Deep Learning-Based Hybrid Multi-Task Model for Adrenocortical Carcinoma Segmentation And Classification

by

Nirjhor Datta
20101540
Md. Hasanur Rashid
23241144
Samiur Rahman
20101147
Naima Tahsin Nodi
20101150
Moin Uddin
20101134

A Thesis Report submitted to the Department of Computer Science and Engineering

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

© 2024. Brac University
All rights reserved.

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:

Nirjhor Datta

Nirjhor Datta
20101540

Md. Hasanur Rashid

Md. Hasanur Rashid
23241144

Samiur Rahman

Samiur Rahman
20101147

Naima Tahsin Nodi

Naima Tahsin Nodi
20101150

Moin Uddin

Moin Uddin
20101134

Approval

The thesis/project titled "Deep Learning-Based Radiomics Analysis for Adrenocortical Carcinoma (ACC) Diagnosis and Staging with Ki-67 Expression" submitted by

1. Nirjhor Datta (20101540)
2. Md. Hasanur Rashid (23241144)
3. Samiur Rahman (20101147)
4. Naima Tahsin Nodi (20101150)
5. Moin Uddin (20101134)

Of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on January 18, 2024.

Examining Committee:

Supervisor:
(Member)

Dr. Muhammad Iqbal Hossain
Associate Professor
Department of Computer Science & Engineering
School of Data and Sciences
Brac University

Co-Supervisor:
(Member)

Mr. Md. Tanzim Reza
Lecturer
Department of Computer Science & Engineering
School of Data and Sciences
Brac University

Program Coordinator:
(Coordinator)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science & Engineering
School of Data and Sciences
Brac University

Head of Department:
(Chairperson)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Ethics Statement

Our research has been made for academic purposes. All the data and resources that have been used here are cited properly with references. Those, who made a contribution to our study, we mentioned their name in the acknowledgment section. Therefore, ethically there is no issue with our work.

Abstract

Adrenocortical Carcinoma (ACC) is a rare but highly lethal cancer that occurs in the adrenal cortex. Accurate diagnosis of ACC are vital in order to determine appropriate treatment strategies and predict patient outcomes. Hence, defining the stages of ACC is a crucial factor for both diagnosis and treatment planning and it is the key aspect that the researchers are still exploring. Our study proposes a novel deep learning-based hybrid Multi-Task model which performs both segmentation to find the exact cancer region and classification based on the cancer stages. Thus our model is resource efficient. In our research, several deep learning-based architectures have been used to segment and evaluate the ACC CT images. Moreover, we have explored how Convolutional Neural Network (CNN) classification models perform on the classification task. This process includes the exploration to find the model based on the Multi-Task learning model's feature extraction perform on classification task.

Keywords: Deep Learning; Convolutional Neural Network; Adrenocortical Carcinoma; Diagnosis; Multi-Task learning; Segmentation; Classification.

Dedication

Dedicating our work to all the parents in the world through whom we come to the earth and experience countless blessings!

Acknowledgement

Firstly, we express our heartfelt gratitude to the Great Almighty for guiding us through the successful completion of our thesis without any major setbacks. We extend our appreciation to our esteemed supervisor, Dr. Muhammad Iqbal Hossain sir, and Co. Supervisor Md Tanzim Reza sir for their invaluable assistance and unwavering guidance throughout the process. Lastly, we are grateful to our parents for their unconditional support, without whom this achievement would not have been possible.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Adrenocortical Carcinoma (ACC), also known as adrenal cortex cancer, originates in the adrenal glands, is a rare and aggressive type of cancer which's survival rate is very low. The adrenal glands, as shown in Figure 1.1, are located on top of both kidneys are responsible for producing hormones that regulate various bodily functions (e.g. metabolism, blood pressure, body's response to stress). ACC appears on the outer layer of the adrenal gland which is known as the adrenal cortex. The Adrenal cortex is responsible for producing hormones such as cortisol, aldosterone, sex hormones, etc. [1].

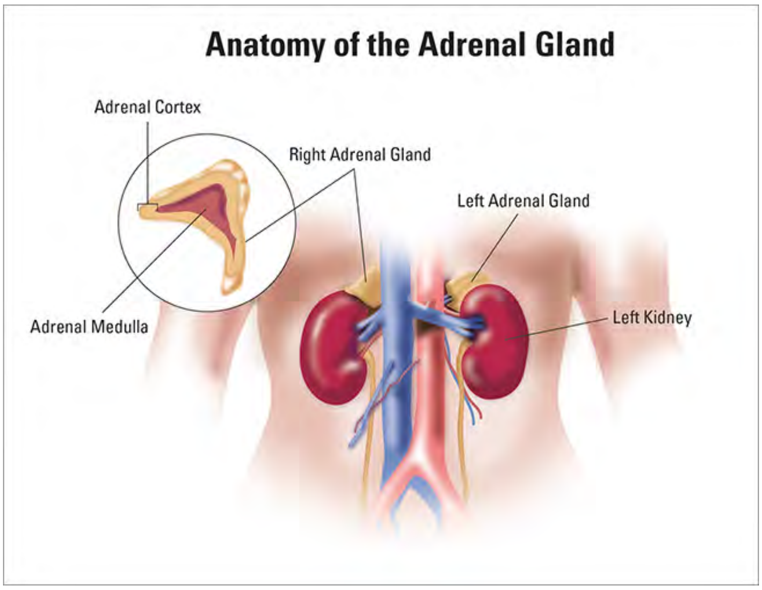


Figure 1.1: Anatomy of the Adrenal Gland [2]

ACC's tendency to grow and spread rapidly makes it aggressive. It often has already metastasized to other organs, such as the liver, lungs, or lymph nodes by the time it is diagnosed [3].

Abdominal pain or discomfort, weight loss, high blood pressure, hormonal changes leading to symptoms like excessive hair growth or virilization (masculinization), and hormonal imbalances are the general symptoms caused by ACC though the symptoms may vary depending on the specific hormones involved [4].

One of the major issues occurring by ACC is that it can produce excessive amounts of hormones which leads to hormonal imbalances. This may result in very compli-

cated situations involving Cushing's syndrome (excess cortisol production), Conn's syndrome (excess aldosterone production), and virilization (excess sex hormone production) [1].
The occurrence of Adrenocortical Carcinoma (shown in Figure 1.2) is rare, with an estimated incidence of 1-2 cases per million people per year [5]. However, if that 1-2 occurrence happens to our beloved ones, what would be the case? Of course, it would be a very concerning issue for us. In addition, the prevalence of ACC is high between the ages of 40-60 although it can occur at any age [6]. Here, the affected case is slightly more common in women (60%) than men (40%) [7]. Moreover, the patient-affected ACC's survival rate is very low. Five years survival rate is for an estimated 22% of people only [8]. This survival rate is even lower for advanced ACC cases. Hence, this cancer is a serious one and there is no way we can ignore it.

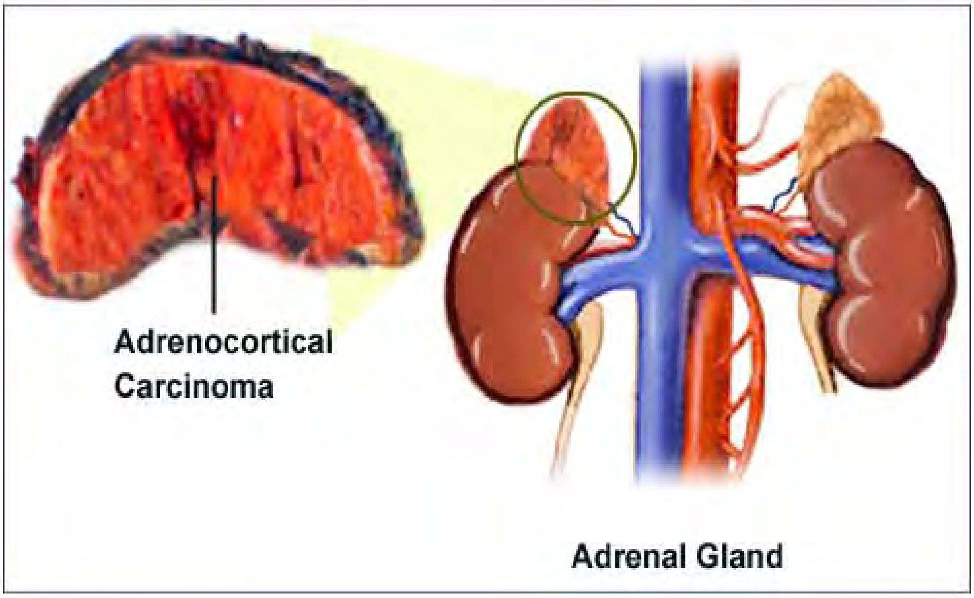


Figure 1.2: Adrenocortical Carcinoma [9]

1.1 Problem Statement

The research for efficient ACC detection and mitigation is limited. Most of the time, ACC diagnosis involves imaging studies such as CT scans or MRI to visualize the tumor which makes it easier to determine its size and extent. To confirm the cancer and assess its aggressiveness, a Biopsy is generally performed. In both ACC diagnosis and treatment planning, staging is usually considered an important part that helps to determine the extent of cancer spread and guides treatment decisions. Removing the tumor by surgery whenever it is possible is one of the major treatments for ACC. However, the surgery approach can not be applied in all stages of ACC. In advanced stages of ACC, other treatments like chemotherapy and radiation therapy are considered. To maintain hormonal imbalances, hormone replacement therapy is often required. Moreover, the stages indicate how far the cancer has spread and allow doctors to decide the way to deal with it based on such things as tumor size and spreading. This manner gives predictive data that facilitates doctors guessing whether a person needs surgery or not at a certain stage and how likely it is that an affected person could survive. Staging additionally enables a look at clinical research, which leads to higher ways to deal with ACC. ACC is staged using the TNM (Tumor, Node, Metastasis) system with four ranges. In stage I, there is a tumor that is either smaller than or equal to 5cm and confined to the adrenal gland. The second stage involves a tumor larger than 5cm within the adrenal gland. Stage III indicates nearby extension, while stage IV implies either extensive invasion (T4) or the presence of distant metastasis (M1). So identifying stages is very important. Notable existing works for the detection and mitigation of ACC are elaborately discussed in the literature review section. However, the issue with the existing methodologies is their effectiveness, especially in defining the stages of ACC which would ease the way to select the proper therapy for certain stages. Hence, in-depth research is needed on enough samples of benign and malignant ACC to visualize its stages by efficient segmentation. Again, to segment and classify the tumor efficiently, Multi-Task learning can be a paramount option. It ensures effective computing which makes the excellent use of computer processes so that it will handle many connected jobs at the same time. It uses techniques like parameter sharing, task-specific modules, and joint optimization to enhance the speed of learning and lower the value of computing in systems that can be doing many tasks immediately. As we are working on both detection and mitigation, it needs a lot of parameters and needs to work independently in both cases if we intend to use current methods. Hence, an efficient hybrid Multi-Task learning model is needed indeed which can suit all the mentioned barriers.

1.2 Research Contribution

The research aims to accurately segment the cancerous regions in CT scan images of Adrenocortical Carcinoma and simultaneously classify them based on their respective classes. Our research goes beyond conventional segmentation and classification approaches by striving to develop a model with reduced parameters. This innovative approach enables the concurrent execution of both tasks within a more streamlined parameter space. Hence, the key contributions of our research can be outlined as follows:

1. We explored the state-of-the-art segmentation models (e.g. U-Net, Attention Res-UNet, DeepLabV3plus) on the first publicly available Adrenocortical Carcinoma dataset.

2. Our research attained superior segmentation results compared to existing models with proper choice of optimizer, loss function and learning rate.

3. We proposed a Multi-Task model that can perform segmentation and classification at the same time with resource and computational efficiency.

4. Presenting the classification outcomes of various state-of-the-art models (e.g. ResNet, DenseNet) and a custom model, we depicted their performance on the dataset through visual representation.

5. Employing Grad-CAM, we visually accentuated the detected regions of interest within an input image. This not only aids in the model's decision-making

process but also enhances interpretability that offers valuable insights into the neural network's decision pathways.

1.3 Thesis Structure

The structure of this paper follows a consecutive chapters approach, each dedicated to a specific aspect of the research. The following is an overview of the chapters:

- Existing literature relevant to the research, highlighting background studies and key theories in the current knowledge base are reviewed in Chapter 2.
- Methodology including dataset description, data pre-processing techniques, model architectures and experiments in detail are outlined in Chapter 3.
- Result and analysis of all the experiments including state-of-the-art models and proposed models are discussed in Chapter 4.
- Concluding remarks and insights into prospective future research are reflected in Chapter 5.

Chapter 2

Literature Review

2.1 Background Studies

U-Net architecture has two paths, a contacting patch, and an expensive path. The architecture of the contacting path is not different from a typical convolutional neural network (CNN) architecture. These architectures have four encoders and four decoder blocks. These encoder and decoder have a connection via a block. The encoder network half the spatial dimensions, also doubles the filter. On the other hand, a decoder network doubles the spatial dimensions and half the number of feature channels. A basic U-Net architecture is shown in Figure 2.1.

