NECROSIS-ML: AN ENCODER-DECODER APPROACH FOR SEMANTIC SEGMENTATION AND QUANTIZATION OF TUMOR NECROSIS USING HISTOPATHOLOGY IMAGES

A Thesis Submitted to

UNIVERSITY OF CALICUT

in partial fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

Under the Faculty of Science

By

SALEENA T S

Under the guidance of

Dr. MUHAMED ILYAS P

Principal, Sullamussalam Science College, Areekode

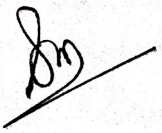
SULLAMUSSALAM SCIENCE COLLEGE

P.G & RESEARCH DEPARTMENT OF COMPUTER SCIENCE
Sullamussalam Science College, Areekode - 673639
Affiliated to University of Calicut
Malappuram Dist., Kerala, India
February 2025

DECLARATION

I, Saleena T.S, hereby declare that this thesis entitled "Necrosis-ML: An encoder-decoder approach for semantic segmentation and quantization of tumor necrosis using histopathology images" is based on the original work done by me under the supervision of Dr. Muhamed Ilyas P., Principal, Sullamussalam Science College, Areekode, Kerala. I confirm that,

- The work presented in this Thesis has not been submitted previously for the award of any degree either to this University or to any other University or Institution.
- I have followed the guiding principles given by the University in organizing the Thesis.
- Whenever I have used materials (theoretical analysis, data, figures, and text) from other sources, I have given due credit to them by citing them in the thesis and giving their particulars in the references.



SALEENA T.S

Place: Areekode

Date: 10/02/2025

SULLAMUSSALAM
SCIENCE COLLEGE
AREEKODE

Phone :0483-2850700
Mob :+919895811093
Website : www.sscollege.ac.in
E-mail : mail@sscollege.ac.in

Aided College | Affiliated to the University of Calicut | Re-accredited by NAAC at A Grade

Ref.

Date :10-02-2025

CERTIFICATE

This is to certify that the thesis entitled **"Necrosis-ML: An Encoder-Decoder Approach for Semantic Segmentation and Quantization of Tumor Necrosis using Histopathology Images"**, submitted by **Ms. SALEENA T.S** to the University of Calicut in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy (PhD) in Computer Science**, is a record of bonafide research work carried out by her under my supervision and guidance in the PG and Research Department of Computer Science, Sullamussalam Science College, Areekode, Malappuram, in accordance with the University of Calicut's code of academic and research ethics. The contents of this thesis, in full or in part, have not been submitted for the award of any other degree or diploma at this or any other Institute or University.

The thesis is revised as per the modifications and recommendations reported by the adjudicators. Soft copy attached is the same as that of the revised copy. The thesis is submitted as such to the University of Calicut with reference to the letter No. 126749/RESEARCH-C-ASST-1/2024/Admn Dated 04.12.2024.

Dr. Muhamed Ilyas P.
Research Guide & Principal,
Sullamussalam Science College,
Areekode, Kerala

Place: Areekode
Date: 10/02/2025

ACKNOWLEDGMENTS

With immense pleasure and a deep sense of gratitude, I wish to express my sincere thanks to my supervisor **Dr. Muhamed Ilyas P.**, Principal, Sullamussalam Science College Areekode. Without his motivation and continuous encouragement, this research would not have been successfully completed. I express my gratitude to the domain expert **Dr. Sajna V.M. Kutty**, Senior Pathologist, MVRCCRI, Calicut, who was with me throughout this research to give guidance and support in the domain side.

I express my sincere gratitude to the members of the Departmental Research Committee (DRC) and Research Advisory Committee (RAC), **Dr. Lajish V.L**, Associate Professor, Department of Computer Science, University of Calicut, **Dr. Vasudevan. T.M**, Professor, Department of Library and Information Science, University of Calicut, **Dr.Mohamed Basheer. K.P**, Associate Professor and Research Nodal Officer, PG and Research Department of Computer Science, Sullamussalam Science College, Areekode, **Dr.Binu P Chacko**, Principal, Prajothi Nikethan College, Puthukad and **Dr.Shameem Kappan**, Associate Professor and Head, PG and Research Department of Computer Science, Sullamussalam Science College, Areekode for their constructive criticism and valuable suggestions which inspired me to widen my research from various perspectives.

I would like to acknowledge the support rendered by **Dr. Syed Ibrahim S.P.**, **Dr. Rajeev R R**, **Dr. Fazna N.**, **Dr. Abdul Jabbar P.**,**Ms. Liji S.K**, **Ms. Anju A.T**, **Dr. Shabina Bhaskar**, **Dr. Sarath V.S**, **Mr. Muthumanickavel Swaminathan**, **Mr. Sabeerali K.P.**, **Mr. Fahd Ambadi**, **Mr. Bahalul Haque**, **Dr. N Asiq Sideeque**, **Dr. Deepak P J**, **Mr. Ajay Sharma**, all the research scholars and faculties of our Department for supporting me in several ways throughout my research work. I would also like to acknowledge the technical staff **Mr. Ameer Ajwad N.V**, and **Mr. Zainul Abid** and office staff **Mr. Lukman Kollathodi**, and **Mr. Abdussalam. M** for their support in technical and office-related matters.

I wish to extend my profound sense of gratitude to my parents **Sulaiman T.P**,

Nusriya M.A, siblings, and in-laws **Abdul Kareem K, Kadeeja K.** for providing me with moral support and encouragement whenever required.

Last but not the least, I would like to thank my husband **Mr. Sheril Kareem**, my daughter **Fathima Zoya**, and my son **Aman Sheril** for all the sacrifices they made during my research and for their constant encouragement and moral support along with patience and understanding.

Place: Areekode

Date: 10/02/2025

SALEENA T.S

ABSTRACT

Clinical pathology is one of the finest diagnostic techniques for all types of cancer, and its outcome determines the treatment plan for a patient. In this research, two types of cancers are taken into account, Osteosarcoma and Renal Cell Carcinoma, where one of the major prognostic factors is the amount of tumor necrosis created due to Neo-adjuvant chemotherapy. Osteosarcoma is a high-grade malignant bone tumor and Renal Cell Carcinoma is the most common type of Kidney Cancer. Tumor necrosis is a condition where the tumorous tissues cannot perform their normal metabolic functions and gradually die. Neo-adjuvant chemotherapy is the treatment given to a patient's body before starting the main treatment. The proposed study aims to develop an automated tool for quantitative image analysis of digital histopathology images of post-neoadjuvant resection specimens. Post-neoadjuvant chemotherapy resection specimens refer to the tissue samples collected from cancer patients who have undergone Neo-adjuvant chemotherapy. Even though so many tools are available for cancer diagnosis, there is no specific tool to calculate the percentage of necrosis available in post-neoadjuvant chemotherapy resection specimens of Osteosarcoma and Renal Cell Carcinoma.

Necrosis-ML is an algorithm that has been developed in this work to perform image-level segmentation and quantization. The two major tasks involved in this tool are the semantic segmentation and quantization of segmented masks. In this algorithm, the input image will undergo a patchification process and segmentation will be performed at patch-level. The segmentation model is set up with U-Net++ architecture using ResNet101 as the feature extractor. The mask of each segmented patch will be merged to get a single binary mask and it will be quantified to get the final result. Pathologists can feed the histopathology images captured from the digital microscope into this tool and the output will be the segmented image and the total area of this segmented part.

Different other methods have been proposed in this research for the segmentation task as this is the only AI-enabled part of this study. One among them is a segmentation model using Autoencoder where the training of the model is done with a single histopathology image. Another proposed model is designed using SegFormer, which is a transformer-based, encoder-decoder architecture. The SegFormer model also can be

used to develop Necrosis-ML by replacing the U-Net architecture mentioned above.

Another major contribution of this research is the dataset created in this study, named as NecrosisDB. This dataset contains 900 patches of images that include various morphologies of necrosis, tumors, fibrosis, and other frequently occurring tissue elements. The major hurdle in any research in the medical domain is the unavailability of the annotated datasets, but in this study, we managed to create a fully supervised annotated dataset that has been annotated using experienced pathologists. This dataset has been made publicly accessible for research or study purpose based on some terms and conditions. It can be accessed from the link,

https://github.com/saleenaareekode/Necrosis-ML.

Keywords: *Tumor Necrosis*, *patchification*, *Neoadjuvant chemotherapy*, *Autoencoder*, *U-Net*, *DeepLab*, *ResNet*, *image segmentation*, *image quantification*, *semantic segmentation*, *SegFormer*.

സംഗ്രഹം

എല്ലാത്തരം കാൻസറുകൾക്കുമുള്ള ഏറ്റവും മികച്ച രോഗനിർണയ സാങ്കേതിക വിദ്യകളിൽ ഒന്നാണ് ക്ലിനിക്കൽ പാത്തോളജി, അതിന്റെ ഫലം ഒരു രോഗിയുടെ ചികിത്സാ പദ്ധതിയെ നിർണ്ണയിക്കുന്നു. ഈ ഗവേഷണത്തിൽ, രണ്ട് തരം കാൻസറുകൾ കണക്കിലെടുക്കുന്നു, ഓസ്റ്റിയോസർകോമയും, റീനൽ സെൽ കാർസിനോമയും. നിയോ-അഡ്ജുവന്റ് കീമോതെറാപ്പി മൂലം ഉണ്ടാകുന്ന ട്യൂമർ നെക്രോസിസിന്റെ അളവാണ് പ്രധാന രോഗനിർണയ ഘടകങ്ങളിലൊന്ന്. ഓസ്റ്റിയോസർകോമ ഒരു ഉയർന്ന ഗ്രേഡ് മാരകമായ അസ്ഥി ട്യൂമറാണ്, കൂടാതെ വൃക്കയിലെ ക്യാൻസറിന്റെ ഏറ്റവും സാധാരണമായ തരമാണ് റീനൽ സെൽ കാർസിനോമ. ട്യൂമർ നെക്രോസിസ് എന്നത് ട്യൂമർ കലകൾ അവയുടെ സാധാരണ ഉപാപചയ പ്രവർത്തനങ്ങൾ നിർവഹിക്കാൻ കഴിയാതെ ക്രമേണ മരിക്കുന്ന ഒരു അവസ്ഥയാണ്. പ്രധാന ചികിത്സ ആരംഭിക്കുന്നതിന് മുമ്പ് രോഗിയുടെ ശരീരത്തിന് നൽകുന്ന ചികിത്സയാണ് നിയോ-അഡ്ജുവന്റ് കീമോതെറാപ്പി. പോസ്റ്റ്-നിയോഅഡ്ജുവന്റ് റിസെക്ഷൻ മാതൃകകളുടെ ഡിജിറ്റൽ ഹിസ്റ്റോപാത്തോളജി ചിത്രങ്ങളുടെ ക്വാണ്ടിറ്റേറ്റീവ് ഇമേജ് വിശകലനത്തിനായി ഒരു ഓട്ടോമേറ്റഡ് ഉപകരണം വികസിപ്പിക്കുക എന്നതാണ് നിർദ്ദിഷ്ട പഠനം ലക്ഷ്യമിടുന്നത്. നിയോ-അഡ്ജുവന്റ് കീമോതെറാപ്പിക്ക് വിധേയരായ കാൻസർ രോഗികളിൽ നിന്ന് ശേഖരിച്ച ടിഷ്യു സാമ്പിളുകളെയാണ് പോസ്റ്റ്-നിയോഅഡ്ജുവന്റ് കീമോതെറാപ്പി റിസെക്ഷൻ സാമ്പിളുകൾ എന്ന് പറയുന്നത്. കാൻസർ രോഗനിർണയത്തിനായി നിരവധി ഉപകരണങ്ങൾ ലഭ്യമാണെങ്കിലും, ഓസ്റ്റിയോസർകോമയുടെയും, റീനൽ സെൽ കാർസിനോമയുടെയും പോസ്റ്റ്-നിയോഅഡ്ജുവന്റ് കീമോതെറാപ്പി റിസെക്ഷൻ മാതൃകകളിൽ ലഭ്യമായ നെക്രോസിസിന്റെ ശതമാനം കണക്കാക്കാൻ പ്രത്യേക ഉപകരണമൊന്നുമില്ല.

ഇമേജ്-ലെവൽ സെഗ്മെന്റേഷനും, ക്വാണ്ടിഫിക്കേഷനും ചെയ്യുന്നതിനായി ഈ പ്രവർത്തനത്തിൽ വികസിപ്പിച്ചെടുത്ത ഒരു അൽഗോരിതമാണ് നെക്രോസിസ്-എംഎൽ. ഈ ഉപകരണത്തിൽ ഉൾപ്പെട്ടിരിക്കുന്ന രണ്ട് പ്രധാന ജോലികൾ സെമാന്റിക് സെഗ്മെന്റേഷനും, സെഗ്മെന്റഡ് മാസ്കുകളുടെ ക്വാണ്ടിഫിക്കേഷനുമാണ്. ഈ അൽഗോരിതത്തിൽ, ഇൻപുട്ട് ഇമേജ് ഒരു പാച്ച്ഡിക്കേഷൻ പ്രക്രിയയ്ക്ക് വിധേയമാക്കുകയും പാച്ച്-ലെവലിൽ സെഗ്മെന്റേഷൻ നടത്തുകയും ചെയ്യും. ResNet101 ഫീച്ചർ എക്സ്ട്രാക്റ്ററായി ഉപയോഗിക്കുന്ന U-Net++ ആർക്കിടെക്ചർ ഉപയോഗിച്ചാണ് സെഗ്മെന്റേഷൻ മോഡൽ സജ്ജീകരിച്ചിരിക്കുന്നത്. ഓരോ സെഗ്മെന്റഡ് പാച്ചിന്റെയും മാസ്ക് ലയിപ്പിച്ച് ഒരൊറ്റ ബൈനറി മാസ്ക് ലഭിക്കും, അന്തിമ ഫലം ലഭിക്കുന്നതിന് അത് ക്വാണ്ടിഫൈ ചെയ്യപ്പെടും. ഡിജിറ്റൽ മൈക്രോസ്കോപ്പിൽ നിന്ന് പകർത്തിയ ഹിസ്റ്റോപാത്തോളജി ചിത്രങ്ങൾ പാത്തോളജിസ്റ്റുകൾക്ക് ഈ ഉപകരണത്തിലേക്ക് ഫീഡ് ചെയ്യാൻ കഴിയും, ഔട്ട്പുട്ട് സെഗ്മെന്റഡ് ഇമേജും ഈ സെഗ്മെന്റഡ് ഭാഗത്തിന്റെ ആകെ വിസ്തീർണ്ണവുമായിരിക്കും.

ഈ പഠനത്തിലെ ഏക AI പ്രാപ്തമാക്കിയ ഭാഗം ആയതിനാൽ, സെഗ്മെന്റേഷൻ ടാസ്ക്കിനായി ഈ ഗവേഷണത്തിൽ വ്യത്യസ്തമായ മറ്റ് രീതികൾ നിർദ്ദേശിച്ചിട്ടുണ്ട്. അതിലൊന്നാണ് ഓട്ടോഎൻകോഡർ ഉപയോഗിച്ചുള്ള ഒരു സെഗ്മെന്റേഷൻ മോഡൽ, ഇവിടെ മോഡലിന്റെ പരിശീലനം ഒരൊറ്റ ഹിസ്റ്റോപാത്തോളജി ഇമേജ് ഉപയോഗിച്ച് നടത്തുന്നു. മറ്റൊരു നിർദ്ദിഷ്ട മോഡൽ രൂപകൽപ്പന ചെയ്തിരിക്കുന്നത് SegFormer ഉപയോഗിച്ചാണ്, ഇത് ഒരു ട്രാൻസ്ഫോർമർ അധിഷ്ഠിത, എൻകോഡർ-ഡീകോഡർ ആർക്കിടെക്ചറാണ്. മുകളിൽ സൂചിപ്പിച്ച U-Net++ ആർക്കിടെക്ചർ മാറ്റിസ്ഥാപിച്ചുകൊണ്ട് നെക്രോസിസ്-എംഎൽ വികസിപ്പിക്കുന്നതിനും സെഗ്ഫോർമർ മോഡൽ ഉപയോഗിക്കാം.

ഈ ഗവേഷണത്തിന്റെ മറ്റൊരു പ്രധാന സംഭാവന ഈ പഠനത്തിൽ സൃഷ്ടിച്ച ഡാറ്റാസെറ്റാണ്, ഇതിന് NecrosisDB എന്ന് പേരിട്ടു. ഈ ഡാറ്റാസെറ്റിൽ നെക്രോസിസ്, ട്യൂമറുകൾ, ഫൈബ്രോസിസ്, പതിവായി സംഭവിക്കുന്ന മറ്റ് ടിഷ്യു ഘടകങ്ങൾ എന്നിവയുടെ വിവിധ രൂപാന്തരങ്ങൾ ഉൾപ്പെടുന്ന 900 പാച്ചുകളുടെ ചിത്രങ്ങൾ അടങ്ങിയിരിക്കുന്നു. വൈദ്യശാസ്ത്ര മേഖലയിലെ ഏതൊരു ഗവേഷണത്തിലും ഏറ്റവും വലിയ തടസ്സം വ്യാഖ്യാനിച്ച ഡാറ്റാസെറ്റുകളുടെ ലഭ്യതക്കുറവാണ്, എന്നാൽ ഈ പഠനത്തിൽ, പരിചയസമ്പന്നരായ പാത്തോളജിസ്റ്റുകളെ ഉപയോഗിച്ച് വ്യാഖ്യാനിച്ച ഒരു പൂർണ്ണ മേൽനോട്ടത്തിലുള്ള വ്യാഖ്യാന ഡാറ്റാസെറ്റ് സൃഷ്ടിക്കാൻ ഞങ്ങൾക്ക് കഴിഞ്ഞു. ചില നിബന്ധനകൾക്കും വ്യവസ്ഥകൾക്കും വിധേയമായി ഗവേഷണത്തിനോ പഠനത്തിനോ വേണ്ടി ഈ ഡാറ്റാസെറ്റ് പൊതുജനങ്ങൾക്ക് ലഭ്യമാകുന്നതാക്കി മാറ്റിയിരിക്കുന്നു. https://github.com/saleenaareekode/Necrosis-ML എന്ന ലിങ്കിൽ നിന്ന് ഇത് ആക്സസ് ചെയ്യാൻ കഴിയും.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Terms and Abbreviations

H and E Hematoxylin and Eosin

NAC Neoadjuvant Chemotherapy

NAT Neoadjuvant Therapy

CT Computed Tomography

PET Positron Emission Tomography

WSI Whole Slide Imaging

RCC Renal Cell Carcinoma

AL Active Learning

TCGA The Cancer Genome Atlas

CAE Convolutional Autoencoder

IoU Intersection over Union

mIoU Mean Intersection over Union

CNN Convolutional Neural Network

GPU Graphics Processing Unit

MLP Multi Layer Perceptron

ViT Vision Transformer

ILSVRC ImageNet Large-Scale Visual Recognition Challenge

ResNet Residual Net

CHAPTER 1

Introduction

1.1 Background

1.1.1 Osteosarcoma and Renal Cell Carcinoma

Osteosarcoma is a rare but aggressive form of bone cancer that primarily affects the cells that form bones. As per the study, it is called rare because it makes up less than 1% of all malignancies diagnosed in the US(Gorlick and Khanna 2010). It usually develops in the long bones, such as the arms or legs, but it can occur in any bone. Bones around the knees and upper arm bone closer to the shoulder are the most common sites for cancers. Osteosarcoma often affects adolescents and young adults, and it tends to grow rapidly (Czarnecka et al. 2020). The exact cause of osteosarcoma is not well understood, but some genetic factors and pre-existing conditions may increase the risk. The most often mutated gene in this cancer is TP53. Common symptoms include bone pain, swelling, and, in some cases, a noticeable lump.

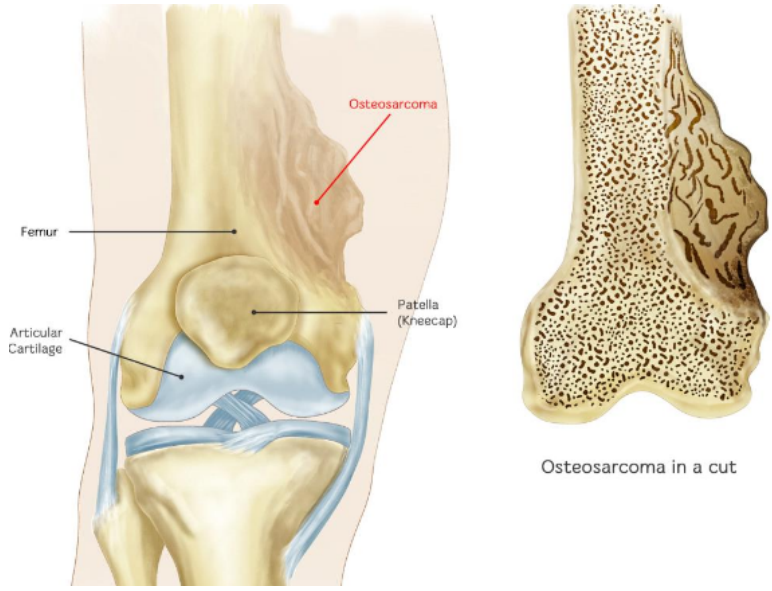


Fig. 1.1 A case of Osteosarcoma

Renal cell carcinoma (RCC) is the most common type of kidney cancer, accounting

for about 90-95% of cases. It originates in the cells lining the small tubes (tubules) of the kidney and is often detected in the renal cortex, which is the outer part of the kidney. Renal cell carcinoma can manifest as a single tumor in one kidney or as multiple tumors in one or both kidneys. The risk factors for this cancer include smoking, obesity, high blood pressure, and a family history of kidney cancer. Additionally, certain genetic conditions, such as von Hippel-Lindau disease, are associated with an increased risk of developing renal cell carcinoma. It has a high chance for metastasis as it will not show any early symptoms in the patients (Motzer et al. 1996) but still some symptoms may include blood in the urine, back pain or pain in the side, a mass or lump in the abdomen, unexplained weight loss, fatigue, and fever.

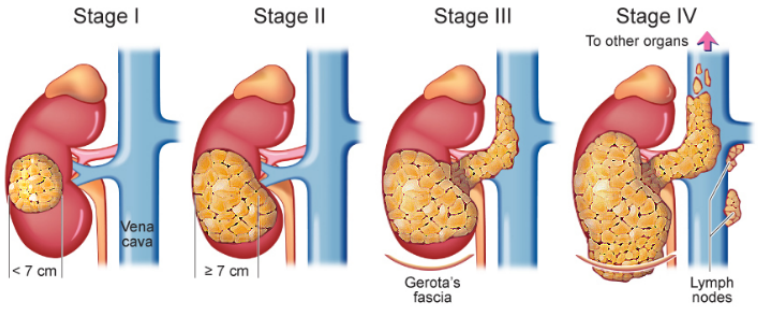


Fig. 1.2 Different stages in Renal Cell Carcinoma

Treatment options for renal cell carcinoma depend on various factors including the stage of the cancer, the patient's overall health, and the specific characteristics of the tumor. Common treatments include surgery to remove the tumor or the entire kidney (nephrectomy), targeted therapies, immunotherapy, and sometimes radiation therapy. Early detection and appropriate treatment contribute to better outcomes for individuals with renal cell carcinoma. Treatment typically involves a combination of surgery to remove the tumor and chemotherapy. In some cases, radiation therapy may also be used. Early detection and a multidisciplinary approach to treatment are crucial for better prognosis and increased chances of successful management.

1.1.2 Clinical Pathology as the Decision-maker in Treatment-plan

Clinical diagnosis is the process of identifying a medical condition or disease of a patient by evaluating his/her signs, symptoms, medical history, and physical examination findings. This process will be usually done by a healthcare professional, such as a physician or nurse practitioner, who has received specialized training in diagnosing and treating various medical conditions (NCI 2022). They may add more laboratory tests or imaging techniques to confirm the diagnosis. The medical professionals will determine the patient's treatment strategy based on the findings of the appropriate tests. As a

result, we can conclude that the clinical diagnosis' findings decide the selection of the patient's treatment plans. Also, the most noticeable part of this is that all the patients with the same disease will not be treated with the same treatment plan even though they are in the same stage of the disease.

Even though clinical diagnosis has different modalities, *Clinical Pathology* is the most trusted one among them in most diseases. In this particular study, we are concentrating on clinical pathology for cancer diagnosis and therapeutic decision-making followed by that. We can more precisely call this branch Diagnostic Pathology. Here, the diagnosis will be done through the analysis of tissues through microscopes. Morphological features are mainly used to do the analysis here. The same procedure will be followed for Biopsy tissue samples, which is the supreme diagnostic method for cancer. Medical Imaging techniques like CT, X-ray, PET, etc can be used for initial screening but the confirmation can be done only with the biopsy results.

The entire procedure of diagnostic pathology will include different stages namely,

1. Tissue(specimen) preparation
2. Staining
3. Microscopy
4. Molecular testing
5. Identification of different tissues.

Each of these steps has several subsections that should be done very carefully so that the entire procedure preserves the tissue architecture, molecular components, and cellular morphology. Each step has been explained in detail as follows:

Tissue preparation: (Tatyana S. Gurina 2022)The tissue preparation is the preliminary and important step in every pathology analysis process. During this stage, the tissue samples will be collected from the suspected part of the body of the patient. In this stage, very thin but high-quality tissue samples will be mounted on appropriate glass slides. These steps are defined as fixation, processing, embedding, sectioning, and sometimes antigen retrieval. During fixation, required chemicals will be added to preserve the tissue architecture and protect it from degradation. Ethanol will be used to do the dehydration process from the sample. Embedding means placing the sample in paraffin wax or a plastic resin to enhance the process of extracting cellular structures. During sectioning, the sample will be mounted on a microtome and will be divided into different sections. Antigens will be retrieved in the last step.

Staining: Staining is the process of giving color to the elements present in the specimen under consideration. Staining applied on the histopathology specimens will help to highlight the features, especially morphology-kind of things and it also enhances

the tissue contrast. (Alturkistani et al. 2016)A literature review has been conducted by the authors in this paper that explains what all changes have happened to histological staining procedures. Hematoxylin and Eosin are two prominently used dyes for staining histopathology specimens. Hematoxylin is used to stain nuclei and give a bluish color to them whereas Eosin stains the cell's nucleus and gives a pinkish color to them. During tumor diagnosis, staining will be applied on the posterior and anterior borders of the sample tissues. Care should be taken while choosing the staining process as some of them may lead to the denaturalization of tissues and cells which inhibit effective histological studies.

Microscopy: Microscopy is a process of viewing and analyzing micro-level objects, that can not be done through the naked eye. The glass slides that has been prepared with biological specimens will be viewed by the pathologist under the microscope to analyze the specimen. Before the emergence of digital pathology, only the glass-keeping mode was available. But now the data can be kept and transferred through digital mode which is a very efficient, effortless, and time-saving process. Now microscopes mounted with digital cameras are available through which pathologists can capture the digital images of the desired part of the slide. Among multiple types of scanners, Whole Slide Imaging(WSI) scanners are the most popular ones that can take images with atmost clarity.

Molecular testing: Here the diagnosis of disease will be done by examining the molecules present in tissues, body fluids, or organs. (Netto et al. 2003) This is an efficient method for determining the prognosis of a disease and gauging the effectiveness of treatment. Thus it can guide physicians to make correct therapeutical decisions to save the life of a patient. In cancer diagnosis especially in lung cancer kind of things, the identification of subtypes is too difficult by looking at the morphology. In such cases, molecular testing is the must-do test to distinguish the type. In the case of treatment plans like targeted therapy, the accurate identification of bio-markers is the most important factor. This can be done properly with molecular testing.

(Strimbu and Tavel 2010)The bio-marker can be a biological molecule, gene, protein, or other measurable characteristics that will be present in blood or other body fluids, whose presence or absence will indicate the normal/abnormal condition of a patient's body. It can be used to measure the response of a patient's body towards some therapeutics during the disease diagnosis. Biomarkers can be used to diagnose diseases, monitor disease progression, evaluate the effectiveness of treatment, and identify individuals who may be at increased risk for developing certain diseases. Some examples are blood glucose levels in diabetes, cholesterol levels in heart disease, and tumor markers in cancer. It can also be called as signature molecule or molecular marker.

But these methods have some drawbacks also. (Smeltzer et al. 2020)A survey has been conducted on molecular testing in lung cancer for this task. In this study, they have identified that the wide use of molecular testing has barriers like the expertise needed

to conduct the test, the cost of the test, its turnaround time, the raw materials needed to conduct the test, etc. So they have found a big gap in the implementation of molecular testing in the real world across the globe.

Identification of different tissues: This is the last but not the least step in any pathology analysis, which is the result part. Plenty of tissues will be present in each slide/sample and each of them has its own features. The pathologists identify the tissues present in the sample using the most appropriate techniques among microscopic analysis, molecular testing, Immunohistochemistry(IHC), or medical imaging techniques. The presence or absence of the tissues will let them conclude the process with correct diagnosis results. IHC is a laboratory test applied on Biopsy samples for precise diagnosis. Antibodies are used here to identify the presence of certain biomarkers or antigens in the tissue samples. Even though they are well-known for their accuracy, their value heavily depends on the quality of the antibodies used in the test, the expertise of the laboratory technician, etc. Also, it is highly expensive as compared with histopathology image analysis.

After the identification of tissues from the tissue samples, the diagnosis part will be over, and based on its results the further treatment plan can be decided for the patient.

1.1.3 Therapeutical Decision-making for Osteosarcoma and Renal Cell Carcinoma

Once the disease has been diagnosed, the proper treatment plan should be decided for that patient as soon as possible. However, this decision will be taken based on some prognostic factors. Prognosis is a condition of a patient that can be used to predict whether he/she will recover from a disease or whether the disease will return or recur. In the case of Osteosarcoma and Renal Cell Carcinoma, Neoadjuvant chemotherapy-induced tumor necrosis(dead tissues) is one of the most important prognostic factors. (Davis et al. 1994, Picci et al. 1994)Prognostic factors are the key elements in the selection of treatment plans for a patient in clinical practice. The term ”prognosis” refers to a forecast of what outcome will be produced by a medical condition.

Tissue necrosis is a state in which the death of cells or tissues in the body of a patient occurs. This is a condition where the tissues are unable to perform their normal metabolic functions and gradually die. Necrosis can occur in tissues of any body part of the patient. Its intensity can be measured based on the location and area affected. (Syntichaki and Tavernarakis 2002)The term has derived from the Greek Kernel, where ’necros’ means ’dead’. Necrosis may occur due to several causes. One reason can be the imbalance in nutritional supply and demand that occurred due to the outgrowth of the tumor. Some other factors are trauma, injuries, infections, exposure to toxins, stroke, and lack of oxygen or blood supply. Proper treatment should be given to this situation, or else it will spread to neighboring tissues. Despite of all these complications, necrosis

is purposefully created in the above-mentioned situation.

Neoadjuvant Therapy (NAT) is a therapeutic agent applied to a patient's body prior to the main treatment like surgery or radiation. This will help to reduce the extent of the main treatment by shrinking the tumor volume(Mauri et al. 2005). Studies have proved that it has a high impact on patients with inoperable conditions and locally advanced disease. NAC is one of the Neoadjuvant Therapy (NAT) methods during which therapeutic agents are applied on a patient's body before the main treatment like surgery or radiation. According to the American Cancer Society, NAC along with surgery shows better survival rates in patients affected with these cancers. The chemotherapy is injected through the vein and it causes the death of tumorous tissues. The amount of tumor tissues that got killed due to this chemotherapy will be accounted for and its total count will be taken as the prognostic factor of that patient. This quantity will reflect the response of the patient's body toward some therapeutics. The further treatment plan will be decided based on this response.

Post-neoadjuvant chemotherapy resection specimens will be used by the pathologists to do the diagnostic process. The term *post-neoadjuvant chemotherapy resection specimens* means, the biological specimen will be collected from a patient who has undergone NAC. The specimen has been fixed in formalin and cut up. Tissue blocks were taken from representative areas, processed and the slides were stained with Hematoxylin and Eosin stain. Then the image has been captured using a camera-mounted microscope such that it contains all the tissues required for this study. This slide may contain different kinds of tissues like tumors, tumor necrosis, fibrosis, mucous, bone tissues, etc. Among them, they have to find out the area spread by the tumor necrosis tissues. The quantification of necrosis will be done by the pathologists manually by observing the digital images of the slides. The percentage obtained by this process will be taken as one of the major prognostic factors in deciding further treatment plans.

1.2 Application of AI in digital pathology

The advent of digital pathology opens up new possibilities for creating algorithms and software tools to aid pathologists in clinical diagnosis and support researchers in investigating disease mechanisms. In the realm of computer vision, digitized pathology slides are often referred to as images and can leverage various image analysis algorithms. Computer vision, a subset of AI, encompasses algorithms capable of efficiently comprehending and interpreting visual data. For instance, tasks such as pathologists identifying and recognizing tissue components can be accomplished using image segmentation and recognition algorithms. The segmentation of cells and nuclei stands out as a crucial initial step in the automated analysis of digitized microscopy images(Saleena, Kutty et al. 2023).

