital to keep in mind that figures relating to renal cancer patient survival rates are simply estimates. Cancer may or may not shorten a person's life, but they cannot predict this for a specific individual. Instead, these figures show patterns in populations of people who have already received the same diagnosis, including details about the disease's various phases.

Chapter 3

Research Objectives

Every year more than 1500 cases are found in Bangladesh and more than 1000 death cases of renal cancer. As the detection and diagnosis of kidney cancer are difficult at the early stage it can not be detected, and the death rate and new case rates are increasing. If we can detect these diseases at an early stage, then we can treat the disease for more time. Although, kidney cancer is hard to deal with at the last stage, which can result in the complete removal of the whole kidney. If renal cancer can be detected much earlier small surgeries can save someone's life. In our country, the tests cost a lot of money and a lot of time which a cancer patient does not have. However, as time passes cancer grows bigger and spreads to other organs and affecting those organs too. The time doctors detect cancer and start the treatment it becomes too late. That is the reason the death rate is high for cancer patients. Moreover, Every year diseases are mutating into more powerful diseases. But we are lacking behind. Our technologies for medical science are not capable of detecting diseases at an early stage. Therefore, some early adjustments should be made to testing processes, spreading awareness of the value of early detection, and ease of access to medical treatments. Besides, research is concentrated on creating more effective diagnostic techniques to identify kidney cancer in its earliest stages. With the use of these techniques, it should be easier to spot kidney irregularities or tiny tumors that could be cancerous. It is being pursued to improve the sensitivity of imaging technologies including computed tomography (CT) scans, magnetic resonance imaging (MRI), and ultrasound to identify kidney cancer in its early stages. This entails developing novel imaging agents or contrast agents as well as enhancing image resolution and imaging methods. As renal cancer is detected in the images from the CT Scan or MRI, we will use Deep learning Models and frameworks to learn the pattern of Renal cancer. Furthermore, Models based on deep learning have proven successful in a number of industries, including picture categorization and identification. These algorithms can recognize the patterns and characteristics indicating the illness by being trained on a sizable dataset of kidney carcinoma pictures. After being equipped, these models can potentially be used to spot possible kidney cancer patients in just emerging images.

Then it can detect small cells from the images from the CT Scan or MRI Scan which will result in decreasing the death rates around the world. Renal Cancer will not be a hardly detectable cancer. We will be using transformers which are models of deep learning to classify renal cancer at the early stage.

Chapter 4

Literature Review

Cancer classification has gained a lot of interest from researchers in the computer vision field as a critical challenge in digital pathology investigation. Recently, a number of Convolutional Neural Network (CNN)-based models for various cancers of various organs, such as breast [14], lung [13], and kidney [21], have been developed. The main goal of these models is to categorize these diseases into subgroups based on histological characteristics that CNN can readily extract, such as tumor architecture.

In an earlier study on the classification of renal cancer subtypes, Kocaka et al. [16] used SVM and ANN to classify the subtypes of renal cancer. Three radiologists performed a reproducibility study as the initial step in feature selection, followed by a wrapper-based classifier-specific method. The model was optimized and features were chosen using layered cross-validation. Artificial Neural Networks (ANN) and Support Vector Machines were the main classifiers (SVM). To enhance generalizability performance, base classifiers were additionally merged with three additional methods. The following categories were used for classification: (i) clear cell RCC (cc-RCC) vs papillary cell RCC (pc-RCC) versus chromophobe cell RCC; and (ii) non-clear cell RCC (non-cc-RCC) versus clear cell RCC (cc-RCC) (chc-RCC). The Matthews correlation coefficient served as the primary performance parameter for comparisons (MCC). The best method for differentiating non-cc-RCCs from cc-RCCs using corticomedullary phase pictures was an ANN with an adaptive boosting algorithm (MCC = 0.728), which had external validation accuracy, sensitivity, and specificity of 84.6%, 69.2%, and 100%, respectively. However, the effectiveness of QCT-TA is pretty subpar when it comes to differentiating the three main subtypes. For differentiating pc-RCC from other RCC subtypes, the SVM with bagging algorithm performed best (MCC = 0.804), with accuracy, sensitivity, and specificity for external validation of 69.2%, 71.4%, and 100%, respectively [27]. Iodine mapping and dual-energy CT were utilized by Mileto et al. [6] to help radiologists distinguish clear cell renal cell carcinoma (ccRCC) from papillary renal cell carcinoma (pRCC) visually. Using color-coded iodine maps, five readers who were blind to the pathologic diagnosis independently assessed each case's lesion iodine concentration. To determine the best threshold for separating clear cells from papillary renal cell cancer, the researchers used receiver operating characteristic curve analysis. Leave-one-out cross-validation was used to confirm the correctness of the results.An intraclass correlation coefficient helped to find the interobserver agreement. Inves-

tigators looked at the relationship between tumor iodine content and tumor grade. Tumor iodine concentrations of 0.9 mg/mL were found to be the ideal threshold for differentiating between clear cell and papillary RCC renal cell carcinoma, with the following findings: sensitivity of 98.2%, specificity of 86.3%, a predictive value of 95.8%, the negative predictive value of 93.7%, and overall accuracy of 95.3% with an area under the curve of 0.923. The measured tumor iodine content was found to be very well agreed upon by the five readers (intraclass correlation coefficient, 0.9990). A strong relationship between tumor iodine content and tumor grade was discovered for both clear cell and malignant tumors.