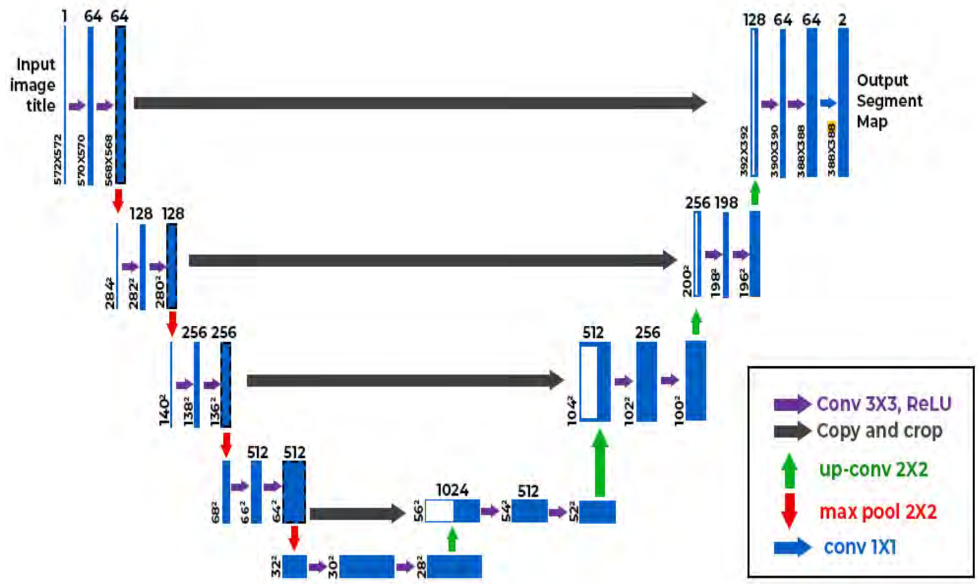


Figure 2.1: Basic U-Net Architecture [10]

The decoder has 13 layers similar to the encoder Each layer includes two 3x3 transposed convolutional layers. As we go deeper the number of filters decreases. The output layer is a convolutional layer with a single filter and sigmoid activation.
Res-UNet or Deep Residual UNet is an advanced encoder-decoder neural network architecture for semantic segmentation tasks. It has found applications in various domains such as brain tumors, human image, and polyp segmentation. It is an enhancement over the traditional UNet architecture that leverages both UNet and

Deep Residual Learning. Pre-activated residual blocks are the main innovation since they allow for the creation of a deeper network while addressing problems such as vanishing or ballooning gradients. Res-UNet comprises an encoding network, a decoding network, and a bridge connecting them. The encoder employs pre-activated residual blocks in three encoder blocks with abstract representation learning. The bridge connects the encoder and decoder. The decoder utilizes skip connections to improve the flow of information between layers. Also, it generates a segmentation mask through a series of decoder blocks, upsampling, concatenation with skip connections, and pre-activated residual blocks. It culminates in a 1×1 convolution with sigmoid activation for pixel-wise classification. The architecture's strengths include effective gradient flow, reduced parameter count, and superior performance in semantic segmentation tasks.
Attention Res-UNet, a combination of Res-UNet and Attention UNet has become a powerful fusion model for medical image segmentation. The system of this model incorporates both residual connections and Attention mechanisms. U-Net Architecture, Residual Blocks, and Attention Gates are the key components of this model. The architecture of Res-UNet is more complex compared to the traditional CNNs. Firstly, it contains a contracting path that downsamples input images through repeated convolution and pooling blocks. Each block contains convolutional layers for feature extraction, residual connections to bypass layers for better gradient flow, and Attention gates to selectively focus on informative features. Secondly, there is a bottleneck which is the narrowest part that connects contracting and expansive paths. Bottleneck processes high-level, semantic features. Thirdly, it contains an expansive path that upsamples feature maps through transposed convolutions. The expansive path also upsampled features with corresponding feature maps from the contracting path through skip connections. Attention gates refine feature maps at each stage. Finally, there is an output layer which contains a final convolutional layer with sigmoid activation for pixel-wise segmentation.
By using the strengths of U-Net, Residual Networks, and Attention Mechanisms the Attention Res-UNet model performs to do medical image segmentation. It uses and utilizes the U-Net for acquiring global and local features, uses residual connections to improve gradient flow, and enhances feature maps by placing Attention gates carefully. We can imagine that there is a U-shaped network where the information given flows downward and gets processed in a central bottleneck, and then by using Attention-guided precision, it is constructed again. Hence, segmentation accuracy is improved, making it a valuable tool for brain tumors, skin lesions, organs, cells, and also potentially many other image segmentation tasks outside the medical field.
ResNet50 encompasses a Residual Network with 50 layers in the context of the ImageNet Large Scale Visual Recognition Challenge. The fundamental innovation of ResNet50 lies in the incorporation of residual learning blocks or residual units which allow the network to efficiently train extremely deep architectures. These residual units mitigate the vanishing gradient problem by introducing shortcut connections through enable the direct flow of information across layers. The 50 layers are organized into several building blocks and each contains multiple residual units. The architecture is divided into a convolutional stage and followed by four identity blocks and a final fully connected layer for classification. In order to extract hierarchical features from input images, a sequence of convolutional and pooling layers is used at the convolutional stage. For classification, the global average pooling layer is

needed to form the final feature representation and also the identity blocks help to preserve the spatial resolution of the input. As the structure is deep and has residual connections, ResNet50 has become a very important basis in deep learning in image classification and various computer vision tasks.
DeepLabV3plus is a sophisticated neural network architecture. It helps in proper semantic segmentation in image processing tasks. The Google Research Brain Team's creation, with the help of an improved decoder module and a feature pyramid (FPN), expanded the usability of DeepLabV3. It collects multi-scale information by making use of atrous(dilated) convolutions with a customized backbone network. These are based on ResNet or as well as Xception. By tying together the many layers of the backbone network, FPN improves feature representation while Spatial Pyramid Pooling (SPP) is utilized to collect contextual data at multiple scales. The decoder module refines segmentation findings by merging high-level semantic information with low-level descriptive data by using skip links to help information flow between the encoder and decoder. The network is optimized during training by employing a suitable loss function such as cross-entropy. Pixel-wise segmentation masks are being generated by the final classification layer. The DeepLabV3plus due to its extensive design becomes an excellent tool for semantic segmentation tasks and also in various scales it can segment objects perfectly. The general architecture of DeepLabV3Plus is provided in Figure 2.2.

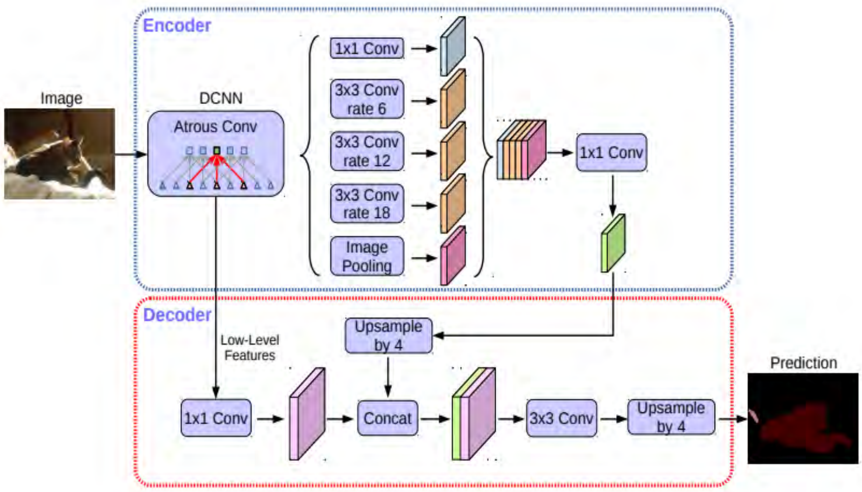


Figure 2.2: DeepLabV3plus Architecture [11]

Convolutional neural networks (CNNs) have problems with information flow and parameter efficiency. DenseNet is a deep learning architecture that aims to solve these problems. DenseNet's dense connectivity pattern is its primary innovation. DenseNet links every layer in a feedforward manner to every other layer, in contrast to conventional CNNs where each layer is only connected to the layer after it. In addition to allowing feature reuse and enhancing the network's capacity for recognizing complex patterns, this extensive connectivity promotes an extremely effective information flow across the system. DenseNet is divided into dense blocks, where every layer gives its own feature maps to every layer after it and receives feature maps from all layers before it in the block. Compared to conventional architectures, this design promotes feature reuse and makes it easier to train very deep networks with fewer parameters. Because of its better parameter efficiency, less vanishing gradient issue, and increased feature propagation, DenseNet is particularly useful for tasks like object detection and image classification. A basic DenseNet architecture is shown in Figure 2.3.

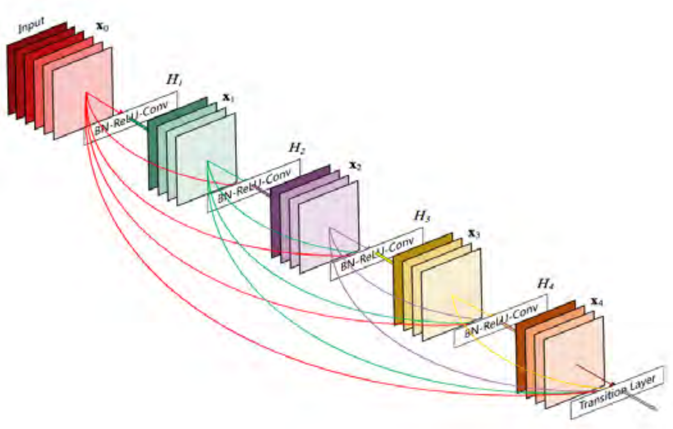


Figure 2.3: General DenseNet Architecture [12]

DenseNet-121 is a convolutional neural network architecture introduced in 2017 by Gao Huang, Zhuang Liu, and Laurens van der Maaten [12]. It belongs to the family of DenseNets and it is characterized by densely connected convolutional layers. The key innovation lies in the dense connectivity pattern. Each layer receives direct input from all preceding layers and helps feature reuse and enhance gradient flow. Comprising multiple dense blocks, DenseNet-121 promotes efficient learning by reducing the number of parameters. Transition blocks, interspersed between dense blocks include batch normalization, 1x1 convolution, and 2x2 average pooling layers for aiding in spatial dimension reduction. This architecture employs global average pooling instead of traditional fully connected layers for a fixed-size representation. Bottleneck layers with 1x1 convolutions contribute to computational efficiency. While the network may incorporate dropout layers for regularization, its 121 layers are specifically tailored for image classification tasks. DenseNet-121 has demonstrated competitive performance with fewer parameters compared to conventional architectures like ResNet. Hence, it is widely adopted in diverse computer vision applications for its efficiency and improved learning capabilities.
Multi-task learning (MTL) is an effective machine learning method that can handle not one but several associated jobs at the same time. It doesn't make separate models for each process; as an alternative, it makes use of the natural connections between them, which improves speed and economy. You may think about it as learning several skills without delay, in which one ability facilitates you to learn the others. In MTL, a neural network is made to do more than one thing by improving each task's objective feature at the same time. The tasks can be very similar and have a lot in common, or they can be very different and pressure the version to study robust and portable representations. Some of the high-quality things about MTL are that it can help make parameters more efficient, reduce overfitting, and use information from different jobs to make generalizations on new data better. Multitask learning (MTL) has many advantages that make machine learning models work better and more efficiently. One massive benefit is higher accuracy, due to the

fact the shared representation stores knowledge that can be utilized in different tasks, making each activity a higher standard. In addition, MTL improves data efficiency by way of needing less data for training than separate models. that is in particular beneficial when running with small datasets. Overfitting can manifest when the model focuses an excessive amount on the details of each job. MTL stops this from happening by acting as an herbal regularizer. Moreover, MTL's shared information makes transferring and gaining knowledge easier, which shall allow skills learned in one challenge to be used in exceptional tasks, making the model more flexible.

The ability of multitask learning (MTL) to deal with multiple tough tasks at once is shown by way of the wide range of areas in which it is used. In the discipline of Natural Language Processing (NLP), MTL is very useful for jobs like identifying how human beings feel about something, translating text automatically, and sorting texts into groups. It's very good at computer vision tasks like locating objects, recognizing images, and figuring out what a person's status or sitting is like. MTL is also very vital in Speech reputation, supporting such things as transcription, language recognition, etc. In addition to those it makes use of, MTL additionally has an effect on robotics, in which it allows problems like transferring on multiple legs, grabbing, and moving objects. Lots of specific uses for Multitask learning show how beneficial and flexible it is in a huge range of cutting-edge settings.

In the fields of computer vision and deep learning, Grad-CAM, or Gradient-weighted magnificence Activation Mapping, is a cutting-edge method that makes convolutional neural networks (CNNs) less difficult to understand. Due to the fact deep neural networks are quite a whole lot a black field, Grad-CAM was created to help us recognize how they make decisions with the aid of pointing out the parts of a picture that might be most critical for a sure classification. To use Grad-CAM, you need to locate the gradients of the goal class score in comparison to the feature maps of CNN's ultimate convolutional layer. Grad-CAM makes a grid that truly suggests which parts of the input picture had been maximum vital to the community's end choice by giving those gradients specific quantities of weight. This localization no longer only facilitates us to recognize how the version works, but it additionally builds agreement with and openness in AI systems, which may be very vital for medical pics, self-driving cars, and different high-stakes areas.

The purpose of the use of a Convolutional Neural network (CNN) interpretability method is to get a better sense of why the model makes predictions. that is performed by means of finding important functions inside the raw picture. In this step, the gradients of the goal class over the final neural feature maps are checked out. higher gradients suggest that the elegance has a larger impact on the version's estimate. The end result looks as if a grid on top of the original picture, with hot areas displaying that they're extra vital to the version's preference. This approach to interpretability has many blessings, including making the model's behavior extra clear and understandable, boosting trust in predictions, and opening up new approaches to repair and analyze errors. Still, it is vital to be aware of its flaws: it may not paint flawlessly for scenes with plenty of information or scenes with a variety of complicated scenes; it'd also be tough to compute as it might be high-priced.

GradCAM has been proven to be a flexible tool that may be used with many CNN designs without changing the shape of the network itself. GradCAM is useful for students and practitioners who need to understand how deep getting-to-know models work and make sure they're used responsibly as it is straightforward to apply, works nicely, and can supply targeted visible descriptions.

2.2 Related Works

In this section, we survey recent developments in Adrenocortical Carcinoma (ACC) research spotlighting the integration of radiomics and machine learning. Studies emphasize the potential of quantitative image analysis in discerning adrenal tumor malignancy, with genomic insights providing valuable predictive factors. Advanced imaging techniques and innovative therapeutic approaches such as microwave ablation, underscore the interdisciplinary efforts to enhance ACC diagnosis and treatment. The review also addresses machine learning applications for predicting Ki-67 expression providing existing limitations in each section provided below.
A study [13] explored the application of radiomic feature extraction and machine learning approaches to differentiate between indeterminate adrenal lesions observed in contrast-enhanced CT scans, distinguishing between benign and malignant cases. Analyzing 40 indeterminate adrenal lesions, the study extracted 3947 radiomic features, ultimately selecting 62 discriminative features through recursive feature elimination. A binary classification model is used and also with the cross-validation, the machine learning approach got the results with a specificity of 71.4%, sensitivity of 84.2%, and AUC of 0.85. Nevertheless, the study is limited by a small number of patients, which hinders the generalization of results and accuracy. Additionally, the study's retrospective nature is another limitation.