The field of medicine, particularly the area of oncology has seen significant transformation due to artificial intelligence. The journey of AI has been clearly explained by Claudio Luchiniet al. from the beginning up to the current scenarios and the future perspectives (Luchini et al. 2022). It touches every stage of cancer treatment, including the initial screening, diagnosis, detection, classification, segmentation, bio-marker detection, strategies for follow-up, and up to the drug discovery (Kann 2019, Luchini et al. 2022) or even in biomarker identification (Wang et al. 2019).In the aforementioned stages, digital pathology has received a lot of attention in the last few years. In another study, the authors have performed an analysis of deep learning methods in selected digital pathology image case studies (Janowczyk and Madabhushi 2016). By name, they have selected nuclei segmentation, epithelium segmentation, tubule segmentation, mitosis detection, invasive ductal carcinoma detection, and lymphoma classification. They have used over 1200 digital pathology images for their evaluation. They have raised the issue of nuclei segmentation, where understanding all potential variations in morphology, texture, and color appearances is necessary before developing a nucleus segmentation method. Here the authors have developed a deep convolutional neural network (DCNN) that showed 71% accuracy in lung cancer classification (adenocarcinoma, squamous cell carcinoma, and small cell carcinoma) which was on par with that obtained by pathologists (Teramoto et al. 2017). In the case of lung cancer, subtype classification is the most tedious task, where pathology images are never suited for manual classification. So normally pathologists are using the Immunohistochemistry(IHC) test in the laboratory to do this task. If the pathologists managed to do this task using pathology images, the cost, time and effort could be reduced a considerable amount.

However, the persistent challenge in such scenarios is the insufficient availability of data, particularly pathology images. Various algorithms, such as one-shot or few-shot algorithms designed to operate effectively with minimal data, exist to mitigate this obstacle. Nevertheless, their effectiveness diminishes in pathology image segmentation and classification due to the intricate nature of morphology, making it challenging to discern accurately. Hence, meticulous data preparation becomes indispensable in studies of this nature.

1.3 Problem Statement

The amount of tumor necrosis created due to NAC can be used to measure the response of the patient's body towards the treatment and helps to measure the severity of the disease. This will eventually lead to decide further therapeutic decisions. So measuring the quantity of such dead tissues is very crucial in such scenarios. But there is no standard automatic method to do this task. The quantization of tumor necrosis is now performed manually by the pathologist with the help of histopathology images. They

are taking the slides with the help of a microscope or they will take digital images and manually calculate the percentage of necrosis present in it. The specimens they are taking for the procedure will be post-neoadjuvant chemotherapy resection specimens. This will contain different types of tissues namely, tumors, necrosis, fibrosis etc. Different challenges that the pathologists face during the clinical pathology examination are:

- Inter-class similarity
- Intra-class variability
- Inter-observer disagreement
- Time taking
- Labour intensive
- The number of features selected for the identification of tissues is very less in manual process

Inter-observer variability, or disagreement, refers to the disparities in results obtained by different pathologists examining the same input. This discrepancy can arise from the subjective nature of pathologists, variations in experience, workload differences, or manual errors. The morphological characteristics of tissues may not be consistent on every occasion and can sometimes exhibit overlapping features. Typically, various types of tissues may be dispersed throughout the specimen, with different classes sharing similarities in color, shape, etc. This can pose challenges for classification. In our study, we focus on HandE stained specimens, where Haematoxylin imparts a purplish-blue color to cell nuclei, Eosin provides a pink stain to the cytoplasm and extracellular components, and the remainder of the tissue appears as a combination of both. However, the color alone may not be sufficient for cell identification, as variations can occur based on the pigment used (Saleena, Ilyas, Sajna and Haque 2023).

Yet another significant challenge is the absence of an appropriate dataset for conducting research on tumor necrosis quantification or segmentation. Even when obtaining data, the annotation of pathology images proves to be an even more arduous task. Properly identifying tissue specimens and providing corresponding labels requires meticulous effort, ideally under the guidance of experienced pathologists. However, the heavy workload of pathologists leaves them with limited time to dedicate to research endeavors.

Digital pathology serves as a highly efficient substitute for the conventional microscopic examination of slides, proving advantageous for future reference and the seamless transfer of data. This technology has the potential to decrease the time required to report clinical cases, consequently minimizing labor efforts. However, the challenge associated with digital images lies in the clarity required for investigative purposes. This

clarity is contingent upon multiple factors, including staining issues, inherent noise artifacts from image acquisition tools, variations in tissue preparation, and differences in how the image is captured by various technical hands.

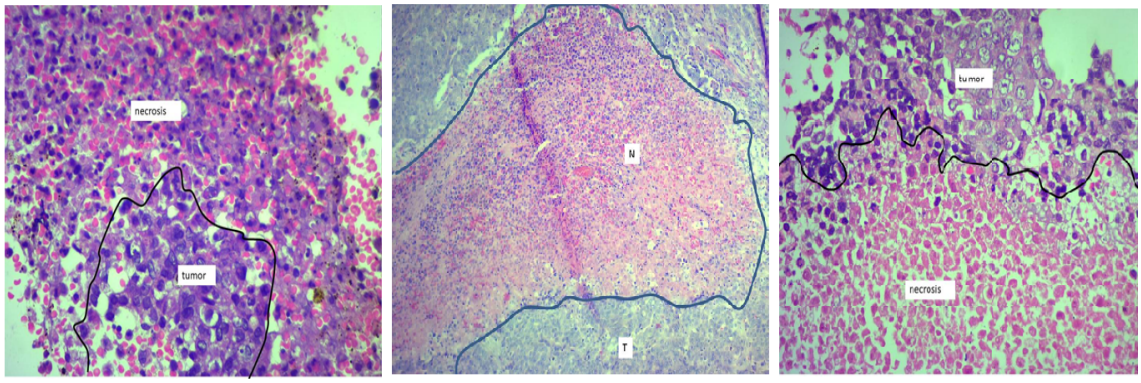


Fig. 1.3 Some of the samples from the dataset

1.4 Proposed System

To find the quantity of dead tumorous tissues created as part of Neoadjuvant chemotherapy from a given digital histopathology image, we have to identify that specific region only. Due to the complex morphology, it is a tedious and error-prone task for the human eye. Nonetheless, the application of deep learning algorithms presents an opportunity to develop automated tools for necrotic region segmentation. Particularly, neural networks within deep learning algorithms serve as effective alternatives to the conventional manual processes of analysis, segmentation, or classification. This not only alleviates the workload for pathologists but also allows them to focus on more critical tasks. Addressing inter-observer disagreement, a common challenge in medical analyses can be significantly mitigated through automation, achieved by training the model with sufficient and relevant data.

This work has introduced an AI-based automated system that can evaluate the response of a patient's body towards Neoadjuvant Chemotherapy (NAC). The digital histopathology images of post-neoadjuvant chemotherapy resection specimens are used in this study as the training data. The specimen has been collected from cancer patients who underwent neoadjuvant chemotherapy. The specimen will be taken from representative areas, and processed and the slides will be stained with Hematoxylin and Eosin stain. Then the image has been captured using a camera-mounted microscope such that it contains all the tissues required for this study. The dataset preparation is explained in detail at Chapter 3.

An AI-based system has been developed in this work to do image-level segmentation and quantization. The entire algorithm includes multiple sections to cut the image

into patches, segmentation of individual patches, merging multiple masks, and finally quantification. To cut the high-resolution image into patches of size 512x512, this work has used a Python library, 'patchify'. The output of the function will be multiple patches or sub-images of the original image in a specific size and each patch can be processed independently. The output will be in numpy array format that can be processed using suitable image-processing techniques. The function accepts 3 input parameters, input image as a numpy array, patch size as tuple of (width, height) and step-size which is the stride for getting non-overlapping patches. Usually, the patch size will be given as the step-size. Each patch can be undergone segmentation by the model created and will get the corresponding mask. A list of masks will be created from this and will be merged into a single output by preserving the order.

The proposed system includes an AI-based segmentation model that can automatically segment the area of necrotic tissues from a histopathology image. Here used U-Net++ for the semantic segmentation process with ResNet101 as the feature extractor. U-Net is one of the foremost neural network architectures which is the most suitable one for medical imaging inputs. The U-Net consists of a contracting or encoder path for feature extraction followed by a symmetric expanding or decoder path for precise localization. ResNet101 is used at the encoder side which has proved its excellency in feature extraction. The feature maps obtained can be used for creating a segmentation map on the decoder side.

The percentage of the segmented area present in the merged image should be calculated from the entire tissues present in the image and that value can be used to grade the tumor. The output from this tool can be used as a second opinion in the further treatment plan. We have created a dataset as part of this study that can be used to segment necrotic tissue areas in a digital histopathology image.

1.5 Objectives

This work aims to develop an AI-based solution for segmenting and quantizing tissue necrosis from post-neoadjuvant chemotherapy resection specimens. The output that they obtain from the software will be a prominent prognostic factor in the case of Osteosarcoma and Renal Cell Carcinoma. The result of this solution will help the pathologists and Oncologists to make critical decisions regarding the treatment plan of the patient.

This particular study is not limited to the solution making but we are also creating a dataset suitable for this study and making this publicly available. This dataset will be very helpful for those who are doing research in this area in the future, as data collection is one of the most tedious parts of deep learning.

The main objectives of this study are,

- To synthesize a dataset of histopathology images of post-neoadjuvant chemotherapy resection specimen and their ground truth or labels. This dataset deploys various morphologies of necrosis, tumors, fibrosis, and other frequently occurring tissue elements while quantizing the tissue necrosis in the case of post-neoadjuvant chemotherapy resection specimen
- To develop an AI-based segmentation and quantization tool to segment the tissue necrosis from histopathology images thereby to grade Osteosarcoma and Renal Cell Carcinoma patients, offering a second opinion for pathologists and doctors in critical treatment decisions.

1.6 Relevance of Necrosis-ML

Even though necrosis is created in every type of cancer, in Osteosarcoma and RCC the percentage of necrosis is very important to assess the prognosis of the patient. In other words, how the disease has responded to chemotherapy can be identified based on the percentage of necrosis created, and ultimately this will help to understand whether the patient will survive or not. Necrosis-ML is an encoder-decoder approach for semantic segmentation and quantization of tumor necrosis using histopathology images. In the manual process, the microscopic view of the specimen will be converted into a digital image. The amount of necrosis created due to NAC will be identified by looking at those images. When pathologists measure the necrosis percentage manually, inter-observer variability will be there and it is very time-consuming also.

There is no specific tool to measure the necrosis of cancer. But almost similar types of tools are owned by multi-national companies which is very expensive. Even though they are used to diagnose the cancer, their working principle is different from Necrosis-ML. They are not specifically meant for necrosis quantization which is crucial for some prognosis factors. The input and output of Necrosis-ML are explained in Figure 1.4. It takes histopathology images of post-neoadjuvant chemotherapy resection specimens as input and outputs the necrosis-segmented masks along with the area of the segmented part. The stakeholders of this tool are the pathologists who are now quantizing the amount of necrosis created through manual analysis. They can use this tool to make the analysis more easier and error-free.

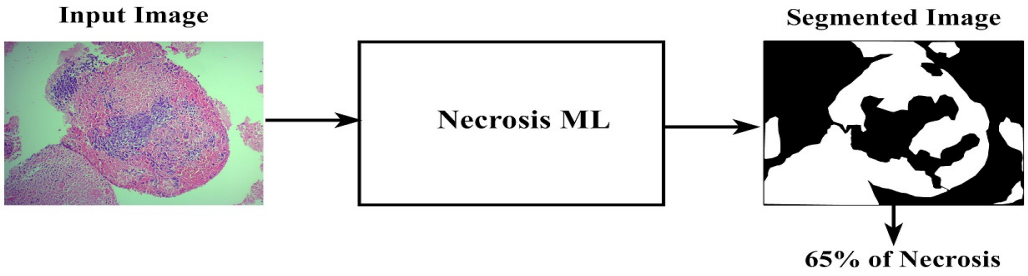


Fig. 1.4 How Necrosis-ML works

There is currently no adequate dataset accessible to conduct any research on the quantification or segmentation of tumor necrosis. And even if we were able to acquire such data, annotating pathology images would pose significant challenges. It's a laborious task to identify appropriate tissue samples and label them correctly, requiring the guidance of experienced pathologists. However, their workload is already substantial, they have insufficient time to devote to research. This study will be a pioneer to this particular case of research.

1.7 Outline of the thesis

The thesis has two important sections, one is dataset preparation and another is the software solution for the problem above mentioned. In the solution section, there is patch-level and image-level segmentation and quantification methodologies.The Chapters, 4 through 6 is about different segmentation methodologies used in this research, that can be incorporated with Necrosis-ML. The chapters of the thesis have been organized as follows:

Chapter 2: This section includes a detailed review of the prominent works that have been conducted in the relevant areas, so far. We explored the background and significance of tumor necrosis, emphasizing its clinical relevance. It examines various methods for dataset generation in pathology stream and image annotation tools critical for creating high-quality data. The review delves into different semantic segmentation techniques, highlighting their importance in analyzing tissue samples. It also investigates works done on segmentation of pathology data using autoencoders, pre-trained networks, and transformer models.

Chapter 3: For every deep learning project, dataset preparation is the most tedious and time-consuming process. Chapter 3 of this thesis gives the details about how the dataset has been generated in this study. It also includes meta-data for the dataset created and the overall workflow of Necrosis-ML for the better understanding of readers.

This chapter describes from where the data have been collected, who are the domain experts, how the data have been converted to trainable format, what functions, tools, and libraries are used for this synthesis, etc.

Chapter 4: Different deep learning methodologies are applied to the dataset that we have created, to get precise output. The first methodology is discussed in this chapter, in which we have used an autoencoder model that has been trained with a single histopathology image. Rather than taking the entire dataset, a model has been built using a single image. Convolutional autoencoder has been preferred for this particular study. This model can be used in the segmentation part of Necrosis-ML.

Chapter 5: Second methodology is discussed in this section, which is the patch-level segmentation of histo-pathology images. An encoder-decoder architecture has been used in this method which is called SegFormer. The encoder side is set up with the Transformer and the decoder side contains additional convolutional layers. Specifically,*nvidia/mit-b5* has been used as the pre-trained model in this work. This model can also be used in the segmentation part of Necrosis-ML.

Chapter 6: This chapter describes the working of Necrosis-ML in detail with both segmentation and quantification modules. Transfer-learning technique has been used in this study for segmentation and image processing methods have been used for the quantification. The histopathology images of post-neoadjuvant chemotherapy resection specimens have been used as the input images for the system. A comparison of two pre-trained networks namely DeepLabv3+ and U-Net++ has been performed initially and U-Net++ has chosen for later system building due to its relatively better performance. This is an end-to-end work that can be utilized as a real-world solution.

Chapter 7: Different proposed segmentation methods, their results, and analyses have been explained in this chapter. Each method has pros and cons in its own way. Along with Necrosis-ML, we can use any of the segmentation methods that show better performance but in this study, we have ended up with the decision to select the pre-trained network as it showed good results with this dataset.

Chapter 8: The conclusion of this entire work is explained in this chapter.

Chapter 9: The research is a never-ending process. In most of the works, the future scope of the study will be there. Necrosis-ML and its associated dataset, have some limitations. Its advancements and further scope have been explained in this chapter.

CHAPTER 2

Review of Literature

2.1 Introduction

A literature review is a write-up of the previous research and works performed in an area of interest. Reading of literature will help us to identify the research gap and thereby the research opportunities in any area. A detailed review of the literature in the area of this research has been given in this chapter. Identification of the most relevant papers is also one of the important parts of this section. As this work is a solution for finding out a particular kind of tissues from histo-pathology images using deep learning algorithms, this literature review has included the papers from both the domain side and technical side.

Cancer has a leading position among the dreadful diseases in the world and it is the second leading cause of death in the world. It is the abnormal condition of body tissues where abnormal cells develop and divide uncontrollably. Such a situation will cause damage to normal cells as they are cramped inside the patient's body. This abnormality of cells is due to the mutation occurring to DNA inside such cells. DNA is the component of the cell that decides what function the cell should do(Clinic 2022). Every type of cancer will have distinct treatment plans based on different prognostic factors. The subsequent sections in this chapter describe various studies regarding these factors and deep learning solutions for such cases.

2.2 Review on Tissue Necrosis as a Prognostic factor

Necrosis is a stage of uncontrolled death of body cells. In the olden days, Apoptosis or programmed cell death and necrosis were the only two reasons known for cell death. However recent studies proved that there are different causes to this phenomenon. Electric shock, exposure to radiation, biological and chemical agents, extremely high temperatures, hypoxia, immunological reactions etc are some of them. Apoptosis is the morphological changes to the structure of cells and it causes minimal damage to surrounding tissues. But the necrosis results in spillage of the contents of the cell and also results in severe damage to the surrounding tissues. A detailed review about various

cell deaths (necrosis, apoptosis, oncosis, pyroptosis and autophagy) has been given by Mark S D'Arcy in his study (Golstein and Kroemer 2007, Ruffolo 1964, D'Arcy 2019).

Neoadjuvant Therapy (NAT) is therapeutic agents applied on a patient's body prior to the main treatment like surgery or radiation. This will help to reduce the extent of main treatment by shrinking the tumor volume (Mauri et al. 2005). Studies have proved that it has a high impact on patients with inoperable conditions and locally advanced disease. A meta-analysis has been performed by Davide Mauri et al. to express how the NAT made responses in breast cancer patients. They have compared the samples from patients who underwent preoperative therapy (Neoadjuvant) and post-operative therapy (adjuvant).In the case of breast cancer, complete pathological response to NAT is a good prognostic factor (Caparica et al. 2019). Neoadjuvant chemotherapy (NAC) is one of such treatment methods.

2.2.1 Neoadjuvant Chemotherapy response in Osteosarcoma Patients

According to the American Cancer Society, NAC along with surgery shows better survival rates in patients affected with Osteosarcoma. The chemotherapy is injected through the vein and thus it helps to reduce the chances for metastasis than performing surgery alone. This NAC will be injected into the patient's body for almost 10 weeks. Adjuvant chemotherapy for almost one year also, prescribing for such patients. Disease-free survival has been achieved in 60-70% patients through this kind of multi-modal approach (Carrle and Bielack 2006). It also reduces the chances for local occurrence of the disease. Poor response to NAC is one of the major negative prognostic indicators. Two groups of Osteosarcoma patients have been compared to find the clinical efficacy of NAC (Zhu et al. 2019). One group has been subjected to chemotherapy and then performed surgery after 3 weeks. Another group receives no chemotherapy. The study has been concluded that the first group had an improved survival rate, progression-free condition and improved limb salvage rate. A meta-analysis has been performed by Marnie Collins et al. to find out the pros and cons of receiving NAC in different age groups and sex groups (Collins et al. 2013). The study revealed that age is not a significant factor in this case and chemotherapy-induced tumor necrosis ends up with higher survival rate. But the survival rate of females is greater than that of male, the same way children have a higher rate than adults. For the study, they have included the specimens of the year between 1979 and 2005. In another study, the authors claim that the histologic response of Osteosarcoma towards chemotherapy is one of the inevitable prognostic factors. Also, they have proved that a threshold value of 90% for tumor necrosis contributes much more to the survival rate of the patient than the rest. In that paper, they have provided all the confidence intervals of survival rates properly (Richardson et al. 2022).

2.2.2 Neoadjuvant Chemotherapy response in RCC Patients

In advanced Renal Cell Carcinoma cases, Nephrectomy is the reference standard therapy. But in unresectable cases, neoadjuvant cabozantinib is a good choice to avoid the risk (Roy et al. 2020). Neoadjuvant cabozantinib is a chemotherapy drug that is given to advanced RCC patients and studies proved that it can make the tumor volume to shrink greater than 50% within 4 months and then can permit potentially curative surgical resection. Cabozantinib is a tyrosine kinase inhibitor (TKI) which is approved by FDA and European Commission and characterized by improved progression-free and overall survival rate and good objective response rate (Escudier et al. 2016). It is considered to be the most efficient first-line and second-line therapy in RCC patients. A review has been performed by Al-Salama et al. on the importance of Cabozantinib in advanced RCC patients. They concluded that it is an inevitable option in patients who have received prior antiangiogenic therapy(Al-Salama and Keating 2016). It showed better accuracy than another inhibitor everolimus(Tannir et al. n.d.). Even though NAT plays a prominent role in cancer treatment it has some disadvantages also, like potential delay in actual treatment, increased cost and toxicity, obscuring of pathological staging etc (Trimble et al. 1993).

2.3 Review on digital histopathology dataset generation

The primary challenge in every deep learning algorithm is the need for an extensive amount of data for model training. With a limited number of experienced pathologists and an overwhelming workload assigned to them, the process of data collection in pathology becomes particularly strenuous. While various online repositories, such as TCGA, offer free access to data, the majority of the available data consists of Whole Slide Imaging (WSI) images. These WSIs, characterized by very high dimensionality (potentially ranging from 100k to 50k resolutions), cannot be directly input into neural networks. A solution to this problem lies in synthetic data generation methods, with patchification being a particularly valuable approach. Numerous studies have been conducted on the patchification process to address this issue.

Pathologists are differentiating the cancer subtypes by looking at the cellular-level morphology in patch-scale level of digital pathology images(Hou et al. 2016). So it is better to train the model in patch-scale rather than at image level, if the image is of higher dimensions. In this study, Le Hou et al. have collected the WSI images and have created around 1000 patches per image per scale each with size 500X500 and with 20X and 5X magnifications. Figure 2.1. shows what the WSI image and its patches look like(Hou et al. 2016).

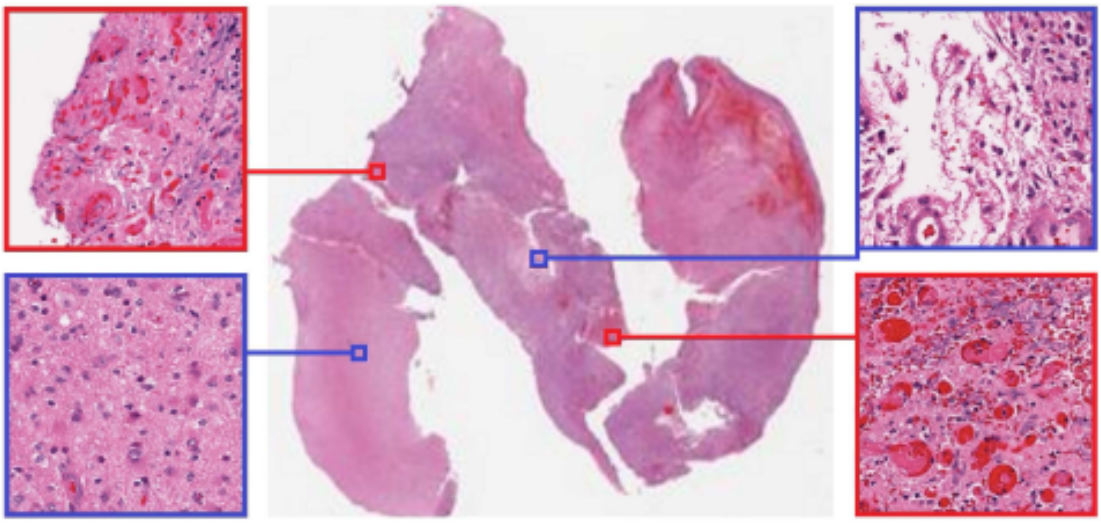


Fig. 2.1 WSI image is at the center and its patches at both sides.

In cancer diagnosis, morphology and distribution of nuclei are the essential markers. In this study, the authors have created a dataset of segmented nuclei from digital histopathology images which are stained with hematoxylin and eosin(Hou et al. 2020).

2.4 Image Annotation and Associated Tools

Image annotation is the process of labeling images for training deep learning or machine learning models. Through the labeling, the region of interest of the image can be correctly identified by the algorithm. This is far more accurate than the classification task. Most of the pathology analysis will be done based on the number, morphology, and position of individual cells present in the given sample. However, the annotation of individual cells, which may be thousands in number in each of the images is a herculean task.

Using different annotation tools we can do the manual labeling of the cells. If different types of cells are there, we can mark them with different colors. Nowadays different open-source tools, libraries, and packages are available to make this task easier. Some of them are given below:

Computer Vision Annotation Tool(CVAT): CVAT is an open-source web-based annotation tool that is available at ”www.cvat.ai”. This tool has been meant for image and video annotation or labeling.

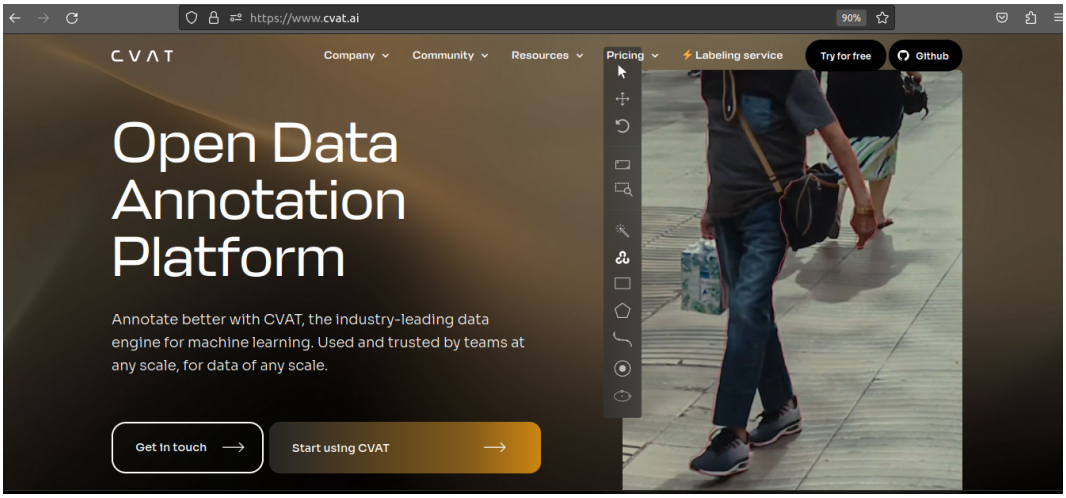


Fig. 2.2 Interface of CVAT

CellProfiler Analyst: CellProfiler Analyst is an example of this kind of tool. This tool is available at the link https://cellprofiler.org/. They provide packaged applications for any OS like Windows, Mac, and Linux(Dao et al. 2016).

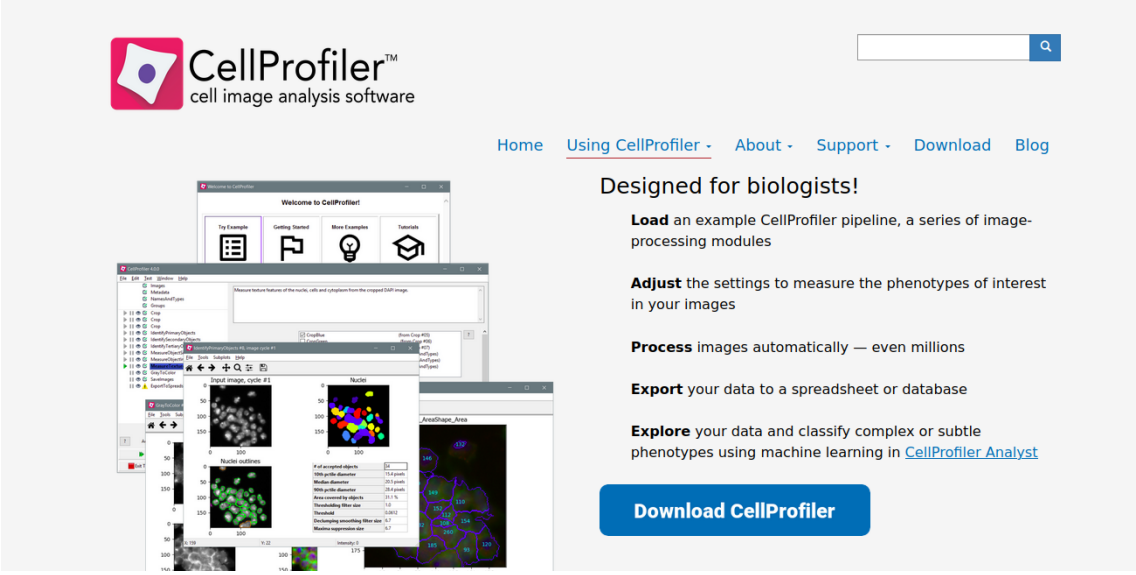


Fig. 2.3 Interface of CellProfiler Analyst

Ilastik: Another tool Ilastik, is a free and open source image segmentation and classification tool. It is mainly used for medical and biological images but can be made use in materials science, geology, and engineering. It also provides support for machine learning tasks along with annotations. It is coming under the terms of the GNU General Public License and available at www.ilastik.org (Sommer et al. 2011).

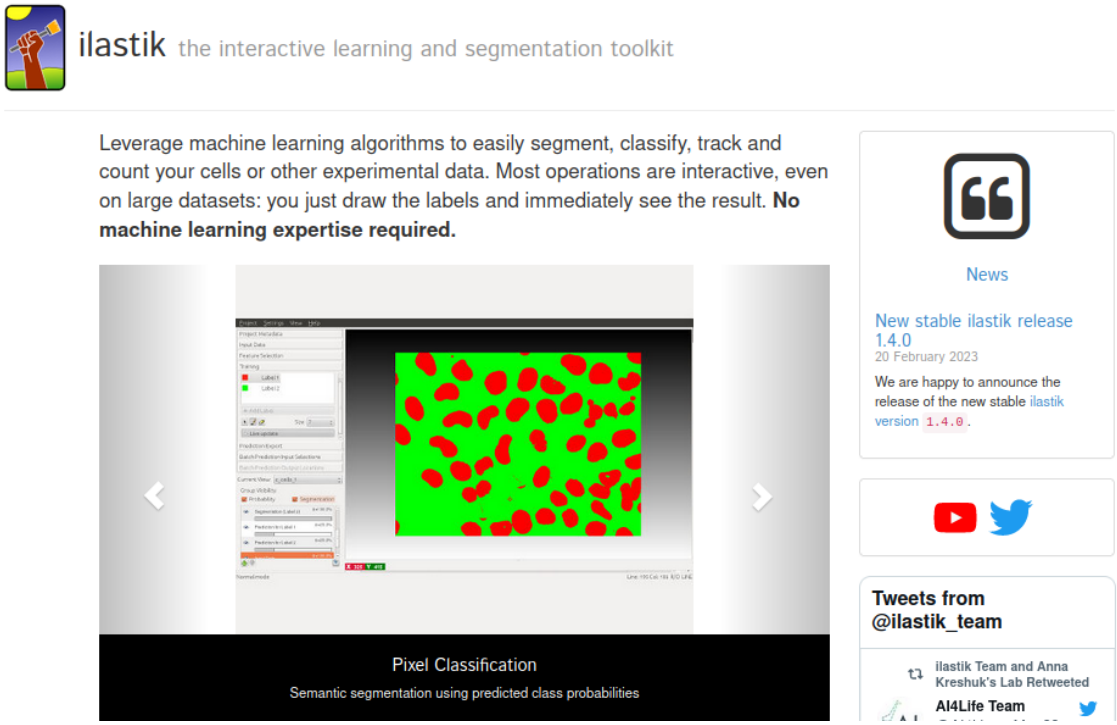


Fig. 2.4 Interface of Ilastik

QuPath is an open-source image analysis tool mainly for digital pathology and whole slide images (Bankhead et al. 2017). The most striking feature of whole slide scanners is their ability to capture microscopic images in at most clarity. It is becoming increasingly apparent that acquiring high-resolution digital scans of entire microscopic slides using high-resolution whole slide scanners is transforming tissue biomarker discovery and companion diagnostics by automating, quantifying, and objectively screening tissue samples utilizing digital image analytics.

Advanced Cell Classifier(ACC): Advanced Cell Classifier(ACC) is a data analysis software tool by which biological cells can be classified automatically from cell based or tissue based specimens(Piccinini et al. 2017). Using advanced machine learning methods, this tool aims to provide very accurate analysis with minimal user interaction. It can be used for a wide range of applications in cell biology, including drug discovery, toxicology testing, and cancer research. ACC is a versatile and customizable tool that can be used to define the features and the classification criteria for users, as well as to train and validate classifiers based on the data sets that they provide. The source code of ACC v2.0 is freely available at ww.cellclassifier.org.