The high-throughput extraction of many picture characteristics from radiographic images is known as radiomics. It has been suggested to use radiomics to extract quantitative information from radiographic pictures and create models that link image aspects to outcomes [2]. As solid cancers are diverse in both time and space. This restricts the use of expensive biopsy-based molecular tests but opens up a world of possibilities for medical imaging that can non-invasively capture intratumoral heterogeneity. Some radiomics models have been put out in recent years to categorize kidney cancers. Texture analysis was utilized by Hodgdon et al. [9] to distinguish angiomyolipoma (AMLs) from renal cell cancer (RCC). The DeLong approach helped to differentiate a receiver operating characteristic curve's regions under(AUC) between subjective heterogeneity evaluations and textural attributes. An AUC of 0.89 was obtained using a model that used many texture characteristics. SVM accuracy for textural characteristics varied from 83% to 91% on average (10-fold cross-validation). A random forest was used by Raman et al. [7] to forecast the pathophysiology of kidney malignancies. External validation of the model was performed on a different group of 19 unidentified cases. Oncocytomas and clear cell RCCs were properly classified by the random forest model in 89% (sensitivity = 89%, specificity = 99%) and 91% (sensitivity = 91%, specificity = 97%) of the cases, respectively. To distinguish between several kinds of tiny renal tumors, Feng et al. [15] used quantitative texture analysis based on machine learning on CT images. However, the characteristics used in these investigations, such as their form, intensity, texture, and wavelet textures, were specifically developed or made by hand [3]. In preoperative three-phase CT scans, texture characteristics were manually segmented from the biggest tumorous areas of interest (ROIs). A preliminary selection of characteristics was made using the Mann-Whitney U test and interobserver reliability. Then, using support vector machines with recursive feature elimination (SVM-RFE) and the synthetic minority oversampling method (SMOTE), discriminative classifiers were created, and their performance was assessed. The SVM-RFE+SMOTE classifier showed the best performance by differentiating between microscopic angiomyolipoma without visible fat (AMLwvf) and RCC with the highest accuracy, sensitivity, specificity, and AUC of 93.9%, 87.8%, 100%, and 0.955, respectively. The potential of radiomics may be limited as a result of the selection of these low-throughput traits based on the professional knowledge of radiologists. Convolutional neural networks (CNNs) have recently made significant advances in computer vision capabilities as a result of the introduction of graphics processing units and massive training datasets [5]. A CNN may automatically extract high-throughput features and forgo the laborious artificial feature extraction method when a sizable training dataset is provided [4]. CNN has performed admirably in the medical domains. By

using 2000 dermatological photos and the associated pathological findings to train a CNN model, Esteva et al. [10] were able to distinguish between benign and malignant skin malignancies utilizing only the inputs of pixels and disease labels. By contrasting the results of the procedure with those reached by 21 board-certified dermatologists, the method's effectiveness was evaluated. The dermatologists concentrated on two crucial categorization tasks: identifying keratinocyte carcinomas from benign seborrheic keratoses and separating malignant melanomas from benign nevi. They combined biopsy data with photos that had been clinically verified. The outcomes showed that some convolutional neural networks (CNNs) are capable of classifying skin cancer with a degree of accuracy comparable to dermatologists. On each task, the CNN fared just as well as the tested pros. Additionally, Arevalo et al. [8] used a CNN to categorize mammography mass lesions. They employed a hybrid strategy in which the representation was supervised and learned using CNNs. In other words, they directed the feature learning process using the annotations and had outstanding results with values ranging from 79.9% to 86.0% as measured by the ROC and AUC curve. Feature learning for mammography mass lesions using convolutional neural networks is evaluated here before being fed to a classification step. It has not yet been thoroughly investigated if such a method may help distinguish properly between benign and malignant kidney cancers based on CT scans.

In the classification of kidney diseases, a variety of deep-learning techniques are used. In the paper [11], They begin by using the median filter, Gaussian filter, and un-sharp masking to improve the image. To analyze kidney stone pictures, they first employed morphological procedures including erosion and dilation. Then, to determine the region of interest, they used entropy-based segmentation. For both the original picture and the segmented image, they computed many metrics, including the standard deviation, entropy, thresholding, energy, and homogeneity. Finally, they used KNN and SVM classification approaches. Subsequently, The K-nearest neighbor (KNN) classifier and principal component analysis (PCA) are used to extract information from the pictures. They proposed two types of classification here. KNN was found to be 89% accurate, whereas SVM was shown to be 84% accurate. This study [25] uses convolutional neural networks and other machine learning techniques to categorize individuals as either healthy or patients based on the presence or absence of kidney stones in medical photographs (CNN). These included CNN-based deep neural networks, Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbor (kNN), and Naive Bayes (BernoulliNB). According to the experiments, the Decision Tree Classifier (DT) provides the best classification performance. With an S+U sample success rate of 85.3%, this approach has the greatest F1 score rate. According to the experimental findings, the Decision Tree Classifier (DT) is a workable technique for differentiating renal x-ray pictures. The automated categorization of B-mode renal ultrasound pictures is suggested in this study [24] and is based on a group of deep neural networks (DNNs) that use transfer learning. The quality selection in ultrasound pictures is based on the perception-based image quality assessor score, and speckle noise often affects the images. The support vector machine is used for classification after the pre-trained DNN models extract features from three different datasets. The majority voting method is used with multiple pre-trained DNNs, including ResNet-101, ShuffleNet, and MobileNet-v2, to produce final recommenda-