2.2.1 Adrenocortical Carcinoma Staging and Prognostic Evaluation

The researchers aimed to evaluate the performance of the 7th and 8th AJCC staging systems for adrenocortical carcinoma using SEER data from 2010-2015 in study [14]. Their findings indicated minimal differences in prognostic performance between the two systems, with C-statistics of 0.726 for the 7th version and 0.745 for the 8th version. They proposed subdividing stage IV into stages IVA and IVB based on the number of metastatic sites for future staging considerations. However, limitations included a small sample size and the absence of data on key prognostic factors such as resection status and hormonal secretion, emphasizing the need for further refinements in staging criteria.
A study [15] talks about Adrenocortical carcinoma (ACC) as the subject of this investigation, and additional driver genes (PRKAR1A, RPL22, TERF2, CCNE1, and NF) were found. It shows that severe disease progression is associated with whole-genome duplication (WGD) after DNA loss. Evidence for this comes from changes in telomeres and the initiation of cell-cycle events. A classification technique based on a DNA methylation signature comprised of 68 CpG probes was suggested. after the identification of 3 ACC sub-types with varied outcomes and molecular features. Possible data obsolescence, lack of functional validation, and a limited immediate impact on patient care due to incomplete biological understanding are just some of the limitations of the study on ACC genomic characterization.

2.2.2 Molecular Insights and Predictive Factors in Adrenocortical Cancer (ACC)

A research [16] talks about inaccurate prediction and therapy of adrenal cortical carcinoma (ACC) are hampered by intratumoral variability and inconsistent quantification methodologies, which prevents the identification of active regions (hotspots) and the measurement of Ki67 Labeling Index (LI). In digital pathology pictures, hotspot discovery and Ki67 LI measurement are automated using a free program called ASH. For quantitative hotspot ranking, it uses ImmunoRatio and NDPI format conversion. Quantitative values for ranking hotspots are included in the result. By automating hotspot discovery, an open-source program called ASH aids histopathologists in locating crucial regions in adrenocortical cancer, through a virtual Galaxy machine accessible. However, the limitation of this study is that this model may not give proper or higher accuracy in a large ACC set.

Researchers examined [17] 45 ACCs and discovered recurring changes in both newly discovered and prominent driver genes, such as ZNRF3 and DAXX. The possible tumor suppressor ZNRF3 was often changed. There are major molecular variations between aggressive and indolent ACCs, as shown by the identification of two ACC subgroups: C1A (bad result) with mutations and methylation alterations, and C1B (excellent prognosis) with microRNA cluster dysregulation. Along with noncoding RNAs and miRNAs in ACC, ZNRF3 looks to be a potential tumor suppressor gene connected to the catenin pathway and needs further research.

The ENSAT stage classification and pathology reports with the Weiss score and resection status are used to make prognosis and therapy options [18]. Complete tumor resection and mitotane are the cornerstones of treatment, with cytotoxic medicines added in for metastatic instances. Mitotane with EDP has been investigated as a potential first-line therapy in recent research. Due to the lack of success with targeted therapeutics like IGF-1 receptor inhibitors, new treatment methods are desperately needed; omic techniques and next-generation sequencing may provide some of these answers. This study's [19] overarching goal is to identify factors that predict whether or not adrenocortical carcinoma (ACC) will return after being surgically removed. An additional 250 participants from three European nations joined the 319 participants from the German ACC registry. Clinical, histological, and immunohistochemical markers are analyzed to determine their association with RFS and OS. Recurrence-free (RFS) and overall survival (OS) were shown to be significantly predicted by age, tumor size, venous tumor thrombus (VTT), and Ki67 in a multivariate analysis of ACC patients. The highest predictive value was found for Ki67. Clinical outcomes varied widely amongst patients with Ki67 values below, between, and beyond 10%. The predictive accuracy of combining tumor size, VTT, and Ki67 was not substantially better than that of Ki67 alone.

In study [20], the authors compared manual analysis (MA) as well as DIA of Ki-67 LI in ACC digital images. Intratumoral heterogeneity was observed, with higher Ki-67 LI in "hot spots." Weiss criterion was associated with Ki-67 LI and patient outcomes. MA in hot spots best reflects ACC features, while DIA needs refinement to avoid overestimation due to non-tumorous cell inclusion and nuclear segmentation. The study's shortcomings include the intratumoral heterogeneity of the Ki-67 labeling index (LI) in adrenocortical carcinoma (ACC), with greater values in "hot spots." Despite the value of Ki-67 LI, nuclear segmentation problems and the inclusion of non-tumor cells in DIA may cause it to be overestimated. The DIA has to be improved further. Curative therapy often consists of surgical excision in addition to mitotane [21]. Due to the low success rate and significant toxicity of systemic medicines, it is crucial to determine which patients would benefit from them. Compared to adrenocortical adenomas (ACAs), ACCs have a plethora of anomalies, making altered pathways including IGF, TP53, and Wnt signaling attractive therapeutic targets. However, only a subset of ACC patients benefit from newer medications, underscoring the critical need for new methods of therapy. Targeting IGF, TP53, and Wnt signaling pathways in the therapy of cancer has a number of drawbacks, including complexity, adverse effects on normal tissue, tumor heterogeneity, development of resistance, limited clinical effectiveness, side effects, high costs, and difficulties in patient selection.

2.2.3 Advanced Segmentation in Tumor and Therapeutic Approaches for Adrenal Gland

In a study [22], a combined deep-learning method has been used for Kidney and tumor segmentation. It actually separates kidneys and tumors for segmentation in CT scans accurately. They use two methods, one of them Faster R-CNN which is used to locate kidneys with tumors. This method helps to reduce computational resources by limiting the area. The second method is V-NET, which uses the extracted volumes taken from R-CNN and carefully examines it pixel by pixel, hence it uses its U-shaped architecture's capacity to capture the detailed outlines of kidney and tumor. However, Faster R-CNN cannot detect tumors that are small and that are hidden or they are outside the spotlight. On the other hand, V-NET depends on pre-cropped volumes which can result in the loss of information for tumors that are slightly beyond the identified zone. But, despite showing great results it has chances of improvement. In the future, they could limit the restrictions on tumor size and also they could reduce the loss of information when cropping. However, we do not need to do cropping if we use 3D segmentation on the whole CT scans directly. This could improve accuracy.
In another study [23], an innovative deep-learning multi-stage method has been explored for precisely segmenting kidney tumors in CT scans. It uses the ROI segmentation with Res-Unet which separates the regions that are cropped in both kidney and tumor segmentation. This method achieves high Dice scores of 0.96 for kidneys and 0.74 for tumors for unknown data. But, due to the complexity of the architecture, this method has weaknesses. For example, this model fails to understand some deep learning models and also cannot distinguish between benign cysts and malignancies. Therefore, the method can automatically segment kidney tumors which is its advantage. But, on the other hand, it fails to improve shortcomings and thus it needs to be more robust so that it can be applied to a wide range of cases.
A strong hybrid V-Net model that segments the kidneys and renal tumors in CT scans has been examined by a group of researchers [24]. This unique model surpasses V-Net architectures that go outside the V-net architectures by strengthening both the encoder and decoder phases. By making the feature extraction more sharper and also sharpening the information flow the encoder gets benefitted. It is done by residual connections and squeeze-and-excitation blocks. At the same time, the decoder precisely tunes the spatial localization and also preserves the crucial tumor

details. This hybrid V-Net model is very impressive as it outran most of the models with increasing dice coefficients for the same dataset for both kidneys and tumors. However, it does not show great results on other datasets. Furthermore, in future research, the computational efficiency can be improved further.
A two-stage cascaded deep neural network for accurate adrenal gland segmentation in abdominal CT scans was introduced by a group of researchers [25]. In the first stage, a localization network was used to find possible adrenal gland volumes, addressing class imbalance and computational load. In the second stage organ boundaries within the chosen volumes were refined using a Small-organNet model with a new border attention focal loss. Their method demonstrated superior accuracy compared to existing deep learning methods, achieving an accuracy rate of approximately 87.42%. However, it's worth noting that the method was not entirely end-to-end, involving two rounds of training, although they could be conducted in parallel. Additionally, while efficiently reducing GPU memory usage without compromising segmentation accuracy, further fine-tuning was necessary for specific organs and scan resolutions.
A research [26] discusses an approach to treat aldosterone-producing adenomas (APAs) using microwave ablation (MWA). To assess the ablation profiles of several microwave ablation applicators and parameters in different patient-specific scenarios, image-based computational models were used. The study suggests that microwave ablation has the potential to be an effective and minimally invasive treatment for adrenal tumors. 70% of the adrenal tumor volumes received an ablative thermal dose according to the authors' findings while limiting thermal damage to non-target structures and thermally sparing over 80% of the normal adrenal glands. However, the computational models used in the study are based on idealized assumptions about the tissue properties and the electromagnetic field distribution which actually varies remarkably from patient to patient in reality. The electromagnetic field distribution can be affected by other factors as well (e.g. the presence of implants or metal objects). In addition, the clinical outcomes of microwave ablation for adrenal tumors were not evaluated by the study. Hence, it is important to conduct further clinical studies to validate the findings including their safety.
A study [27] proposed a new automatic kidney segmentation approach in Autosomal Dominant Polycystic Kidney Disease patients. An end-to-end trained convolutional neural network was used to develop the method. The convolutional neural network is trained on a dataset of 244 CT acquisitions of patients exhibiting a variety of stages of Autosomal Dominant Polycystic Kidney Disease. The Dice Similarity Coefficient (DSC) between CNN kidney segmentations and ground truth kidney segmentations from clinical specialists was overall 0.86 ± 0.07 (mean±SD) when the suggested method was examined by the authors on a test set of 79 CT acquisitions. With segmented kidney volume measurements across the full test set, the approach likewise showed good agreement with the genuine TKV, with a concordance correlation coefficient of 0.98 (p¡0.001). However, the proposed method was trained and examined on a relatively small dataset of 244 CT acquisitions and hence the method may not generalize well to new data, especially if the data originates from different scanners or contains different image quality. The model did not take into account any other clinical information (e.g. the patient's age, gender, or medical history) that could potentially improve the method's performance.
A machine-learning approach is suggested by research [28] based on radiomic char-

acteristics taken from contrast-enhanced CT images to predict the Ki-67 index in adrenocortical carcinoma. The texture, shape, and size of the tumor are all described by the radiomic features, which are quantitative features. 53 patients with adrenocortical carcinoma were included in the dataset used by the paper's authors to train the model. The patients' pre-resection contrast-enhanced CT scans were used to extract the radiomic characteristics. The histological analysis of the resected tissue included calculating the Ki-67 index. The researchers demonstrated that the model has an accuracy of 80% in predicting that index. In addition, the model also has a sensitivity of 83% and a specificity of 75% for detecting patients with elevated Ki-67 expression. The model might be used to assist clinicians in choosing the best course of action for ACC patients. However, the model was trained on a single dataset and it would have been better to have more samples for training.
A new approach for automatic kidney segmentation in ultrasound images was proposed [29] where they used boundary distance regression and pixel-wise classification networks. At first, for each pixel in the images, the method learns to predict the distance to the kidney boundary. A pre-trained deep neural network on a large dataset of natural images was used to do this. Here the pre-training helps the network architecture to learn general features that are useful for segmenting images. Next, the method learns to classify whether each pixel in the image is kidney or non-kidney. A deep neural network that is trained on the kidney boundary distance maps predicted in the initial step was used to assess this classification. However, the proposed method has several limitations. Firstly, The authors used a dataset of 289 ultrasound images although it requires a large amount of training data. This is a relatively small dataset for training deep learning models which may overfit the training data and not generalize well to new data. Moreover, it is not robust to artifacts in ultrasound images as ultrasound images are often noisy and contain artifacts such as shadows and speckle noise. It is difficult for the model to accurately predict the kidney boundary distance maps and classify pixels as either kidney or non-kidney because of these artifacts.
In a study, the researchers proposed prognostic models for adrenocortical carcinoma using a machine learning approach [30]. The authors used 825 ACC patient's data and the Backpropagation Artificial Neural network, the random forest, and the support vector machine were trained. The models were trained to forecast the overall survival of ACC patients. The area under the curve (AUC) for predicting overall survival was 0.921. The Backpropagation Artificial Neural network performed the best where AUC scores for Random Forest and Support Vector Machine were 0.885 and 0.86 respectively. The authors came to the conclusion in their analysis that prognostic models for ACC could be created using machine-learning approaches. The study suggests a machine learning approach to predict the Ki-67 index in adrenocortical carcinoma based on radiomic characteristics taken from contrast-enhanced CT images. The quantitative features like texture, shape, and size of the tumor are all described by the radiomic features. However, the study was limited to a single dataset and a large number of samples would have been better.
In a study [31], the effect of different parameters in deep learning has been discussed where the authors try to deal with the challenge of precise tumor delineation in T1-weighted MR images using a customized U-Net. The method is most effective when parameters and preprocessing methods like batch sizes, DWT, CLAHE, image fusion, and k-fold values are optimized in order to maximize accuracy. The model

achieved a Dice score of 0.631 for DWT-preprocessed T1-weighted images which outperforms the original U-Net. However, the usage of multi-modality could improve accuracy more instead of using only T1-weighted data. In addition, sensitivity and specificity along with the standard Dice score would provide a more nuanced view. Moreover, comparison with other deep learning architectures would solidify the model's validity for a wider context.
A study proposed [32] a novel encoder-decoder network with a Transformer module for accurate 3D segmentation of adrenal tumors in CT scans. This method leverages Deep Separable Convolutions for efficiency and employs the Dice Score Coefficient loss function to handle class imbalance. It has achieved state-of-the-art performance on a small dataset and exceeded comparable algorithms in metrics like DSC and IOU. However, it has exhibited some outliers for complex cases with small lesions. Despite this limitation and the need for broader validation, the integration of Transformers and the promising initial results suggest potential for improved diagnosis and treatment in adrenal tumor patients. Further research with larger datasets and comparisons to dedicated small organ segmentation techniques could solidify the method's effectiveness and generalizability.