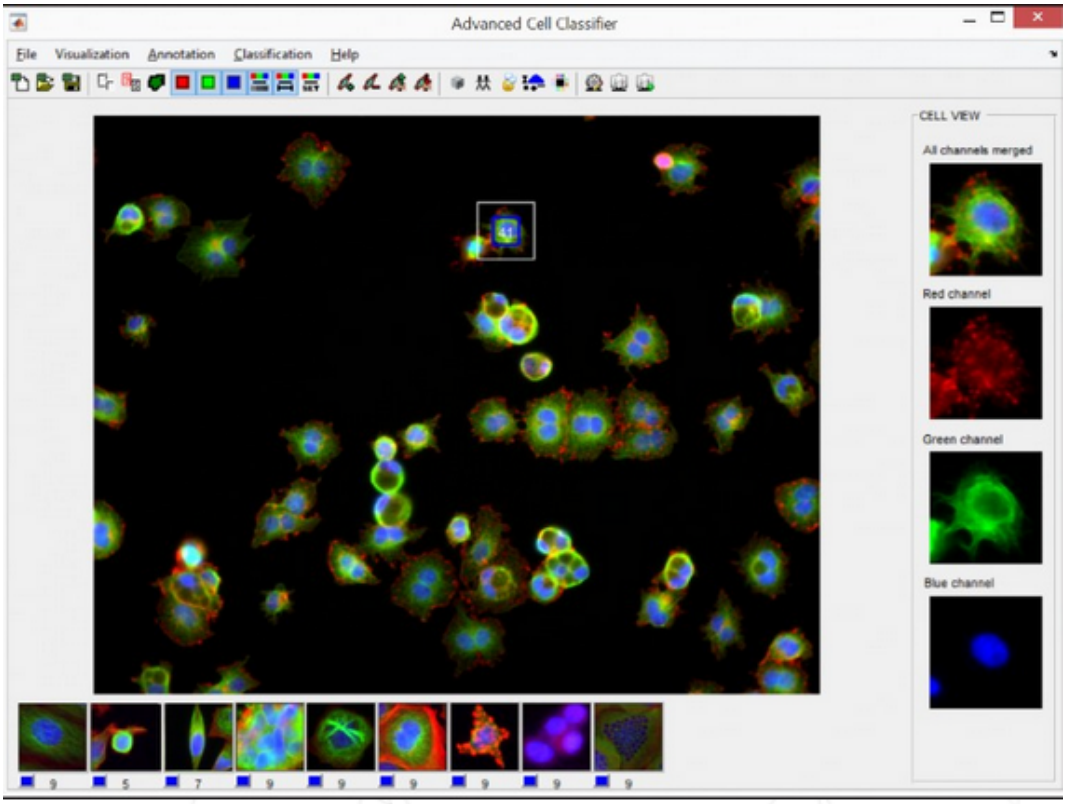


Fig. 2.5 Interface of Advanced Cell Classifier

The authors have introduced a Python library known as 'Keras R-CNN', for detecting the individual cells in digital pathology images using neural networks(Hung et al. 2020). They have demonstrated the working of this library on two case studies, nucleus detection and malaria stage classification. In the latter case, the result obtained from this computation was benchmarked with human annotation and got comparable results. This tool identifies the cells automatically for brightfield, fluorescence images, and large image sets. The package is freely available in GitHub and the Broad Bioimage Benchmark Collection.

2.5 Review on Semantic Segmentation

Image segmentation is the process by which we can grasp the region of interest from a particular image. We can use it for identifying and classifying different categories of an image (Walia 2022). It can be object detection, boundary or shape detection, morphology detection etc. Figure 2.6 shows how an image will be segmented (Zhang 2022*a*).

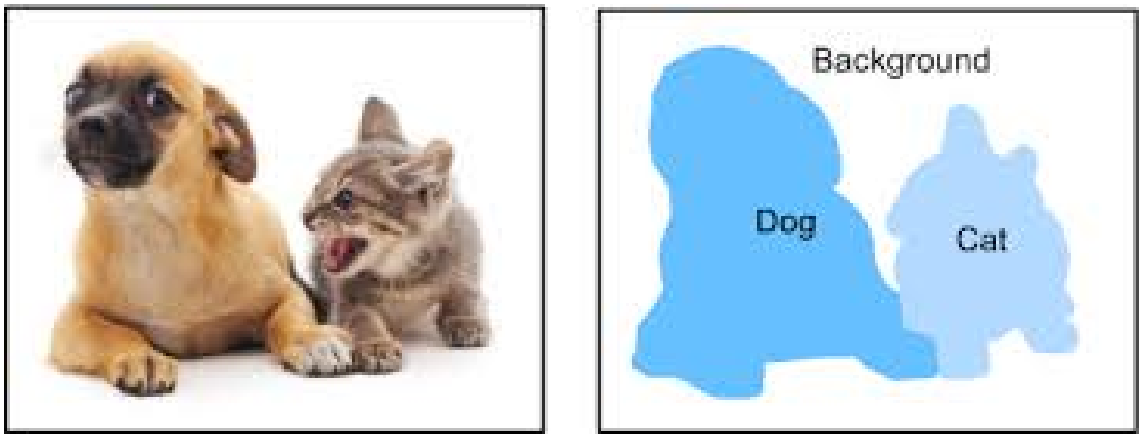


Fig. 2.6 Each of the objects in the original image gets segmented on the right picture.

Three major types of image segmentation are,

- Semantic segmentation
- Instance segmentation
- Panoptic Segmentation

Semantic segmentation makes use of deep learning algorithms for labeling every pixel with its own category. The grouping up of pixels will be based on, which semantic class a particular pixel belongs to. In another sense, semantic segmentation is a classification task itself but at the pixel level.

Whereas, the instance segmentation identifies which instance of the object is that, along with the category. so additional processing of finding the instance along with pixel-wise classification will occur there. So there may be multiple objects with the same label in this case. This procedure will help us to identify each instance separately and we can count the number of instances present in the image in category wise. This case is depicted in Figure 2.7 (Walia 2022).

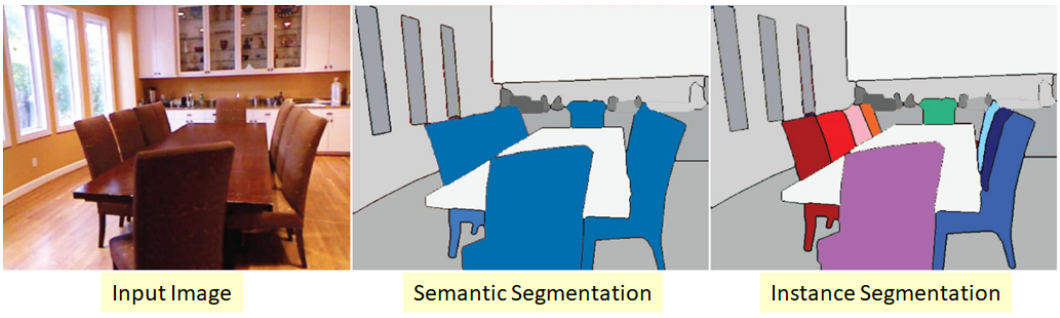


Fig. 2.7 How an image gets segmented using semantic and instance segmentation

Panoptic segmentation is a complex computer vision method that simultaneously handles both semantic and instance image segmentation (Kirillov et al. 2019). It provides a comprehensive understanding of scenes by assigning semantic labels to stuff (e.g., background) and instance labels to things (e.g., objects).

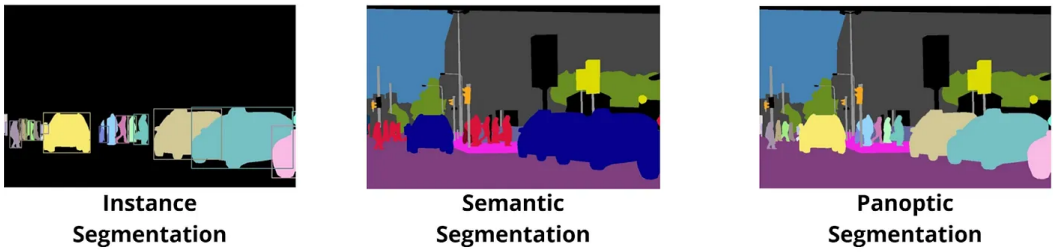


Fig. 2.8 Example for different types of segmentation

The importance of semantic segmentation has been increased due to its wide use in medical image segmentation. In the case of digital pathology, it is the process of grouping the voxels that belong to the same tissue structures according to their morphology.

Segmenting images, particularly histopathology images with a large number of dimensions, is a highly time-consuming and computationally intensive task. When it comes to Whole Slide Imaging (WSI) images, it is impossible due to their extremely large gigapixel size. Therefore, patch-level processing is the most suitable approach in such instances. In clinical pathology, analyzing tissue at a microscopic level is critical for accurate diagnosis. Furthermore, as the number of dimensions in the images increases, the neural networks' computational requirements become more substantial.

A patch-level tissue segmentation model has been introduced by Han et al., that has very little annotation and computation effort. Through this work, they have raised the issues in fully-supervised segmentation despite of their good accuracy. This is a two-step model where a classification part is followed by the segmentation process. In the first phase, pseudo labels at patch level will be created. This is very effortless and time-saving as compared with manual labeling. In the second phase, they used Multi-Layer Pseudo-Supervision. They have also developed a weakly supervised dataset of lung adenocarcinoma(LUAD) (Han et al. 2022).

Since the cancer subtypes are distinguished based on cellular-level morphological factors, the patch-level classifier will perform better than the image-level classifier. But the complication that comes with classification is that the normal downsampling process may cause the loss of several pieces of information from the image and also the ground truth labels of individual patches are unknown, as only the image-level ground truth label is given. So a novel Expectation-Maximization (EM) based method has been introduced by Le Hou et al. in which discriminative patches will be automatically located. The network was trained using CAFFE toolbox on a single NVidia Tesla K40

GPU (Hou et al. 2016).

A state-of-the-art method, named as 3D U-Net has been introduced for the segmentation of volumetric images (Çiçek et al. 2016). Volumetric images include common medical images in 3D modality, like CT, MRI, PET etc. Such kind of images will be more effective for diagnosis purpose. Usually the annotation of 2D images is not so difficult as it contains only 2 dimensions. However, in volumetric data, it will be very difficult to create annotated data. To overcome this situation, 3D U-Net has used sparsely annotated volumetric images to train the model. In sparse annotation, only a bunch of pixels will be get annotated. This is a best alternative for fully annotated dataset creation which is a more expensive, time-consuming, and laborious task. Such kind of annotation using attention-guided active learning has been introduced by Zhenxi Zhang et al. which is specifically done for volumetric medical images(Zhang et al. 2019). Through their study, they have proved that only 15-20% annotation on slides is required to do the brain extraction task whereas 30-35% for the tissue segmentation task. Densely annotated and sparsely annotated examples are given in Figure 2.9 (Bokhorst et al. 2019).

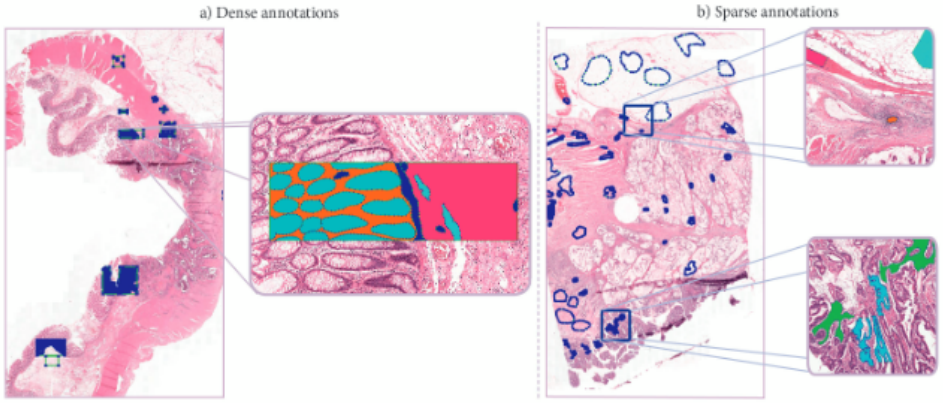


Fig. 2.9 Densely annotated image(left) and sparsely annotated data.

An alternative for the time-consuming process of annotating WSI images has been introduced in this paper as a region-based active learning (AL) approach(Qiu et al. 2023). AL involves training the model on a limited number of annotated image regions, iteratively selected to optimize model performance while minimizing annotation efforts. The conventional method selects informative square regions of a specified size, but its efficiency depends on the AL step size, which can lead to redundant annotations or increased computation costs. The paper introduces a novel adaptive technique that dynamically determines informative areas and their optimal bounding boxes instead of predefined shapes, resulting in higher sampling efficiency across various AL step sizes. The method is evaluated on breast cancer metastase segmentation and consistently achieves better efficiency than the standard approach, reducing annotation costs

significantly. Breast cancer is a significant cause of death for women worldwide, regardless of the economic status of the country(Zeb et al. 2023). While histological diagnosis remains the gold standard for cancer detection, it can be time-consuming and labor-intensive. To address this, researchers have proposed an automated and precise histopathology image analysis system using deep learning. They utilized CNN-based architectures to extract features from images and employed a fully connected layer for classifying malignant and benign cells, along with their subclasses. The effectiveness of the system was tested using benchmark datasets, resulting in an impressive accuracy of 94% and an F1 score of over 90%.

2.6 Nuclei Segmentation

One of the main challenges in clinical pathology is the identification of nuclei from the tissue segments. Many of the pathological evaluations, including the identification of numerous biomarkers, depend on the ability to quantify cell nuclei in histological images(Höfener et al. 2018). Accurate nuclei segmentation is the pre-requisite for the cancer diagnosis, prognosis, grading, cell classification etc. As each image contains millions of nuclei, the manual segmentation will be beyond the bounds of possibility. But on the computational side, different approaches are there to do this task like thresholding, graph-based method, Watershed segmentation, machine learning and deep learning based approaches etc. Some of the examples for nuclei segmentation is given in Figure 2.10 (Veta et al. 2013).

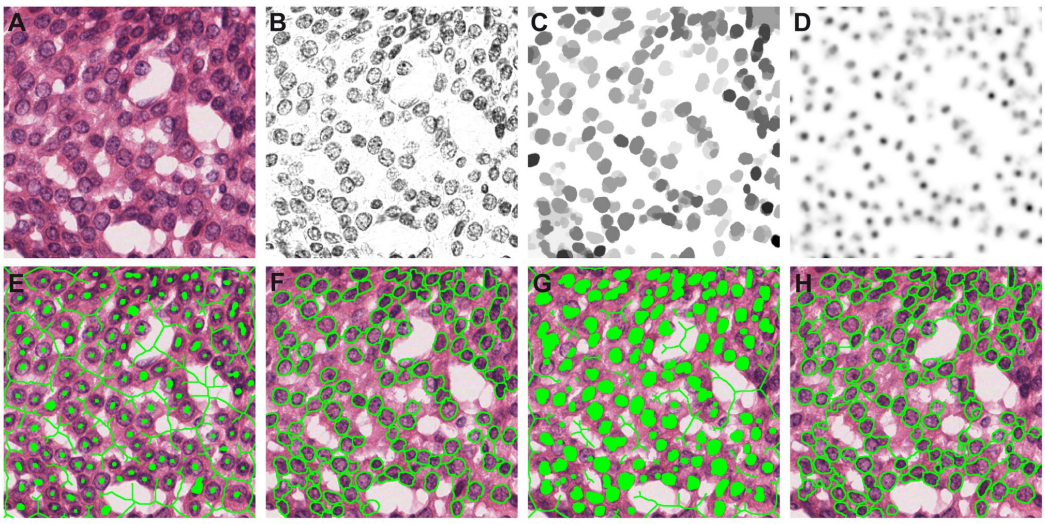


Fig. 2.10 How the nuclei get segmented.

The automatic detection of nuclei using normal machine learning and image processing techniques couldn't perform well due to variations in the appearance of nuclei

as the staining and tissue preparation changes and its unpredictable morphology. But the emergence of deep neural networks brought changes to this situation. A fully-automated deep learning approach for nuclei segmentation has been introduced by the authors and they have published their related dataset(Naylor et al. 2017). A marker imposition and watershed segmentation method for the nuclei segmentation in Hand E stained histopathology images of breast cancer has developed in 2013(Veta et al. 2013). Another work on the same domain has reported in a work where a wrapper feature extraction method has been proposed for better discrimination of breast cancer tissues from normal one(Wang et al. 2016). Hofener et al. has used PMap approach to find the presence of the nucleus center(Höfener et al. 2018). They have identified different essential parameters that can boost up the basic PMap approach and made a state-of-the-art method for identifying the nuclei. Post-processing of the PMap was one parameter among them that can improve the detection quality. They have configured two network architectures whose F1-measures were 0.816 on colorectal cancer and 0.819 on breast cancer respectively.

2.6.1 Segmentation of Whole Slide Imaging(WSI) images

The major obstacle in every deep learning algorithm is the need for a substantial amount of data for effective model training. Given the scarcity of experienced pathologists and their overwhelming workload, the collection of data in pathology has become increasingly challenging. However, this challenge can be significantly alleviated through the use of synthetic data generation methods, with patchification emerging as a particularly valuable approach.

Public repositories are somewhat of a blessing, as direct data collecting is the most time-consuming and nearly impossible phase in AI, especially in the case of medical data. WSI images are a good source of digital pathology images that are captured using WSI scanners. Along with its wide use in the field of digital pathology, it enables the digitization of pathology workflows, enabling remote consultation and collaboration. As the name suggests, they used to take the image of an entire slide. They are well-known for their high-resolution images. Such scanning method has several pros and cons as follows. Some of the advantages are given below:

- Improved efficiency: WSI images allow for rapid and efficient scanning of pathology slides, reducing the time required for manual analysis.
- Enhanced accuracy: Due to the high picture quality, it helps the algorithms to improve its accuracy where this images can be used as the input.
- Increased accessibility: The electronic storage and sharing of WSI images with

other pathologists and medical staff allows for remote consultation and collaboration.

- Reduced storage and retrieval costs: Digital storage of WSI images is more cost-effective than traditional slide storage and can be easily retrieved for future reference.
- Improved patient outcomes: The use of WSI images in digital pathology can lead to faster and more accurate diagnoses, leading to improved patient outcomes.

Along with several pros, there are some cons also:

- High resolution: Due to the high resolution, this kind of image can't be used directly in the deep learning algorithms. So they have to be divided into small patches so that they can be processed using neural networks.
- Cost: Scanning and storing large WSI images can be expensive and require specialized equipment.
- Technical difficulties: Interpreting WSI images can be challenging, and there may be technical difficulties in viewing or sharing images due to differences in software and hardware.
- Training and expertise: Pathologists and healthcare professionals may require additional training and expertise to effectively interpret and use WSI images.
- Quality control: Ensuring the quality and accuracy of WSI images requires careful attention to detail and a rigorous quality control process.
- Regulatory barriers: There may be regulatory barriers to the use of WSI images in some countries, which can impact the widespread adoption of this technology.

By neglecting all these cons, WSI is used worldwide in digital pathology. The main cause for this is the scarcity of digital pathology images and the deficiency of experienced pathologists. In most of the pathology-related research, the data are taken from online repositories. One of the main sources of WSI images is TCGA or The Cancer Genome Atlas. It was launched in 2005 as a joint venture by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). It serves as a sizable open resource for biospecimens to link researchers around the globe. It includes specimens of about 33 types of cancer.

In this study, the authors have introduced an online platform called The Cancer Digital Slide Archive (CDSA) for the integration and visualization of Whole Slide Images (WSIs), which can be accessed publicly at the URL http://cancer.digitalslidearchive.net (Gutman et al. 2013). Alongside the images, the platform provides metadata such as

age of diagnosis, mutation status, and more. Given that a WSI can have 40X magnification, specialized software is necessary for processing, and several options include open-source libraries like those from the Open Microscope (OMERO) environment, Zoomify, BIRN pathology workbench, DeepZoom, etc. The dataset has been acquired from TCGA in SVS format, converted to TIFF format, and thumbnails are generated using the VIPS library. Human annotations have been performed using the Aperio ImageScope client.

A detailed review of segmentation methods applied to WSI images has been done by Shidan Wang et al. (Wang et al. 2019). They have explained the underlying processes of segmentation of such images. When an image has been taken as the input to the model, it will be divided into patches of convenient size and after segmentation, all these patches combined to become a single image. The images that we have taken don't need to be of the same size, but the patches that we are making from it should be of the same size to speed up the algorithm. For the segmentation process, we have to create the labels of the image by manually annotating the image. So much software is available for doing this task.

A total of 5,060 Whole Slide Images (WSI) representing 10 distinct cancer types were gathered from TCGA in another study(Hou et al. 2020). Nuclei segmentation was conducted in this study, involving the annotation of 5 billion nuclei and the manual segmentation of 1,356 image patches from these datasets. The distribution, appearance, shape, and morphology of nuclei play a pivotal role in cancer diagnosis. For this dataset, both Mask R-CNN and manual annotation by three pathologists were employed, and the results were compared using the Dice-coefficient metric. The outcome closely resembled the inter-annotator variability observed in manual segmentation.

For the pathology image segmentation process, a model that performs either semantic or instance segmentation should be selected based on our needs along with proper loss function(Wang et al. 2019). The authors of this study had curated a novel dataset, LC25000, comprising images of lung and colon carcinoma. Each of the five classes within the dataset consists of 25,000 images. These images were generated from a pool of 750 lung tissues and 500 colon tissues(Borkowski et al. 2019). The original images, sized at 1024 x 768 pixels, were resized to 768 x 768 pixels. Data augmentation was applied using the Augmentor software package, involving left/right rotations and vertical/horizontal flips to enhance the dataset.

FCN or Fully Convolutional Network is a pixel-to-pixel semantic segmentation neural network and it is the first in its kind(Long et al. 2015). In this architecture, they are making use of pre-trained classification networks for the segmentation process. The last fully connected layer in CNN is replaced with a deconvolutional layer to efficiently classify each pixel. In any kind of deep neural network, the input image will be given to the first layer which is the input layer(Long et al. 2015). The last layer or output

layer will calculate the output for the whole model. The layers in between input and output are called hidden layers because they are not getting any values directly from the user and not provide output directly to the user. Features will be extracted from each layer and output of each layer will be fed as the input to the next layer. The deep neural networks are so-called because of the depth of the network or the number of layers present in it. The idea of CNN has derived from the concept of the visual cortex, where instead of feeding with entire parameters from the previous layers a receptive area or a convoluted feature matrix will be taken. Depending on the number of parameters we require, we can adjust the size of the convolution kernel. As CNN is choosing the features based on spatial factors, the model's performance won't be declined based on the staining variations or hand-crafted features.

The subtypes of cancers will be differentiated by analyzing the cellular-level visual features obtained from image patch scale(Hou et al. 2016). This work has proved that the accuracy of the model is higher when the input is taken at patch-level rather than image-level. Due to the very high dimensionality nature of WSIs (maybe of 100k to 50k resolutions), it can't be directly fed to the CNN. This study clearly describes the difficulties of using WSI images directly as the input to the CNN. Extensive loss-less downsampling is needed in such cases. Le Hou et al. have created around 1000 patches per image per scale each with a size of 500X500 and with 20X and 5X magnifications. They have proposed a novel method, patch-level CNN, which is a decision fusion model to aggregate patch-level predictions.

Convolutional Neural Networks can be used in the processing of images, video, speech, and audio whereas Recurrent Neural Nets are for text and speech only(LeCun et al. 2015). 1634 WSIs from TCGA have been taken for the creation of a dataset that included Lung Adenocarcinoma(LUAD), Lung squamous cell carcinoma (LUSC), and normal tissues(Coudray et al. 2018). 512X512 non-overlapping patches have been taken from each slide ranges tens to thousands of patches from each slide. In this work, they have made a deep neural network based on *inception v3* and they have achieved an average area under the curve (AUC) of 0.97. The most highlighted part of this work is that this model can predict six mutated genes of LUAD, which are STK11, EGFR, FAT1, SETBP1, KRAS and TP53.

2.7 Review on Autoencoders

Autoencoder is one of the finest unsupervised neural networks. They were initially used for image reconstruction but are now widely used for noise reduction, segmentation of text/image, domain adaptation, image/video compression, feature extraction, recommendation system, dimensionality reduction, etc. There are different types of autoencoders like Convolutional Autoencoder, Variational Autoencoder, Denoising au-

toencoder, Sparse Autoencoder, Deep Autoencoder, Contractive Autoencoder, Undercomplete Autoencoder etc. The basic working system behind this architecture is that it compresses and encodes the data and then reconstructs or decodes the data whose output is more or less similar to the original data(Badr 2019). This principle is applied in noise reduction from images or videos. The architecture of the autoencoder makes it possible to use in LSTM, feed-forward network and convolutional neural networks.

The first application of an autoencoder was in the 1980s when it was just like a dimensionality reduction and feature extraction tool (Bandyopadhyay 2022). Even though principal component analysis(PCA) is one of the most commonly used dimensionality reduction methods, autoencoders provide more features in this area than PCA. Some of them are, that autoencoders can make use of convolutional layers or even transfer learning methods for feature extraction, they can facilitate linear as well as non-linear transformations on images, etc. But now its reach is very high and it can be used for any kind of machine learning application like image reconstruction, feature extraction, dimensionality reduction, information retrieval, etc. It has proved to be accurate and efficient in several unsupervised machine learning algorithms.

The performance of a network can be improved by increasing the depth of the network which can be done by stacking multiple autoencoders together(Jégou et al. 2017). The classical stacked autoencoder has been introduced in 2006(Hinton and Salakhutdinov 2006). Even though gradient descent can be used for fine-tuning the weights of the model, it was found to be very efficient only when the initial weights are close to the optimum weights. This particular work describes how these weights can be initialized in a way such that it will be suitable for dimensionality reduction in a proper way.

A multi-layer or stacked Convolutional Autoencoder (CAE) has been implemented in Caffe using MNIST dataset in 2017(Turchenko and Luczak 2017). This model does not have pooling and unpooling layers but still shows good accuracy in dimensionality reduction as compared with a simple autoencoder. The following principles have been used to build this model:

- Both the encoder and decoder components' total feature map size and the number of neurons in each hidden layer.
- Instead of pooling and unpooling layers, include a sigmoid activation function after each convolutional and deconvolutional layer.
- Use sigmoid cross-entropy loss and euclidean loss

Segmentation of digital rock images using different algorithms is explained by Karimpouli et al..Stochastic algorithm has been used to train the model here and it showed an accuracy of 96%(Karimpouli and Tahmasebi 2019). This result is better as compared with the conventional thresholding and watershed algorithm, methods of

segmentation. The main reason for this is that conventional methods perform segmentation based on the color contrast while the digital rock images are in a grey-scale image format. The output of the model is also compared with the manual segmentation output.

Semantic segmentation is a process in which each pixel of the image gets assigned with one of the possible class labels (Yusiong and Naval 2019). SAsiANet, a novel nested autoencoder network, was introduced by Yusiong et al. for semantic image segmentation of outdoor scenes. They didn't use any balancing techniques for loss function, cross-entropy loss. The model has been trained on the datasets, the CamVid urban scenes dataset and the Freiburg forest dataset. They had achieved 72.40% mIoU, and 89.90% mIoU accuracy for segmentation on the CamVid urban scenes dataset and the Freiburg forest dataset respectively. A state-of-the-art method, the One Hundred Layers Tiramisu has been introduced by Simon Jegou et al. for semantic segmentation(Jégou et al. 2017). This is a densely connected Convolutional Neural Network or DensNet where each layer is directly connected to every other layer in a feed-forward fashion. They have used the benchmark datasets, CamVid and Gatech without applying any pre and post-processing methods.

Variational autoencoders are neural networks that can generate new data by training the system using existing data, unlike normal autoencoders(An and Cho 2015). In normal autoencoders, re-construction is the main task where the output will be the same as the input with reduced noise. Variational encoders have this generative property due to the regularized training process which can avoid over-fitting and ensures the latent space. A state-of-the-art approach has been proposed by Suman Sedai et al. where a semi-supervised method for the segmentation of images is used(Sedai et al. 2017). The model is meant for the segmentation of Optic Cup in Retinal Fundus Images. Even though the segmentation of optic cup is very difficult due to its irregular boundaries, this model has managed to do this task successfully with a limited number of labeled data. It is a combined approach where a generative variational autoencoder(VAE) will first be trained with a large number of unlabelled images and the feature embedding will be the result of that. This result will be used by segmentation VAE which will be trained using a small number of labeled data.

A novel semantic segmentation architecture that has 13 convolution layers in the encoder part and its corresponding reverse convolution layers in the decoder part has been contributed by Badrinarayanan et al.(Badrinarayanan et al. 2017). The architecture of the encoder network is identical to the VGG16 pre-trained network. The layers of the network in the encoder and decoder side are symmetrical. The indices that have been computed in the max-pooling layers of the encoder part will be used by those layers in the decoder part and no need of extra learning. This work mainly concentrates on SegNet for image segmentation and HYPPS data augmentation. The same method will be effective in pathology images also, as the staining color changes from image to

image.

A boolean autoencoder has been introduced by Pierre Baldi, which is a framework for handling non-linear autoencoders(Baldi 2012). It is a mathematical model that can be used to learn about both linear and non-linear autoencoders. This framework has a resemblance to clustering techniques that will be difficult to handle when the number of clusters becomes large.

2.8 Review on Pre-trained neural networks

Pre-trained neural networks are those which are trained on very large datasets and saved for various purposes. We can make use of these saved models for our use, either classification or segmentation. For this task, we can use the model as is or customize the entire model through a transfer learning technique. Among these, transfer learning is the most popular and computationally efficient one. The idea behind transfer learning is that if a model is taught on a sizable enough dataset, no matter it is text or image or speech or video, with enough generality, it can act as a generic model for that specific domain(TensorFlow 2022). The parameters of the model like weights, biases etc will be adjusted such that it will be capable of extracting the features of a given input data. The feature maps that the model has learned can be re-used in such works without starting from scratch.

Transfer learning technique has been used in almost every area of AI including text, image and speech. The journey of transfer learning has now reached at Generative Pre-trained Transformer(GPT) models, which is the most trending networks now. GPT models are pre-trained language models developed by OpenAI. It has been trained on massive amounts of text data and they use unsupervised learning methods to train the model. It is the foremost network available now for any kind of NLP applications including chatbots, text summarization, language translation, question answering, sentimental analysis, content creation, etc.

Different AI communities like Hugging Face (a French-American company) are giving widespread accessibility to different pre-trained models. Hugging Face's Transformers library has become widely acclaimed in the research, development, and data science communities due to its user-friendly interface and the diverse selection of pre-trained models it offers, some of which are built on OpenAI's GPT architecture(Face 2023).

As this research is related to images, this work is considering only the pre-trained networks trained on image datasets. Plenty of pre-trained models are there for image classification and image segmentation tasks. There are various datasets for training such models as ImageNet, CIFAR, MNIST, COCO dataset, etc. Among them, ImageNet is the prominent one. This dataset is available at the website, https://image-

net.org/download. It contains 1000 object classes and 1,281,167 images for training, 50,000 for validation, and 100,000 for testing. ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is the most popular and widely used subset of ImageNet. Since 2010, every year a competition in computer vision and machine learning has been conducted based on ImageNet, and the winners of each competition will be taken as the pre-trained networks that we use and this challenge is named as ILSVRC. This challenge will include image classification, detection, localization, etc. The first winner of this competition was AlexNet.

A pre-trained network has been developed by Atsushi Teramoto et al., for the classification of cytological images of NSCLC into 3 subtypes Adeno, Squamous and Small Cell, with 3 convolution layers, 3 pooling layers, and 2 fully connected layers(Teramoto et al. 2017). This was a novel method of NSCLC classification using cytology images. The feature map created in each layer was 256X256X32, 128X128X32 and 64X64X32 respectively and 3-fold cross-validation was used for performance evaluation. The accuracy rate was adeno - 89.0%, squamous - 60.0%, and small cell - 70.3%.

2.8.1 U-Net as an architecture for image segmentation

U-Net is one of the foremost neural network architectures which is the most suitable one for medical imaging inputs. The architecture is named for its U-shaped layout. This architecture was introduced in 2015 to train convolutional networks with very few annotated images (Ronneberger et al. 2015). It was the winner of the SBI challenge for segmentation of neuronal structures in electron microscopic stacks by outperforming the prior best architecture, a sliding-window convolutional network. It consists of a contracting or encoder path for feature extraction followed by a symmetric expanding or decoder path for precise localization. U-Net is so fast that, on GPU it took only a fraction of a second to segment a 512x512 image. U-Net is a popular neural network design that works exceptionally well with medical imaging inputs.

Among multiple revisions of U-Net, U-Net++ is the latest one that improves its performance by introducing nested and dense skip connections between the encoder and decoder networks. These connections enable the model to extract more detailed information at multiple scales. The dense skip connections combine features extracted from different levels of the network, while the nested skip connections connect feature maps that have the same scale.

A state-of-the-art method, named as 3D U-Net has been introduced for the segmentation of volumetric images(Çiçek et al. 2016). Volumetric images include common medical images in 3D modality, like CT, MRI, PET etc. Such kind of images will be more effective for diagnosis purposes. Usually, the annotation of 2D images is not so difficult as it contains only 2 dimensions. However, in volumetric data, it will be

very difficult to create annotated data. To overcome this situation, 3D U-Net has used sparsely annotated volumetric images to train the model. In sparse annotation, only a bunch of pixels will be annotated. This is the best alternative for fully annotated dataset creation which is a more expensive, time-consuming, and laborious task. Such kind of annotation using attention-guided active learning has been introduced by Zhenxi Zhang et al. which is specifically done for volumetric medical images(Zhang et al. 2019). Through their study, they have proved that only 15-20% annotation on slides are required to do the brain extraction task whereas 30-35% for the tissue segmentation task.

U-Net and U-Net++ have also evinced their performance in areas other than medical imaging(Bousias Alexakis and Armenakis 2020). It has been used to find the change detection in high-resolution satellite images to track human activities and their interaction. Another modified version, mUnet has been introduced to apply on satellite images itself for Land-use-land-cover classification (LULC)(Garg et al. 2019). It is a process of assigning labels that describe the physical characteristics of a land area and it is automated using mUnet to determine how the land is being utilized. Several sectors such as telecommunications, utilities, hydrology, etc. require land use and land cover information obtained from remote sensing images. This information provides an understanding of the geographical distribution of a region, including low-level features like vegetation quantity, building area, and geometry, as well as higher-level concepts such as land use categories.