tions. By combining the predictions from several DNNs, the ensemble model outperforms the individual models in classification[28]. When contrasted with traditional and DNN-based categorization approaches, the given method clearly demonstrated its advantages. The established ensemble model divides the normal, cyst, stone, and tumor classes for the kidney ultrasound pictures. The authors achieved the maximum accuracy of 95.58% using ultrasound pictures there for the categorization challenge. In order to reliably and quantitatively detect chronic kidney illnesses, this research [26] concentrated on using deep learning techniques for the segmentation of CT images. First, renal cysts in CT images were automatically segmented using the residual dual-attention module (RDA module). As research participants, 79 individuals with renal cysts were chosen, of whom 52 instances served as the training group and 27 cases served as the test group. The segmentation results for the test group were evaluated using the Dice similarity coefficient, recall, and precision (DSC). The experimental results demonstrated that the RDA-UNET model's loss function value quickly converged and declined, Additionally, the segmentation outcomes of the study's model were nearly identical to those of hand labeling, confirming the model's high level of picture segmentation accuracy as well as its capacity to precisely segment the kidney's shape. By obtaining 96.25% DSC, 96.34% precision, and 96.88% recall for the left kidney and 94.22% DSC, 95.34% precision, and 94.61% recall for the right kidney, the RDA-UNET model beat earlier methods. The results showed that the algorithm model utilized in this study outscored other algorithms in each assessment index. The authors in [18] suggested a lesion identification algorithm based on morphological cascaded convolutional neural networks that use multiple intersections over union (IOU) thresholds (CNNs). To improve the detection of small lesions (1–5 mm) and boost network stability, we proposed two morphological convolution layers, updated feature pyramid networks (FPNs), and four IOU threshold cascade RCNNs. PyTorch was used to train the modified CNN for this lesion detection task. The research was done using DeepLesion kidney CT pictures supplied by hospital picture archiving and sharing systems (PACSs). The findings showed that our suggested detector is an excellent tool for detecting lesions in CT and outperformed the dataset, with our technique achieving an AP of 0.840 and AUC of 0.871. In [12], the scientists built a fully automated system for detecting renal cysts that are backed by reliable kidney segmentation done by a fully convolutional neural network. Initial candidates for cysts are provided by an integrated 3D fluid and kidney distance map around them. The final step is to classify the candidate's status as cysts or non-cyst objects as a second convolutional neural network. 52 abdomen CT images with more than 70 cysts that were randomly picked from a genuine radiological workflow and annotated by a skilled radiologist were used to evaluate performance. When the minimum cyst diameter was set to 10 mm, the system detected 59/70 cysts with a true-positive rate of 84.3% and an average of 1.6 false positives per case.

Chapter 5

Working Plan

5.1 Methodology

Vision transformers (Vit) are a type of neural network architecture which is a transformer-based neural network and were developed by Google back in 2017 and were mostly used for natural language processing. But throughout the years it has been proven to be quite effective in computer vision applications, especially in the area of object detection and image segmentation. To use a vision transformer for renal cancer image classification we need to follow some necessary steps. These steps or techniques are crucial to train the model well and ensure the expected outcome from our model.

5.2 Data Collection

The used image dataset is of the KMC kidney histopathology dataset. This Image dataset comes in the form of five categories (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4), starting from normal cells which are Grade-0, and ending with Grade 4 which is heavily affected by Renal Cell Carcinoma. All of the images used are collected from Kasturba Medical College(KMC) Mangalore and from Manipal Academy of Higher Education (MAHE), Karnataka, India. The images were produced during a clinical study at the Department of Pathology from October 2020 to December 2022 at Kasturba Medical College(KMC). The images were produced by following the 1964 Helsinki Declaration ethical standards, hence human participants were informed and consent from all the patients was obtained during the period of the whole experiment and the personal details of each individual were secured and protected[29].

5.3 Resizing Image Data using Open CV

All of the images were captured by a camera with a pixel size of 3.69*3.69 and for that reason, the total pixel number will be 1920*1440*3. All images must be of a constant size before they can be used, so all the raw un-resized images were resized to an image size of 72 * 72. There were 144 patches per image and the patch size was 6 by 6.

5.4 Creating Training Dataset

Images come in five categories (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4) as discussed earlier. The following steps are taken to create a training dataset.

Step 1:
Images from each of the five categories (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4) are split into 2 parts training and testing, 80% of the images are in the first part which is used for training.

Step 2:
Some traditional techniques which include vertical and horizontal flipping and data augmentation techniques were used to increase the diversity of our training part of the dataset.

Step 3:
After completing the above-mentioned steps we obtained a total of 693 patches for our Grade 0(non-cancerous), followed by 708 patches for Grade 1, 648 parches for Grade 2, for Grade 3 735 patches, and finally 648 patches for Grade 4 kidney cancer images.

Step 4:
A portion of training data is used for validation so that the model can be fine-tuned to produce appropriate output. While creating the validation set non-overlapping regions of training patches were used.

5.5 Creating Test Dataset

Previously 80% of the images from the five categories (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4) were allocated for training the dataset, after this the rest 20% of the images were allocated for testing purposes. The testing dataset also consists of five categories, each of the categories consists of an x-axis and a y-axis. Where x is the features and y is labeled Grade 0 to Grade 4.

5.6 Configuring the Hyperparameter

Before we run our MLP model we need to configure the hyperparameter to ensure our MLP model is optimized. Firstly the hyperparameter that is optimal for this model is identified, this may include learning rate, batch size, and number of layers. The search space is defined and a search method is selected. The result is evaluated and the best hyperparameter is selected.

5.7 Data Augmentation

Data augmentation is conducted to increase the diversity of the dataset artificially, using common straightforward operations like rotations, translations, flips, and dis-

tortions. Overfitting is reduced by exposing a broad range of variations so that the model becomes more robust.

5.8 Multi-Layer Perceptron

A two-layer classifying network is an MLP using GELU at the end. The final MLP block, commonly known as the MLP head, works as the output of the transformer. Lastly, to this output, softmax can be used to apply categorization label.

5.9 The Vision Transformer Model

Vision transformers (Vit) are a type of neural network architecture that is a transformer based neural network and transformers were developed by Google back in 2017 and were mostly used for natural language processing. But throughout the years it has been proven to be quite effective in computer vision applications, especially in the area of object detection and image segmentation.

Patch Embeddings & Positional Encodings
Firstly, each individual image is turned into 16 by 16 size patches by dividing into a grid of non-overlapping patches. Before using these as input for our ViT transformer model each of these patches is linearly embedded so that a sequence of flattened vectors is created, this introduces a sense of relative positioning of each patch. Hence, to represent relative positions in the patches position all encodings are used.

Patch Encoding Layer
The previous patch embeddings and the positional encodings both are combinedly and used as the input for our vision transformer model. These input embeddings interact with the first layer of the transformers.

Core Architecture of Vision Transformer
The transformer encoder serves as the core component for our vision transformer. Self-attention and feed-forward network layers are the core components of the encoder of our transformer architecture, these two layers are crucial as they allow the model to capture the infractions between image patches and to capture the dependencies between them.