2.2.4 Multi-Task Learning for Enhanced Medical Image Segmentation and Classification

In a study [33], Multi-Task Attention-Based Semi-Supervised Learning (MASSL) for medical image segmentation has been discussed where Attention mechanisms integrate with Multi-Task learning. This method leverages the vast trove of unlabeled data which potentially outperforms traditional baselines. One of the key and secret weapons of this approach is that it contains an Attention-guided reconstruction task that cultivates highly discriminative feature representations within the deeper layers of the encoder. However, a few major areas remain ripe for exploration to amplify further its potential. Firstly, examining the Attention mechanism's complex inner workings would provide a more clear view of its function in feature extraction. Secondly, a thorough head-to-head comparison with other semi-supervised techniques such as graph-based methods would be required to firmly establish MASSL's distinct advantages and disadvantages. Finally, visualizing the learned features would provide solid evidence of the transforming potential of the Attention process. These refinements would enhance MASSL's contribution and pave the way for future advancements in medical image segmentation.
In another study [34], Multi-Task deep learning for image segmentation has been examined using recursive approximation tasks. This approach deals with the resource-intensive dilemma of pixel-level annotations in image segmentation with an ingenious Multi-Task framework. The problem is divided here into a sequence of tasks requiring less demanding labels to address the cost constraint. The core component of this method is the recursive approximation task, which gradually gets closer to correct object areas by iteratively improving user-provided, frequently rough partial masks. To form a collaborative learning loop, this task interacts with a traditional segmentation task trained on pixel-level labels. This mutual learning between the approximation job and the segmentation task, which simultaneously directs the latter's comprehension of complex object bounds, is the key to the framework's efficiency. To enhance the learning further, optional supplementary tasks that concentrate on

particular visual elements can be added. This framework significantly reduces annotation cost by utilizing easily accessible coarse labels including its flexible support for a wide range of initial labels and uncertainties.
A different study has introduced an advanced Multi-Task learning framework designed for the segmentation of brain tumors, specifically addressing segmentation challenges in MRI scans [35]. In another investigation, a deep Multi-Task learning framework was created for the segmentation of brain tumors, focusing on overcoming segmentation challenges in MRI. The strength of this study is to design a parallel decoder where a distance transform decoder smooths out rough borders and improves mask prediction, which is a common flaw in conventional methods. Using a unique loss function that combines category-focused loss and mean squared error, class imbalance could be resolved further, by concentrating on the BraTS datasets. Although this idea seems promising, testing on a variety of datasets should improve its generalizability. Better results could be acquired by examining additional weighting mechanisms inside the loss function to address class imbalance.
In another study [36], a novel Multi-Task learning approach has been proposed named Cerberus for simultaneous histology image segmentation and classification which addresses shortcomings of traditional single-task deep learning in computational pathology. The method leverages shared features between tasks which improves efficiency and reduces data hunger. This single-model technique simplifies training and inference compared to independent models and the large training dataset guarantees robust performance. To ensure the generalizability across tasks of this model, the authors initiate transfer learning for nuclear classification and signet ring cell detection additionally. However, the black-box nature of deep learning persists which limits interpretability despite feature sharing. Furthermore, the dependence on large datasets may constrain applicability to resource-limited settings which should be addressed in future work.
A RMTL (Regional-Attentive Multi-Task Learning) framework has been discussed in research [37] for breast ultrasound image segmentation and classification. This technology with the help of ultrasound images helped doctors to analyze breast cancer. It performs classification and segmentation together resulting in a Multi-Task learning model. This model helps to find the exact location of the tumor and also tells us whether it is malignant or benign. It's interesting because it uses a regional attention (RA) module to focus on tumor, peritumoral, and background regions based on maps that are made by probability. However, it does not work with larger datasets. It requires a wide range of images to function effectively in a variety of scenarios. Furthermore, it focuses on whether the tumor is terrible or not, but it ignores distinctions within each group. It also takes time and requires more computer power to complete and that is why it is not suitable in real-life scenarios. But still, RMTL-Net helps to detect breast cancer earlier and more reliably. The tests need to be done in more places so that they can become useful in real life.

Chapter 3

Methodology

3.1 Work Plan

In Figure 3.1, the step-by-step approximate processes that we have followed to conduct our research are presented.

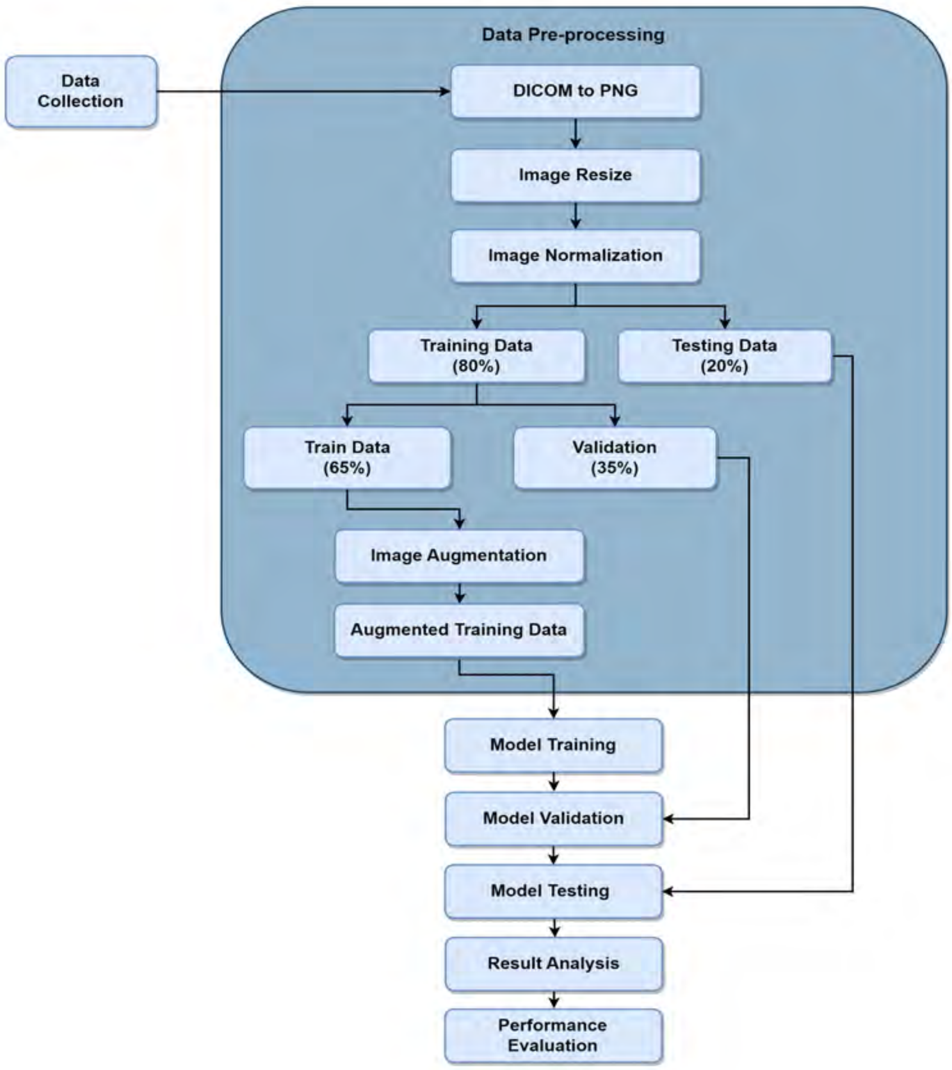


Figure 3.1: Workflow Diagram

The work plan for our research follows a systematic workflow outlined in the provided workflow diagram. Beginning with the collection of DICOM-format medical images of Adrenocortical Carcinoma (ACC), the data undergoes conversion to PNG format for compatibility with deep learning models. Subsequent preprocessing steps involve resizing and normalization of pixel values to ensure standardized inputs. The dataset is then divided into Training Data (80%) and Testing Data (20%). The Training Data (80%) is divided into further subdivisions for Training (65%) and Validation (35%). Augmentation techniques are applied to the training data to enhance model robustness. The proposed hybrid Multi-task model is then trained using the augmented data and validated on the Validation (35%) set. Testing is performed on the Testing Data (20%) to assess overall accuracy. Result analysis (discussed in Chapter 4) includes an in-depth examination of the model's segmentation and classification outcomes followed by a comprehensive performance analysis considering several metrics. This structured approach ensures a rigorous evaluation of the deep learning-based model's efficacy in addressing the challenges posed by ACC and contributes to accurate segmentation and classification for improved diagnosis and treatment planning.

3.2 Dataset Description

In this research, we have collected the Adrenocortical Carcinoma dataset from the Cancer Imaging Archive [38]. This is the first publicly-available library for adrenal lesions. The dataset consists of a cohort of 52 patients, with each patient contributing both DICOM-format CT (Computed Tomography) images and its associated segmentation DICOM images. The data consists of information from 2006 to 2018 and includes comprehensive clinical and pathological information. All participants met specific inclusion criteria, encompassing a histopathological confirmation of ACC, and surgical resection of the tumor as part of the standard evaluation. The segmentation of DICOM images explains distinct regions of interest (ROIs). Each ROI is outlined through binary annotations, where a pixel value of 1 means the presence of an ROI, on the other hand, 0 indicates its absence. All the images, including CT and segmentation DICOM images, were maintained in the DICOM format, ensuring adherence to industry standards for medical imaging. The Demographic Distribution of the ACC Dataset is available in Table 3.1.

Parameter	**Statistic**
Gender - Male	22
Gender - Female	31
Age (Min/Max)	22/82
Mean	53.00
Median	54
Mode	56
SD	13.47421
Data Collection Period Years	2006–2018

Table 3.1: The Demographic Distribution of the ACC Dataset

Domain experts meticulously reviewed each CT image to identify significant anatomical ROIs that warrant segmentation. They have already annotated manually delineated the ROIs within the segmentation DICOM images. These annotations were associated with binary values. The value 1 signifies ROI presence and 0 for absence. However, despite the significance of this information, the primary dataset presented a challenge for deep learning-based analysis due to its inherent structure in a purely medical format. We embarked on a comprehensive process to transform and augment the dataset. Through meticulous steps and careful considerations, we have successfully adapted the data to a format conducive to deep learning methodologies. This strategic modification ensures that the dataset is now not only amenable to sophisticated analysis but also aligns seamlessly with the intricacies of deep learning algorithms. Still, this primary data was not suitable for any deep learning-based analysis since it was in purely medical format. That is why we have taken the necessary steps to make this dataset suitable for analysis.

3.3 Careful Data Splitting

We can not apply direct random split for our task. There can be situations when some data of a patient may go to training and some data may go to testing or validation. It will create bias. Since the model will see the data during training it can create bias. That is why we have made sure during data splitting the same patient's data does not go to multiple sections.

3.4 Data Pre-Processing

Data pre-processing is a crucial phase in the data analysis and deep learning pipeline, where raw data is transformed, cleaned, and organized to enhance its quality and usability. This preparatory step is essential to ensure that the data is suitable for our analysis or model training, as raw datasets often contain noise, inconsistencies, and missing values. In the next subsections, we will discuss how the Adrenocortical Carcinoma Dataset is pre-processed sequentially in our research.

3.4.1 Selection Criteria

For our analysis, only segmentation DICOM images that contain at least one annotated ROI (pixel value of 1) were considered. Corresponding CT images aligned with these positive segmentation images were retained for contextual correlation.

3.4.2 Conversion to PNG Format

Both CT and segmentation images were used for the conversion process of converting from DICOM to the widely used PNG format. We have converted these images using the Pydicom library [39], designed for manipulating and converting DICOM images.

3.4.3 Correlation Preservation

We have considered the preservation of spatial correlation between CT images and corresponding segmentation annotations. This process helped us to ensure accurate alignment during our segmentation.

3.4.4 Conversion Process with Pydicom

Using Pydicom we have extracted pixel data and metadata from DICOM images. Extracted data was utilized to generate PNG images, retaining important metadata that is essential for clinical context. This is to make sure we can extract radiomic features later. An example of CT image and its corresponding aligned Mask is shown in Figure 3.2.

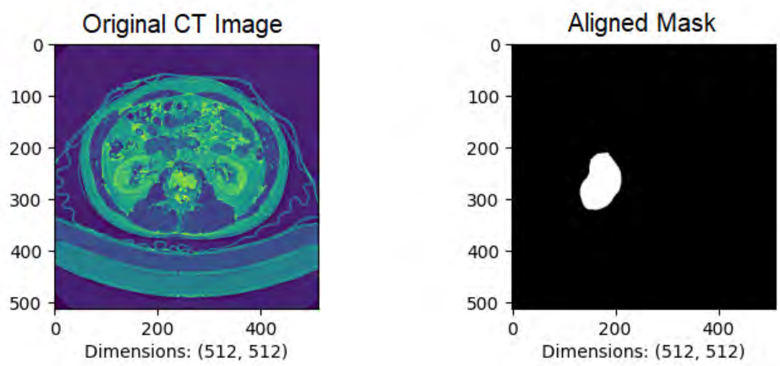


Figure 3.2: CT and Mask after converting to PNG

3.4.5 Resizing

The resizing process involves two steps. The dimensions of both the CT images and their corresponding masks were 512x512. Both the CT images and the masks were resized to (224, 224) pixels while preserving their aspect ratios. Later the CT images were converted to 3 channels in order to make them compatible for the deep learning models.

3.4.6 Normalization

During the normalization process, we rescaled the pixel's intensity value within the range [0, 1]. Each pixel's value was subtracted from the minimum pixel and then divided by the range. This process ensures that all images and masks have the same consistency scale for better analysis.

3.4.7 Data Augmentation

The Data Augmentation of our dataset is divided into the given subtasks:

Horizontal Flip

Horizontal flip means flipping the image horizontally, like looking at it in a mirror. This process mirrors the image from left to right. It can be useful for segmentation since the orientation of objects doesn't affect the outcome. We can see a Horizontally Flipped CT image and its Mask in Figure 3.3.