2.8.2 DeepLab as an architecture for image segmentation

DeepLab is a series of deep learning models meant for the semantic segmentation of image data. The goal of semantic segmentation is to understand the content of an image at a pixel level, assigning a label to each pixel based on its semantic category, such as 'person', 'car', 'tree', etc. The DeepLab series was developed by Google Research, and it has seen several iterations over the years. Each iteration introduces improvements and modifications to enhance the model's performance. These models have been used in a variety of applications like satellite image analysis, medical image analysis, scene understanding of robots, segmentation of autonomous vehicles, augmented reality, etc.

Some of the key features and innovations in the DeepLab series include the use of deep convolutional neural networks (CNNs), atrous (dilated) convolutions, and spatial pyramid pooling (Chen, Papandreou, Kokkinos, Murphy and Yuille 2017). Liang-Chieh et al. presented three main contributions through their work. First, it emphasizes the effectiveness of 'atrous convolution,' a technique allowing explicit control over feature resolution within Deep Convolutional Neural Networks (DCNNs) without increasing parameters or computation. Second, it introduces 'atrous spatial pyramid pooling (ASPP)' to segment objects robustly at multiple scales by employing filters

with various sampling rates and effective fields-of-views. Third, the work enhances object boundary localization by combining DCNNs with probabilistic graphical models, specifically a fully connected Conditional Random Field (CRF). The DeepLab system achieved state-of-the-art results on the PASCAL VOC-2012 semantic image segmentation task and improved performance on other datasets, including PASCAL-Context, PASCAL-Person-Part, and Cityscapes. The system attained a 79.7% mean Intersection over Union (mIOU) on the PASCAL VOC-2012 test set.

Some of the variants of DeepLab are as follows,

- DeepLabv1: The original DeepLab model introduced atrous convolution to capture multi-scale context information.
- DeepLabv2: This version incorporated the use of atrous spatial pyramid pooling (ASPP) to capture multi-scale information more efficiently.
- DeepLabv3: DeepLabv3 further improved upon the previous versions by combining the advantages of atrous convolutions and image-level features. It introduced the use of a powerful backbone network (like ResNet or MobileNet) for feature extraction.
- DeepLabv3+: DeepLabv3+ extended the DeepLabv3 architecture by including a decoder module to refine the segmentation results. This version also explored different backbone networks for better performance.

An extended architecture of DeepLabv3 has been used in a work that shows comparable performance with an ensemble of two different DeepLabv3 models(Yurtkulu et al. 2019). Another work with an improved version of the DeepLabv3 model has been used in road extraction from remote sensing images (Quan et al. 2021).

2.8.3 ResNet and its variations as pre-trained network for image classification and feature extraction

Neural networks are working on the principle that "The deeper the better". From various studies, it is evident that, as network depth deepens, accuracy also improves(Boulch 2017). But in some cases, even though the depth of the network can increase the model's performance, at some point it is found that the training error increases as the number of layers increases. For each layer, the number of parameters and memory required will be almost doubled. A Vanishing gradient is a problem that occurs as the depth of a neural network increases beyond a limit(Roodschild et al. 2020). ResNet or Residual Network is renowned for its residual blocks that can address the problem of vanishing gradient. It is one of the most efficient pre-trained networks for image classification. It is trained on the ImageNet dataset and was the winner of the ILSVRC 2015 classification task

with 152 layers which is 8 times deeper than the prior best model(He et al. 2016*a*). Even though deeper networks are performing better than the shallow ones, training accuracy starts degrading after a certain point. A *deep residual learning framework* has been included in ResNet to solve this problem. This feature helps this model to skip some layers while training and directly connects to the output(skip connection)(He et al. 2016*a*). ResNet is in different variants like ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152.

ResNet has achieved the finest accuracy in many applications. It is one of the best pre-trained image classification models. Its variants ResNet-18 and ResNet-50 have been used for the detection of colorectal cancer using colon gland images (Sarwinda et al. 2021). A state-of-the-art model, TCNN(ResNet-50) has been tested on three datasets of data-driven fault diagnosis(Wen et al. 2020). The network acts as a feature extractor in this case. They have achieved prediction accuracies of 98.95%±0.0074, 99.99%±0, and 99.20%±0 that outperformed all other traditional methods. Another work has been done on the COVID-19 dataset using ResNet50 itself(Farooq and Hafeez 2020). It has an accuracy of 96.23% for the screening of Radiographs.

A novel approach ShaResNet, has been introduced to tackle this problem by removing the redundant information of the network. As the name ShaResNet sounds, it shares the convolutional layer weights among residual blocks that operate on the same spatial scale. This case has been demonstrated with an example where a 152-layer-deep residual network can be condensed to 106 convolutional layers, resulting in a parameter gain of 39% with a loss of accuracy on ImageNet of less than 0.2%.

ResNet101 is a variation of ResNet that has 101 layers. More feature extraction tools and increased network complexity will be added by the successive convolutional layers. This additional feature will be helpful because the morphology of pathology images is more challenging than other images. To train the model, there are around 42 million parameters. A modified ResNet101 model with dense dilated convolution has been invented for liver tumor classification (Zhang 2022*b*). Its deeper facility helps to reduce the noise data from the images and bring out the required region only. The dense dilated convolution module is intended to capture broader and deeper semantic information. Also, the up and downsampling module of the network is enhanced, and numerous convolution kernels with various sizes are cascaded to broaden the network, which can effectively prevent feature loss, to tackle the frequent feature loss problems in the classification network.

2.9 Review on Transformers

Transformers are a type of neural network that was initially meant for solving sequence-to-sequence problems but later extended to other fields like computer vision and speech

data. It was introduced by Vaswani et al. in their paper with the title, "Attention is All You Need" in 2017(Vaswani et al. 2017). The ability to capture long-range dependencies and their promising performance helped transformers to become more well-received than CNN architecture even though they have achieved adequate accuracy in medical image segmentation (Huang et al. 2021). Xiaohong Huang et al. introduced a segmentation tool for medical images, named Medical Image Segmentation tranSFormer(MISSFormer) in 2021. A feed-forward network and an Enhanced Transformer Context Bridge made MISSFormer capable of capturing discriminating dependencies and this outperforms every other pre-trained network trained on ImageNet. The code for this model is available in the GitHub link, https://github.com/ZhifangDeng/MISSFormer.

A simple but powerful segmentation framework named SegFormer has been introduced by Enze Xie et al. (Xie et al. 2021). This is an encoder-decoder architecture itself, where a Transformer is used as an encoder and a lightweight multi-layer perception(MLP) is used as the decoder. The model is scaled from SegFormer-B0 to SegFormer-B5, demonstrating significantly improved performance and efficiency compared to previous models. For instance, SegFormer-B4 achieves 50.3% mIoU on ADE20K with 64M parameters, being 5× smaller and 2.2% better than the prior best method. The top model, SegFormer-B5, attains an 84.0% mIoU on the Cityscapes validation set and exhibits excellent zero-shot robustness on Cityscapes-C.

A lightweight SegFormer has been developed to cut down the high computational cost while deploying the model in edge devices(Bai et al. 2022). For this, some of the layers are selectively pruned by removing the least informative set of neurons based on the input instance. Additionally, to enhance the dynamically pruned SegFormer, they have incorporated a two-stage knowledge distillation process, facilitating the transfer of knowledge from the original network to the pruned network. This model has saved 60% computation resources with only a marginal drop of 0.5% in mIoU. SegFormer has been used in Panoptic segmentation also, which is a computer vision task used to solve both semantic and instance segmentation simultaneously. Panoptic SegFormer has been made with three additional components, an efficient deeply-supervised mask decoder, a query decoupling strategy, and an improved post-processing method. Sketch-Segformer is another invention for the semantic segmentation of sketch images with a transformer-based framework. In this model, the sketches will be treated as stroke sequences rather than pixel maps. The code for this is available at https://github.com/PRIS-CV/Sketch-SF (Zheng et al. 2023).

The widespread use of Transformers in the area of computer vision has started with the arrival of Vision Transformers or ViT. It is an image classification model in which the input image is divided into fixed-sized patches, then both features and position of each patch will be embedded and the resultant sequence of vectors will be redirected to the transformers(Dosovitskiy et al. 2020). ViT has proved its excellence as com-

pared to convolutional neural networks with lower computational resources and higher accuracy. ViT can be used for applications like image segmentation and image recognition. Both Segformer and Vision Transformer are based on transformer architecture and share the fundamental idea of applying self-attention mechanisms to process visual data. However, Segformer is designed with modifications to suit the requirements of semantic segmentation tasks, while ViT is more focused on image classification.

2.10 Conclusion

This chapter summarizes various studies performed in the fields that this study make use of, especially the role of neo-adjuvant chemotherapy in RCC and OSteosarcoma, various types of open-source tools for image annotation, and the segmentation and quantization of tissues from pathology images using different image processing and deep learning algorithms.

CHAPTER 3

Synthetic Dataset Generation from Histopathology Images of Post Neoadjuvant Chemotherapy Resection Specimen

3.1 Introduction

Data is integral to every Artificial Intelligence system, serving a pivotal role in both training and evaluating the model. Generating a proficient AI model necessitates a significant quantity of data. Deep learning models exhibit enhanced performance with escalating data volumes, surpassing the efficacy of traditional methods. Analogous to humans drawing insights from past experiences, AI systems acquire knowledge from the data they are exposed to during the training process.

Here are some reasons why datasets are important in AI research:

- Training machine learning models: Dataset is the backbone of AI systems as it is used to train the model. The system will learn only those things that we have given in the form of data. The models learn to recognize patterns in the data and use this knowledge to make predictions or decisions.
- Evaluating model performance: Dataset is also used to evaluate the performance of machine learning models. By testing a model on a dataset that it has never seen before, researchers can get an idea of how well it generalizes to new data.
- Improving accuracy: The more data a model is trained on, the better its accuracy will be. Therefore, having a large and diverse dataset is crucial for achieving high performance in AI systems.
- Advancing AI research: Dataset is important for advancing AI research. By creating new datasets, researchers can test new hypotheses, develop new algorithms, and push the boundaries of what is possible with AI.

3.1.1 Dataset Generation in Digital Pathology

In the conventional method, pathologists manually analyze the slides with the help of a microscope and performs the diagnosis. The emergence of digital pathology provides

new opportunities to develop algorithms and software tools that can assist pathologists in clinical diagnosis and researchers in studying disease mechanisms. The slide-keeping method had several pitfalls related to physical storage, transportation, time, labor, accessibility, risk of damage etc. It can reduce the turnaround time for reporting the clinical cases and thereby reducing the labor time.

One of the main challenges associated with digital pathology is the clarity of the images obtained during collection. This clarity hinges on various factors, such as staining issues, inherent noise artifacts from image acquisition tools, variations in tissue preparation, and differences in how the image is captured by different individuals. Deep-learning algorithms have the potential to address and mitigate these challenges to a certain extent. In the realm of computer vision, digitized pathology slides are commonly referred to as images and can leverage numerous image analysis algorithms. For instance, tasks commonly performed by pathologists, like locating and identifying tissue components, can be accomplished using image segmentation and recognition algorithms. The segmentation of cells and nuclei stands out as a crucial initial step in the automated analysis of digitized microscopy images(Saleena, Kutty et al. 2023).

Despite of all these advantages, the under-supply of labeled data is an ever-time obstacle in the health care field especially in pathology. As in every other area, here also we can build the dataset either from the online repository or by our own. We can collect and label the data from clinics or hospitals. The Cancer Imaging Archive (TCIA), The Open Database of Digital Pathology (ODDP), The National Institutes of Health (NIH) ImageJ repository, The Human Protein Atlas (HPA) and The Whole Slide Imaging Repository (WSIR) are some of the repositories. Even though we can access the data from them, the labeling of the images according to our research problem is a back-breaking task. This should be done under the supervision of experienced pathologists. This points to the main obstacle in digital pathology dataset creation.

Another challenge in histopathology dataset creation is the cumbersome in annotation. To overcome this issue most of the recent works are performed using sparse annotation or weekly supervised or semi-supervised methods. A novel weakly supervised annotation method, self-attention multiple instance learning (SA-MIL) has been introduced to pixel-level segmentation in histopathology images(Li et al. 2023). The main intention of this work is to improve the performance of the weekly supervised segmentation process. Another method sparse annotation has been used in this work to do the semantic segmentation of histopathology images (Bokhorst et al. 2019). An attention-guided active learning has been introduced in another work which is specifically done for volumetric medical images (Zhang et al. 2019).

3.2 Data Collection

In this research, the dataset has been created using real-time images that are collected from MVR Cancer Center and Research Institute(MVRCCRI), Calicut, Kerala, India. It is one of the best and leading cancer centers in Kerala and it is well-known as a Cancer Research Institute also. It was established as a project of the Cancer and Allied Ailments Research (CARE) Foundation. This foundation, a non-profit charity organization, was founded by the Calicut City Service Co-operative Bank in 2011 in remembrance of MV Raghavan, who was famous as a revolutionary leader of Kerala's Communist movement. Sullamussalam Science College, Areekode has signed a Memorandum of Understanding(MoU) with MVRCCRI for this research. Two ethical committee meetings (internal to the hospital and external) also cleared to ensure that there is no privacy breach in patient's personal data while collecting the data.

Two experienced pathologists were actively involved all through this work. Dr. Sajna V.M. Kutty, Senior Pathologist from MVRCCRI supervised the domain side including data collection and annotation throughout the study. She has collected the entire images for this study from the hospital and helped in the annotation. Dr. Fazna N., Pathologist, KMCT Medical College, Calicut, Kerala also helped to annotate the images. Both of them have given guidance in the domain-side things.

Usually, Renal Cell Carcinoma is a disease of elderly people that affects people aged 50 years and above and it is not seen in children and adolescents. The Osteosarcoma will be seen in two types of people mostly children and adolescents. So the specimens also will be collected in those respective ages.

The images have been taken using the digital camera mounted on a microscope of Meiji Japan, with low power magnification of 10x. The low power adjustment has been used to get the maximum portion of necrotic tissues from the specimen and each image is of resolution 2592 x 1932. From the entire slide, only the representative sections have been captured such that it avoids maximum vacant potions and captures maximum required tissues like necrosis, fibrosis, tumors, etc. Even though we require only necrosis for our study, in the real situation the specimen will contain all the above mentioned type of tissues. So training data should contain all such things.

The dataset created for this study has been made publicly available with a name NecrosisDB. This can be accessed only if the end user agreed to some terms and conditions.

3.3 Synthetic Dataset Generation from Histopathology Images of Post Neoadjuvant Chemotherapy Resection Specimen

In this research, the training data are digital histopathology images of post-neoadjuvant chemotherapy resection specimens. The specimen has been collected from cancer patients who underwent neoadjuvant chemotherapy. The sample will be collected from representative areas, processed, and Hematoxylin and Eosin stain was applied to the slides.

We have created a synthetic dataset from 60 images which are captured and manually annotated by experienced pathologists. The data collection part has been explained in the above section. Each image captured is of dimensions 2592 x 1932, which are split into patches of size 512x512. So we can have 15 patches from each image and thereby we created a dataset of 900 samples. Different augmentation techniques can be applied to this dataset to increment the number again. For the augmentation technique, we have used the library, 'albumentation', so that we can have a variety of options to create multiple mirrored images of the original one.

As we are performing segmentation tasks, image annotation is a mandatory thing. Image annotation is a technique that is used to label or identify the required object from the entire image given. Because in segmentation we have to identify only a specific region or some objects within the image. We have to train the system with images and their corresponding ground truth(labeled image) in which the region of interest is labeled correctly. This method can considerably improve the accuracy of the model.

But in image classification annotation is not a mandatory thing. We can label the entire image into a single name from where the model itself identifies the objects in it. But still, if we add annotation in it, the accuracy of the classification will be greater. But why people are hesitating to do this is, the cost and time required for this. Figure 3.1 shows how the annotated image looks like, where the region of interest portion in the mask can be of any color as per our choice

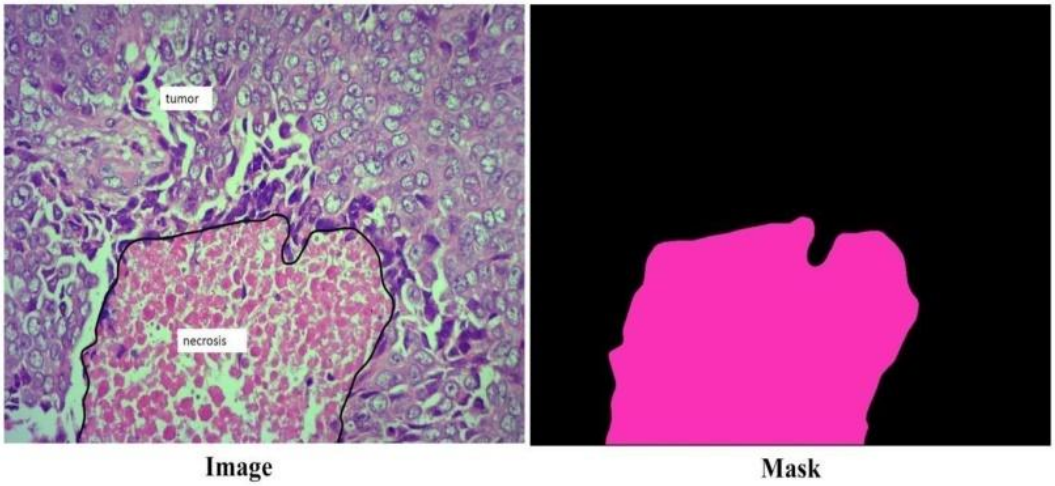


Fig. 3.1 The original image and the annotated image

The image annotation has several techniques. Among them, we have selected an open-source, web-based tool named CVAT, which is very easy to use and very accurate.

3.3.1 CVAT- Computer Vision Annotation Tool

CVAT is an open-source web-based annotation tool previously which was available as ”cvat.org”. But it has been migrated to ”www.cvat.ai”. This tool has been meant for image and video annotation or labeling. CVAT is made to be simple to use and may be tailored to meet a project's unique requirements. It allows users to export annotations in these forms for use in other computer vision applications and supports a number of annotation formats, including Pascal VOC, COCO, and YOLO. For a variety of computer vision applications, such as object identification, image classification, and semantic segmentation, CVAT is extensively utilized in both industry and research. It is a well-liked option among researchers and practitioners due to its open-source nature, flexibility, and simplicity of usage.

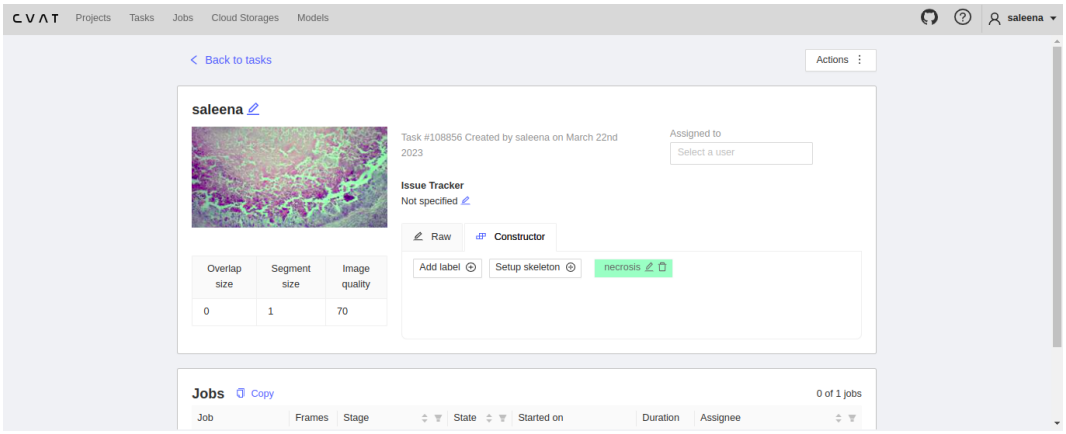


Fig. 3.2 How different projects look like in CVAT

As shown in Figure 3.2, we can include multiple tasks in the CVAT. In each task, we can add any number of images. When we click on the specific job, we will get the workspace for annotation as shown in Figure 3.3. The workspace provides multiple tools that can be used as per our requirements. We can have various shapes like polygons, rectangles, polylines, ellipses, points etc. In our project, our region of interest will be scattered around multiple places and we need to draw in different shapes as the tissues scattered around. So we are using polygons in our project.

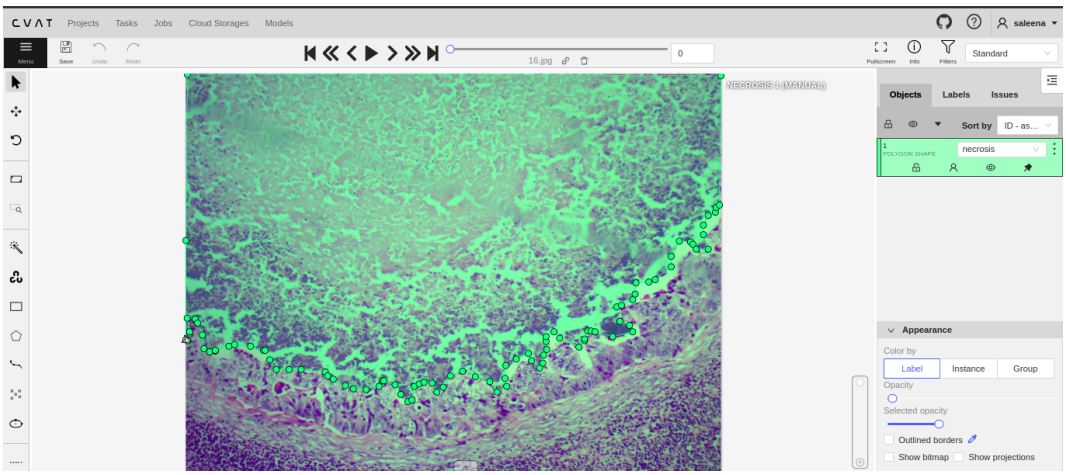


Fig. 3.3 The workspace of CVAT and how the required region has been marked using polygon

We can download the segmented output as '.png' file that can be directly used in training the model. In some other annotation tools, the output may be in other formats like .json or .csv file.

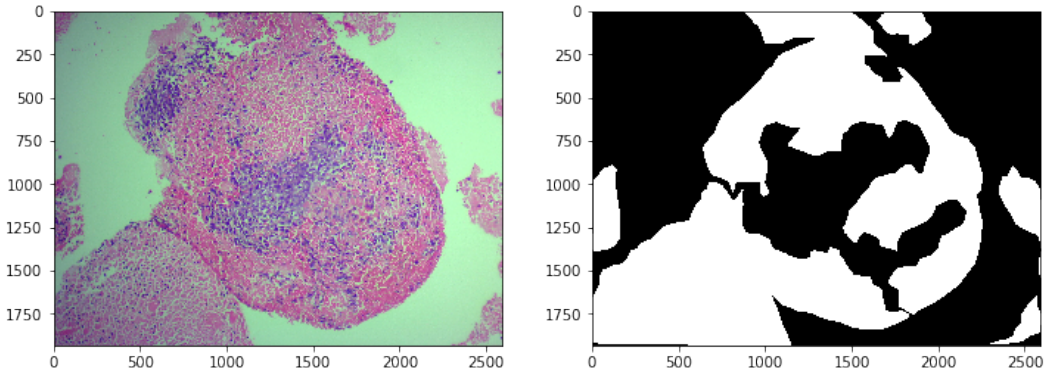


Fig. 3.4 How the image's label has been created using CVAT.

In this work, we need to identify the tissue necrosis only. So a single label is enough in this case and we have labeled it as 'necrosis'. CVAT allows multiple labeling in the same image. We can specify all these labels in it. The color of the label can be selected as per our choice. If multiple labeling is needed, we can represent each label with different colors.

3.3.2 OpenCV: A Python Library

In this research, the synthetic dataset generation has been done with the help of OpenCV, or Open Source Computer Vision Library, which is a widely used open-source computer vision and machine learning software library. It was originally developed by Intel in 1999 and later supported by Willow Garage and Itseez (which was later acquired by Intel). OpenCV is designed to provide a common infrastructure for Image processing, and computer vision applications and accelerate the development of vision-based projects(Bradski et al. 2000). It is a cross-platform software and can be operated in Linux, Mac, Android, and Windows. It has a rich set of functions to work on images such that it can have 2D and 3D operations. It is very easy to use as it has an interface in Python. This library has strong community support also.

3.3.3 Creating patches from the labeled images

Each image captured is of dimensions 2592 x 1932, which are split into patches of size 512x512. Patchification can be done on each image and its corresponding mask, in a non-overlapping manner using the library called 'patchify'. So we can have 15 patches from each image and thereby we created a dataset of 900 samples. The same method will be applied to the corresponding masks of the images without losing the order. In this example, the patchify function is used to divide the image 'img' into patches of size (512, 512) and with a step size of 512. The step size determines the overlap between the patches. A larger step size results in less overlap between the patches, while a

smaller step size results in more overlap. The images are split up in such a manner that the dimensions, 2592 and 1932 has undergone integer division such that we will get patches in 3 rows and 5 columns(2592/512 = 5 and 1932/512 = 3).

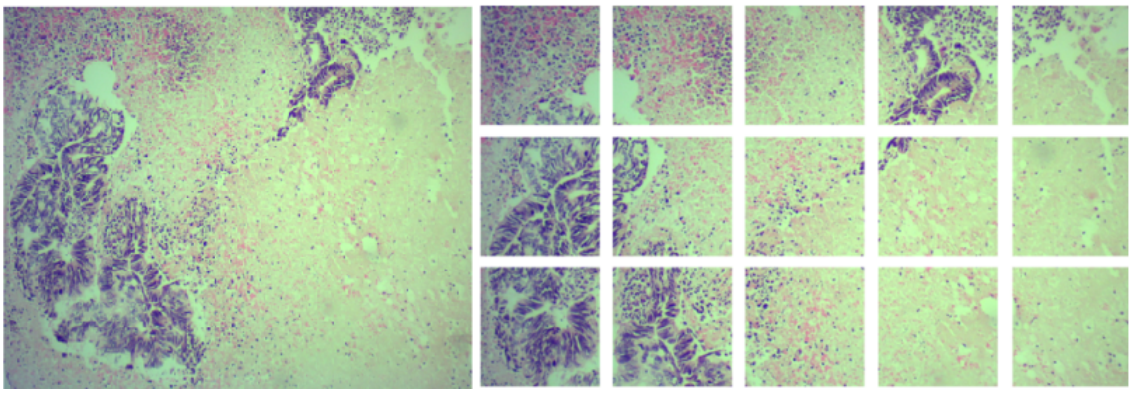


Fig. 3.5 Original image(left) and patches created from it(right)

Pseudo-code 1: To make patches from images and masks

1. Import the libraries opencv and patchify
2. Arrange the image and its corresponding mask in the same order in multiple folders
3. Cut the images using patchify() with 512X 512 which will end up to 15 images in 3 rows and 5 columns
4. Each patch will be saved into a new folder
5. Repeat steps 3 and 4 with masks also.

The dataset obtained can be converted into NumPy arrays for further segmentation. The U-Net should have 4D shape of the image and mask where the four dimensions will represent (batch size, height, width, and channel). Images should be 3-channel(RGB) and masks should be 1-channel. The patches variable is a 4D NumPy array, where the first dimension is the patch index, the second and third dimensions are the patch height and width, and the fourth dimension is the number of color channels.

3.4 Meta-data for the dataset

Metadata of a dataset is ”data about data” and helps users to understand about what the dataset contains, how it was collected, and how it can be utilized etc. It is a structured information that provides context, description, and essential details about the data, enabling better understanding, organization, and usage. The Table 3.1, shows the metadata about NecrosisDB, that we have created in this study.

Table 3.1 Meta-data for the dataset

Parameters	**Values**
Type of image	JPG file
Type of mask/label	PNG file
No. of images collected	60
Image resolution	2592 x 1932
Magnification	10X
Dataset name	NecrosisDB
No. of patches in dataset	900
Patch resolution	512x512
Tissue type segmenting	Tumor necrosis
Annotation Type	Semantic Mask
Annotation tool	CVAT (https://www.cvat.ai/)
Dataset link	https://github.com/saleenaareekode/Necrosis-ML

3.5 The overall workflow of Necrosis-ML

The dataset generated here can be used for training the segmentation model used in the Necrosis-ML. Different models tried for the same, namely Autoencoder, transfer learning methods, transformer model etc. These methods are explained in the Chapters from 4 through 6. The working of the Necrosis-ML is shown in Figure 3.6. It depicts how this tool can be used by a pathologist during their clinical diagnosis. When an image has been fed into this tool, it will output the percentage of necrosis present in the image along with the segmented image. Along with the segmented image, the percentage of the segmented area also will be outputted. This value is the principal result of this research. The architecture shown in the center of the figure is the segmentation model, which can be any of the above mentioned trained models. From the segmented image, the quantification can be done.

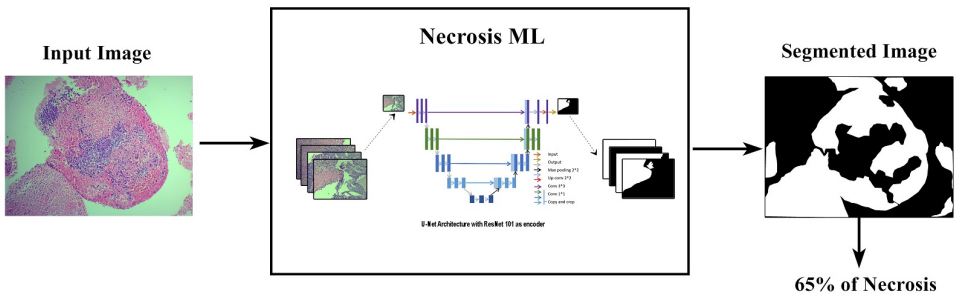


Fig. 3.6 Necrosis-ML

3.6 Conclusion

In this research, we have managed to develop a fully supervised annotated dataset which has been annotated using experienced pathologists. We have collected and processed real-time data obtained from the Cancer Research Center. The data collected has been annotated using an open-source tool under the supervision of experienced pathologists. The dataset created using the above-mentioned methods is used to create deep neural networks to segment the histopathology images of post-neoadjuvant chemotherapy resection specimens to estimate the tumor necrosis present in them. The later Chapters describe the different image processing and deep learning methods using this dataset.

CHAPTER 4

Semantic Segmentation of Necrotic Tissues in Post Neoadjuvant Chemotherapy Resection Specimens using Autoencoder Network

4.1 Introduction

The identification of a proper treatment plan is the most important factor in clinical practices for any disease. The patients who are affected with the same disease, may not be suggested for the same treatment plans. Prognosis in medical terms indicates the chance for recovery of a patient. There are different prognostic factors for every disease. In the case of cancer, especially Osteosarcoma and Renal Cell Carcinoma(RCC), the volume of tumor necrosis created due to Neoadjuvant chemotherapy is one among them. The term necrosis means the uncontrolled death of body cells. The percentage of these tissues can be used as a decision factor in treatment plans.

However, there is no proper method for this quantization task. The pathologists are now doing this task manually. So there is considerable inter and intra-observer variability and labor/ time consumption in the manual estimation of necrosis generated. This is mainly due to the complex morphology of tissues.

The process is making use of Hematoxylin and Eosin(HandE) stained resection specimens as the input. As the necrotic tissues are in scattered foci and not accumulated in a single place in an image, it is usually very difficult for the pathologist to accurately segment all these portions and calculate the exact percentage. This leads to significant inter-observer variability in the manual segmentation process. The manual process is also heading with intra-observer disagreement, where the same pathologist may have different impressions on the same specimen in multiple observations.

In this chapter, we are distinguishing necrotic tissues with the help of autoencoders from the given specimen. The necrotic tissues that we are focusing are originated due to neoadjuvant chemotherapy in the case of patients who are the victims of Renal Cell Carcinoma and Osteosarcoma. The dataset creation has been explained in Chapter 3 in detail.