Multi-Head Attention and Fine-Tuning
Multiple self-attention heads are used to increase our model‘s capacity to interact with more diverse patterns and relationships, in this step, each head must perform its own self-attention computation. The data which was previously separated for validation are now used during this training process, the primary objective is to fine-tune the model using these validation images.

Feed-Forward Networks
The output of the self-attention layers is taken, then feed-forward into a neural network as a result a feed-forward neural network is applied independently to each embedding. A group of fully connected non-linear activation functions is present in

the feed-forward neural network.

Layer-Stacking
This process consists of repeated implementation of the entire process of self-attention and feed-forward computations. To allow the ViT model to capture more detailed and increasingly complex features each layer has its own set of learnable parameters.

Classification of the head
The classification head is the last layer for our Vision Transformer, this layer receives output from all the previous layers and receives encodings from the last encoder. The objective of this layer is to perform classifications by using the softmax activation functions and fully connected layers, hence this layer is responsible for creating probability distribution over the possible classes of categories.

5.10 Training & Evaluating the Model

This is the final step of our working plan and it involves repeating the necessary steps from the above. To begin with, typically a vision transformer is pre-trained on a large-scale dataset, in the first step the objective is to teach the model to predict the correct order of shuffled patches. After that, our pre-trained model needs to be fine-tuned on a specific downstream of tasks, this means predicting the probability distribution over the possible classes of categories using distinct features of the images. Because of this, our fine-tuned model can be used for image classification and object detection for our specific labeled data. Hence this model becomes effective to classify the input data into five categories (Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4).

When test data is fed into the model, our model outputs the probability of each input falling under the five categories, this result is compared against the actual checked true outcome to measure the accuracy of our model.

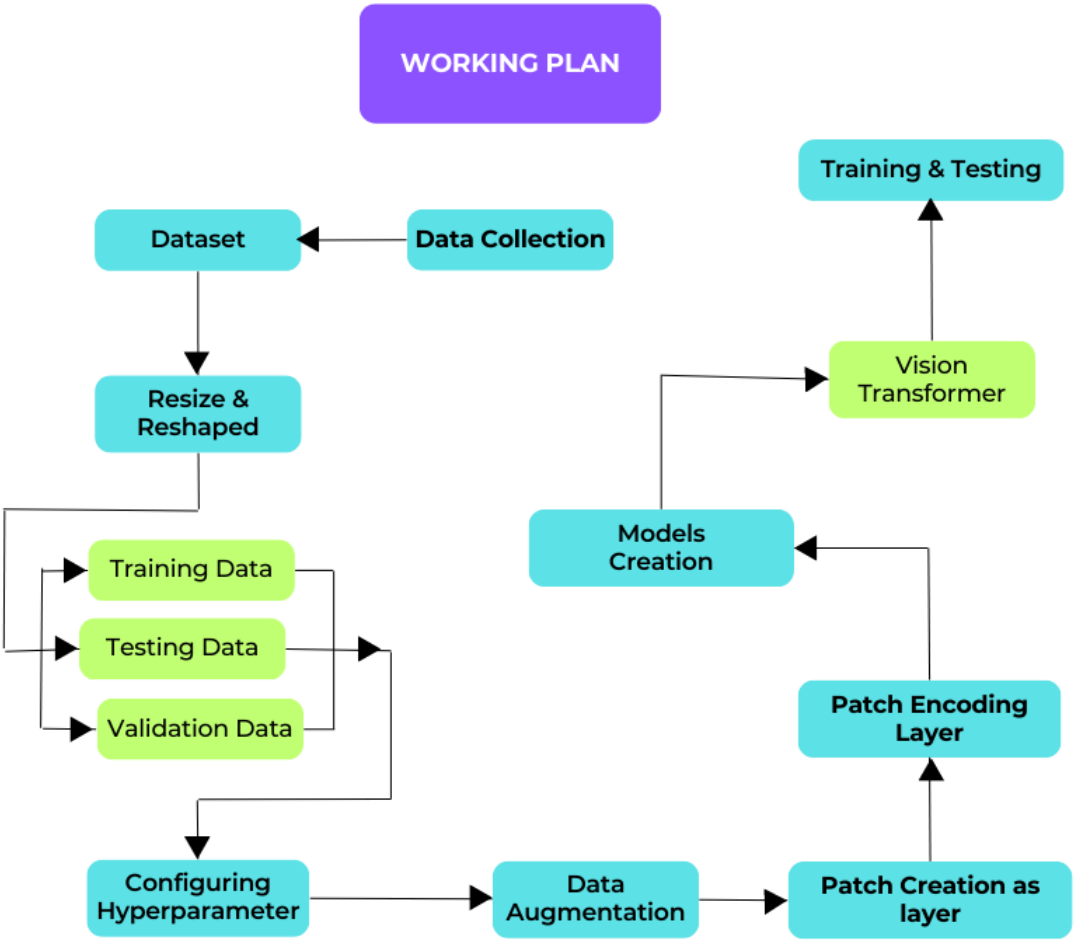

Figure 5.1: Working Plan

Chapter 6

Description Of The Model

6.1 Vision Transformer

Although transformers were developed by Google back in 2017, The vision transformers (ViTs) are relatively new and were first introduced in a paper titled "An image is worth 16 by 16 words: Transformer for image recognition at scale"[20]. Transformers were originally developed for natural language processing but in recent years they are being used in computer vision, the vision transformer model uses transformer architecture. Vision transformers work by dividing an image into patches and then flattening them in a linear array, their relative position to each other is also noted. These patches are then fed into a layer of transformer encoder so that global dependencies and relationships between the patches are captured. When a large enough data set is used, a vision transformer model can outperform traditional CNN models.

6.2 Multi-Layer Perceptron

The Multi-Layer Perceptron is a type of neural network where information flows in one direction and is hence known as a feedforward artificial neural network. The main components of MLP are the input layer, the hidden layers, and the output layer. Each layer consists of neurons which are interconnected nodes, and each neuron takes input from the previous neuron and forwards the output to the next neuron. Multi-Layer Perceptrons are used in image classifications and regression.