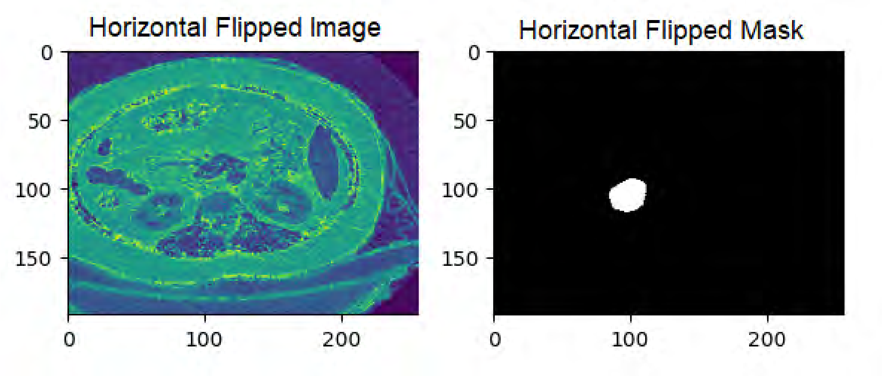


Figure 3.3: Horizontally Flipped CT Image and Mask

Vertical Flip

Vertical flip means flipping the image vertically, like turning it upside down. This process mirrors the image from top to bottom. Similar to horizontal flip, vertical flip can also be useful in segmentation tasks. A Vertically Flipped CT image and its corresponding Flipped Mask is shown in Figure 3.4.

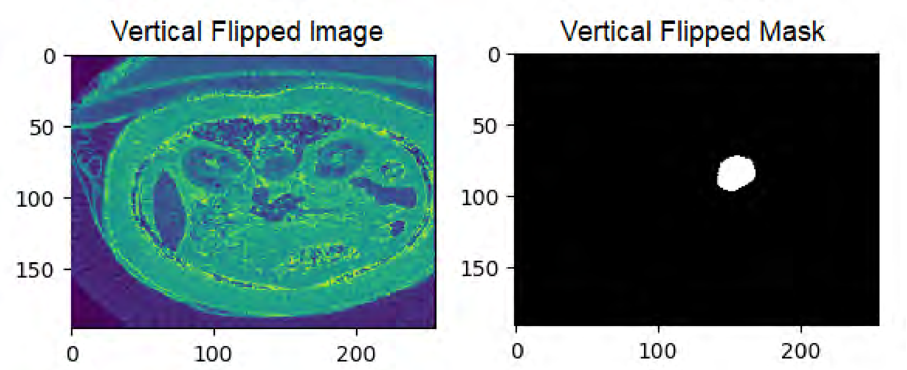


Figure 3.4: Vertically Flipped CT Image and Mask

Both (Horizontal and Vertical) Flip

This augmentation combines both horizontal and vertical flips, effectively rotating the image by 180 degrees. As a result, we get a more extensive set of transformations. These transformations can be helpful for training models that need to be robust to various orientations. Both Horizontally and Vertically Flipped CT Image and Mask example is shown in Figure 3.5.

Random Rotation

Random rotation involves rotating the image at a random angle. It's often used to make the model more robust to variations in the orientation of objects in the images. The rotation angle is typically chosen randomly within a specified range. For our analysis, we have chosen -60 degrees to +60 degrees. An example of randomly rotated CT Image and Mask is provided in Figure 3.6.

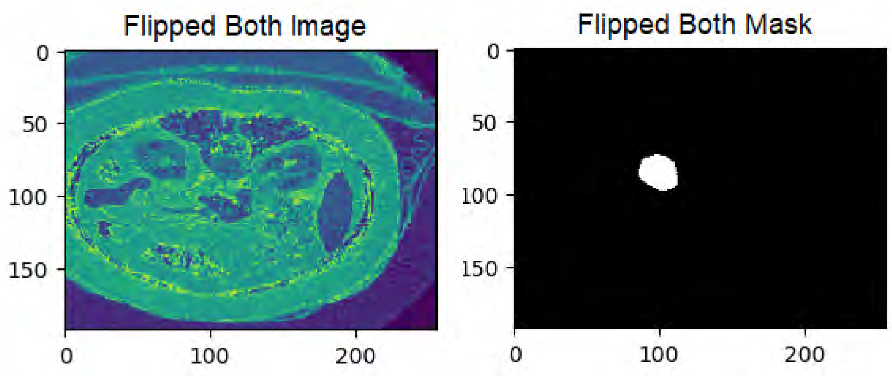


Figure 3.5: Both Horizontally and Vertically Flipped CT Image and Mask

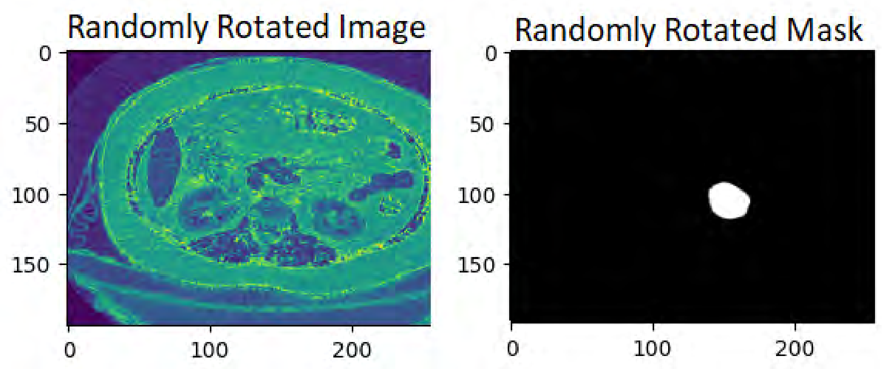


Figure 3.6: Randomly Rotated CT Image and Mask

All these steps are applied to both CT image and their corresponding masks in order to ensure consistency. After augmentation, we get more data which makes the deep-learning models more suitable.

3.5 Experimental Setup of Segmentation Model

In this section, we will explain the necessary steps and configuration of the segmentation models. We have used U-Net, ResAttentionUNet, and DeepLabV3plus for segmentation. Adam optimizer and a learning rate of 0.003 are used. We have chosen dice loss as our loss function. Our choice for dice loss is because of class imbalance. Compared to T2 and T3, the number of images in the T1 and T4 class are less. It might affect the segmentation. Another reason we use dice loss is because we are focusing on the dice coefficient as one of the metrics. First, the segmentation is predicted by the model and the dice coefficient is calculated. We are using a batch size of 8 for training all the models. Moreover, we have set the epoch to 400 and also added two early stopping and reduced learning rate callbacks. Early stopping focuses on validation loss. We have set the factor to 5 for early stopping which means when the validation loss does not improve for 5 consecutive epochs the training process stops. Reducing the learning rate is to train the model a bit longer to see if the model's performance improves at a different learning rate. We have set the factor to 5 for reduced learning rate too.

3.6 Experimental Setup of Classification Model

3.6.1 Division to 2 groups

Our dataset consists 4 stages of ACC - T1, T2, T3, T4 where T1, T2 are low and T3, T4 are high severity. We have grouped these 4 classes into 2 groups. One reason for combining 2 groups together is that there are less number of images in T1 and T4 than in T2 and T3. Since T1 images are closer to T2 and T3 images are closer to T4 we have grouped them together. So there are basically 2 groups. Class0 - (t1, t2 stages), Class1 - (t3, t4 stages).

3.6.2 Basic CNN Architecture

In this section, we are going to present a CNN architecture It starts with 4 ConvPool blocks. The CNN architecture starts with 4 ConvPool blocks. Then we added a Flatten layer. After that, we have 4 dense layers which have 1024, 512, 256, and 128 neurons respectively. In the final layer, we have the Sigmoid function with 1 neuron. A detailed architecture of the model is presented in Figure 3.7.

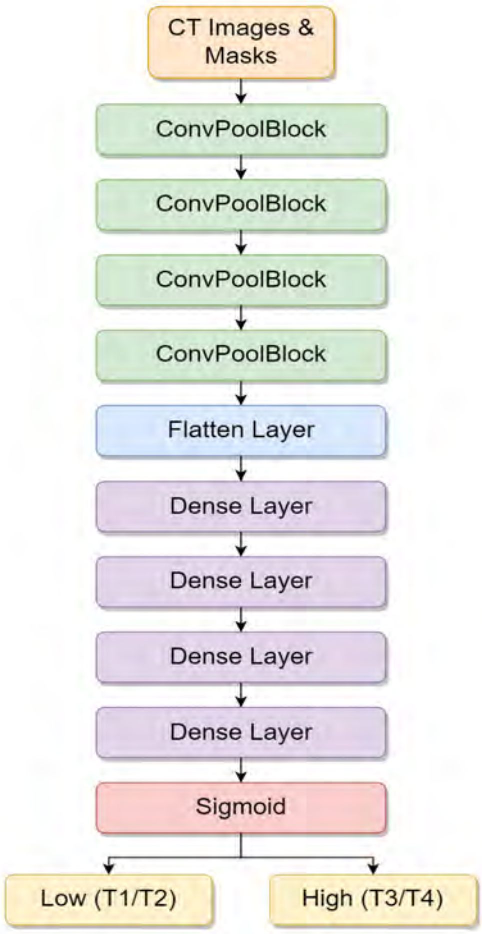


Figure 3.7: Basic CNN Architecture Flow

3.6.3 Transfer Learning in Classification Task - DenseNet121 And DenseNet201

We have used DenseNet121 and DenseNet201 as feature extractors. We have fed our (224, 224, 3) sized images into the pre-trained model. We have frozen all the trainable layers. After that, we added a 2*2 MaxPool layer followed by Flatten layer. Then we added a Dropout layer of 50 percent to prevent over-fitting. Then there is a fully connected layer of 1024 neurons. Later we added a Dense layer of 1 neuron with a sigmoid activation function. The goal of using the Sigmoid function is to produce output between 0 and 1s since we have 2 classes. The architecture is shown in Figure 3.8.

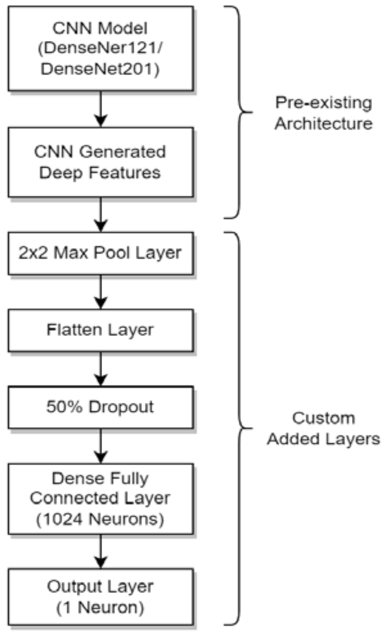


Figure 3.8: Transfer Learning in Classification Task

Hence, leveraging DenseNet121 and DenseNet201 as feature extractors, the frozen layers, augmented with a MaxPool layer, Dropout, and a fully connected layer, culminate in an effective image classification model with a final Dense layer enabling precise 2-class predictions.

3.6.4 Custom Model

We have proposed a custom model that basically uses the feature extraction part of the DeepLabV3plus model as the feature extractor.
Since our goal is to propose a Multi-Task Learning model and the DeepLabV3plus model performed better than the other models in the segmentation task we have used the feature extraction part of DeepLabv3plus. One of the reasons for this is it uses multi-level feature fusion for the feature extraction process. After that, we added a MaxPooling layer followed by a Flatten layer. Then there is a Dropout of 0.5. Then we added two dense layers of 2048 and 1024 neurons for learning the patterns. It is followed by a dropout layer of 50 percent. The goal of using the dropout layer was

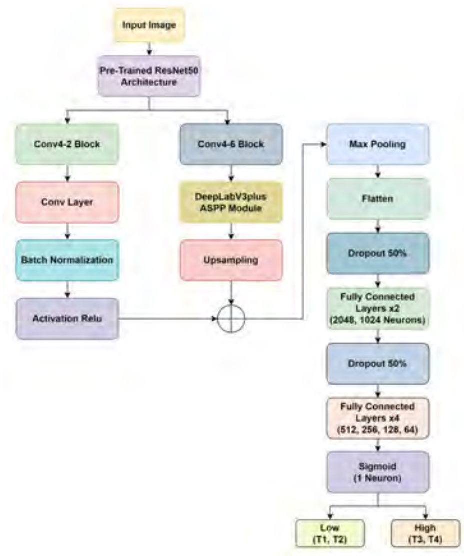


Figure 3.9: Custom Model Architecture

to make sure the model was not overfitting. Then there are 4 fully connected layers of 512, 256, 128, and 64 neurons. Since it is a binary classification we have used only 1 neuron in the Dense layer and it is using the Sigmoid activation function. The architecture is as follows in Figure 3.9.

3.7 Multi-Task Learning

Multi-task learning (MTL) involves training a model to perform multiple tasks simultaneously. MTL can lead to shared representations, where features learned for one task may benefit the other. Training on multiple tasks can act as regularization, improving the model's generalization. Instead of training separate models for segmentation and classification, MTL allows joint training, potentially reducing overall training time. Multi-task learning on edge computing is an emerging research area that aims to leverage the benefits of both techniques.

In our research, we have proposed a Multi-task learning system that completes segmentation and classification at the same time. The goal of this process is efficient computing. Usually, powerful algorithms take many parameters. As a result, performing many tasks one by one can be time-consuming along with the use of too many resources. When a low resource is available main goal usually becomes performing the required tasks with the limited resources even with the slightest performance trade-off. We know for any segmentation model first part consists of feature extraction. On the other hand feature extraction is the most important part of image classification. So this can be a common feature in both tasks. Usually, in a multitask system, multiple tasks share something in common. In this case, it is feature extraction. Now, coming to the feature extraction part, it creates a concern about which model should we use for feature extraction. Since we are performing both tasks at the same time it is very common to get a little bit less value than the actual segmentation model. Since, in our analysis, DeepLabv3Plus performed better than all the other models (Discussed in Chapter 4), we have used it for our multi-task learning. A graphical representation of the Proposed Multi-task learning approach is presented in Figure 3.10.