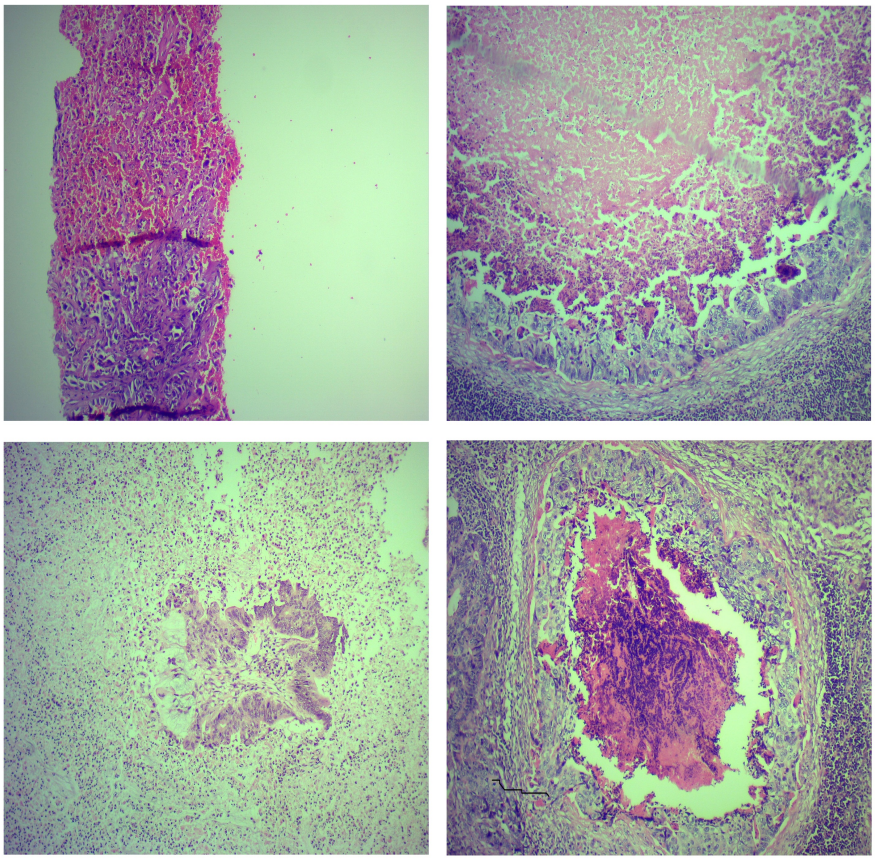


Fig. 4.1 Some of the specimens collected for the study

4.2 Proposed Model for Semantic Segmentation of Necrotic Tissues using Autoencoders

Identification of tissues based on morphology and staining color is a tedious task that every pathologist has to face during their routine work. The proposed work uses Autoencoders for model building. (Bandyopadhyay 2022)The main aim of the autoencoder is to enable the network to capture the most important features of the input image by encoding the higher-dimension data into lower-dimension data. The pre-trained networks like (Ronneberger et al. 2015) U-Net are a better choice for semantic segmentation of images than autoencoders, but they require highly expensive training resources as compared to autoencoders. Since we are training the model with a single image it requires very less computation resources. Autoencoders are initially used for applications like image reconstruction, where the output of the model will be the same as the input. But obviously, the output will not be exactly the same as the input when we have a pixel-wise analysis, due to the reconstruction loss. This architecture is leveraged in applications like noise reduction, anomaly detection, domain adaption, colorization, etc.

In the normal image reconstruction process, the same input image will be recon-

structed in the output with reduced noise. But in the domain adaptation process, we can reconstruct a totally different image from an input image. Here from the original image, its mask is reconstructed. A Bi-shifting Autoencoder network (BAE) which has been introduced by Kan et al. (Kan et al. 2015), and domain adaptation in the classification of remote sensing images explain the applications of this method (Elshamli et al. 2017). It can also be used in changing the resolution of an image from one intensity to another.

The proposed method introduces a customized autoencoder model specifically for segmenting the necrotic tissues from histopathology images. By applying pixel-wise segmentation, autoencoders can automatically segment such regions using digitized images The main attraction of this work is that the model has been trained using a single gold-standard image. Due to this, the time and effort required to collect and annotate the data are much less than compared with other neural networks.

4.2.1 Convolutional Autoencoder

Autoencoder is one of the foremost unsupervised neural networks for the past several decades. Its journey started in 1986 when D.E. Rumelhart et al. invented this architecture with the motto of "back propagation without a teacher" (Rumelhart 1986). It was a solution for the inability of a system to learn new features from the input given. It only converts the input image to the same output pattern. Linear and Boolean autoencoders have been explained with a solid mathematical framework by Baldi et al. which elucidates how the network learns from input data, how it resemblance with clustering, and its horizontal and vertical composability in deep architectures, etc.(Baldi 2012). Some of the renowned autoencoders are Convolutional Autoencoders, Under-complete autoencoders, Sparse autoencoders, Contractive autoencoders, De-noising autoencoders, and Variational autoencoders(Mohamed and Ali n.d.).

Convolutional Autoencoder(CAE) showed better results than conventional autoencoders, especially in image processing tasks(Zhang 2018). In this study, we have used CAE which will be more suitable for our data. The image we have is of dimension 2592 x 1932. As this is not affordable for the convolutional network, we have reduced the dimension to 512x512 and did the same with its mask also. The autoencoder has three main elements: encoder, decoder, and bottleneck or latent space.

- Encoder part: This part is meant for feature extraction and dimension reduction. Convolutional neural networks are used to perform the earlier one and pooling for the latter.

- Bottleneck: This is the encoded or compressed version of the input image and it will be fed into the decoder part as the input data. This is the lowest dimension of the image obtained from this architecture. It delimits the flow of information or parameters passing from encoder to decoder.

- Decoder part: In this part, the model learns to reconstruct the image with minimized reconstruction loss using the features obtained from the encoder part.

We have made use of the power of CAE in our proposed study to train the model. Domain adaptation technique using convolutional autoencoders is used in this work for the semantic segmentation of histopathology images. A single image is used as input for training and while fitting the model, the input will be the image and the output will be the mask. The architecture of the model is shown below.

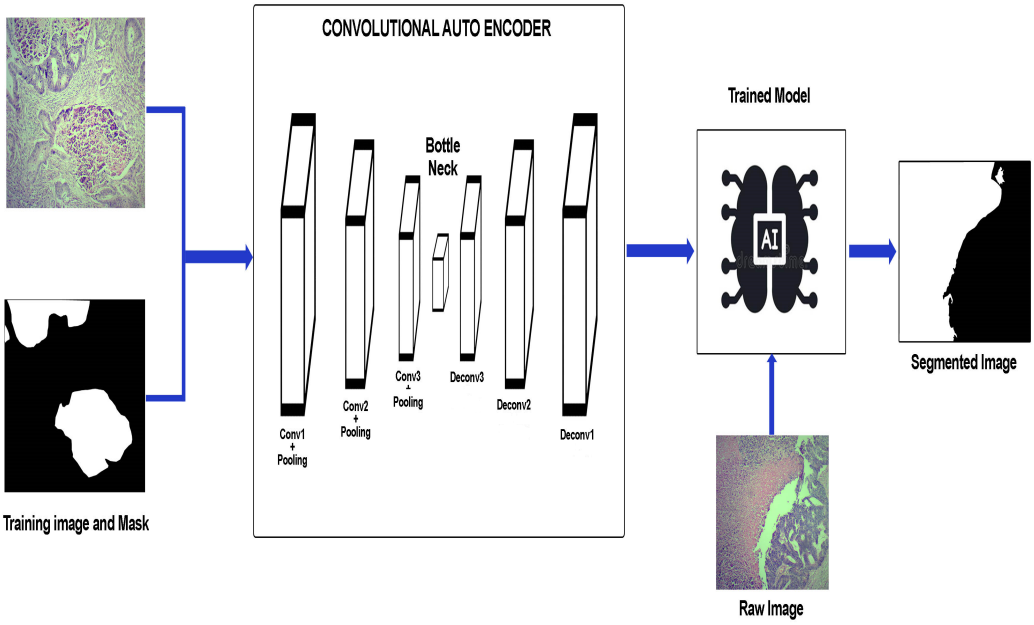


Fig. 4.2 Architecture of CAE model trained with a single image and its mask

In this model, we have used three convolution layers followed by the max-pooling layer at the encoder part and the same number of layers along with upsampling at the decoder part. The entire model summary has been given in Figure 4.4. While encoding, the image is reduced into lower dimensions, the weights and biases in each layer will be clearly recorded and it will be used in the decoding process. Max-pooling layers at the encoding side are responsible for reducing the dimensions, whereas the Upsamlping layers at the decoder side do the exact reverse process. The loss represents the reconstruction loss created while reconstructing the image.

If X is the input image, we can represent loss as,

$$Loss = decoder(encoder(X))–X \tag{4.1}$$

Pseudo code: model building

1. Read an image and its mask

2. Reshape them with a resolution of 512x512
3. Convert it into a floating point array
4. Normalize it by dividing it with 255
5. Build a sequential model with the configuration specified in Figure 4.3.
6. Compile the model with Adam optimizer, Mean-squared error as loss, and accuracy
7. Train the model with X= image array and Y=mask array for 1000 epochs

Once the model has been trained, it can be used as the segmentation part of the Necrosis-ML framework. The Figure 4.3 shows the overall workflow of the system using CAE. The quantization part will be explained in detail in the Chapter 6.

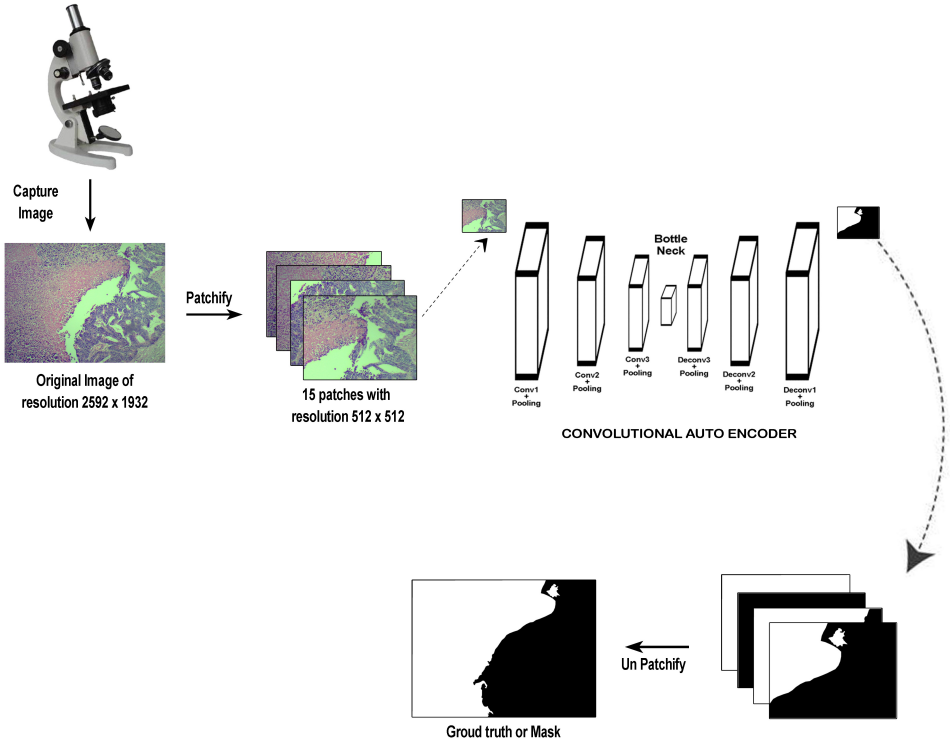


Fig. 4.3 The overall framework of Necrosis-ML using CAE

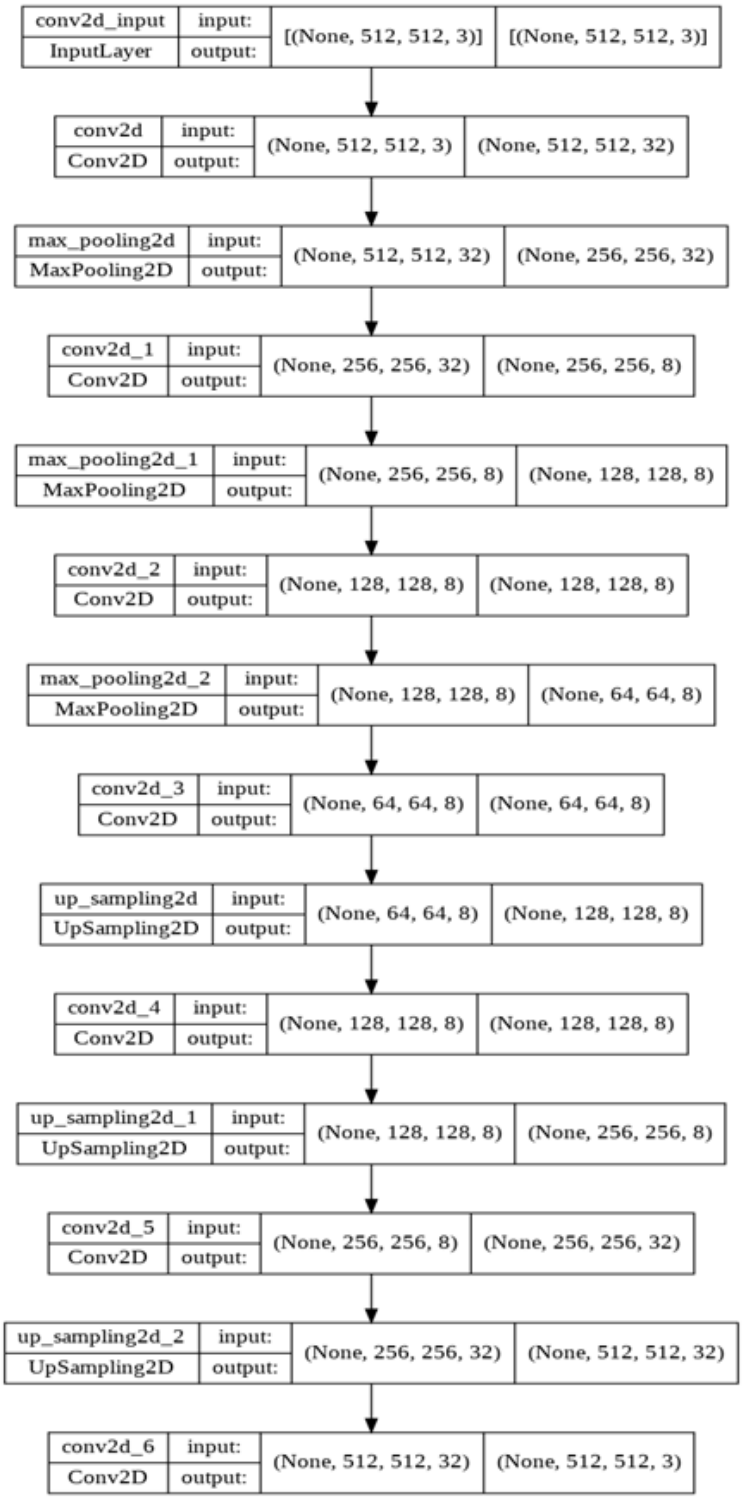


Fig. 4.4 CAE Model summary

4.3 Experimental Analysis and Results

In this study, we have developed a segmentation model, that can separate tumor necrosis from histopathology images. Out of different types of tissues, we are separating one

type of tissue only. A single gold standard image has been selected for training the model. The word gold standard means that it contains all types of tissues like necrosis, tumor, fibrosis, etc. This kind of segmentation is an extension of image reconstruction and it is generally called domain adaptation.

The model has been trained here with a single image and its mask. All the available 8,163 parameters were trainable. It took only 10 minutes to run these 1000 epochs in a 4GB RAM system, even though the image is of size 512x512. The Median filter and Gaussian filter results showed bad accuracy. So we didn't use any pre-processing methods as we took the image with maximum clarity and noiselessness. We have used TensorFlow Keras for implementing the code. The accuracy and loss obtained during model building are shown in Table 4.1.

Table 4.1 Accuracy and loss for the CAE model

Parameters	**Values**
Training Accuracy	0.8228
Training Loss	0.1096
Testing Loss	0.3240
Testing Accuracy	0.7258

Graph plots are very helpful to know about the nature of training performance. We have plotted a graph in Figure 4.5, with data points obtained during the training process. The X-axis represents the epochs(1000 here) and the Y-axis represents loss or accuracy. From the graph, we can have an assumption that the loss decreases as the epochs increase, whereas the accuracy is increasing.

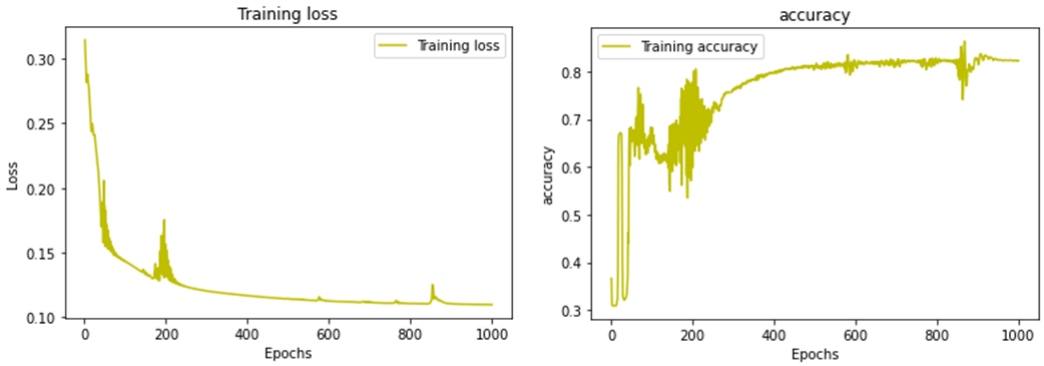


Fig. 4.5 Graph showing the loss and accuracy of the model trained with a single image

We have saved the model and using the same we have segmented some other images also. Figure 4.8 shows some of the images. The prediction is more or less similar in most of the cases but it shows wrong prediction for unknown types of morphology or blurred images.

4.3.1 Hyperparameter Tuning

Hyperparameter tuning is a method of finding optimal values for parameters whose combination leads to the best result for neural network algorithms. Learning rate, number of epochs, number of neurons(image size), number of hidden layers, optimization algorithms, etc are the different hyperparameters that can be tuned in such a manner that we get a model with maximum accuracy. This combination is mainly depending upon the data that we are giving for training the model. The combination of values that we have selected for the hyperparameters is shown in Table 4.2.

Table 4.2 The hyperparameter values we have applied in the CAE model

Parameters	**Values**
Number of epochs	1000
Optimizer	Adam
Learning rate	0.001
Loss	Mean-squared error

We have performed this fine-tuning process here and some of the results are shown below. Figure 4.6 narrates a graph that indicates how the loss and accuracy change with respect to the learning rate. We have tested with three values, which are 0.01,0.001, and 0.0001. The first one is indicated with blue color, second with orange and last with green. The optimal value for learning rate is indicated as 0.001.Whilst, Figure 4.7 depicts the predicted output with respect to the number of epochs, where we have selected three cases, 500, 1000, and 1250. The first row is with epochs=500, second row with epochs=1000 and last one with epochs=1250.Among the three, the second one shows good accuracy. As we increase the number of epochs to more than 1000, it started showing blurred results.

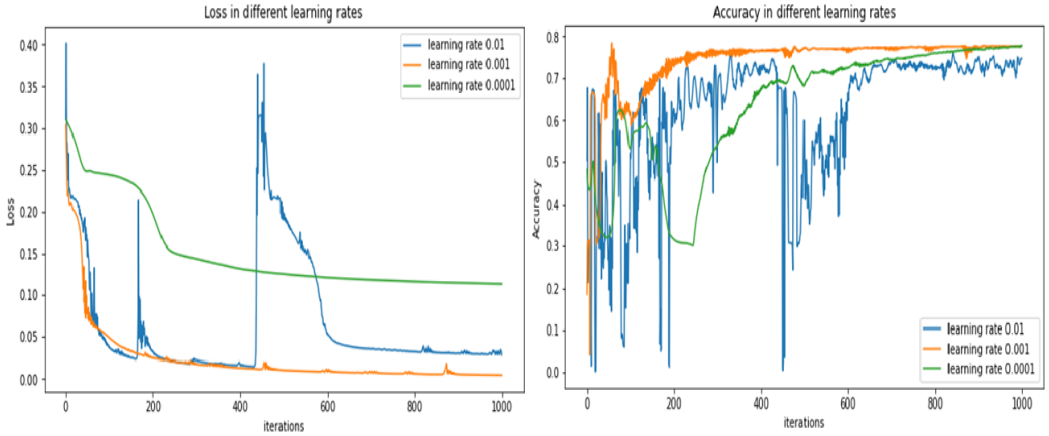


Fig. 4.6 Rate of change of loss and accuracy with respect to the learning rate

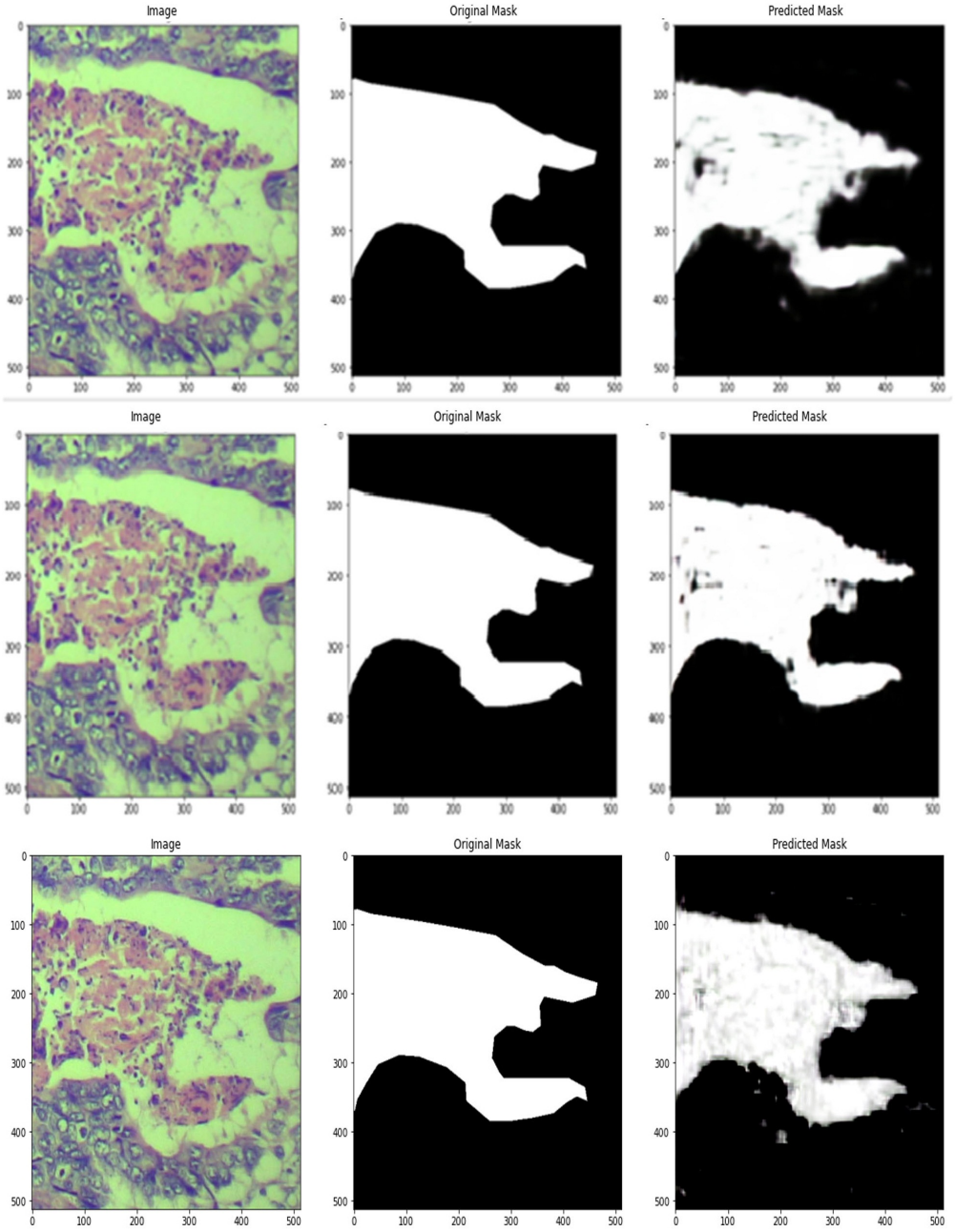


Fig. 4.7 How the prediction changes with respect to the number of epochs.The first row is with epochs=500, second row with epochs=1000 and last one with epochs=1250.

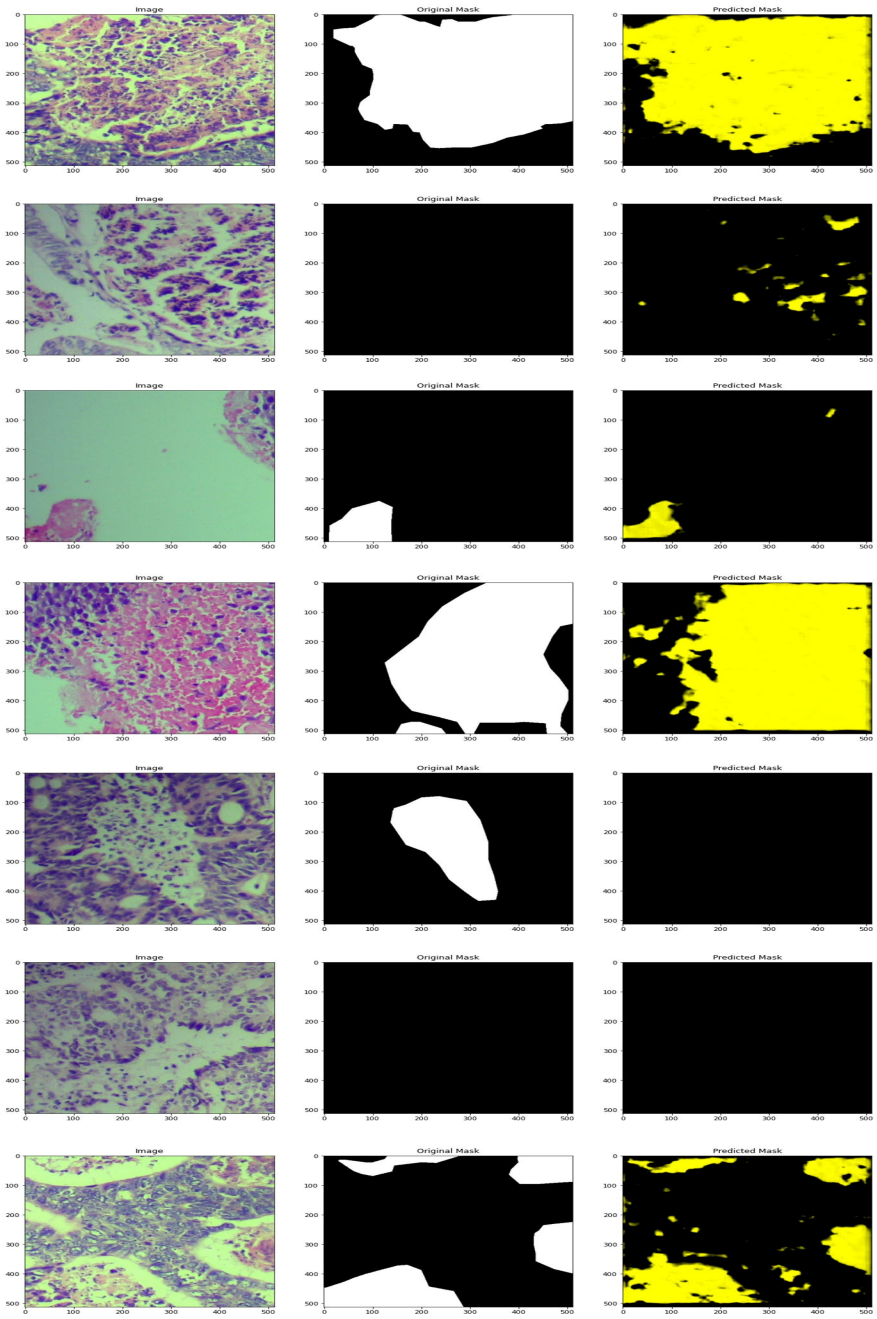


Fig. 4.8 Test results of CAE model. Each row shows the image, original mask and predicted mask respectively.

4.4 Conclusion

Segmentation and quantization of tissue necrosis from a given histopathology image is a crucial prognostic factor for Renal Cell Carcinoma and Osteosarcoma. But data collection and annotation are the main hurdles in any digital pathology-related research as it requires great time and effort from the pathologist's side. The normal clustering

algorithms also won't work with the pathology images as their morphology is complicated to distinguish. We made use of a convolutional autoencoder-based model where the time and effort required are much less compared with other segmentation methods. Even though the main intention of the autoencoder is image reconstruction, we can adopt it for semantic segmentation.

The proposed model has been trained with a single histopathology image and its mask. The image has been selected wisely such that it contains almost all relevant kinds of tissues and selects the necrotic tissues only from the input. It showed better results which are almost similar to that done by manual segmentation. The purpose of this work is to reduce the workload of pathologists by automating their work, as the number of pathologists is very low compared with their work. This tool can be extended and can be used as a second opinion for their analysis. The proposed model has been trained with a single image and its corresponding mask. Even though the model shows a fine accuracy in similar kinds of images, it has failed in segmenting some other images which have different staining intensities, and also in less cohesive morphology where the area of necrosis is seen scattered in between other tissue elements. Hence there is a need to further train the model with images of divergent morphology and staining intensities. For this, more annotated images are required to train and test the model.

CHAPTER 5

Patch-level Segmentation of Tumor Necrosis in Post Neoadjuvant Chemotherapy Resection Specimens using SegFormer Model

5.1 Introduction

Necrosis is the state of cell death which may be caused by several reasons. In some situations, tumor necrosis or death of cancerous tissues is purposefully created with the help of some treatments that include radiation therapy, chemotherapy, and hormone therapy and they are collectively called Neoadjuvant treatment (NAT). This is an initial step to lessen the malignant area before the primary treatment and it will increase the survival rate times. Different treatment regimens should be used for different cancer patients. It will depend on the patient's physical state, how their body reacts to the treatments, and many prognostic variables. Prognostic scores are the key elements in the selection of treatment plans in clinical practice (Davis et al. 1994, Picci et al. 1994). In the case of Osteosarcoma and Renal Cell Carcinoma, chemotherapy-induced tumor necrosis is one of the most important prognostic factors.

The main objective of this study is to evaluate the patient's body's therapeutic response to neoadjuvant chemotherapy or NAC. The NAC can lower risk stratification, reduce tumor volume, and increase surgical resection success rate (Wang et al. 2011, Davis et al. 1994). Initially, neoadjuvant chemotherapy (NAC) has been introduced to reduce the volume of the cancerous region, where the cancer is almost in the unresectable stage. But now, it is proclaimed by experts that NAC followed by de-bulking surgery, is the best treatment for advanced cancer(Elies et al. 2018). Several studies are reported to evaluate the clinical and pathological response after NAC on various cancers (Burcombe et al. 2002, Alawad 2014). Nevertheless, all these are done either through laboratory tests or medical imaging modalities like PET.

This chapter introduced an AI-powered automated system designed to assess a patient's body response to NAC using digital histopathology images. Our focus is specifically on Osteosarcoma and Renal Cell Carcinoma, where the pivotal prognostic factor is identified as cell death resulting from NAC. The model efficiently delineates the necrotic tissue area from the input histopathology image and employs image processing techniques to quantify this segmented region. This quantified value serves as a critical

determinant in shaping subsequent treatment plans for the patient. The tool's output can be utilized as a secondary opinion for guiding further steps in the treatment process. To facilitate this research, a dataset comprising 900 images and their corresponding masks was meticulously curated with the assistance of an experienced pathologist. The dataset comprises images that vary in texture due to differences in staining for various specimens. The creation of the dataset used for building the model is described in Chapter 3 in detail. The image and its corresponding mask in this study will look like as shown in Figure 5.1.

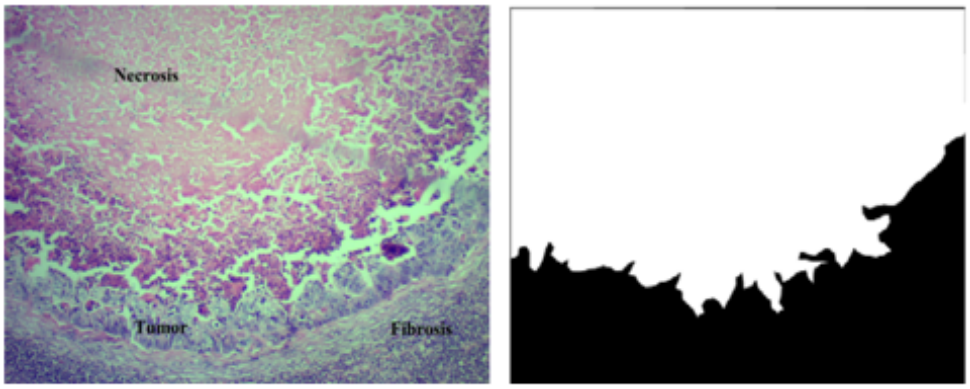


Fig. 5.1 Image(left) and its mask obtained from CVAT(right)

This chapter details the semantic segmentation of histopathology images, particularly the post-neoadjuvant chemotherapy resection specimens. Semantic segmentation is the process of pixel-by-pixel classification of the image. In this method, a segmentation model has been developed using an architecture called SegFormer(Xie et al. 2021). The SegFormer is a Transformer-based semantic segmentation framework. It is inspired by the success of transformers in natural language processing tasks and has been adapted for computer vision tasks. SegFormer and Vision Transformer(ViT) (Dosovitskiy et al. 2020) are the extensions of transformers that made breakthroughs in this area. Both of them divide an input image into fixed-size patches, linearly embed them, and process the embeddings using transformer layers to capture global context information. ViT was primarily intended for image classification and SegFormer for image segmentation problems. Different variants of SegFormer models, from SegFormer-B0 through SegFormer-B5 have been compared in a study and SegFormer-B5 showed better accuracy in it(Xie et al. 2021). It was 84.0% mIoU on the Cityscapes dataset's validation set. A graph plotted based on this analysis has been given in Figure 5.2 (Xie et al. 2021).