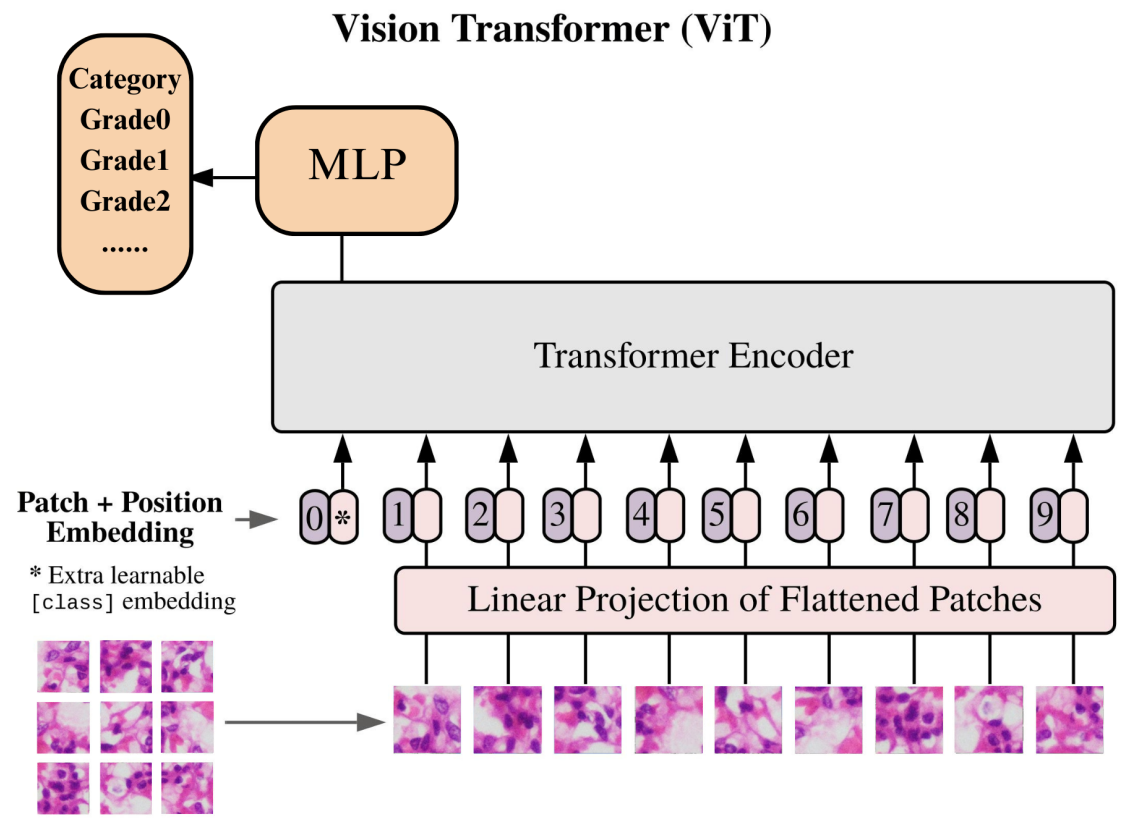

Figure 6.1: Vision Transformer Architecture

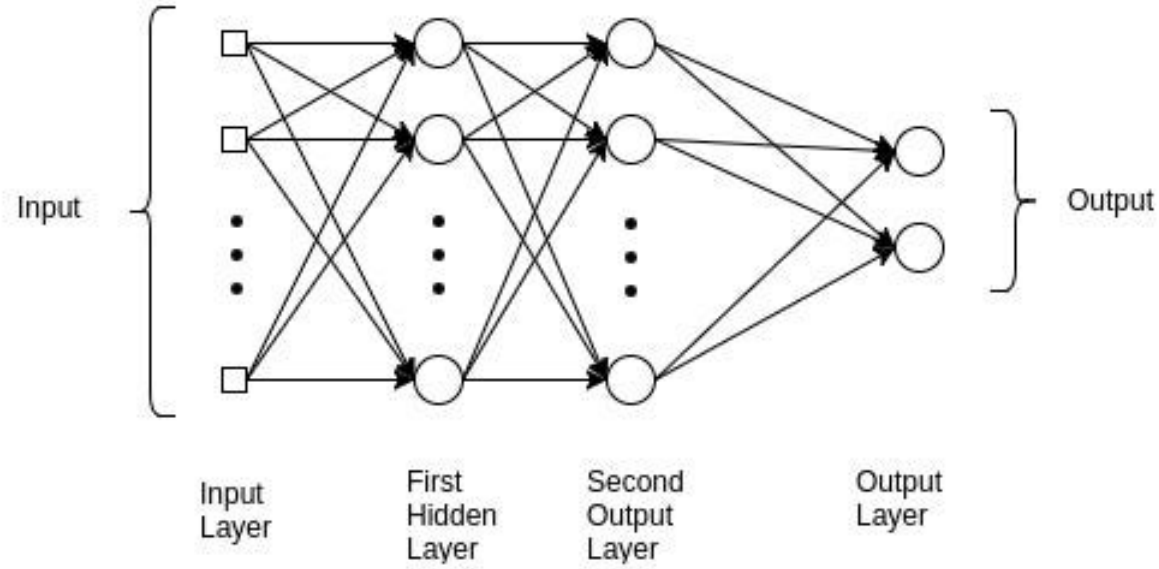

Figure 6.2: MLP Architecture

Chapter 7

Description Of The Data

The Renal Clear Cell Carcinoma images from the KMC kidney histopathology dataset used in this study include both non-cancerous (Grade 0) and malignant (Grades 1 to Grade 4) samples. These pictures were gathered as part of a clinical study that was carried out at the Department of Pathology, Kasturba Medical College(KMC) Mangalore, Karnataka, India, between October 2020 and December 2022. Under protocol number, IEC KMC MLR 02/2022/57, Kasturba Medical College(KMC), Mangalore's institutional ethics committee granted the project ethical approval. With respect to the data, the experiments were carried out after obtaining informed consent from each patient, and strict confidentiality was maintained to protect patient information[29].