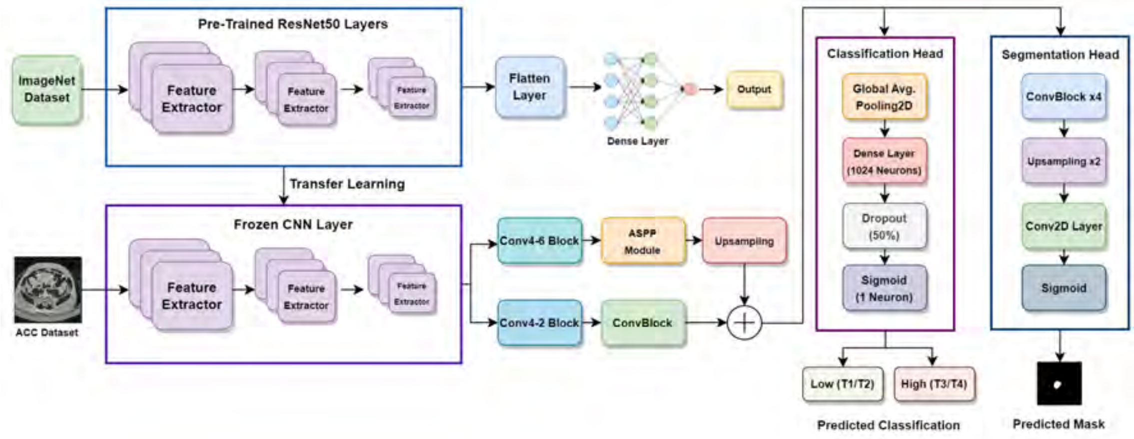


Figure 3.10: Graphical Representation of The Proposed Multi-Task Learning Approach

We have used ResNet50 as the backbone architecture. ResNet50 is a pre-trained architecture that uses transfer learning for image feature extraction from our images. We have chosen ResNet50's backbone because the custom model with ResNet50 performed better in the classification task than any other model. So building our Multi-Task learning model we have considered both segmentation and classification performance in order to reduce the loss in accuracy as less as possible. Now, ResNet's conv4 block6's output is passed to the ASPP module. A detailed description of the ASPP module is given in Figure 3.11.

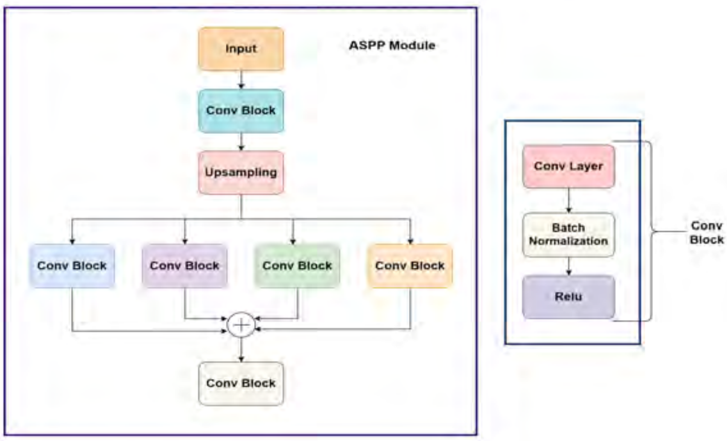


Figure 3.11: ASPP Module

The output of the ASPP module is passed to an upsampling layer. On the other hand, ResNet50 architecture's Conv4block2 which is more of a lower-level feature extractor is passed to a conv block. One part consists classification head which consists of a GlobalAveragePooling2D, Dropout, Flatten, Fully connected layer, and a final classification layer. Another connected network is for segmentation. Later it goes through 4 Conv blocks and 2 Upsampling blocks. The architecture of the Conv block and Upsampling block is presented in Figure 3.10 respectively. If we run independently for the segmentation task we need an amount of trainable parameters and for classification, we need the amount of trainable parameters. So we will have the amount of trainable parameters if we run them independently. On the other hand, through multi-task learning, we will need only amount of parameters. So it saves lots of computing resources. Now coming to the experimental setup for this task we have used Adam optimizer and a learning rate of 0.003. Since it is a multi-task algorithm we have used 2 loss function. We have used binary cross-entropy for classification and dice loss for segmentation. The evaluation metrics for each task for each task is still the same. But now comes a concern is how we are going to use early stopping. Previously we have used early stopping of patience 5 but now we have 2 loss function. One for segmentation and another for classification. So, the question arises of which cost function we should focus on. So we have implemented a Custom Early Stopping. Pseudo-code for the Custom Early Stopping algorithm is given in Algorithm 1.

Algorithm 1 Custom Early Stopping

```
procedure CustomEarlyStopping(patience)
    Input: patience - Number of epochs with no improvement to wait before stopping
    Output: Stops training if no improvement in the validation loss and classification loss

    Initialization:
    best_v_loss ← ∞
    best_map10 ← ∞
    best_weights ← None
    patience_count ← 0

    for each epoch do
        v_loss ← segmentation validation loss at the end of the epoch
        map10 ← classification validation loss at the end of the epoch
        if v_loss is not None and map10 is not None then
            if v_loss < best_v_loss and map10 < best_map10 then
                best_v_loss ← v_loss
                best_map10 ← map10
                best_weights ← current model weights
                patience_count ← 0                    ▷ Reset patience count
            else
                patience_count ← patience_count + 1
            end if
            if patience_count > patience then
                Stop training
                print("Restoring model weights from the end of the best epoch.")
                Set model weights to best_weights
            end if
        end if
    end for
end procedure
```

Chapter 4

Result Analysis

4.1 Segmentation and Classification Metrics

Dice Coefficient

The Dice Coefficient is a metric used to measure the similarity between two sets, usually a predicted set and a ground truth set. It is commonly used in tasks like image segmentation. The Dice Coefficient is calculated using Equation 4.1.

$$\text{Dice Coefficient} = \frac{2 \times \text{Intersection of Predicted and Ground Truth}}{\text{Total Predicted} + \text{Total Ground Truth}} \tag{4.1}$$

The Dice Coefficient ranges from 0 (no overlap between predicted and ground truth) to 1 (perfect overlap).

Precision

Precision is a metric that measures the accuracy of positive predictions made by a model. It is commonly used in binary classification tasks. The Precision is calculated using Equation 4.2.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{4.2}$$

Precision focuses on the fraction of relevant instances among the predicted positive instances.

Recall (Sensitivity or True Positive Rate)

Recall is a metric that measures the ability of a model to identify all relevant instances. It is also commonly used in binary classification tasks. The Recall (Sensitivity) is calculated using Equation 4.3.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4.3}$$

Recall focuses on the fraction of relevant instances that have been successfully retrieved by the model.

Intersection over Union (IoU)

Intersection over Union is a metric used for evaluating the accuracy of object detection and image segmentation tasks. It measures the overlap between the predicted and ground truth regions. The Intersection over Union (IoU) is calculated using Equation 4.4.

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \tag{4.4}$$

IoU values usually range from 0 (no overlap) to 1 (perfect overlap). A commonly used threshold for considering a prediction as correct is IoU ≥ 0.5.

4.1.1 Comparative Result Analysis

The comparative analysis of the three segmentation models in the Training, Validation, and Testing Dataset is shown in Table 4.1, Table 4.2, Table 4.3 respectively.

Model	**Dice Coefficient**	**IoU**	**Precision**	**Recall**
Attention Res-UNet	0.7274	0.9655	0.7125	0.8104
U-Net	7130	0.9597	0.8395	0.7811
DeepLabV3Plus	0.8828	0.9876	9001	0.8827

Table 4.1: Performance Metrics on Training Data

Model	**Dice Coefficient**	**IoU**	**Precision**	**Recall**
Attention Res-UNet	0.8046	0.9830	0.8262	0.7916
U-Net	0.8053	0.9833	0.8395	0.7811
DeepLabV3Plus	0.8946	0.9903	0.9087	0.8832

Table 4.2: Performance Metrics on Test Data

Model	**Dice Coefficient**	**IoU**	**Precision**	**Recall**
Attention Res-UNet	0.8192	0.9128	0.8362	0.8109
U-Net	0.8156	0.9826	0.8358	0.8050
DeepLabV3plus	0.8940	0.9896	0.9027	0.8867

Table 4.3: Performance Metrics on Validation Data

In the Training Data, Attention Res-UNet exhibits a Dice Coefficient of 0.7274, IoU of 0.9655, Precision of 0.7125, and Recall of 0.8104. U-Net and DeepLabV3Plus show their own metrics. In the Testing Data, Attention Res-UNet and U-Net display almost similar performance, while DeepLabV3Plus outperforms with higher metrics. Validation Data results are consistent with Testing Data trends. These metrics, including Dice Coefficient, IoU, Precision, and Recall, offer insights into the models' effectiveness in image segmentation. Higher values indicate better performance in capturing spatial overlap and accuracy of positive predictions.

4.2 Mask Generation of the Segmentation models

In this section, we have presented some samples of the generated mask along with the CT image with the corresponding mask. Mask generation is shown in Figure 4.1, Figure 4.2, and Figure 4.3 for U-Net, DeepLabV3plus, and Attention Res-UNet respectively.

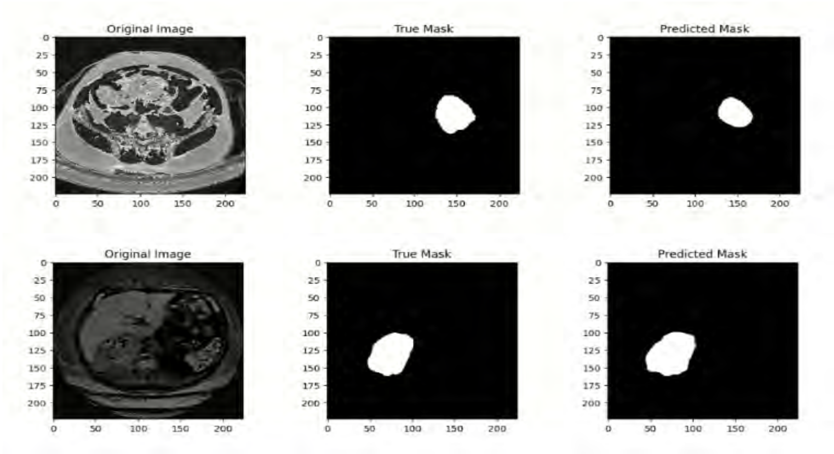


Figure 4.1: True Mask Vs Predicted Mask for U-Net

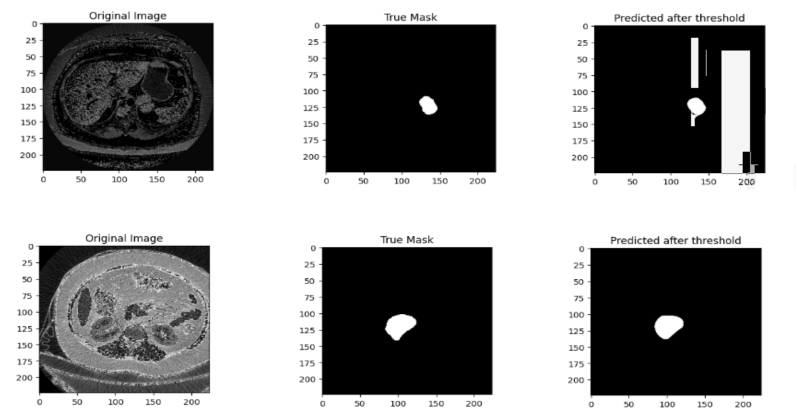


Figure 4.2: True Mask Vs Predicted Mask for DeepLabV3plus

Here we can see DeepLabV3plus and Attention Res-UNet models gave a better prediction than the original U-Net model.

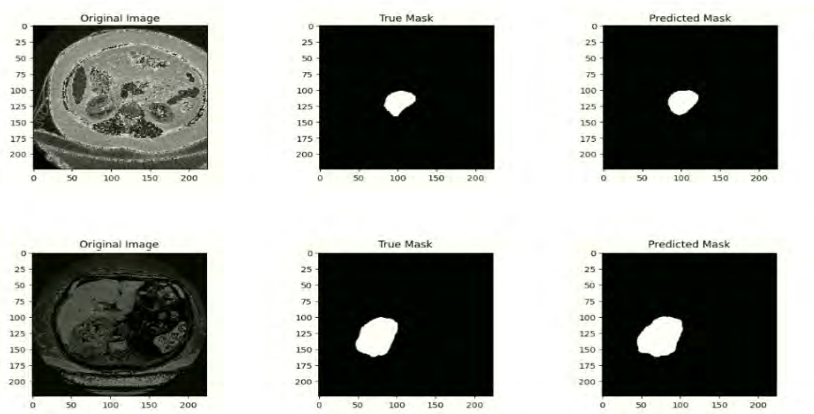


Figure 4.3: True Mask Vs Predicted Mask for Attention Res-UNet

4.3 Experimental Result Analysis of Classification Models

First, we have trained our model on a CNN model we have built from scratch. We have explained the architecture of this CNN model in Chapter 4. Then we used DenseNet121 and DenseNet201 in order to see the use of transfer learning in our classification task. Later we explored a custom model that basically uses the feature extraction part of our proposed Multi-Task learning system. Training loss vs validation loss, Training Accuracy vs Validation Accuracy, Classification Report, and AUC-ROC graph have been presented in the following subsections.

4.3.1 Classification Report of The Classification Models

The precision, recall, and F1 score of the basic CNN model along with Test accuracy, Pre-trained architecture, and classification model with Multi-Task learning are provided in Table 4.4 and Table 4.5.

Model	Class	Precision	Recall	F1
Basic CNN Architecture	Class0	0.92	0.68	0.79
	Class1	0.65	0.91	0.76
DenseNet121	Class0	0.87	0.97	0.92
	Class1	0.91	0.70	0.79
DenseNet201	Class0	0.83	0.95	0.89
	Class1	0.98	0.68	0.80
Custom Model	**Class0**	**0.87**	**0.97**	**0.92**
	Class1	**0.95**	**0.77**	**0.85**

Table 4.4: Classification Report of All The Models on Test Data

The table presents a comparison of precision, recall, and F1 scores for three different models: Basic CNN Architecture, DenseNet121, and DenseNet201, along with a custom model, each applied to a binary classification task involving two classes (Class0 and Class1). DenseNet121 and DenseNet201 exhibit similar class-wise metrics, with variations in precision, recall, and F1 score. Additionally, the Custom

Model	Test Accuracy
Basic CNN Architecture	0.77
DenseNet121	0.80
DenseNet201	0.87
Custom Model	**0.89**

Table 4.5: Test Accuracy of all the models

Model has shown comparable or improved performance with precision, recall, and F1 score values for both classes. These metrics serve as indicators of the model's ability to accurately classify instances of the respective classes. Higher values suggest better performance in precision, recall, and the overall balance captured by the F1 score.

4.3.2 Training and Validation Loss and Accuracy

In this section, we have presented the training loss vs validation loss and training accuracy vs validation accuracy of all the models. Detailed analysis of the loss and accuracy graph is available in Figure 4.4, Figure 4.5, Figure 4.6 and Figure 4.7.