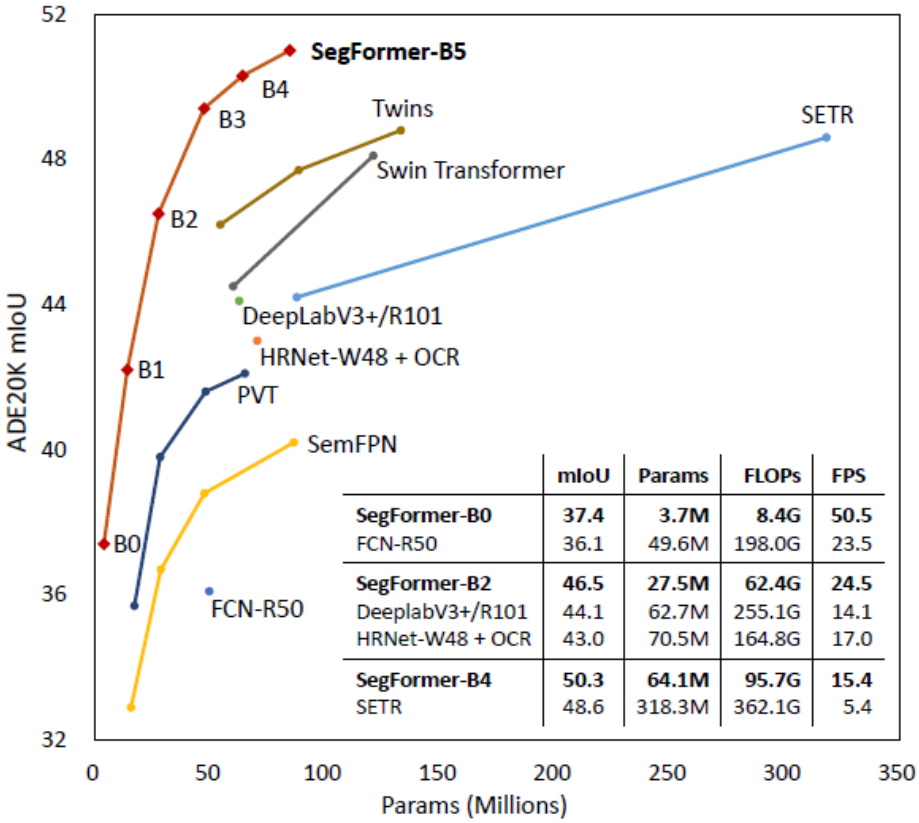


Fig. 5.2 The performance analysis of different models of segmentation on ADE20K dataset (Xie et al. 2021)

5.1.1 The Proposed Methodology

In Necrosis-ML, one of the major parts is the semantic segmentation. This chapter details such a model developed using SegFormer, which is a Transformer-based neural network framework meant for semantic segmentation tasks. Among multiple series of SegForemers, SegFormer-B5 has been used in this model. Initially, the transformers were built only for NLP-related tasks like sequence-to-sequence problems, but later widely used for any kind of data. It is an encoder-decoder architecture with self-attention mechanism. It has a hierarchical Transformer encoder and a lightweight Multi-layer Perceptron(MLP) decoder head. Transformer is an encoder-decoder-based neural network (Vaswani et al. 2017) with an attention mechanism. It allows the model to focus on different parts of the input sequence when making predictions, assigning different weights to different elements. The self-attention mechanism enables the model to consider the relationships between all pairs of words in a sequence simultaneously.

The hierarchical nature of the encoder Transformer extracts multiscale features from the input. In every other segmentation architecture, at the decoder side upsampling will be conducted for the image reconstruction. But in this framework, Multi-layer Percep-

tron is used as the decoder which is faster and more efficient than the others(Bomma 2022). The basic architecture of the SegFormer is shown in Figure 5.3.(Xie et al. 2021).

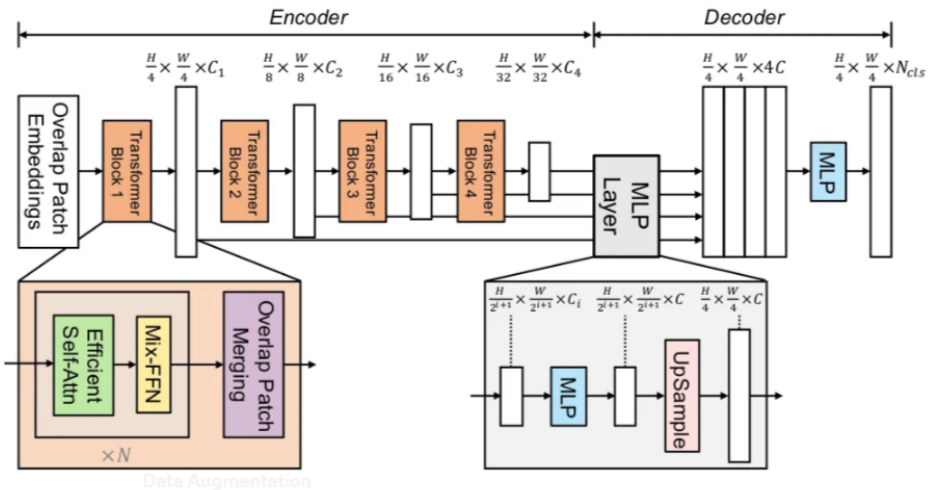


Fig. 5.3 The architecture of SegFormer model (Xie et al. 2021)

Segformer incorporates ideas from both convolutional neural networks (CNNs) and transformers to capture both local and global features in an image. Different steps involved in its working are given below,

- SegFormer divides the input image into fixed-sized, non-overlapping patches.
- These patches are then linearly embedded to acquire a flattened representation, ensuring compatibility with the transformer architecture.
- To obtain the spatial information of these patches, positional encoding is also added to the embedded patches. This will keep track of the position information of patches within the image.
- The transformer layers process these embedded patches, allowing the model to consider relationships between pixels across the entire image.
- Each layer contains self-attention mechanisms, enabling the model to capture both local and global dependencies within the image.
- At the decoder end, it reconstructs the spatial information and produces segmentation maps, where each pixel is assigned a class label.
- Using proper performance metrics, the accuracy of the model will be measured.

Almost a similar kind of operation is performed in Vision Transformer whose architecture is shown in Figure 5.4(Dosovitskiy et al. 2020).

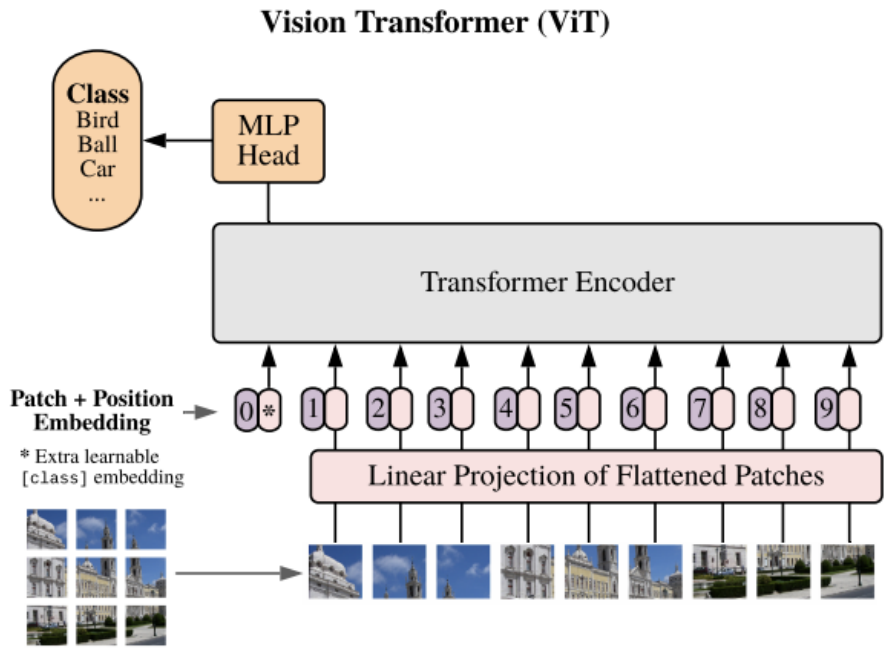


Fig. 5.4 Architecture of vision transformer for image classification(Dosovitskiy et al. 2020)

The encoder Transformer in SegFormer was initially pre-trained on the dataset, Imagenet-1k for a classification purpose. Then the classification layer will be replaced with MLP decode head and will be fine-tuned with different datasets namely, ADE20K, Cityscapes, and COCO-stuff which are good benchmarks for segmentation.

In this particular model, the pre-trained model from NVIDIA NGC (NVIDIA GPU Cloud) model hub, named *nvidia/mit-b5* has been used, where *mit* stands for Mix Transformer encoder backbone. NGC is a repository of GPU-optimized software and deep-learning models provided by NVIDIA. NVIDIA had released several SegFormer models with different model sizes or variants. Among them, 'nvidia/segformer-b0' is the baseline model in the SegFormer series. 'nvidia/segformer-b1 ', 'nvidia/segformer-b2' etc are the subsequent models that have better performance with larger model sizes. 15 GB T4 GPU has been used for training the model.

Once the model has been trained, it can be used as the segmentation part of the Necrosis-ML framework. The Figure 5.5 shows the overall workflow of the system using SegFormer. The quantization part will be explained in detail in the Chapter 6.

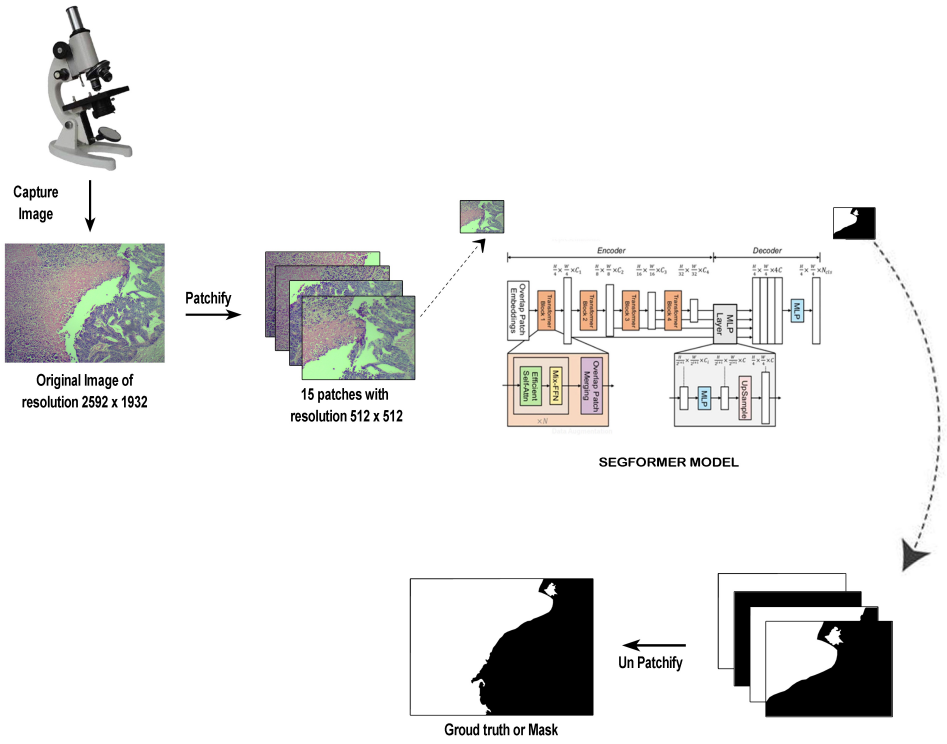


Fig. 5.5 The overall framework of Necrosis-ML using SegFormer

5.2 Experimental Analysis and Results

Usually in SegFormer or Vision Transformer, the computation power required for the image-level segmentation will be much less, as it doesn't require an explicit patchification process. But here, as we have only 60 images in the dataset, this model has been built on patches of the image itself.

In semantic segmentation, accuracy cannot be considered a good performance metric, due to the class imbalance in such cases. In segmentation tasks, the goal is to assign a label to each pixel in an image, outlining the boundaries of objects or regions. So the difference in nature between classification and segmentation tasks often necessitates different loss functions. Out of several performance metrics to measure the performance of the model that we build, we have chosen the Jaccard Index or Intersection over Union (IoU) and Dice loss as they have proven their supremacy in semantic segmentation. Dice loss has been used to measure the loss and the accuracy by IoU score.

Dice loss is a commonly used metric for segmentation loss. It is the measure of similarity between the predicted and ground truth mask, which is calculated as twice the intersection divided by the sum of the pixels in both masks. It was introduced in 2016 for medical image segmentation works. The mathematical formula for binary dice

loss is as follows(ShuchenDu 2020),

$$Diceloss = \frac{\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{5.1}$$

where *pi* is the pixel values of prediction and *gi* that of ground truth. Dice loss gives the measure of overlap between these two sets. As it considers the loss of information both locally and globally, it ensures maximum accuracy for its task.

IoU calculates how much of the predicted segmentation mask overlaps with the ground truth mask (Hasty 2023). The bigger the measure of overlap, the higher will be the accuracy of the model. In other words, it is the ratio between the intersection and union of the two sets, namely pixels of ground truth and predicted mask. It also referred as Jaccard index or Jaccard similarity coefficient. The value of IoU will be almost 1, which is almost ideal case.

$$IoUscore, J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{5.2}$$

where A and B are pixel sets of ground truth and predicted mask. The demonstration of IoU value calculation is given in Figure 5.6(Hasty 2023).

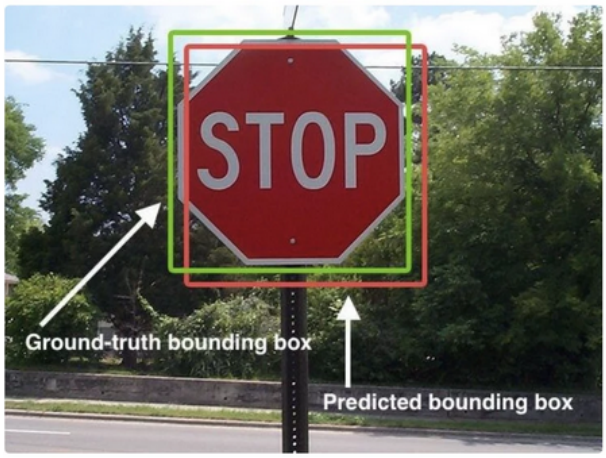


Fig. 5.6 How IoU calculates, the measure of overlap between ground truth and predicted mask

Table 5.1 The hyperparameter values used in the SegFormer model

Parameters	**Values**
Number of epochs	50
Optimizer	Adam
Learning rate	0.00006
Loss	Dice Loss

During model creation, the entire dataset was divided into training, testing and validation set. During training, the training set and validation set will be used and once the model built, its performance will be measured using test set. Validation set is usually used to optimize hyperparameters such as learning rate, batch size, number of layers, regularization techniques etc. Without a validation set, the model might overfit the training data, learning noise or irrelevant patterns. It makes the model more generalized to handle all kinds of external data.

The training, validation, and testing accuracy and loss information is given in below Table.

Table 5.2 Accuracy and loss for the SegFormer model

Parameters	**Values**
Training Accuracy	0.8021
Training Loss	0.0315
Validation Accuracy	0.7581
Validation Loss	0.3652
Testing Accuracy	0.7935
Testing Loss	0.2750

Some of the prediction results using this model are given in Figure 5.7.

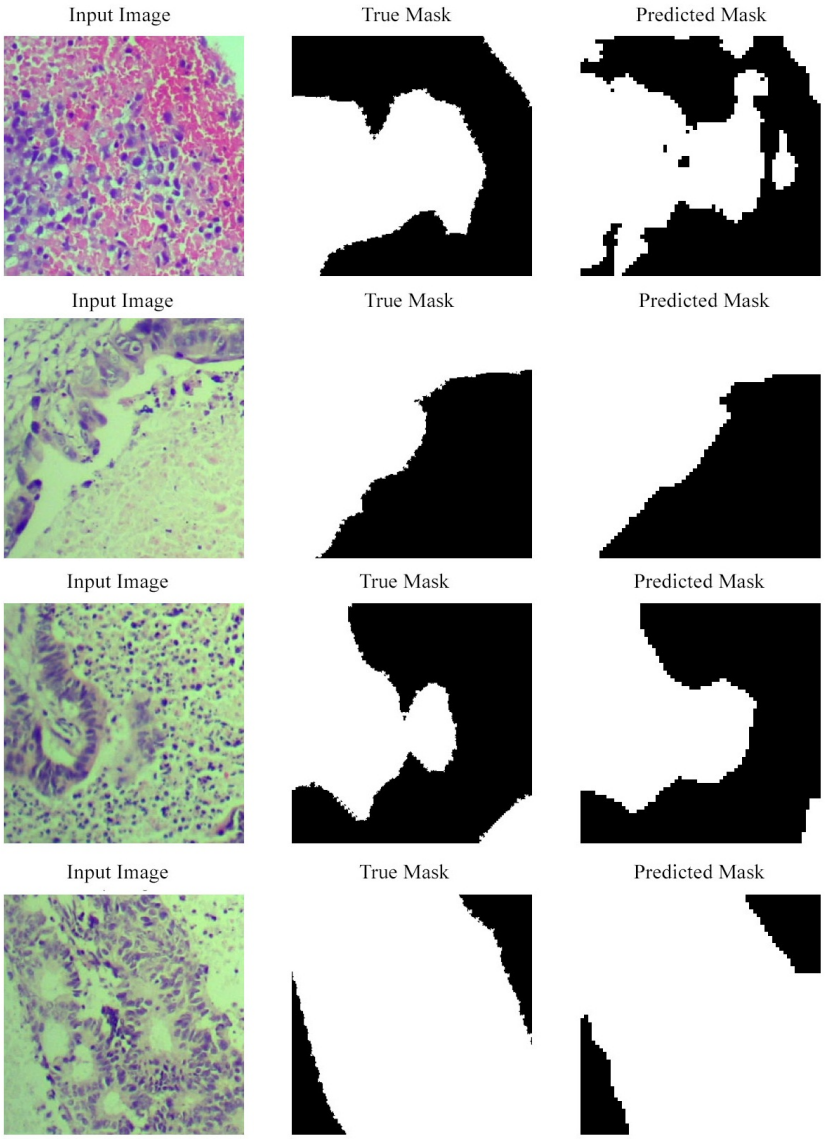


Fig. 5.7 Segmentation results of SegFormer model

5.3 Conclusion

The quantification of necrosis is currently done manually by the pathologists by visualizing tissue sections through the microscope. This is a crude method that can cause significant inter-observer bias. This chapter described the major step towards developing an AI-powered clinical decision-making tool for cancer patients specifically for Osteosarcoma and Renal Cell Carcinoma. A Transformer-based semantic segmentation method has been used in this chapter to do the segmentation of histopathology images.

SegFormer-B5 has been used with a pre-trained model, 'nvidia/mit-b5'. The encoder side is a Transformer and its decoder side is a Multi-layer Perceptron that causes high performance with lesser computation power. The major attraction of this model is that an image will be implicitly converted into patches in this method and the transformer will be working on each of these patches. This will avoid the additional work of patchification and un-patchification. In this study, as we have a very less number of images(60), this model has been applied to patches of the images themselves. This method can contribute to state-of-the-art performance if the dataset size has been increased to some extent.

CHAPTER 6

Necrosis-ML: An AI Tool for Semantic Segmentation and Quantification of Necrotic Tissues from Histopathology Images

6.1 Introduction

Osteosarcoma is a type of bone cancer and Renal Cell Carcinoma(RCC) belongs to kidney cancer. For Osteosarcoma and RCC, one of the crucial prognostic factors is the volume of tumor necrosis created due to Neoadjuvant chemotherapy(NAC). Necrosis is a stage of uncontrolled death of body cells. Usually, it occurs due to electric shock, exposure to radiation, biological and chemical agents, extremely high temperatures, hypoxia, immunological reactions etc. But in this scenario, the cell death is purposefully created by applying NAC to the patients.

In this chapter, an AI-powered segmentation and quantization tool has been introduced named Necrosis-ML, which takes a digital histopathology image as the input and outputs the tumor necrosis volume. It can highlight the necrotic tissues from the entire image even if it is scattered around. In our investigation, we took Haematoxylin and Eosin(HandE) stained resection specimen into account. A detailed description of dataset creation, annotation, and patchification has been given in Chapter 3.

The algorithm of Necrosis-ML is a combination of image processing methods and deep learning algorithms. It mainly contains two tasks: semantic segmentation and quantization of the segmented area of the image. For semantic segmentation, transfer-learning techniques have been utilized. However, as the size of images that we collected from the microscope will be of high resolution, feeding this directly into the neural networks will be a tedious task. So the segmentation model has been developed using patches of the images rather than as the whole.

Some of the challenges associated with training neural networks with high dimension images are:

- High computational cost: The time and space complexity for training the model will be higher
- Data availability and size: The scarcity of getting annotated images is a tedious task

- Overfitting: High-dimensional data increases the risk of overfitting, where the model may learn to memorize the training data instead of generalizing patterns.
- Gradient Vanishing/Exploding: Deep neural networks may suffer from the vanishing or exploding gradient problem, especially in high-dimensional spaces.
- Architectural Complexity: Deeper architectures with more complex structures may be necessary, but they also increase the risk of overfitting and training difficulties.

The process of making patches from the images and after the segmentation, the process of merging multiple masks into a single mask has been done with the help of image processing methods. Different modules involved in this tool method has been depicted in Figure 6.1.

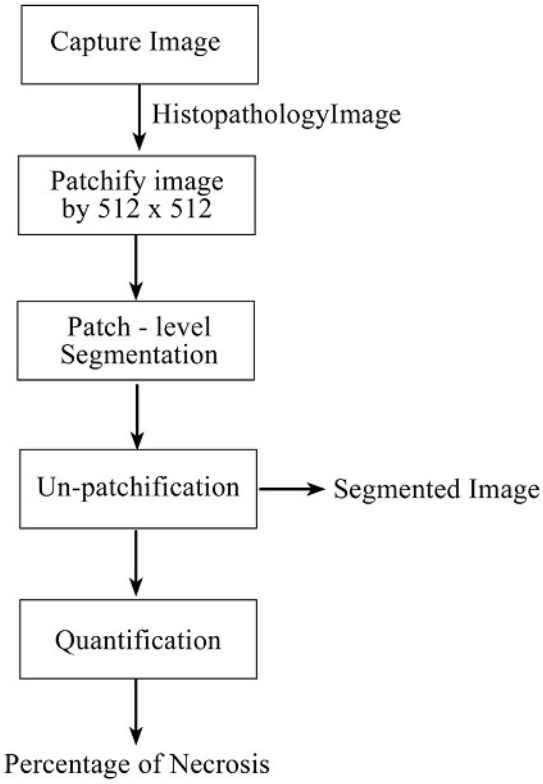


Fig. 6.1 Necrosis-ML modules

6.1.1 Data Pre-processing

In this method, some of the data augmentation techniques have been used in the pre-processing step. The size of the dataset that we have created may not be enough to train a good model. So we can use image augmentation techniques to artificially expand the dataset. It is one of the best regularization methods to reduce the overfitting problem.

Here, a Python library 'albumentations' has been used for this purpose. 'Albumentations' is a Python library for image augmentation, which is capable of effortlessly applying different types of image transformations with just a single line of code (Buslaev et al. 2020). The transformation process is seemingly magical, instantly altering the image as desired. This is an open-source library crafted by Kaggle Masters and Grandmasters with expertise in computer vision that offers more than 70 diverse augmentation techniques. Some of them are shown in the Figure.6.2 These techniques are categorized into two types: pixel-level transforms and spatial-level transforms.

So including different color variations and rotated or cropped copies of original images will increase the accuracy of the model. Moreover, semantic segmentation heavily relies on the pixels of an image, and therefore any augmentation process should avoid modifying the pixel values. This library provides techniques to increase the number of images without altering their pixel values. Different augmentation methods that we have used in our work are CLAHE, RandomBrightnessContrast, Blur, MotionBlur, RandomBrightnessContrast, HueSaturationValue, HorizontalFlip, ShiftScaleRotate, RandomCrop with 512 x 512, and PadIfNeeded with 512 x 512. In the case of pathology images, there is a high chance of staining variations, anatomical variations etc.

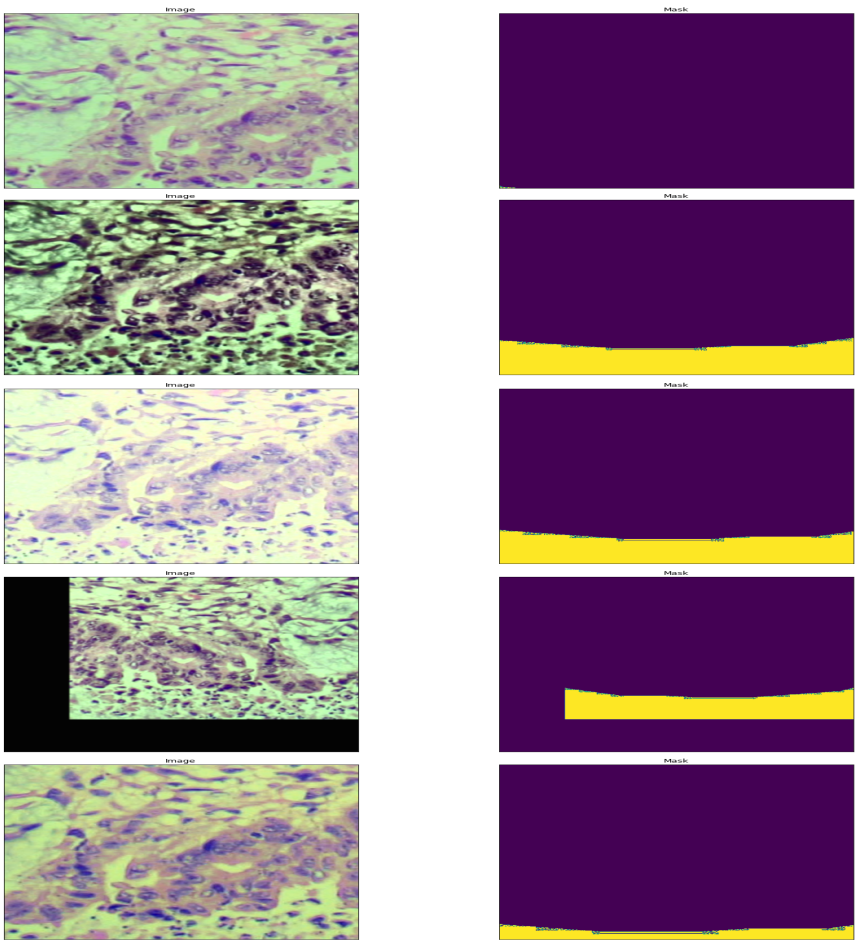


Fig. 6.2 Different augmentation methods applied on an image and its mask(right)

6.2 Proposed Method for Segmentation and Quantization of Histo-pathology Images using Transfer-learning Techniques

The idea behind transfer learning is that if a model is taught on a sizable enough dataset, no matter if it is text or image, speech or video, with enough generality, it can act as a generic model for that specific domain. The parameters of the model like weights, biases, etc will be adjusted such that it will be capable of extracting the features of a given input data (TensorFlow 2022). The feature maps that the model has learned can be re-used in our work without starting from scratch. As the name suggests, it transfers the knowledge in terms of weights from a previously learned task to solve a different but related problem. These models benefit from using very little data as compared with other neural networks for training the model. Most popular pre-trained networks are trained on the ImageNet dataset (Deng et al. 2009). It contains 1000 object classes and 1,281,167 training images, 50,000 validation images, and 100,000 test images. ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is the

most popular and widely used subset of ImageNet.

Necrosis-ML is an AI tool for semantic segmentation and quantification of necrotic tissues from histopathology images. In this work, we train the model that can perform binary segmentation, where the classes can be either necrosis or the background. Here we are using semantic segmentation to separate the region of interest. The input images and their corresponding masks (ground truth) are used to train the model.

6.2.1 Performance comparison of the models built using pre-trained models

A comparison work has been performed with different pre-trained models before selecting the segmentation architecture for Necrosis-ML. Pre-trained neural networks are those which are trained on very large datasets and saved for various purposes. We can make use of these saved models for our use, either classification or segmentation. For this task, we can use the model as is or customize the entire model or through a transfer learning technique. Among these, transfer learning is the most popular and computationally efficient one.

In this comparative study, we employed two pre-trained neural network architectures, specifically U-Net++ and DeepLabv3+ (Saleena, Ilyas, Sajna and Haque 2023). Both models were trained using the dataset we generated. U-Net and DeepLab are convolutional neural networks that adopt the encoder-decoder architecture, forming the foundation for semantic segmentation. It's worth noting that the encoder portion of this architecture can be substituted with pre-trained classification networks such as ResNet, VGG-16, AlexNet, EfficientNet, MobileNet, etc., to enhance feature extraction (Zhang et al. 2020). Even though they are trained for classification tasks, they are getting repurposed for the segmentation process. In this work, we have used ResNet-101 for both U-Net++ and DeepLabv3+. The fully connected layers or dense layers of these networks are transformed into convolution layers (Chen et al. 2018). The backend of each model has been tested with different variations of ResNet and selected the most appropriate one. The illustration of different architectures and pre-trained models used in this comparison study is given below.

6.2.1.1 U-Net+ +

U-Net has been proven as one of the finest segmentation models that had ever been developed for medical image segmentation. They won the ISBI cell tracking challenge 2015 on the dataset of light microscopy images (Ronneberger et al. 2015). Their architecture has an encoder part followed by a decoder part, where the former gradually reduces the dimension of the input image to the lowest one and then the latter does the exact reverse process to get the original image (Ronneberger et al. 2015, Chandhok 2021). The name of the architecture came from its shape, where the expanding path and

the contracting path are more or less symmetric and form a ‘U’ shape.

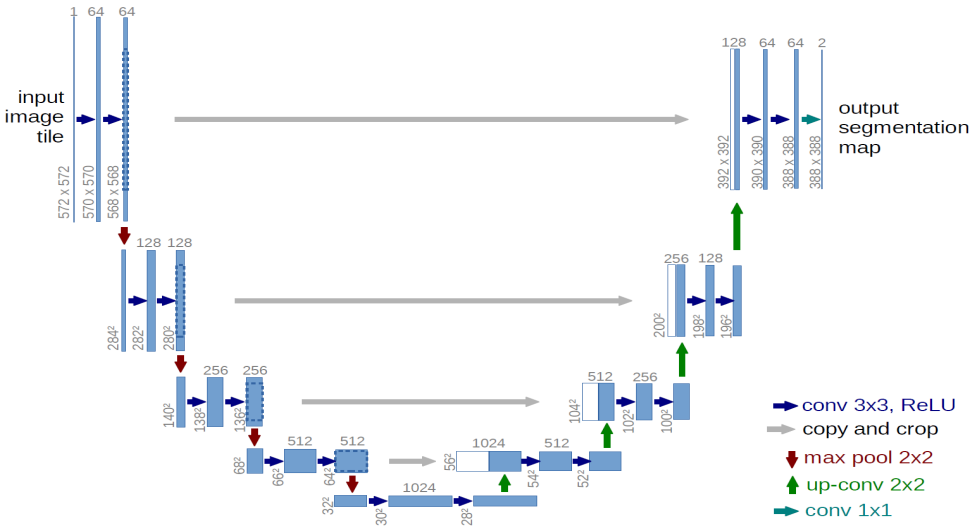


Fig. 6.3 Architecture of U-Net

The basic U-Net consists of a contracting or encoder path for feature extraction followed by a symmetric expanding or decoder path for precise localization. The feature maps obtained can be used for creating a segmentation map at the decoder side. On the contracting side, downsampling of the image occurs with a convolution process using two 3x3 kernels, each followed by a 2x2 max-pooling layer with stride 2 and ReLU activation function. Whilst, at the right side, a symmetric upsampling process with up-convolution will be performed. The entire network contains 23 convolutional layers.

The difference between the U-Net network and the Autoencoder network is the skip connection present in the former which enables the flow of information from layers of the encoder part to corresponding layers in the decoder part that causes good predictions. Skip connection helps the model to retain the information from previous layers. Without a skip connection, it is just a neural network itself. Zongwei Zhou et al. have introduced a novel architecture for bio-medical segmentation which is U-Net++ (Zhou et al. 2018). Its enhanced features from U-Net are,

- The dense convolution layers on the skip pathways
- Dense skip pathway
- Deep supervision

(Zhou et al. 2018)UNet++ enhances the U-Net architecture by introducing nested and dense skip connections between the encoder and decoder networks, which enable the model to extract more detailed information at multiple scales. The dense skip connections combine features extracted from various levels of the network, while the nested

skip connections connect feature maps that have the same scale. The semantic gap between the feature maps of the encoder and decoder sub-networks can be reduced by this re-designed skip connection. In the experiments done on various datasets, UNet++ shows an average IoU gain of 3.9 and 3.4 points over U-Net and its other variants.