The surgical biopsy used kidney tissue stained with Hematoxylin & Eosin (H&E). This open surgical approach was used to acquire the biopsy samples. The pathology division of KMC subsequently analyzed and categorized the kidney tissue slides based on their histological features. The grades were decided after a thorough clinical study and include Normal/Non-cancerous(Grade 0) to cancerous(Grade 1, Grade 2, Grade 3, and Grade 4). The pathologists processed the tissue samples using the conventional paraffin procedure, which includes a number of processes including fixation, dehydration, cleaning, infiltration, embedding, and trimming. The pathologist next used a microscope to visually identify the tumor-affected regions on the produced slides. The pathologists at KMC, Manipal have provided annotations of this dataset. All other parameters, with the exception of selecting the area, were automatically maintained throughout all images. Non-overlapping square patches were chosen, resized to 50x50, and normalized to have a mean and variance of 0 and 1, respectively, in order to get them ready for training and testing with a deep learning architecture.

To make distinct training and test sets, the entire dataset was divided. The remaining 80% of photographs were utilized for training, with about 20% set aside for testing. Cropping non-overlapping areas from the training patches produced the validation set.

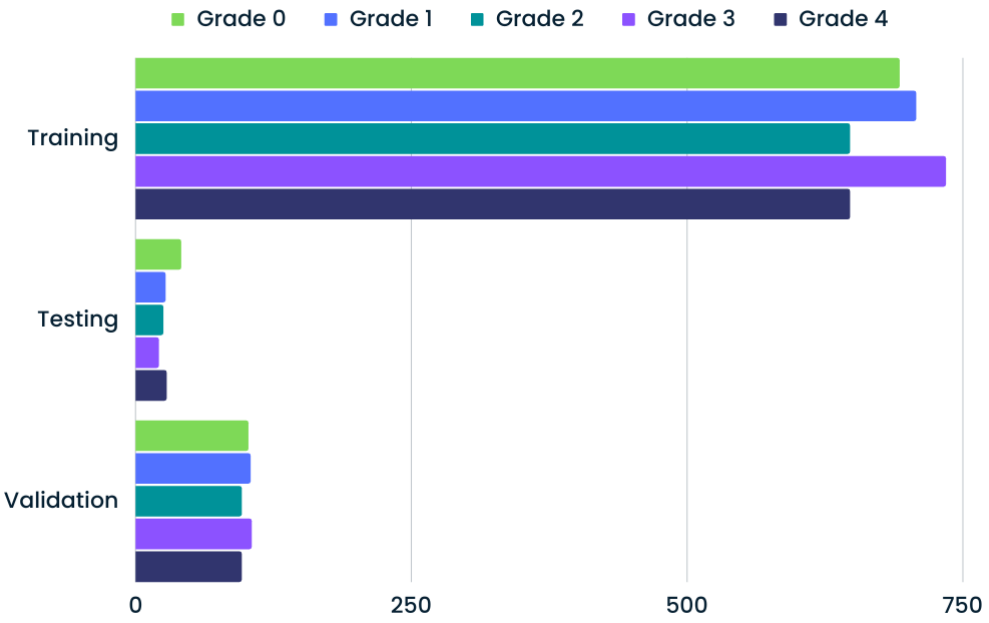

Figure 7.1: Graph of Dataset

Data augmentation techniques like horizontal and vertical flipping were used to increase the training set's variety. A total of 3442 patches from five distinct classes were produced after the use of data augmentation and random cropping, and these patches served as the training set. These patches included 693 patches with a grade of Normal/Non-Cancerous, 708 patches with a grade of Grade 1 Cancer, 648 patches with a grade of Grade 2 Cancer, 735 patches with a grade of Grade 3 Cancer, and 648 patches with a grade of Grade 4 Kidney Cancer. It's important to note that the model assessment was limited to the original test set and that the test set was not exposed to data augmentation.

Type	Training Patch	Test Patch	Validation Patch
Grade 0	693	41	102
Grade 1	708	27	104
Grade 2	648	25	96
Grade 3	735	24	105
Grade 4	648	28	96
Total	3432	145	503

Table 7.1: Proposed Dataset Grades Distribution

Chapter 8

Preliminary Analysis

First of all, the image dataset we used is categorized into 5 grades (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4), starting from the normal cell which is Grade 0, and ending with Grade 4 which is heavily affected by Renal Cell Carcinoma. Besides, our primary objective is to classify renal cell carcinoma(renal cancer), we labeled the dataset into 5 categories and merged the whole dataset of training to find if it is 'non-cancerous' or 'cancerous'. Besides, we also merged the dataset of testing and validation as x for the image data and y for the labels of the grades of that image.

Our dataset consists of a total of 4077 images of cancer-affected kidneys. Therefore, 84.2% of the dataset is used for training the models of Deep learning, 3.5% of the dataset is used for testing the models and 12.3% of the dataset is used for validation. Although, there is 84.2% of the dataset is for training, 20% of the training dataset is merged with the testing dataset. Besides, the Deep Learning ViT model is used for training and testing. For the ViT model, Multi-Level Perceptron is implemented. The MLP is the two-layer classification network with GELU at the end and the MLP head is the out of the transformer. Now, the image data are resized and patches are created as layers. A total of 144 patches are made from a single image data. Then, the patch will be encoded by the patch encoder layer which will transform the patch linearly into a vector size of projection. The projected vector also gains a learnable position embedding. We used epoch 100 for the ViT algorithm.

After running the architecture of models, the training was completed after 100 epochs. In the Vision Transformer model, we received an accuracy of 84.51% and the rate of data loss is 0.5178 at 100 epochs.