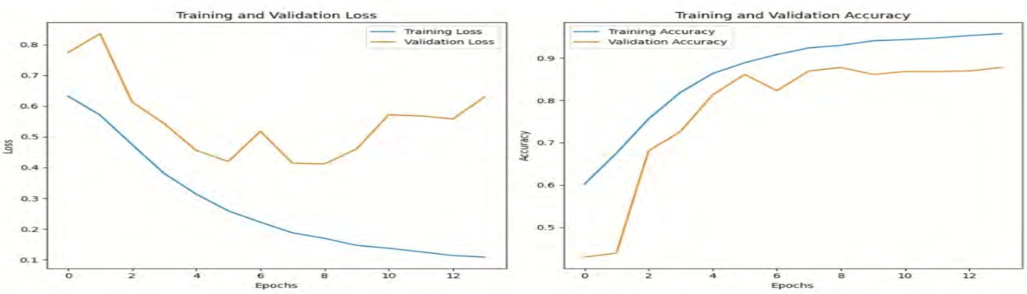


Figure 4.4: Basic CNN Training and Validation Loss and Accuracy Graph

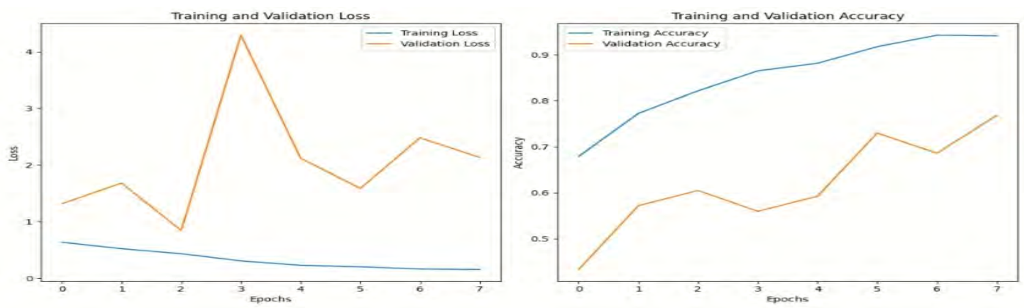


Figure 4.5: DenseNet121 Training and Validation Loss and Accuracy Graph

Here we can see the training and validation accuracy and training and validation loss of our 4 models. From the model, we can see that these models do not take training to 50 epochs even though epoch was set to 50. The models converge before reaching that epoch. Understanding the nuances of early convergence provides a valuable lens into the models' ability to grasp underlying patterns swiftly.

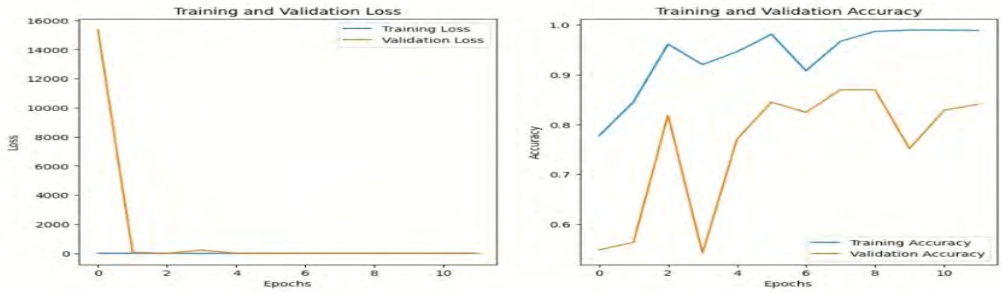


Figure 4.6: DenseNet201 Training and Validation Loss and Accuracy Graph

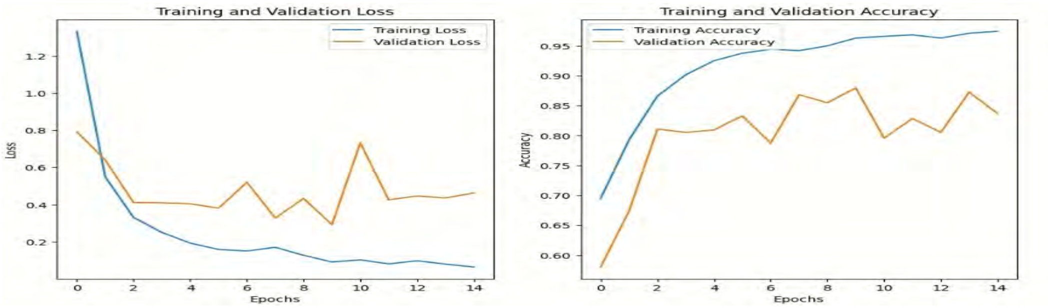
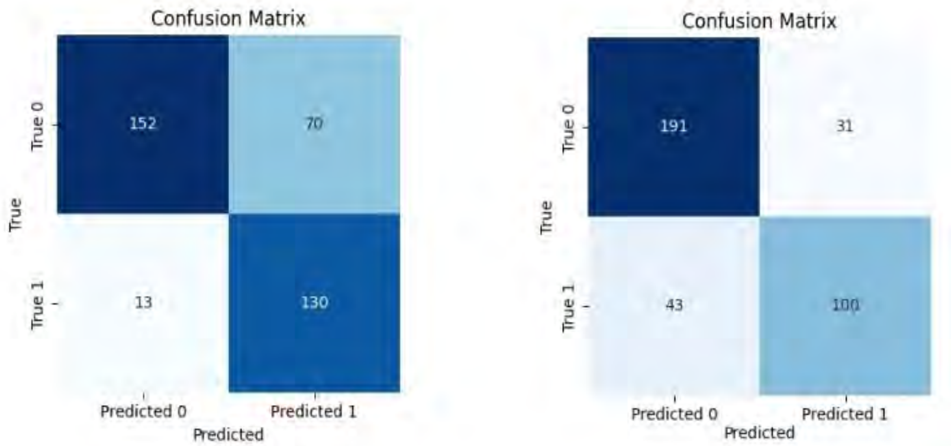


Figure 4.7: Custom Model Training and Validation Loss and Accuracy Graph

4.3.3 Confusion Matrix of The Classification Models

Figure 4.8 and Figure 4.9 show the Confusion Matrix of all the classification models - Basic CNN model, DenseNet121, DenseNet201, and Custom Model. Confusion Matrix shows that for class 0 (T1/T2 stage) DenseNet201 model performs better than all the other models. Similarly, for class 1 (T3/T4) basic CNN model outperforms other models.



Basic CNN

DenseNet121

Figure 4.8: Confusion Matrix for Basic CNN And DenseNet121

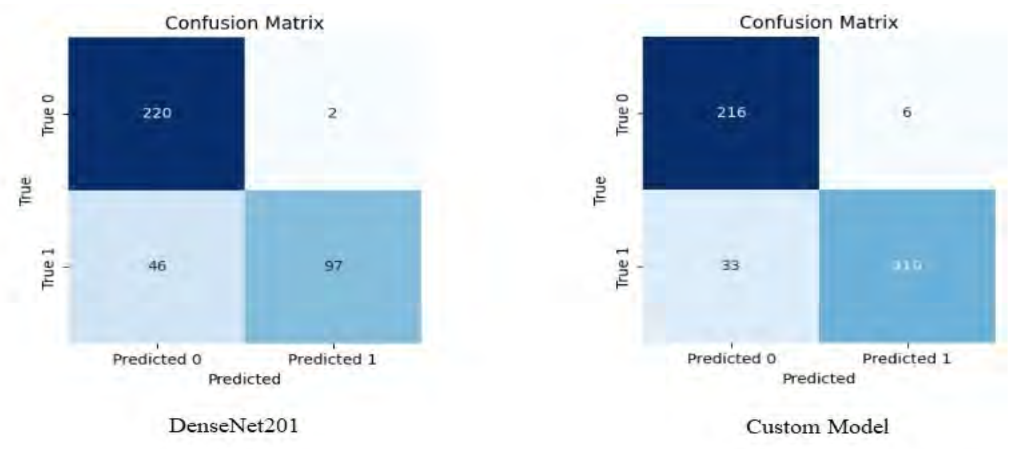


Figure 4.9: Confusion Matrix for DenseNet201 And Custom Model

4.3.4 AUC-ROC Curve of The Classification Models

Here we have shown the AUC-ROC graph model. From AUC-ROC graph we can say how well the graph is. From AUC-ROC graph we can say how well the model is. The AUC serves as a measure for assessing the effectiveness of a binary classification model. The ROC curve visually depicts the balance between sensitivity (true positive rate) and specificity (1-false positive rate) across different threshold values. This graph gives us an idea about the model. In this case, custom model performs better than all the other models. The AUC-ROC Curves for all the models are provided in Figure 4.10 and Figure 4.11.

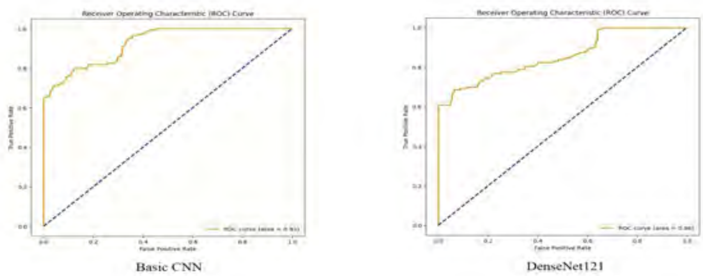


Figure 4.10: AUC-ROC Curve for Basic CNN And DenseNet121

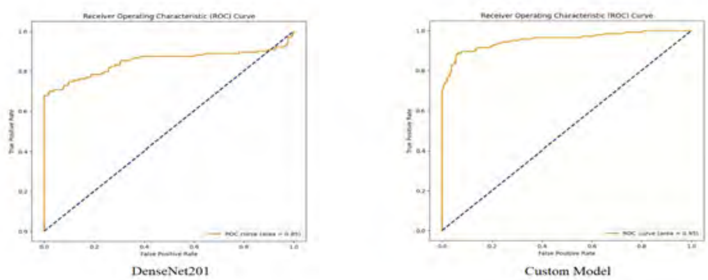


Figure 4.11: AUC-ROC Curve for DenseNet201 And Custom Model

4.4 Result Analysis of the proposed Multi-Task Learning Model

In this section, we are going to explain the result of the proposed Multi-Task learning model. Since we have performed two tasks there will be 2 types of results. One for a segmentation task and another for a classification task.

4.4.1 Segmentation Task

Segmentation metric for the Multi-Task learning system is presented in Table 4.6.

Data	Dice Coefficient	Precision	Recall	IOU
Train	0.8267	0.8872	0.8769	0.9820
Test	0.8522	0.8633	0.8459	0.9725
Valid	0.8606	0.8592	0.8641	0.9650

Table 4.6: Performance Metrics for Multi-task learning

Training loss vs validation loss and Training accuracy vs validation accuracy is presented in 4.12.

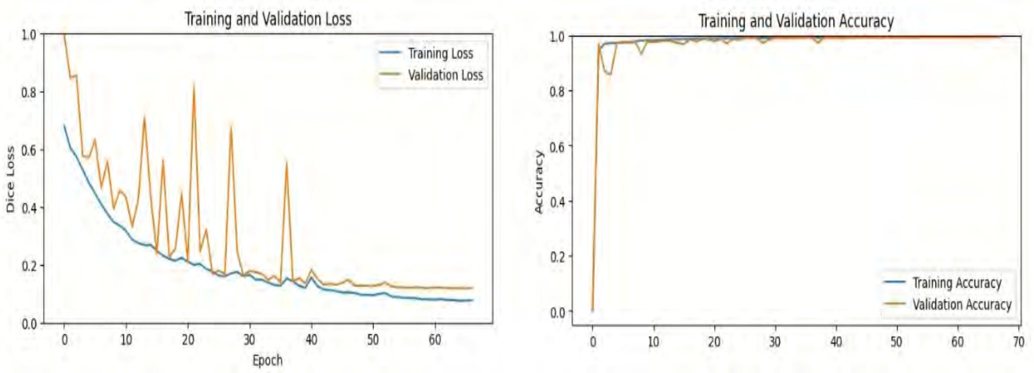


Figure 4.12: Multi-Task Learning Training and Validation Loss and Accuracy Graph

We can visualize the mask generation of the multitask model in Figure 4.13. Additionally in order to find the model's generated ROI section within the CT image we have taken the help of GRADCAM. Very few research have focused on GRADCAM visualization of the ROI within the CT image of the segmentation task. However, this approach holds significant promise for advancing the understanding and efficacy of multitask modeling in medical image analysis. GRADCAM visualization of the model's generated ROI in the CT scan is presented in Figure 4.14.
Though the dice value decreases a bit in Multi-Task learning, still it is a considerable solution since with this little trade-off we are achieving our goal with fewer resources. Furthermore, the versatility of Multi-Task learning shines through as it allows a model to simultaneously tackle multiple related tasks. This is because it shares knowledge and optimizes overall performance. This collaborative approach streamlines resource utilization as well as fosters a more holistic understanding of

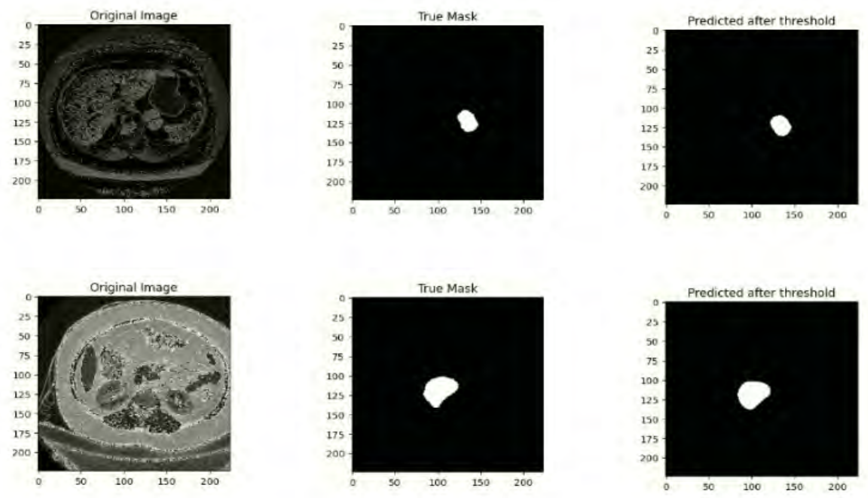


Figure 4.13: True Mask Vs Predicted Mask for Multi-Task Learning

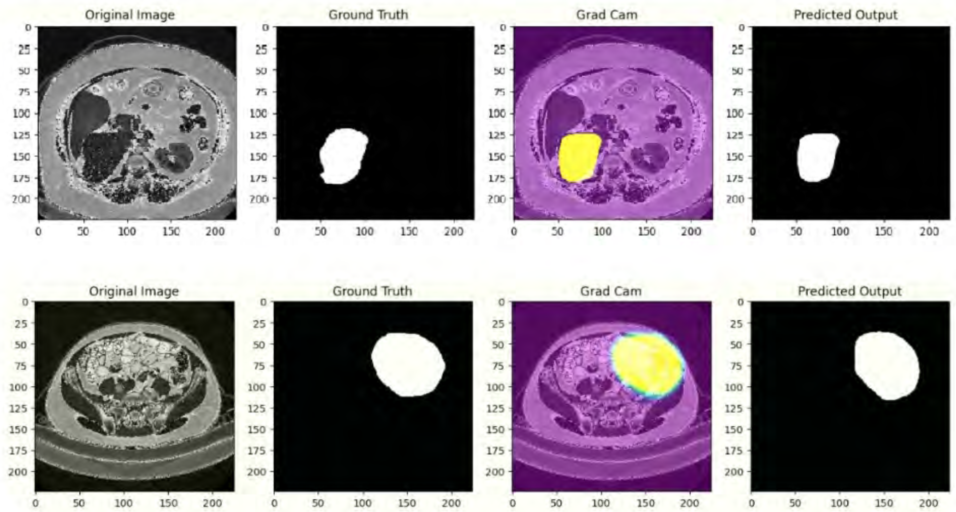


Figure 4.14: GRADCAM Visualization for The Multi-Task Learning Model

the underlying data. This enhances the model's ability to generalize across various domains. In essence, the marginal decrease in dice value becomes a worthwhile compromise when weighed against the substantial benefits of efficiency and broader learning capabilities that Multi-Task learning brings to the table.