The convolution layers within the skip pathway play a crucial role in minimizing the semantic gap between the feature maps of encoder and decoder networks. This is particularly beneficial in the segmentation of medical images, as it simplifies the optimization problem. In this approach, the feature maps from the encoder section are directed into a dense convolution block, contrasting with the U-Net structure where they are directly received at the decoder side. The deep supervision approach has been taken from the idea of Deeply-Supervised Nets or DenseNet by Chen-Yu-Lee et al.(Lee et al. 2015), which allows the model to operate in normal mode and faster mode. In the normal mode, the outputs from all segmentation branches get averaged, whereas, in the faster mode, the final segmentation maps are selected from any one of the segmentation branches. The architecture of this network with the above-mentioned features is shown in Figure 6.3(Zhou et al. 2018).

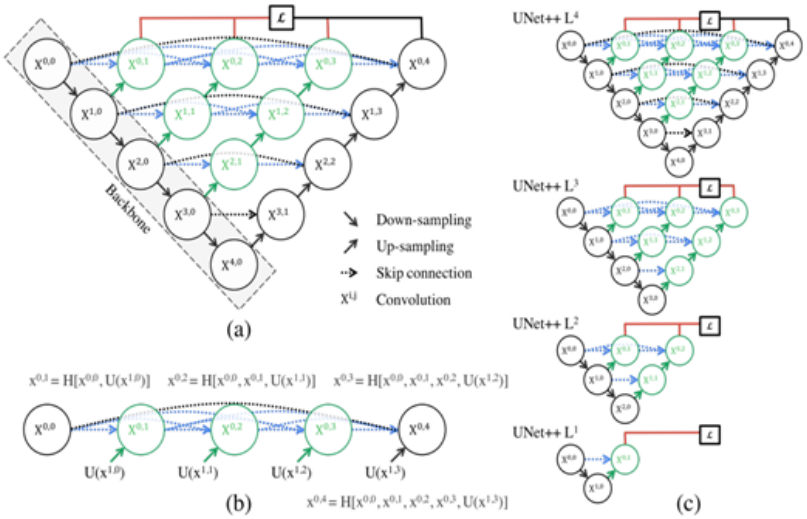


Fig. 6.4 Architecture of U-Net++ where (a)explains the dense convolution layers in skip connection (b) shows the dense skip connection(c) deep supervision.

6.2.1.2 Deeplabv3Plus

DeepLab series is a state-of-the-art semantic segmentation model that has been invented by Google. Among them, Deeplabv3 has proven its excellence as an image segmentation benchmark in PASCAL VOC 2012 with an accuracy of 85.7% without DenseCRF (Chen, Papandreou, Schroff and Adam 2017, Yan et al. 2016). DenseCRF is a post-processing method that is used to refine blurred boundaries in images where we can

add a fully connected CRF (Conditional Random Fields) layer and thereby increasing the efficiency of the algorithm (Yan et al. 2016). Multi-scale segmentation has been made possible in this architecture using atrous convolution in parallel. Atrous Spatial Pyramid Pooling (ASPP) which is proposed in DeepLabv2 is also supported here also. Atrous convolution or dilated convolution allows the networks to extract dense feature maps and thereby removing the downsampling and upsampling overheads. This enables the network to improve feature maps without increasing the number of parameters and eventually the computation power.

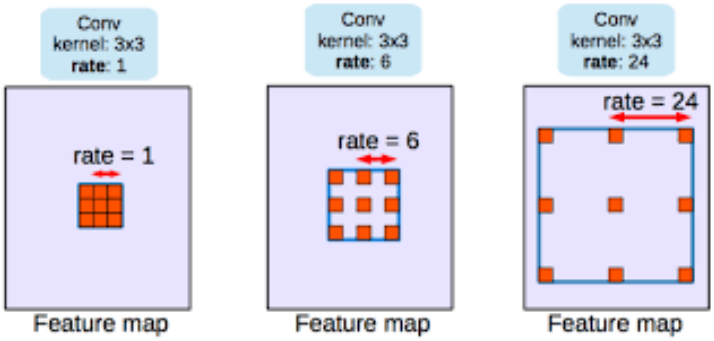


Fig. 6.5 Atrous convolution with kernel size 3x3 and different rates

Figure 6.5. depicts how the atrous rate increases the field of view of the image (Chen, Papandreou, Schroff and Adam 2017). Here, even if the size of the kernel is same everywhere, 3x3, the rates are different. The value of this rate defines the rate by which the dilation by filter occurs. If there are empty values, they are substituted with zeros. A higher rate results in a broader receptive field for the filter. When the rate is set to one, atrous convolution is equivalent to regular convolution.(Gsaxner et al. 2019). The entire architecture of the DeepLabv3+ is depicted in Figure 6.6(Chen et al. 2018).

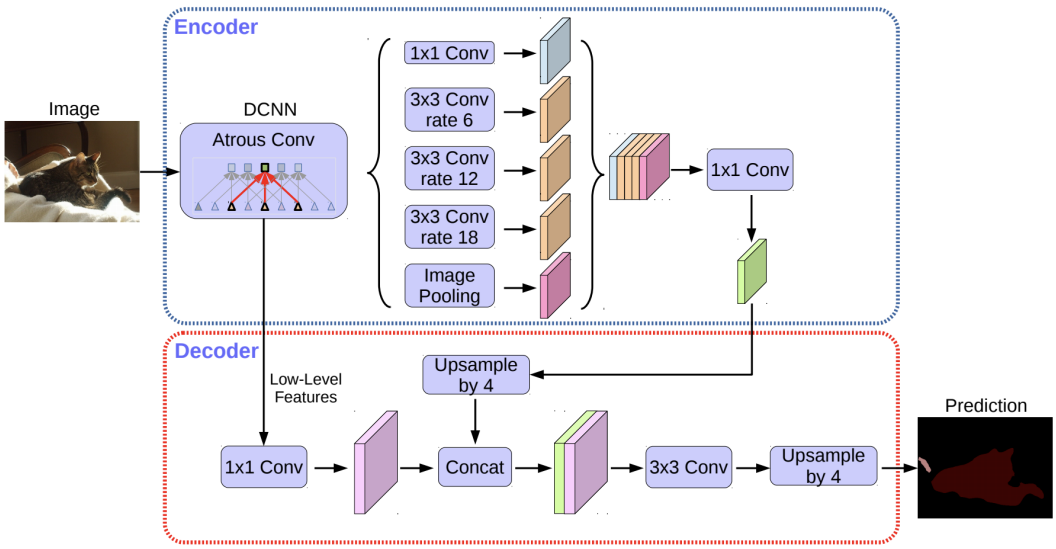


Fig. 6.6 Architecture for DeepLabv3+ with Atrous convolution layers

6.2.1.3 ResNet101

Residual Net or ResNet is one of the most powerful pre-trained classification models available now and was the winner of the ImageNet Large Scale Visual Recognition Challenge 2015 (ILSVRC2015). It was proposed by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in their 2016 paper titled ”Deep Residual Learning for Image Recognition.” As this is a pre-trained model, we can re-use the weights of this network to train our custom model, we don't have to do the training from scratch. So, all the inner layers can be kept frozen. ResNet101 is an extension of ResNet with 101 layers in it. The additional convolutional layers will add more feature extraction facilities along with additional network complexity. As the morphology of pathology images is quite difficult than other images, this added facility will be beneficial.

In CNN, normally there is a convention that the model can be “The deeper the better”. But in some cases, even though the depth of the network can increase the model's performance, at some point it is found that the training error increases as the number of layers increases. This phenomenon is known as the vanishing or exploding gradient. This kind of degradation can be solved using the deep residual learning framework using the technique of skip connections (He et al. 2016*a*).

ResNet is renowned for its residual blocks that can address the problem of vanishing gradient. (Roodschild et al. 2020) Vanishing gradient is a problem that occurs as the depth of a neural network increases beyond a limit. Even though deeper networks are performing better than the shallow ones, training accuracy starts degrading after a certain point. A *deep residual learning framework* has been included in ResNet to solve this problem. This feature helps this model to skip some layers while training and

directly connects to the output(skip connection)(He et al. 2016*a*). The skip connection in each residual block has been explained in Figure 6.7.

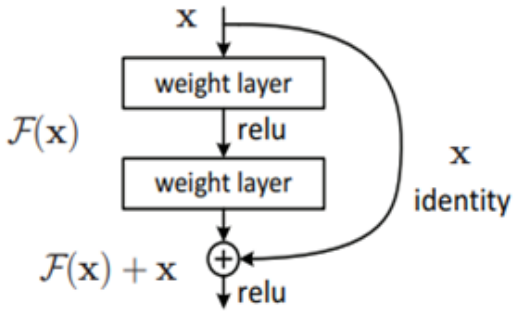


Fig. 6.7 Short-cut connection or skip connection in a residual block, where x is the input for the layer and f(x) +x is meant for short-cut connection which can skip one or more layers

Each of these units can be denoted as follows (He et al. 2016*b*),

$$y_l = h(x_l) + F(x_l, W_l), x_(l+1) = f(y_l) \tag{6.1}$$

where xl and xl+1 are the input and output of the lth unit respectively, F is a residual function, Wl is the set of weights and biases related to the lth unit, h(xl) = xl is an identity mapping and f is Relu activation function. It allows the model to skip some layers while training and directly connects to the output. This feature was introduced in 2015 by Microsoft Research and it has been named ResNet.

Each layer of ResNet is composed of several blocks. How the ResNet architecture differs from other networks are shown in Figure 6.8 (Tsang 2018)

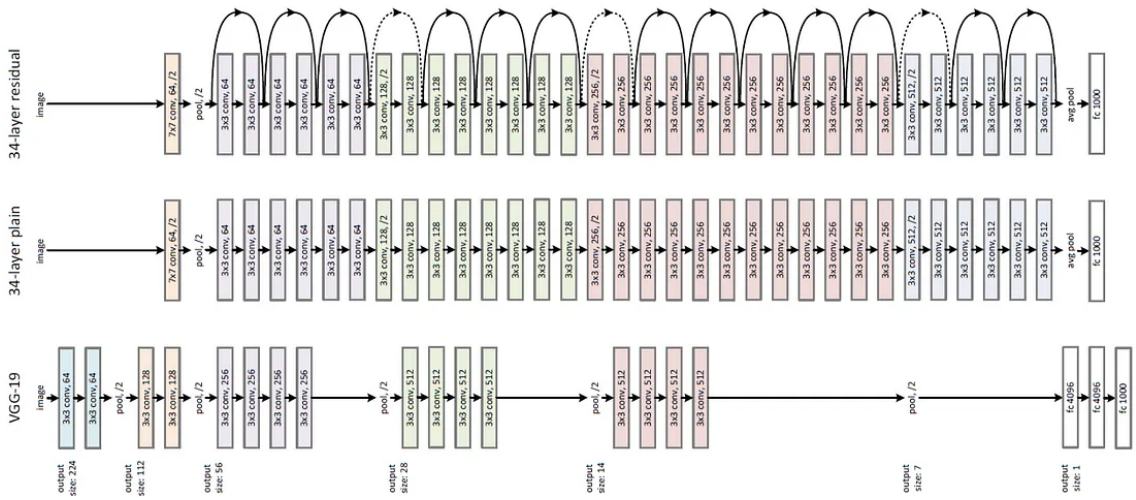


Fig. 6.8 Comparison of the blocks present in variants of ResNet and VGG19

The loss surface will contain several irregular ups and downs if we don't use the skip connection. The loss surface refers to the landscape of the loss function in a high-dimensional space, where each dimension corresponds to a parameter of the neural network. The irregular ups and downs in the loss surface without using skip connections are often associated with the challenges of training deep neural networks. In contrast, a smooth surface will be obtained by using it (Li et al. 2018). The introduction of skip connections helps alleviate these issues by allowing the gradient to flow more directly through the network during back-propagation. With skip connections, the network can more effectively learn and propagate important information through the layers, mitigating the vanishing gradient problem and facilitating the training of deeper architectures. As a result, the loss surface becomes smoother and the optimization process is more stable, leading to improved convergence during training.

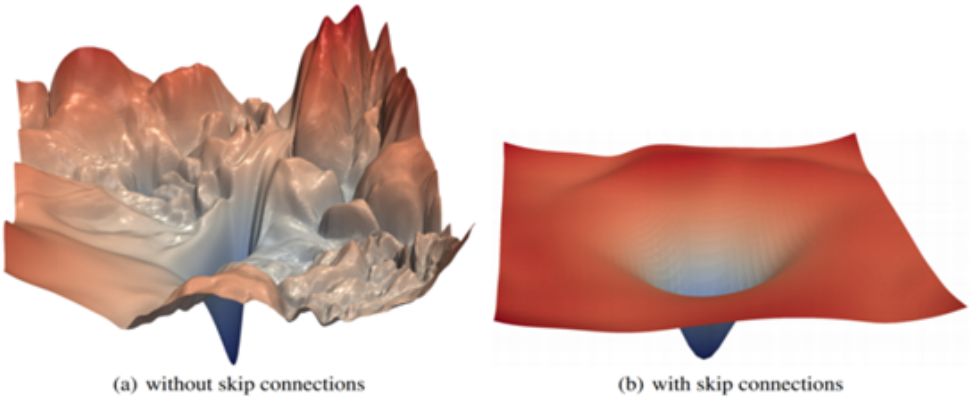


Fig. 6.9 (a) shows the irregularities in the loss surface and (b) is the smooth surface by using skip connection

6.2.1.4 Model Training

In our study, we have used ResNet101 as the feature extractor which is 101 layers deeper as the name suggests. It provides 44.5 million trainable parameters(Boulch 2017). ResNet can act as the downsampling or encoder part of the U-Net. The residual functionality of ResNet makes it a good choice for feature extraction.

Two crucial hyperparameters in this context include the choice of optimizer and the learning rate. In this instance, Adam serves as the optimizer, and the learning rate is specified at 0.0001. Given that this is a binary model, Sigmoid is employed as the activation function. The training process spans 120 epochs, with the learning rate decreasing to 1e-5 (0.00001) after the initial 80 epochs. While a decreasing learning rate extends the training time and, consequently, the GPU cost, it contributes to more accurate predictions. The selection of the best model is based on the highest validation Intersection over Union (IoU) score achieved throughout the training phase(Saleena, Ilyas, Sajna and Haque 2023).

Pseudo code 1: model training

The following algorithms have been used to build the model and quantization process.

1. Set learning rate = 0.0001
2. Set max score = 0
3. Repeat for 120 epochs
    - if validation iou score is greater than max score,
      set max score = validation iou score
      save the model with training and validation parameters of that epoch
    - if epoch = 80 then reduce learning rate = 0.00001

6.2.2 Necrosis-ML: AI tool for semantic segmentation and quantization of histopathology images

Necrosis-ML is an AI tool developed for semantic segmentation and quantification of necrotic tissues from histopathology images of post-neoadjuvant chemotherapy resection specimens. Image processing techniques and deep-learning algorithms have combined in this algorithm and it has been used in this work to do the image-level segmentation and quantization. Different Python libraries like OpenCV, Patchify, numpy etc are used for the image processing tasks. PyTorch has been used to handle the deep-learning(segmentation) part.

The entire algorithm includes multiple sections to cut the image into patches, the segmentation of individual patches, the merging of multiple masks, and finally the quantification. To cut the high-resolution image into patches of size 512x512, we have used a Python library, 'patchify'. The output of the function will be multiple patches or sub-images of the original image in a specific size and each patch can be processed independently. The output will be in numpy array format that can be processed using suitable image-processing techniques. The function accepts 3 input parameters, input image as a numpy array, patch size as tuple of (width, height) and step-size which is the stride for getting non-overlapping patches. Usually, the patch size will be given as the step size. Each patch can be undergone segmentation by the model created and will get the corresponding mask. A list of masks will be created from this and will be merged into a single output by preserving the order.

For the segmentation task, we utilized a combination of two architectures, namely U-Net++ and ResNet101, operating at the patch level. We obtained a mask, which is a binary image from each patch that can be considered as the output for segmentation. In this method, the performance of the model has been evaluated using IoU Once the segmentation was complete, we performed a reverse process known as un-patchification, which involved combining these segmented masks to get the combined output. This resulted in the final output of the image segmentation, where distinct areas were marked by masks indicating the boundaries of the segmented regions.

To quantify the segmented area, we calculated the measure of the segmented region from the combined image. This allows us to determine the extent or quantity of the segmented area within the original image. The segmented region will be quantified using the equation,

$$Percentage = (N/T) * 100 \tag{6.2}$$

where N - Number of pixels of necrosis in mask and T - total number of pixels in it The above section showed that while comparing the performance of the models, the one with U-Net++ architecture has the better accuracy. So the segmentation part of the Necrosis-ML has been designed using this architecture.

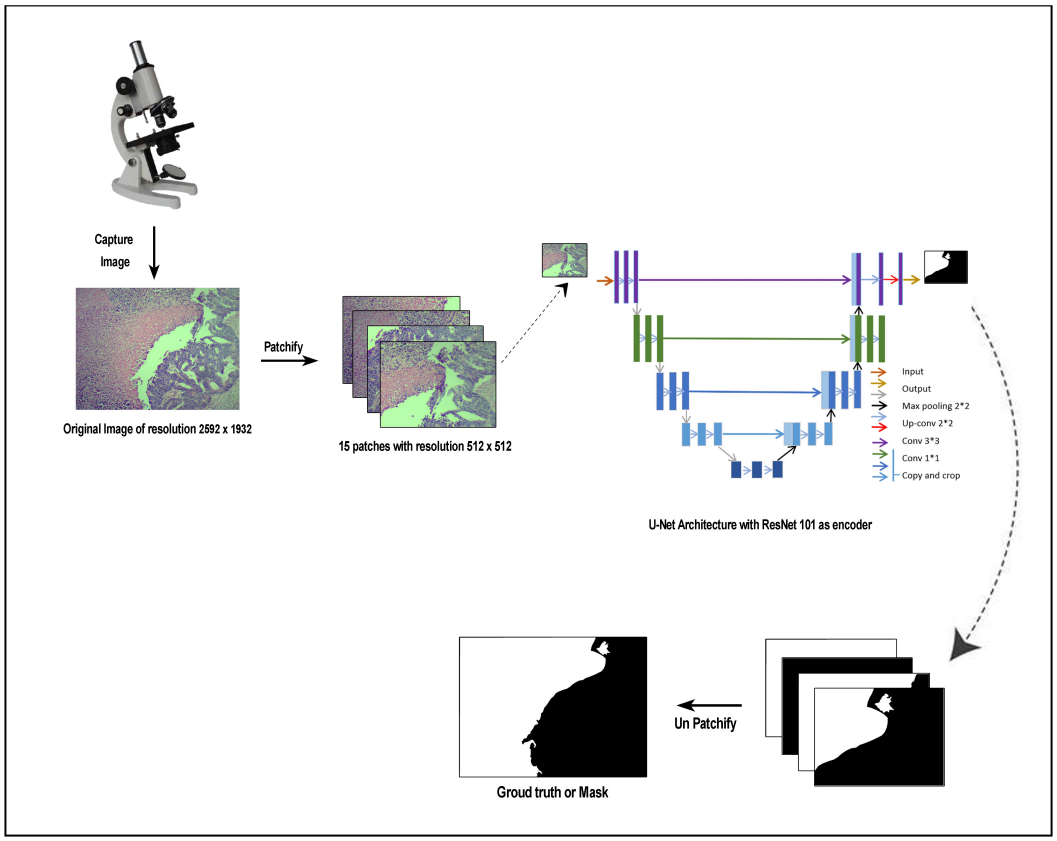


Fig. 6.10 Workflow of the Necrosis-ML with Transfer learning model

6.2.2.1 Patchification and Reverse patchification

As part of this research project, we have collected 60 images with a resolution of 2592 x 1932. However, this quantity of images is insufficient to properly train the model. Moreover, increasing the dimensions of the input to the neural network beyond a limit will lead to various complications.

- When each pixel of an image is processed by a neuron, the number of neurons and the complexity of the model also increase with the image's resolution.
- This will demand high computation time and resources which will lead to more training and validation time.
- This may also lead to overfitting as the model has been over-trained on these high-dimensional data.
- In the case of pathology images, tissue-level analysis is more crucial for the diagnosis. However, if this is done with high-resolution images, there is a significant risk of losing critical information.

(Hou et al. 2016)Apart from all these points, pathologists are differentiating the cancer sub-types by looking at the cellular-level morphology in the patch-scale level of digital pathology images.

All the above-mentioned reasons point out that image-level segmentation and analysis will be more imprecise than patch-level. But when this research becomes a product, patch-level analysis is not possible. The image that has been captured from the microscope will be directly fed to the software. Therefore as the solution, we have applied pathification on the input image and performed patch-level segmentation using U-Net model.

Due to the high dimension of the image we gathered for our research (2592 x 1932), it poses a challenge to process it efficiently using deep neural networks. To overcome this obstacle, we employed a technique where we divided the images into smaller sections called patches, each with a dimension of 512x512.

The masks obtained as the output of this segmentation model, have merged using un-patchification and a single mask will be outputted. The area of the segmented region will be calculated from this. Gatti et al. has described the patchification and un-patchification with real life examples with the help of Python algorithms(Gatti 2022).

Figure 6.11 depicts how the patchification can be applied to a histopathology image.

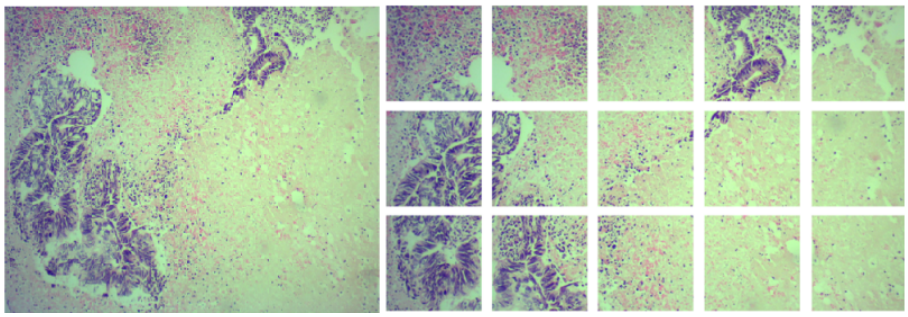


Fig. 6.11 A histopathology divided into patches

Pseudo code 2: How does the Necrosis-ML works

1. Read the image of any size
2. Cut the image using the function patchify() by giving 3 parameters as input image, patch size = 512x512(HxW), and step-size = 512, where H-height, W-width
3. The output of the function will be an n-dimensional array, with dimension (rows, cols, 1, H, W, N), where rows and cols are the number of rows and columns patches obtained, N is the number of channels of patch. We can access each patch for processing.

4. Each patch has fed as the input for the segmentation model and mask will be obtained as output.

5. Load the saved segmentation model and do the following steps on each patch

    - Applying the pre-processing techniques of ResNet101 on the data
    - converting the patch to a Torch Tensor and add an extra dimension to match the expected shape of the input data by the model
    - Use the output model to predict the mask
    - Squeeze the tensor to remove the additional dimension
    - Convert the tensor to a numpy array
    - Append this mask to a list, for merging, lst-patches

6. Create a new blank image for merging the masks of the patches

    - The shape of the original image will be measured.
    - A numpy of zeroes with shape(HxW) where H-height and W-width is the dimension of original image(channel is not needed as we need to create a binary image)

7. Place each patch of lst-patches on the blank image

    - Calculate the position of each patch on the final image, using the number of rows, columns, and total number of patches, which is the same as step 3.
    - Iterate through each patch and place them onto the blank image according to their intended positions.
    - The pixel values from each patch are copied onto the matching area of the empty picture in this method.

8. Save the resultant mask to the disk.

9. Count and display the area of the segmented region in percentage using equation 6.3.

    - nec-pixels =Get the count of pixels having non-zero value, using the count-NonZero function of openCV library
    - total-pixels = total number of pixels in the whole image from the shape of the original image(height*width)
    - Percentage = (nec-pixels / total-pixels)*100.

6.2.2.2 Model training and performance evaluation for segmentation

We have used U-Net++ for the semantic segmentation process with ResNet101 as the backbone. The training of the model has been done using the GPU, Tesla V100-PCIE-32GB with 12 workers for training and 4 workers for validation. A step size of 7 and a gamma value of 0.1 have been used during training to decay the learning rate after each 7 epochs by a multiplicative factor of 0.1. This will help the system to slowly gain maximum accuracy. We have used GPU for training this model, as it is consuming very high computation power. The implementation of model training has been done using the PyTorch library, which provides easy methods to switch between CPU and GPU. PyTorch is a Python-based, open-source library developed by Facebook's AI Research Lab (FAIR). It is widely used for tasks involving deep learning, including neural network implementations, optimization, and building machine learning and deep learning models(PyTorchContributors 2023).

The main challenge in segmenting the necrosis was the color variation in staining. Pre-processing methods that are in-built into U-Net++ and Deeplabv3+ have been used for the data cleaning. The dimension of the image is 512x512x3 and that of the mask is 512x512x1. The entire dataset has been divided into 3 sets for training, validation, and testing. We have used 16 parallel workers, with a batch size of data as 64.

The performance metrics used for classification and segmentation will be different as their method of calculation will be different. Necrosis-ML is a binary class segmentation model and thus we can not use 'accuracy' metric for the evaluation. Jaccard Index or Intersection over Union(IoU) is the more suitable metric here (Hasty 2023).

$$IoU = \frac{Areaofoverlap}{Areaofunion} \tag{6.3}$$

For multi-class segmentation tasks with more than two classes, mIoU or Mean Intersection over Union will be more preferable. For calculating mIoU the following equation can be used,

$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoUi \tag{6.4}$$

Where N is the total number of classes.

In this work, a combination of binary cross-entropy and dice coefficient has been used as the loss function, which can be denoted as (van Beers 2018),

$$L(Y, \widehat{Y}) = -\frac{1}{N}\sum_{b=1}^{N}(\frac{1}{2}.Y_b.log\widehat{Y}_b + \frac{2.Y_b.\widehat{Y}_b}{Y_b + \widehat{Y}_b}) \tag{6.5}$$

Where N is the batch size, Y_b is the ground truth, and $\widehat{Y}_b$ is the prediction value for bth image.

Table 6.1 The hyperparameter values we have applied in this model

Parameters	**Values**
Number of epochs	40
Optimizer	Adam
Learning rate	0.001
Loss	Dice Loss

6.3 Results

Table 6.2 explains the segmentation loss and accuracy during training, validation, and testing for both the models, U-Net++ and DeepLabv3+.

Table 6.2 The Dice loss value and IoU value obtained during segmentation

Parameters	**U-Net++**	**DeepLabv3+**
Training loss	0.0440	0.0477
Validation loss	0.2796	0.2582
Testing loss	0.1777	0.1780
Training IoU	0.9178	0.9109
Validation IoU	0.7513	0.7457
Testing IoU	0.8264	0.8150

The values of loss and accuracy during the training and validation process have collected and drawn graphs for both U-Net++ and DeepLabv3+ as shown in Figures 6.12 to 6.15. Loss or accuracy is taken at the y-axis and epochs at the x-axis.

Fig. 6.12 Training and validation loss of U-Net++ model

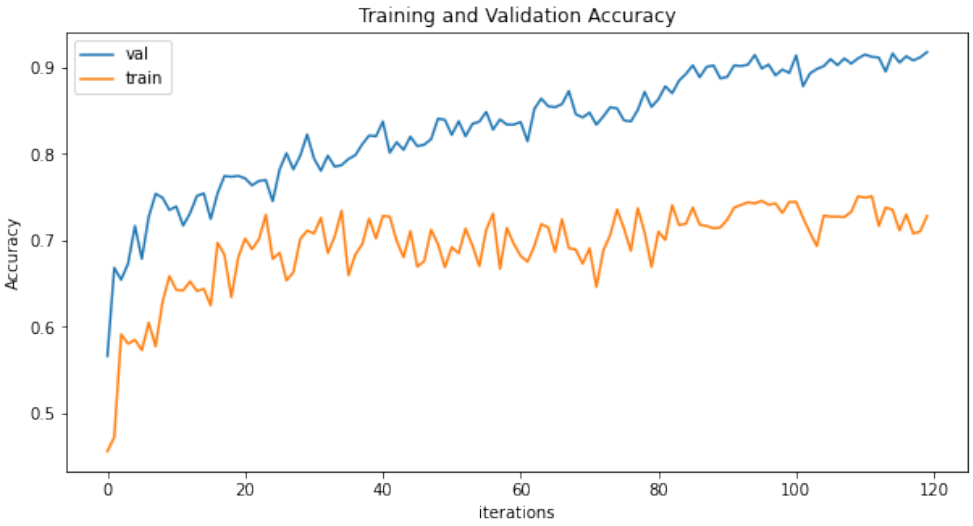


Fig. 6.13 Training and validation accuracy of U-Net++ model

Fig. 6.14 Training and validation loss of DeepLabv3+ model

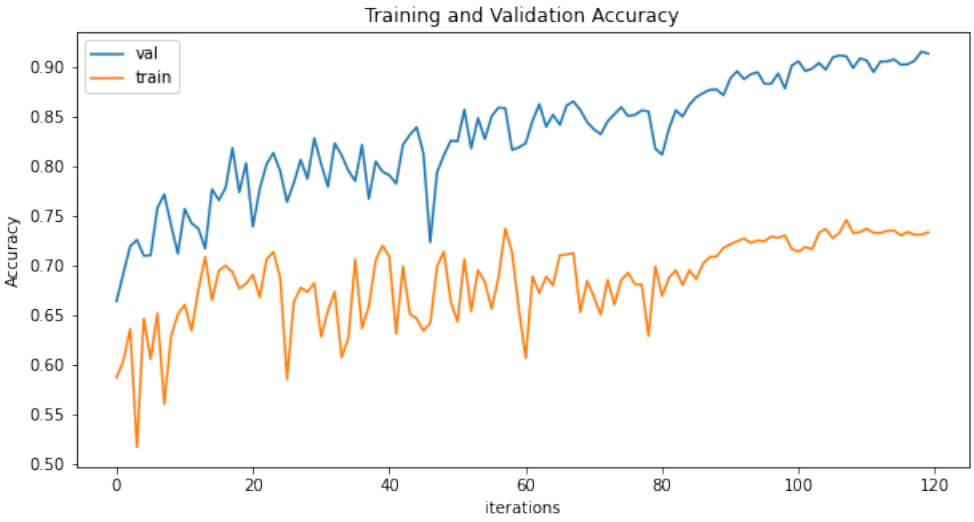


Fig. 6.15 Training and validation accuracy of DeepLabv3+ model

The results show that both the models are performing more or less the same and have many similarities with the manual annotation.

Figure 6.16 depicts how the mask or segmented region originated from our proposed model related to the actual mask. Most of the testing images are showing good likeness with the original images.

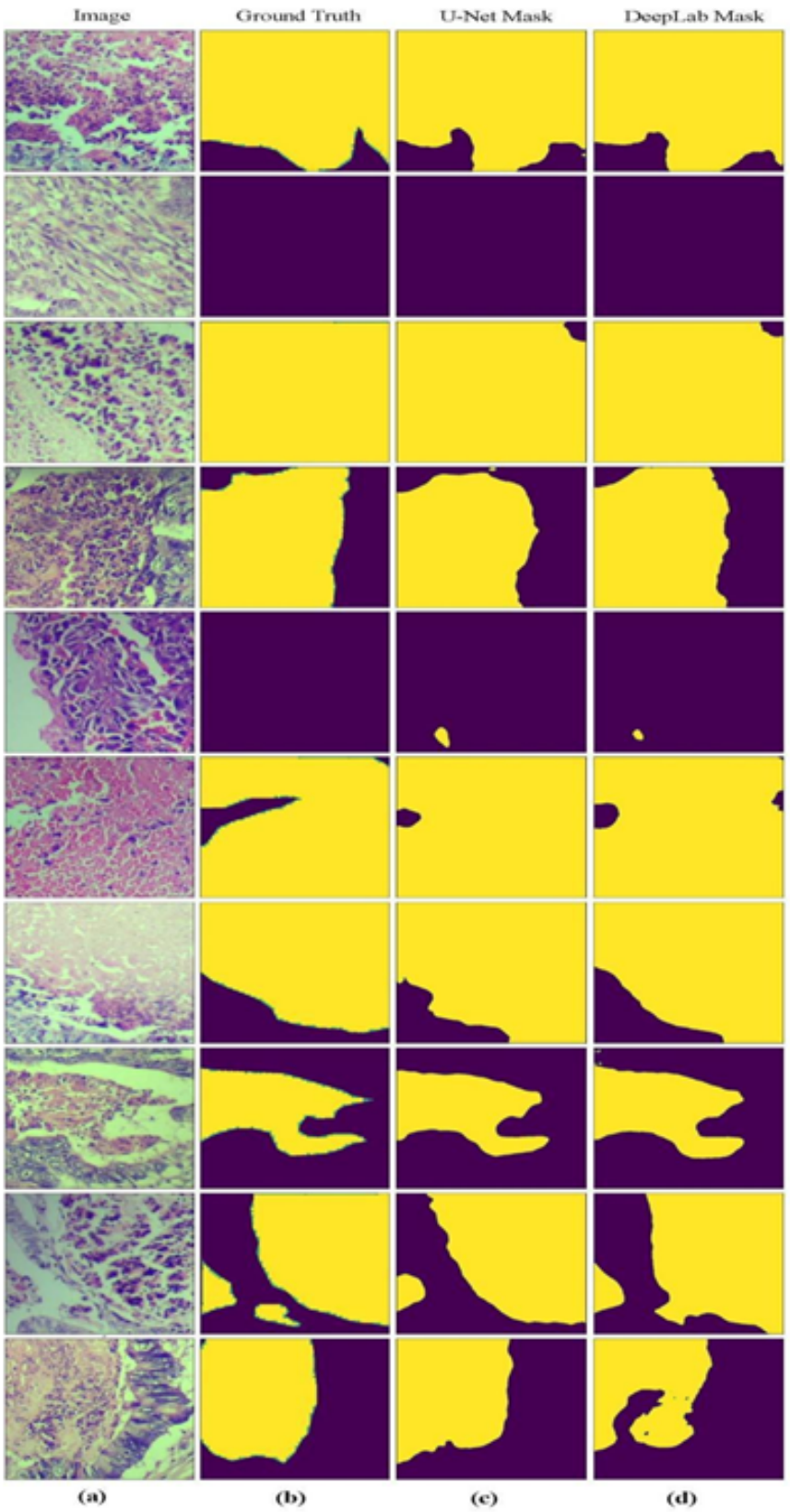


Fig. 6.16 Segmentation result with training set (a) Original image(b)manually annotated ground truth (c)result from the model, U-Net++ (d) result from the model, DeepLabv3+

6.4 Conclusion

For accurate diagnosis, tissue-level analysis is necessary. Usually, the digital histopathology images captured from a microscope will be of higher dimensions which will be difficult to process directly if it has been used as input for the neural networks. This will make image-level processing difficult using deep learning methods. In our study, we have proposed an image-level semantic segmentation followed by a necrosis quantization method. The process starts by cutting the image into 15 patches using proper image processing techniques and semantic segmentation will be applied on each patch. The architecture we have selected for this task is U-Net++ with ResNet101 as the backbone. Before selecting the segmentation architecture, a performance comparison has been done by building two models using U-Net++/ResNet101 and the next is DeepLabv3+/ResNet101, where UNET++ shows slightly higher performance than the other. After the segmentation process, image processing techniques will be used to do reverse patchification of those patches by preserving the order. The area of the segmented region will be calculated from the restored image. This quantity will be taken as the most important prognostic factor for determining the treatment plan for the patient. The result that we obtained was compared with manual segmentation done by two pathologists and we got comparable results. This study also contributed a dataset of 900 images along with their ground truth and made it public.