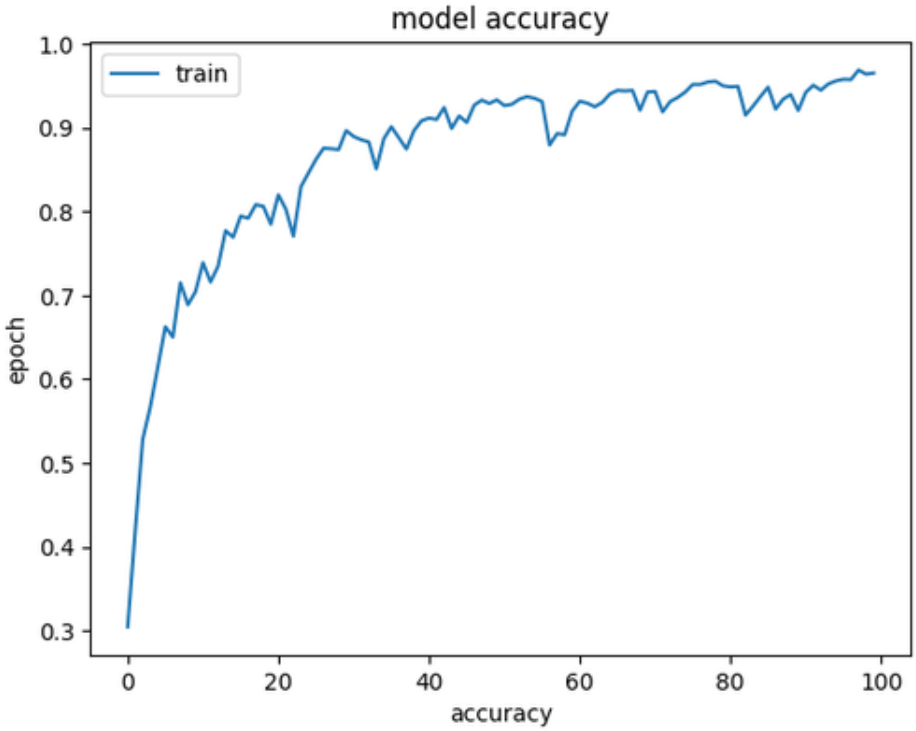

Figure 8.1: Accuracy of ViT Model

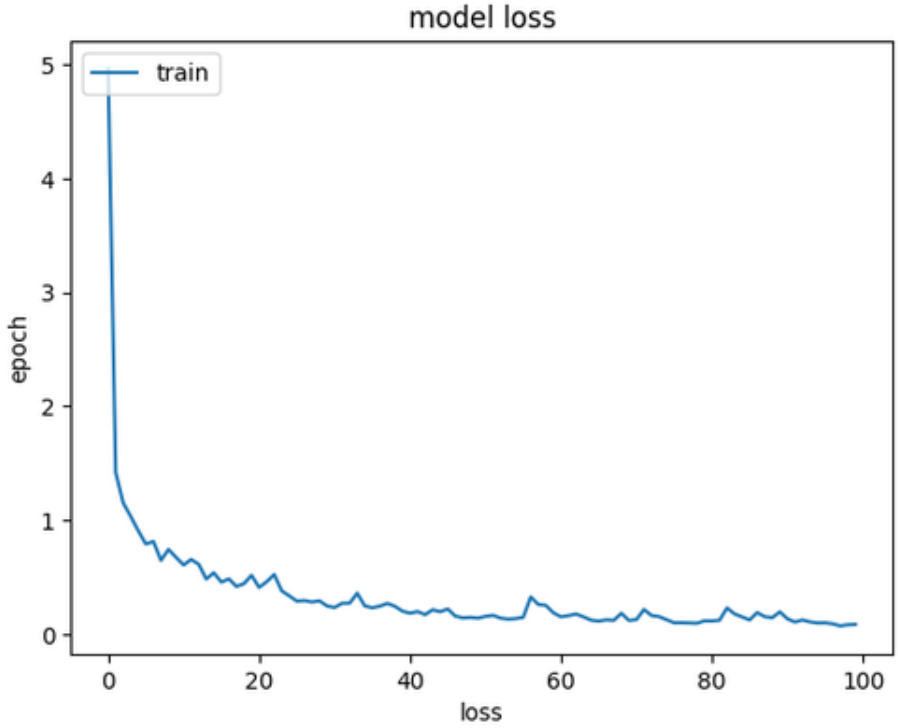

Figure 8.2: Data Loss of ViT Model

Chapter 9

Conclusion

This study developed a comprehensive deep learning framework that operates autonomously to identify the degree of malignancy in kidney histopathology images for renal cell carcinoma (RCC). In this paper, we conduct an analysis of various transformer-based approaches applied to kidney histopathology images, which have not been previously investigated. We know that renal cancer is one of the most deadly and fatal malignancies in the world. Similar to any cancer, early identification and treatment are crucial. There is a good chance of recovery if we can identify renal cancer in its early stages and begin the appropriate therapy. As a result, we can make it possible to predict and detect the signs of renal cancer in humans early on and take appropriate action by combining various algorithms with Deep Learning methods. The majority of the information that we gathered and analyzed came from linked research papers and publications about the identification of renal cancer. We have gathered material from similar papers and nearby publications linked to our study topic for the Literature review and Workflow sections of the work. We reviewed the most current studies on the detection and diagnosis of renal cancer. Moreover, we divided these findings into categories including detection, pre-processing, classification, and so on. We looked into each area and gathered knowledge on the models that were applied to various tasks. Then we evaluated all the transformer based approaches. This evaluation can provide insights into the strengths and weaknesses of each approach and help identify the most effective technique for the given dataset. Conducting a comprehensive comparison allows us to establish a benchmark for transformer-based approaches on the specific dataset. This benchmark can serve as a reference point for future research and provide a standardized evaluation metric for other researchers working on similar tasks. Furthermore, we intend to expand on this in the future to include more pertinent subjects.

References

[1] P. C. G. E. Board, "Hereditary kidney cancer syndromes (pdq®): Patient version," *PDQ Cancer Information Summaries [Internet]*, 2002.