4.4.2 Classification of the Multi-Task Learning Model

A classification report serves as a brief overview offering a thorough assessment of a classification model's performance. It encompasses essential metrics like precision, recall, F1 score, and accuracy which deliver valuable insights into the model's overall effectiveness. Precision is a key metric that signifies the proportion of accurately predicted positive observations among the total predicted positives. It indicates the model's capability to minimize false positives.

We can find the AUC-ROC Curve for Multi-Task learning classification in in Figure 4.15.

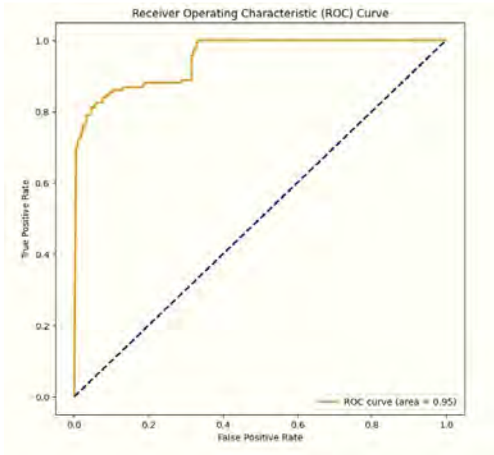


Figure 4.15: AUC-ROC Curve for Multi-Task Learning

Class	**Precision**	**Recall**	**F1**
Class0	0.84	0.89	0.86
Class1	0.86	0.79	0.82

Table 4.7: Performance Metrics for Multi-Task Learning Classification

Model	Test Accuracy
Multi-Task Learning	0.84

Table 4.8: Test Accuracy of Multi-Task Learning

From Table 4.7 and Table 4.8, we can find a comprehensive overview of the classification balanced performance of a Multi-Task Learning model, including precision, recall, F1 scores for individual classes, and overall test accuracy.

4.5 Comparision with the Existing Work

There are only a few segmentations that work on Adrenal tumors or cancer. The analysis of those studies is available in the literature review. With proper analysis of the optimizer, learning rate, and loss function we have got better segmentation value than existing works. The study [31] proposed a segmentation model with a dice value is 0.631. Another paper [32] has proposed a model with a Dice score of 0.858. In our final deep learning model with proper choice of learning rate and optimizer along with loss function, we have got a dice value of 0.8946. Moreover, we have performed classification based on T stages where we have proposed a custom model which surpassed all the state-of-the-art models. In the Multi-Task learning model, we have got dice value of 0.8522. So with the slightest loss in accuracy, we have achieved efficient computing which makes our work the first efficient computing-based highly accurate model in this domain.

Chapter 5

Conclusion

In conclusion, our research addresses the very vital need for quick and powerful diagnosis and treatment of Adrenocortical Carcinoma(ACC). The problems with the strategies used now, especially in terms of describing ACC steps for choosing the right medicine, make our work even more crucial. Inside the literature review, we talked at length about the previous works, mentioning the problems with the manner things are accomplished now when it comes to correctly describing ACC stages. Our observer makes critical additions to the field by means of looking into the fine segmentation techniques, suggesting a segmentation model that works better than current models, and showing how properly one-of-a-kind models classify data on the dataset. We additionally created a Multi-Task model that could do segmentation and classification at the same time with fewer factors, which makes the computer work more effectively. The suggested model does not most effectively improve the accuracy of ACC recognition, but it also accelerates the procedure via doing both segmentation and classification jobs more quickly. While GradCam is used to genuinely highlight important components of incoming images, it makes the neural network's decision-making process easier to understand and follow. Our study indicates how vital it is to work alone on those jobs, as we look into each of the detection and prevention parts. Using that specialization in effective division and classification, we are hoping to provide the medical community with beneficial tools for better making plans and diagnosing ACC. That being said, our work is an all-around try to fix the problems with the way ACC is currently observed and fixed. The recommended Multi-Task model and segmentation algorithms appear to be good methods to enhance accuracy and speed, to assist us in learning more about Adrenocortical Carcinoma and locating better ways to deal with it.

References

[1] P. A. T. E. Board, "Adrenocortical carcinoma treatment (adult)(pdq®): Patient version," *PDQ Cancer Information Summaries [Internet]*, 2002.

[2] *About Adrenocortical Carcinoma - Dana-Farber Cancer Institute — Boston, MA — dana-farber.org*, https://www.dana-farber.org/adrenocortical-carcinoma/about/, [Accessed 18-09-2023].

[3] *Overview of Adrenal Cancer — adrenal.com*, https://www.adrenal.com/adrenal-cancer/overview, [Accessed 15-09-2023].

[4] E. Ruggiero, I. Tizianel, M. Caccese, *et al.*, "Advanced adrenocortical carcinoma: From symptoms control to palliative care," *Cancers*, vol. 14, no. 23, p. 5901, 2022.

[5] S. Puglisi, V. Basile, P. Sperone, and M. Terzolo, "Pregnancy in patients with adrenocortical carcinoma: A case-based discussion," *Reviews in Endocrine and Metabolic Disorders*, vol. 24, no. 1, pp. 85–96, 2023.

[6] X. Ni, J. Wang, J. Cao, *et al.*, "Purpose: Spinal metastasis of malignant adrenal tumor (smmat) is an extremely rare and poorly understood malignant tumor originating from the adrenal gland. the objective of this study is to elucidate the clinical characteristics and discuss surgical management and outcomes of smmat. methods: Included in this study were six smmat patients who received surgical," *Diagnosis and Treatment of Bone Metastases*, vol. 5, p. 68, 2023.

[7] D. Neupane, S. Khadka, P. Upadhyaya, *et al.*, "Primary adrenocortical carcinoma: A case report," *Annals of Medicine and Surgery*, vol. 85, no. 5, p. 1834, 2023.

[8] W. A. El-Dakroury, H. M. Midan, A. I. Abulsoud, *et al.*, "Mirnas orchestration of adrenocortical carcinoma-particular emphasis on diagnosis, progression and drug resistance," *Pathology-Research and Practice*, p. 154 665, 2023.

[9] *Adreno Cortical Carcinoma - Symptoms-Diagnosis-Treatment-Prognosis — medindia.net*, https://www.medindia.net/patients/patientinfo/adreno-cortical-carcinoma.htm, [Accessed 18-09-2023].

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer, 2015, pp. 234–241.

[11] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[13] A. W. Moawad, A. Ahmed, D. T. Fuentes, J. D. Hazle, M. A. Habra, and K. M. Elsayes, "Machine learning-based texture analysis for differentiation of radiologically indeterminate small adrenal tumors on adrenal protocol ct scans," *Abdominal Radiology*, vol. 46, no. 10, pp. 4853–4863, 2021.

[14] O. Abdel-Rahman, "Revisiting the ajcc staging system of adrenocortical carcinoma," *Journal of Endocrinological Investigation*, vol. 45, no. 1, pp. 89–94, 2022.

[15] S. Zheng, A. D. Cherniack, N. Dewal, *et al.*, "Comprehensive pan-genomic characterization of adrenocortical carcinoma," *Cancer cell*, vol. 29, no. 5, pp. 723–736, 2016.

[16] H. Lu, T. G. Papathomas, D. van Zessen, *et al.*, "Automated selection of hotspots (ash): Enhanced automated segmentation and adaptive step finding for ki67 hotspot detection in adrenal cortical cancer," *Diagnostic pathology*, vol. 9, no. 1, pp. 1–9, 2014.

[17] G. Assié, E. Letouzé, M. Fassnacht, *et al.*, "Integrated genomic characterization of adrenocortical carcinoma," *Nature genetics*, vol. 46, no. 6, pp. 607–612, 2014.

[18] R. Libé, "Adrenocortical carcinoma (acc): Diagnosis, prognosis, and treatment," *Frontiers in cell and developmental biology*, vol. 3, p. 45, 2015.

[19] F. Beuschlein, J. Weigel, W. Saeger, *et al.*, "Major prognostic role of ki67 in localized adrenocortical carcinoma after complete resection," *The Journal of Clinical Endocrinology & Metabolism*, vol. 100, no. 3, pp. 841–849, 2015.

[20] Y. Yamazaki, Y. Nakamura, Y. Shibahara, *et al.*, "Comparison of the methods for measuring the ki-67 labeling index in adrenocortical carcinoma: Manual versus digital image analysis," *Human Pathology*, vol. 53, pp. 41–50, 2016.

[21] S. Creemers, L. Hofland, E. Korpershoek, *et al.*, "Future directions in the diagnosis and medical treatment of adrenocortical carcinoma," *Endocrine-related cancer*, vol. 23, no. 1, R43–R69, 2016.

[22] Y. Kamkova, H. Ali Qadir, O. Jakob, and R. Prasanna Kumar, "Kidney and tumor segmentation using combined deep learning method," 2019.

[23] G. Santini, N. Moreau, and M. Rubeaux, "Kidney tumor segmentation using an ensembling multi-stage deep learning approach. a contribution to the kits19 challenge. arxiv 2019," *arXiv preprint arXiv:1909.00735*,

[24] F. Türk, M. Lüy, and N. Barışçı, "Kidney and renal tumor segmentation using a hybrid v-net-based model," *Mathematics*, vol. 8, no. 10, p. 1772, 2020.

[25] G. Luo, Q. Yang, T. Chen, T. Zheng, W. Xie, and H. Sun, "An optimized two-stage cascaded deep neural network for adrenal segmentation on ct images," *Computers in Biology and Medicine*, vol. 136, p. 104 749, 2021.

[26] J. Sebek, G. Cappiello, G. Rahmani, *et al.*, "Image-based computer modeling assessment of microwave ablation for treatment of adrenal tumors," *International Journal of Hyperthermia*, vol. 39, no. 1, pp. 1264–1275, 2022.

[27] K. Sharma, C. Rupprecht, A. Caroli, *et al.*, "Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease," *Scientific reports*, vol. 7, no. 1, p. 2049, 2017.

[28] A. Ahmed, M. Elmohr, D. Fuentes, *et al.*, "Radiomic mapping model for prediction of ki-67 expression in adrenocortical carcinoma," *Clinical Radiology*, vol. 75, no. 6, 479–e17, 2020.

[29] S. Yin, Q. Peng, H. Li, *et al.*, "Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks," *Medical image analysis*, vol. 60, p. 101 602, 2020.

[30] J. Tang, Y. Fang, and Z. Xu, "Establishment of prognostic models of adrenocortical carcinoma using machine learning and big data," *Frontiers in Surgery*, vol. 9, p. 966 307, 2023.

[31] A. Solak, R. Ceylan, M. A. Bozkurt, H. Cebeci, and M. Koplay, "Adrenal tumor segmentation on u-net: A study about effect of different parameters in deep learning," *Vietnam Journal of Computer Science*, 2023.

[32] L. Wang, M. Ye, Y. Lu, *et al.*, "A combined encoder–transformer–decoder network for volumetric segmentation of adrenal tumors," *BioMedical Engineering OnLine*, vol. 22, no. 1, p. 106, 2023.

[33] S. Chen, G. Bortsova, A. García-Uceda Juárez, G. Van Tulder, and M. De Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*, Springer, 2019, pp. 457–465.

[34] R. Ke, A. Bugeau, N. Papadakis, M. Kirkland, P. Schuetz, and C.-B. Schönlieb, "Multi-task deep learning for image segmentation using recursive approximation tasks," *IEEE Transactions on Image Processing*, vol. 30, pp. 3555–3567, 2021.

[35] H. Huang, G. Yang, W. Zhang, *et al.*, "A deep multi-task learning framework for brain tumor segmentation," *Frontiers in Oncology*, vol. 11, p. 690 244, 2021.

[36] S. Graham, Q. D. Vu, M. Jahanifar, *et al.*, "One model is all you need: Multi-task learning enables simultaneous histology image segmentation and classification," *Medical Image Analysis*, vol. 83, p. 102 685, 2023.

[37] M. Xu, K. Huang, and X. Qi, "A regional-attentive multi-task learning framework for breast ultrasound image segmentation and classification," *IEEE Access*, vol. 11, pp. 5377–5392, 2023.

[38] A. W. Moawad, A. A. Ahmed, M. ElMohr, *et al.*, "Voxel-level segmentation of pathologically-proven adrenocortical carcinoma with ki-67 expression (adrenal-acc-ki67-seg)[data set]," 2023. DOI: 10.7937/1FPG-VM46.

[39] D. Mason, "Su-e-t-33: Pydicom: An open source dicom library," *Medical Physics*, vol. 38, no. 6Part10, pp. 3493–3493, 2011.