[2] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *European journal of cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[3] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, no. 1, pp. 1–9, 2014.

[4] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *International conference on machine learning*, PMLR, 2014, pp. 647–655.

[5] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[6] A. Mileto, D. Marin, M. Alfaro-Cordoba, J. C. Ramirez-Giraldo, C. D. Eusemann, E. Scribano, A. Blandino, S. Mazziotti, and G. Ascenti, "Iodine quantification to distinguish clear cell from papillary renal cell carcinoma at dual-energy multidetector ct: A multireader diagnostic performance study," *Radiology*, vol. 273, no. 3, pp. 813–820, 2014.

[7] S. P. Raman, Y. Chen, J. L. Schroeder, P. Huang, and E. K. Fishman, "Ct texture analysis of renal masses: Pilot study using random forest classification for prediction of pathology," *Academic radiology*, vol. 21, no. 12, pp. 1587–1596, 2014.

[8] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Convolutional neural networks for mammography mass lesion classification," in *2015 37th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, 2015, pp. 797–800.

[9] T. Hodgdon, M. D. McInnes, N. Schieda, T. A. Flood, L. Lamb, and R. E. Thornhill, "Can quantitative ct texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced ct images?" *Radiology*, vol. 276, no. 3, pp. 787–796, 2015.

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[11] J. Verma, M. Nath, P. Tripathi, and K. Saini, "Analysis and identification of kidney stone using kth nearest neighbour (knn) and support vector machine (svm) classification techniques," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 574–580, 2017.

[12] N. Blau, E. Klang, N. Kiryati, M. Amitai, O. Portnoy, and A. Mayer, "Fully automatic detection of renal cysts in abdominal ct scans," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 7, pp. 957–966, 2018.

[13] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature medicine*, vol. 24, no. 10, pp. 1559–1567, 2018.

[14] H. D. Couture, L. A. Williams, J. Geradts, S. J. Nyante, E. N. Butler, J. Marron, C. M. Perou, M. A. Troester, and M. Niethammer, "Image analysis with deep learning to predict breast cancer grade, er status, histologic subtype, and intrinsic subtype," *NPJ breast cancer*, vol. 4, no. 1, pp. 1–8, 2018.

[15] Z. Feng, P. Rong, P. Cao, Q. Zhou, W. Zhu, Z. Yan, Q. Liu, and W. Wang, "Machine learning-based quantitative texture analysis of ct images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma," *European radiology*, vol. 28, no. 4, pp. 1625–1633, 2018.

[16] B. Kocak, A. H. Yardimci, C. T. Bektas, M. H. Turkcanoglu, C. Erdim, U. Yucetas, S. B. Koca, and O. Kilickesmez, "Textural differences between renal cell carcinoma subtypes: Machine learning-based quantitative computed tomography texture analysis with independent external validation," *European Journal of Radiology*, vol. 107, pp. 149–157, 2018.

[17] S. Han, S. I. Hwang, and H. J. Lee, "The classification of renal cancer in 3-phase ct images using a deep learning method," *Journal of digital imaging*, vol. 32, no. 4, pp. 638–643, 2019.

[18] H. Zhang, Y. Chen, Y. Song, Z. Xiong, Y. Yang, and Q. J. Wu, "Automatic kidney lesion detection for ct images using morphological cascade convolutional neural networks," *IEEE Access*, vol. 7, pp. 83 001–83 011, 2019.

[19] L. Zhou, Z. Zhang, Y.-C. Chen, Z.-Y. Zhao, X.-D. Yin, and H.-B. Jiang, "A deep learning-based radiomics model for differentiating benign and malignant renal tumors," *Translational oncology*, vol. 12, no. 2, pp. 292–300, 2019.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[21] Z. Gao, P. Puttapirat, J. Shi, and C. Li, "Renal cell carcinoma detection and subtyping with minimal point-based annotation in whole-slide images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 439–448.

[22] S. Safiri, A.-A. Kolahi, M. A. Mansournia, A. Almasi-Hashiani, A. Ashrafi-Asgarabad, M. J. Sullman, D. Bettampadi, M. Qorbani, M. Moradi-Lakeh, M. Ardalan, *et al.*, "The burden of kidney cancer and its attributable risk factors in 195 countries and territories, 1990–2017," *Scientific reports*, vol. 10, no. 1, pp. 1–20, 2020.

[23] R. Siegel, "Cancer statistics 2020/r. l. siegel, kd miller, a. jemal," *Cancer J Clin*, vol. 70, pp. 7–30, 2020.

[24] S. Sudharson and P. Kokil, "An ensemble of deep neural networks for kidney ultrasound image classification," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105 709, 2020.

[25] I. AKSAKALLI, S. KAÇDIOĞLU, and Y. S. HANAY, "Kidney x-ray images classification using machine learning and deep learning methods," *Balkan Journal of Electrical and Computer Engineering*, vol. 9, no. 2, pp. 144–151, 2021.

[26] X. Fu, H. Liu, X. Bi, and X. Gong, "Deep-learning-based ct imaging in the quantitative evaluation of chronic kidney diseases," *Journal of Healthcare Engineering*, vol. 2021, 2021.

[27] K.-H. Uhm, S.-W. Jung, M. H. Choi, H.-K. Shin, J.-I. Yoo, S. W. Oh, J. Y. Kim, H. G. Kim, Y. J. Lee, S. Y. Youn, *et al.*, "Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography," *NPJ Precision Oncology*, vol. 5, no. 1, pp. 1–6, 2021.

[28] A. Abdelrahman and S. Viriri, "Kidney tumor semantic segmentation using deep learning: A survey of state-of-the-art," *Journal of Imaging*, vol. 8, no. 3, p. 55, 2022.

[29] A. Chanchal, D. S. Lal, R. Kumar, J. T. Kwak, and J. Kini, "A novel dataset and efficient deep learning framework for automated grading of renal cell carcinoma from kidney histopathology images," *Scientific Reports*, vol. 13, Apr. 2023. DOI: 10.1038/s41598-023-31275-7.