CHAPTER 7

Comparative Analysis of Various Proposed Methods of Segmentation in Necrosis-ML

In this research, two types of cancers have been considered, Osteosarcoma and Renal Cell Carcinoma, in which tumor necrosis created due to Neoadjuvant chemotherapy determines the treatment plan for a patient. The quantity of dead tumor tissues will indicate the response of the patient's body towards particular therapeutics and thus can measure the prognosis value. Currently, the identification of tumor necrosis and its estimation is done manually by the pathologists. This is a labor-intensive process and is highly prone to error. This research aims to find the solution to this problem along with a proper, annotated dataset that will be beneficial for later researchers in this area.

In this study, three different architectures have been used for the semantic segmentation. The first one uses Autoencoders, the second uses transfer learning techniques, and the last one uses SegFormer. The common factor in all of these architectures is the encoder-decoder component available in it. The encoder part of all these architectures is used for feature extraction and dimension reduction, whereas the decoder part reconstructs the compressed data into its original form by retaining all the relevant features of the input image. The methods used for feature extraction and image reconstruction will be in different manner, based on which the performance of the methods will be drastically different.

In the first method, a Convolutional autoencoder has been used for the segmentation which is a self-supervised neural network. Its encoder part is built using simple convolution layers and pooling layers where five convolution layers along with three pooling layers have been used in it. Symmetric layers of upsampling are there on the decoder side to do the image reconstruction. In the transfer learning model, U-Net++ has been used as the encoder-decoder architecture, whereas ResNet101 has been used as the encoder. A similar number of upsampling layers will be there at the decoder end to do the reverse operation. In this method, a comparative analysis has been performed on U-Net++ and DeepLabv3+. In the third method, SegFormer-B5 has been used in which the encoder part is designed using a Transformer and the decoder part is run by Multi-Layer Perceptron instead of upsampling layers. This makes the model more

faster and efficient than any other model.

Table 7.1 The Loss and Accuracy(IoU) values obtained during segmentation in different models

Parameters	**U-Net++**	**DeepLabv3+**	**SegFormer**	**Autoencoder**
Training loss	0.0440	0.0477	0.0315	0.1096
Validation loss	0.2796	0.2582	0.3652	–
Testing loss	0.1777	0.1780	0.2750	0.3240
Training IoU	0.9178	0.9109	0.8021	0.8228
Validation IoU	0.7513	0.7457	0.7581	–
Testing IoU	0.8264	0.8150	0.7935	0.7258

Table 7.1 shows the comparison of accuracy and loss values obtained from various segmentation models used in this research. All these models have been trained on patches of images. In Necrosis-ML any of these segmentation models can be used but in this study, the U-Net++/ResNet101 has been used. As it is trained on the patches of images, additional image processing methods of patchification and unpatchification have to be included in it to make it practically useful. The SegFormer was a good choice for handling this hurdle due to its methodology. This framework divides the input image into patches before feeding the data into the encoder. But the images we have collected are very limited in number, 60 which is not adequate to get a good result in this architecture. In this study, we have trained the SegFormer with patches of images itself which is almost showing accuracy as of transfer learning methods.

CHAPTER 8

Conclusion

Therapeutical decision-making is the most critical factor in every disease especially cancer, as it decides the survival rate of the patients. Each type of cancer will have its own prognostic factors that lead to this decision-making. In the case of Osteosarcoma and Renal Cell Carcinoma, one of the chief accelerating elements is the volume of tumor necrosis generated due to Neoadjuvant chemotherapy that has been injected into the patient. By looking at the quantity of this, the physicians can assess the response of that patient's body toward particular therapeutics. If this quantity is higher, the greater the response of the patient is. Post-neoadjuvant chemotherapy resection specimens are used for this analysis.

Currently, there is no specific tool to do this kind of analysis in the case of Osteosarcoma and Renal Cell Carcinoma. These evaluations are usually carried out by visual estimation, which is labor and time-intensive and can result in significant inter and intra-observer variability. The emergence of digital pathology brought many changes in this situation. The application of AI in digital pathology is helping the entire medical team to reduce their workload. The proposed system can be an automated tool to guide clinical decision-making, using digital histopathology images.

Through this study, we filled the research gap in this problem by developing an automated tool, called Necrosis-ML that can perform the segmentation and quantification of tumor necrosis from a given histo-pathology image. Necrosis-ML has been developed using different image-processing and deep-learning methods. Due to the high dimension of the image(2592 x 1932) that we collected for our research, it poses a challenge to process it efficiently using deep neural networks. To overcome this obstacle, a technique was employed where we divided the images into smaller sections called patches, each with a dimension of 512x512. This method has made use of deep neural networks for the segmentation task which is equipped with U-Net++ as the segmentation architecture and ResNet101 as the feature extractor, operating at the patch level. A binary image has been obtained from each patch called a mask that can be considered as the output for segmentation. Once the segmentation was complete, a reverse process was performed known as un-patchification, which involved combining these segmented masks to get the combined output. This resulted in the final output of the image seg-

mentation, where distinct areas were marked by masks indicating the boundaries of the segmented regions. To quantify the segmented area, we calculated the measure of the segmented region from the combined image. This allows us to determine the extent or quantity of the segmented area within the original image.

For the segmentation process, some other architectures were also tried namely Autoencoder and SegFormer. Both have encoder-decoder architecture. Initially, a work was done by developing an autoencoder model trained with a single gold-standard image. It showed considerable performance in images that have almost similar morphology to the training image, but it failed in others. One more autoencoder model was built with the entire dataset, but its accuracy was very poor so it did not apply to building a real-time system. The next work was done with a Transformer-based segmentation architecture called SegFormer. Its encoder side was fit with a Transformer and the pretrained model used in it was nvidia/mit-b5. In this work, there is no need to apply explicit patchification methods as it converts its input image into patches on its own. But the main drawback of this method was the lesser number of images in the dataset.

The major contribution of this research is the dataset developed as part of this study with 900 patches of images along with their ground truth. The images are collected and annotated by experienced pathologists and the samples are selected such that they contain the area of necrosis in different color variations, the measure of dispersion, and also other prevalent tissues like fibrosis, tumors, bone tissues, etc. There is no published dataset yet for necrosis calculation of this particular case and thus this dataset will be very helpful to the research community.

The main contributions of this study are:

- Introduced a dataset, NecrosisDB of 900 images and their labels. This dataset deploys various morphologies of necrosis, tumors, fibrosis and other frequently occurring tissue elements while quantizing the tissue necrosis in the case of post-neoadjuvant chemotherapy resection specimen. This dataset has been made publicly available only for the research and study purpose.

- Developed an AI-based, image-level semantic segmentation tool to grade Osteosarcoma and Renal Cell Carcinoma patients and thereby can be used as a second opinion for pathologists and doctors to take critical decisions on treatment plans.

- This tool can be used to quantify how much cancerous tissue has been dead (tissue necrosis) due to Neoadjuvant chemotherapy using pathology images. This value of measurement will be used to identify the response of the patient's body toward the therapeutics.

- The area of necrotic tissues will be highlighted from a given microscopic image.

CHAPTER 9

Recommendations

The following recommendations can be suggested for this research,

- Necrosis-ML is the segmentation and quantization tool for the necrotic tissues from histopathology images. As a future extension of this work, segmentation of other kinds of tissues like fibrosis, oedema, tumors or any other tissue component that may be of therapeutic importance can be attempted.
- The dataset that we have created is based on limited data. One of the reasons is that, before the study can proceed, the pathologist has to annotate the data, and this is a time-consuming manual process that limits the number of annotated images that can be made available. For this study, we only obtained 60 variety samples. Consequently, we can incorporate more pictures with various morphological and staining variances.
- Segmenter and TrasUNet are some of the Transformer based models that we can use which are good alternatives for U-Net++ in both computations and accuracy, but they need more images for training the model. If we use them, the extra work of patchification and un-patchification can be avoided. The problem of segmenting high-resolution images can be solved using these types of architectures.
- After developing the full-fledged software it can be implemented in different hospitals for the automated tissue necrosis quantization process. Also, this particular work can be deployed as a website or as embedded software along with digital image-capturing microscopes so that pathologists can easily make use of it.
- Lastly, as there is no previous work in this particular case, the benchmarking of the dataset is difficult. The dataset provided in this research can be used for further enhancement in this area. Further research that may follow this study can include more variation in data with separate annotations of different tissue elements.

REFERENCES

1. Al-Salama, Z. T. and Keating, G. M. (2016), 'Cabozantinib: a review in advanced renal cell carcinoma', *Drugs* **76**(18), 1771–1778.

2. Alawad, A. A. M. (2014), 'Evaluation of clinical and pathological response after two cycles of neoadjuvant chemotherapy on sudanese patients with locally advanced breastcancer', *Ethiopian journal of health sciences* **24**(1), 15–20.

3. Alturkistani, H. A., Tashkandi, F. M. and Mohammedsaleh, Z. M. (2016), 'Histological stains: a literature review and case study', *Global journal of health science* **8**(3), 72.

4. An, J. and Cho, S. (2015), 'Variational autoencoder based anomaly detection using reconstruction probability', *Special Lecture on IE* **2**(1), 1–18.

5. Badr, W. (2019), 'Auto-encoder: What is it? And what is it used for? (part 1)', https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for- part-1-3e5c6f017726

6. Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017), 'Segnet: A deep convolutional encoder-decoder architecture for image segmentation', *IEEE transactions on pattern analysis and machine intelligence* **39**(12), 2481–2495.

7. Bai, H., Mao, H. and Nair, D. (2022), Dynamically pruning segformer for efficient se- mantic segmentation, *in* 'ICASSP 2022-2022 IEEE International Conference on Acous-tics, Speech and Signal Processing (ICASSP)', IEEE, pp. 3298–3302.

8. Baldi, P. (2012), Autoencoders, unsupervised learning, and deep architectures, *in* 'Proceedings of ICML workshop on unsupervised and transfer learning', JMLR Workshopand Conference Proceedings, pp. 37–49.

9. Bandyopadhyay, H. (2022), 'An introduction to autoencoders: Everything you need toknow', https://www.v7labs.com/blog/autoencoders-guide.

10. Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G. et al. (2017), 'Qupath: Open source software for digital pathology image analysis', *Scientific reports***7**(1), 1–7.

11. Bokhorst, J.-M., Pinckaers, H., van Zwam, P., Nagtegaal, I., van der Laak, J. and Ciompi, F. (2019), 'Learning from sparsely annotated data for semantic segmentation in histopathology images'.

12. Bomma, P. (2022), 'Semantic segmentation with seg-former', https:// medium.com/geekculture/semantic-segmentation-with-segformer-2501543d2be4.

13. Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A. and Mastorides, S. M. (2019), 'Lung and colon cancer histopathological image dataset (lc25000)', *arXiv preprint arXiv:1912.12142* .

14. Boulch, A. (2017), 'Sharesnet: reducing residual network parameter number by sharingweights', *arXiv preprint arXiv:1702.08782* .

15. Bousias Alexakis, E. and Armenakis, C. (2020), 'Evaluation of unet and unet++ architectures in high resolution image change detection applications', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **43**, 1507–1514.

16. Bradski, G., Kaehler, A. et al. (2000), 'Opencv', *Dr. Dobb's journal of software tools* **3**(2).

17. Burcombe, R., Makris, A., Pittam, M., Lowe, J., Emmott, J. and Wong, W. (2002), 'Evaluation of good clinical response to neoadjuvant chemotherapy in primary breast cancer using [18f]-fluorodeoxyglucose positron emission tomography', *European Jour- nal of Cancer* **38**(3), 375–379.

18. Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M. and Kalinin, A. A (2020), 'Albumentations: Fast and flexible image augmentations', *Information* **11**(2). **URL:** *https://www.mdpi.com/2078-2489/11/2/125*

19. Caparica, R., Lambertini, M., Pondé, N., Fumagalli, D., de Azambuja, E. and Pic- cart, M. (2019), 'Post-neoadjuvant treatment and the management of residual disease in breast cancer: state of the art and perspectives', *Therapeutic advances in medical oncology* **11**, 1758835919827714.

20. Carrle, D. and Bielack, S. S. (2006), 'Current strategies of chemotherapy in osteosar- coma', *International orthopaedics* **30**(6), 445–451.

21. Chandhok, S. (2021), 'U-net: Training image segmentation models in pytorch', https://pyimagesearch.com/2021/11/08/ u-net-training-image-segmentation-models-in-pytorch/.

22. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L. (2017), 'Deeplab: Semantic image segmentation with deep convolutional nets, atrous convo- lution, and fully connected crfs', *IEEE transactions on pattern analysis and machine intelligence* **40**(4), 834–848.

23. Chen, L.-C., Papandreou, G., Schroff, F. and Adam, H. (2017), 'Rethinking atrous convolution for semantic image segmentation', *arXiv preprint arXiv:1706.05587* .

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

25. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. and Ronneberger, O. (2016), 3d u-net: learning dense volumetric segmentation from sparse annotation, *in* 'Medical Im- age Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19', Springer, pp. 424–432.

26. Clinic, M. (2022), 'Leading cancer experts', https://www.mayoclinic.org/ diseases-conditions/cancer.

27. Collins, M., Wilhelm, M., Conyers, R., Herschtal, A., Whelan, J., Bielack, S., Kager, L., Kühne, T., Sydes, M., Gelderblom, H. et al. (2013), 'Benefits and adverse events in younger versus older patients receiving neoadjuvant chemotherapy for osteosarcoma: findings from a meta-analysis', *Journal of Clinical Oncology* **31**(18), 2303–2312.

28. Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A. L., Razavian, N. and Tsirigos, A. (2018), 'Classification and mutation pre- diction from non–small cell lung cancer histopathology images using deep learning', *Nature medicine* **24**(10), 1559–1567.

29. Czarnecka, A. M., Synoradzki, K., Firlej, W., Bartnik, E., Sobczuk, P., Fiedorowicz, M., Grieb, P. and Rutkowski, P. (2020), 'Molecular biology of osteosarcoma', *Cancers***12**(8), 2130.

30. Dao, D., Fraser, A. N., Hung, J., Ljosa, V., Singh, S. and Carpenter, A. E. (2016), 'Cellprofiler analyst: interactive data exploration, analysis and classification of large biological image sets', *Bioinformatics* **32**(20), 3210–3212.

31. Davis, A. M., Bell, R. S. and Goodwin, P. J. (1994), 'Prognostic factors in osteosar-coma: a critical review.', *Journal of Clinical Oncology* **12**(2), 423–431.

32. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009), Imagenet: A large-scale hierarchical image database, *in* '2009 IEEE conference on computer visionand pattern recognition', Ieee, pp. 248–255.

33. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. et al. (2020), 'An image is worth 16x16 words: Transformers for image recognition at scale', *arXiv preprint arXiv:2010.11929* .

34. D'Arcy, M. S. (2019), 'Cell death: a review of the major forms of apoptosis, necrosis and autophagy', *Cell biology international* **43**(6), 582–592.

35. Elies, A., Rivière, S., Pouget, N., Becette, V., Dubot, C., Donnadieu, A., Rouzier, R. and Bonneau, C. (2018), 'The role of neoadjuvant chemotherapy in ovarian cancer', *Expert Review of Anticancer Therapy* **18**(6), 555–566.

36. Elshamli, A., Taylor, G. W., Berg, A. and Areibi, S. (2017), 'Domain adaptation using representation learning for the classification of remote sensing images', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10**(9), 4198–4209.

37. Escudier, B., Lougheed, J. C. and Albiges, L. (2016), 'Cabozantinib for the treatment of renal cell carcinoma', *Expert Opinion on Pharmacotherapy* **17**(18), 2499–2504.

38. Face, H. (2023), 'The ai community building the future.', https://huggingface.co/.

39. Farooq, M. and Hafeez, A. (2020), 'Covid-resnet: A deep learning framework for screening of covid19 from radiographs', *arXiv preprint arXiv:2003.14395* .

40. Garg, L., Shukla, P., Singh, S. K., Bajpai, V. and Yadav, U. (2019), Land use land cover classification from satellite imagery using munet: A modified unet architecture.,*in* 'VISIGRAPP (4: VISAPP)', pp. 359–365.

41. Gatti, M. (2022), 'How to split an image into patches with python', https://levelup.gitconnected.com/ how-to-split-an-image-into-patches-with-python-e1cf42cf4f77.

42. Golstein, P. and Kroemer, G. (2007), 'Cell death by necrosis: towards a molecular definition', *Trends in biochemical sciences* **32**(1), 37–43.

43. Gorlick, R. and Khanna, C. (2010), 'Osteosarcoma', *Journal of bone and mineral re- search* **25**(4), 683–691.

44. Gsaxner, C., Roth, P. M., Wallner, J. and Egger, J. (2019), 'Exploit fully automatic low-level segmented pet data for training high-level deep learning algorithms for the corresponding ct data', *PloS one* **14**(3), e0212550.

45. Gutman, D. A., Cobb, J., Somanna, D., Park, Y., Wang, F., Kurc, T., Saltz, J. H., Brat, D. J., Cooper, L. A. and Kong, J. (2013), 'Cancer digital slide archive: an informatics resource to support integrated in silico analysis of tcga pathology data', *Journal of theAmerican Medical Informatics Association* **20**(6), 1091–1098.

46. Han, C., Lin, J., Mai, J., Wang, Y., Zhang, Q., Zhao, B., Chen, X., Pan, X., Shi, Z., Xu, Z. et al. (2022), 'Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels', *Medical Image Analysis* **80**, 102487.

47. Hasty (2023), 'Intersection over union (iou)', https://hasty.ai/docs/mp-wiki/metrics/iou-intersection-over-union.

48. He, K., Zhang, X., Ren, S. and Sun, J. (2016*a*), Deep residual learning for image recog-nition, *in* 'Proceedings of the IEEE conference on computer vision and pattern recogni-tion', pp. 770–778.

49. He, K., Zhang, X., Ren, S. and Sun, J. (2016*b*), Identity mappings in deep residual networks, *in* 'European conference on computer vision', Springer, pp. 630–645.

50. Hinton, G. E. and Salakhutdinov, R. R. (2006), 'Reducing the dimensionality of data with neural networks', *science* **313**(5786), 504–507.

51. Höfener, H., Homeyer, A., Weiss, N., Molin, J., Lundström, C. F. and Hahn, H. K. (2018), 'Deep learning nuclei detection: A simple approach can deliver state-of-the-artresults', *Computerized Medical Imaging and Graphics* **70**, 43–52.

52. Hou, L., Gupta, R., Van Arnam, J. S., Zhang, Y., Sivalenka, K., Samaras, D., Kurc, T. M. and Saltz, J. H. (2020), 'Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of ten cancer types', *Scientific data* **7**(1), 185.

53. Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E. and Saltz, J. H. (2016), Patch-based convolutional neural network for whole slide tissue image classification, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 2424–2433.

54. Huang, X., Deng, Z., Li, D. and Yuan, X. (2021), 'Missformer: An effective medical image segmentation transformer', *arXiv preprint arXiv:2109.07162* .

55. Hung, J., Goodman, A., Ravel, D., Lopes, S. C., Rangel, G. W., Nery, O. A., Malleret, B., Nosten, F., Lacerda, M. V., Ferreira, M. U. et al. (2020), 'Keras r-

cnn: library for cell detection in biological images using deep neural networks', *BMC bioinformatics* **21**(1), 1–7.

56. Janowczyk, A. and Madabhushi, A. (2016), 'Deep learning for digital pathology im- age analysis: A comprehensive tutorial with selected use cases', *Journal of pathology informatics* **7**(1), 29.

57. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A. and Bengio, Y. (2017), The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition work-shops', pp. 11–19.

58. Kan, M., Shan, S. and Chen, X. (2015), Bi-shifting auto-encoder for unsupervised do- main adaptation, *in* 'Proceedings of the IEEE international conference on computer vision', pp. 3846–3854.

59. Kann, B. H. (2019), 'Artificial intelligence in oncology: Current applications and future directions', https://www.cancernetwork.com/view/artificial-intelligence-oncology-current-applications-and-future-di

60. Karimpouli, S. and Tahmasebi, P. (2019), 'Segmentation of digital rock images using deep convolutional autoencoder networks', *Computers and geosciences* **126**, 142–150.

61. Kirillov, A., He, K., Girshick, R., Rother, C. and Dollár, P. (2019), Panoptic segmen- tation, *in* 'Proceedings of the IEEE/CVF conference on computer vision and pattern recognition', pp. 9404–9413.

62. LeCun, Y., Bengio, Y. and Hinton, G. (2015), 'Deep learning', *nature* **521**(7553), 436–444.

63. Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z. and Tu, Z. (2015), Deeply-supervised nets, *in* 'Artificial intelligence and statistics', PMLR, pp. 562–570.

64. Li, H., Xu, Z., Taylor, G., Studer, C. and Goldstein, T. (2018), 'Visualizing the loss landscape of neural nets', *Advances in neural information processing systems* **31**.

65. Li, K., Qian, Z., Han, Y., Eric, I., Chang, C., Wei, B., Lai, M., Liao, J., Fan, Y. and Xu, Y. (2023), 'Weakly supervised histopathology image segmentation with self-attention', *Medical Image Analysis* **86**, 102791.

66. Long, J., Shelhamer, E. and Darrell, T. (2015), Fully convolutional networks for se- mantic segmentation, *in* 'Proceedings of the IEEE conference on computer vision

and pattern recognition', pp. 3431–3440.

67. Luchini, C., Pea, A. and Scarpa, A. (2022), 'Artificial intelligence in oncology: currentapplications and future perspectives', *British Journal of Cancer* **126**(1), 4–9.

68. Mauri, D., Pavlidis, N. and Ioannidis, J. P. (2005), 'Neoadjuvant versus adjuvant sys- temic treatment in breast cancer: a meta-analysis', *Journal of the National Cancer Institute* **97**(3), 188–194.

69. Mohamed, K. M. and Ali, M. H. (n.d.), 'Ultrasound images enhancement using unet- deep learning according to resolution and speckle noise'.

70. Motzer, R. J., Bander, N. H. and Nanus, D. M. (1996), 'Renal-cell carcinoma', *New England Journal of Medicine* **335**(12), 865–875.

71. Naylor, P., Laé, M., Reyal, F. and Walter, T. (2017), Nuclei segmentation in histopathol- ogy images using deep neural networks, *in* '2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)', IEEE, pp. 933–936.

72. NCI (2022), 'clinical diagnosis', https://www.cancer.gov/publications/dictionaries/cancer-terms/def/clinical-diagnosis.

73. Netto, G. J., Saad, R. D. and Dysert, P. A. (2003), Diagnostic molecular pathology: current techniques and clinical applications, part i, *in* 'Baylor University Medical Center Proceedings', Vol. 16, Taylor and Francis, pp. 379–383.

74. Picci, P., Sangiorgi, L., Rougraff, B., Neff, J., Casadei, R. and Campanacci, M. (1994), 'Relationship of chemotherapy-induced necrosis and surgical margins to local recur- rence in osteosarcoma.', *Journal of Clinical Oncology* **12**(12), 2699–2705.

75. Piccinini, F., Balassa, T., Szkalisity, A., Molnar, C., Paavolainen, L., Kujala, K., Buzas, K., Sarazova, M., Pietiainen, V., Kutay, U. et al. (2017), 'Advanced cell classifier: user- friendly machine-learning-based software for discovering phenotypes in high-content imaging data', *Cell systems* **4**(6), 651–655.

76. PyTorchContributors (2023), 'Pytorch documentation', https://pytorch.org/docs/stable/index.html.

77. Qiu, J., Wilm, F., Öttl, M., Schlereth, M., Liu, C., Heimann, T., Aubreville, M. and Breininger, K. (2023), 'Adaptive region selection for active learning in whole slide im-age semantic segmentation', *arXiv preprint arXiv:2307.07168* .

78. Quan, B., Liu, B., Fu, D., Chen, H. and Liu, X. (2021), Improved deeplabv3 for better road segmentation in remote sensing images, *in* '2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)', IEEE,

pp. 331–334.

79. Richardson, S. M., Wurtz, L. D. and Collier, C. D. (2022), 'Ninety percent or greater tumor necrosis is associated with survival and social determinants of health in patients with osteosarcoma in the national cancer database', *Clinical Orthopaedics and RelatedResearch*® pp. 10–1097.

80. Ronneberger, O., Fischer, P. and Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, *in* 'International Conference on Medical image com- puting and computer-assisted intervention', Springer, pp. 234–241.

81. Roodschild, M., Gotay Sardiñas, J. and Will, A. (2020), 'A new approach for the vanishing gradient problem on sigmoid activation', *Progress in Artificial Intelligence* **9**(4), 351–360.

82. Roy, A. M., Briggler, A., Tippit, D., Dawson, K. and Verma, R. (2020), 'Neoadjuvant cabozantinib in renal-cell carcinoma: a brief review', *Clinical Genitourinary Cancer* **18**(6), e688–e691.

83. Ruffolo, P. R. (1964), 'The pathogenesis of necrosis: I. correlated light and electron microscopic observations of the myocardial necrosis induced by the intravenous injectionof papain', *The American Journal of Pathology* **45**(5), 741.

84. Rumelhart, D. E. (1986), 'Learning internal representations by error propagation, in parallel distributed processing', *Explorations in the Microstructure of Cognition* pp. 318–362.

85. Saleena, T., Ilyas, P., Sajna, V. and Haque, A. (2023), 'Deep learning techniques for quantification of tumour necrosis in post-neoadjuvant chemotherapy osteosarcoma resection specimens for effective treatment planning', *Acta Informatica Pragensia* **12**(1), 87–103.

86. Saleena, T., Kutty, S. V. et al. (2023), Synthetic dataset generation from histopathology images for quantizing necrosis in post neo-adjuvant chemotherapy resection specimen, *in* '2023 2nd International Conference on Computational Systems and Communication(ICCSC)', IEEE, pp. 1–5.

87. Sarwinda, D., Paradisa, R. H., Bustamam, A. and Anggia, P. (2021), 'Deep learning in image classification using residual network (resnet) variants for detection of colorectalcancer', *Procedia Computer Science* **179**, 423–431.

88. Sedai, S., Mahapatra, D., Hewavitharanage, S., Maetschke, S. and Garnavi, R. (2017), Semi-supervised segmentation of optic cup in retinal fundus images using variational autoencoder, *in* 'International Conference on Medical Image Computing and Computer- Assisted Intervention', Springer, pp. 75–82.

89. ShuchenDu (2020), 'Understanding dice loss for crisp boundary detection', https://medium.com/ai-salon/understanding-dice-loss-for-crisp-boundary-detection-bb30c2e5f62b.

90. Smeltzer, M. P., Wynes, M. W., Lantuejoul, S., Soo, R., Ramalingam, S. S., Varella- Garcia, M., Taylor, M. M., Richeimer, K., Wood, K., Howell, K. E. et al. (2020), 'The international association for the study of lung cancer global survey on molecular testingin lung cancer', *Journal of Thoracic Oncology* **15**(9), 1434–1448.

91. Sommer, C., Straehle, C., Koethe, U. and Hamprecht, F. A. (2011), Ilastik: Interactive learning and segmentation toolkit, *in* '2011 IEEE international symposium on biomed-ical imaging: From nano to macro', IEEE, pp. 230–233.

92. Strimbu, K. and Tavel, J. A. (2010), 'What are biomarkers?', *Current Opinion in HIV and AIDS* **5**(6), 463.

93. Syntichaki, P. and Tavernarakis, N. (2002), 'Death by necrosis', *EMBO reports* **3**(7), 604–609.

94. Tannir, N. M., Schwab, G. and Grünwald, V. (2017), 'Cabozantinib: an active novel multikinase inhibitor in renal cell carcinoma', *Current oncology reports* **19**(2), 1–8.

95. Tatyana S. Gurina, L. S. (2022), 'Histology, staining', https://www.ncbi.nlm.nih.gov/books/NBK557663/.

96. TensorFlow (2022), 'Transfer learning and fine-tuning', https://www.tensorflow.org/ tutorials/images/transfer_learning.

97. Teramoto, A., Tsukamoto, T., Kiriyama, Y. and Fujita, H. (2017), 'Automated classi- fication of lung cancer types from cytological images using deep convolutional neuralnetworks', *BioMed research international* **2017**.

98. Trimble, E., Ungerleider, R., Abrams, J., Kaplan, R., Feigal, E., Smith, M., Carter, C. and Friedman, M. (1993), 'Neoadjuvant therapy in cancer treatment', *Cancer* **72**(S11), 3515–3524.

99. Tsang, S.-H. (2018), 'Review: Resnet — winner of ilsvrc 2015 (image classi-fication, localization, detection)', https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localizati

100. Turchenko, V. and Luczak, A. (2017), Creation of a deep convolutional auto-encoder in caffe, *in* '2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications

(IDAACS)', Vol. 2, Ieee, pp. 651–659.

101. Van Beers, F. (2018), Using intersection over union loss to improve binary image seg- mentation, PhD thesis.

102. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł.and Polosukhin, I. (2017), 'Attention is all you need', *Advances in neural informationprocessing systems* **30**.

103. Veta, M., Van Diest, P. J., Kornegoor, R., Huisman, A., Viergever, M. A. and Pluim, J. P. (2013), 'Automatic nuclei segmentation in hande stained breast cancer histopathologyimages', *PloS one* **8**(7), e70221.

104. Walia, M. (2022), 'Semantic segmentation vs. instance segmentation: Explained', https://blog.roboflow.com/ difference-semantic-segmentation-instance-segmentation/.

105. Wang, P., Hu, X., Li, Y., Liu, Q. and Zhu, X. (2016), 'Automatic cell nuclei segmen- tation and classification of breast cancer histopathology images', *Signal Processing* **122**, 1–13.

106. Wang, S., Yang, D. M., Rong, R., Zhan, X. and Xiao, G. (2019), 'Pathology image anal- ysis using segmentation deep learning algorithms', *The American journal of pathology***189**(9), 1686–1698.

107. Wang, Y., Wang, G., Wei, L.-H., Huang, L.-H., Wang, J.-L., Wang, S.-J., Li, X. -P., Shen,D.-H., Bao, D.-M. and Gao, J. (2011), 'Neoadjuvant chemotherapy for locally advanced cervical cancer reduces surgical risks and lymph-vascular space involvement', *Chinesejournal of cancer* **30**(9), 645.

108. Wen, L., Li, X. and Gao, L. (2020), 'A transfer convolutional neural network for fault diagnosis based on resnet-50', *Neural Computing and Applications* **32**, 6111–6124.

109. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M. and Luo, P. (2021), 'Seg- former: Simple and efficient design for semantic segmentation with transformers', *Ad- vances in Neural Information Processing Systems* **34**, 12077–12090.

110. Yan, Z., Zhang, H., Jia, Y., Breuel, T. and Yu, Y. (2016), 'Combining the best of con- volutional layers and recurrent layers: A hybrid network for semantic segmentation', *arXiv preprint arXiv:1603.04871* .

111. Yurtkulu, S. C., Şahin, Y. H. and Unal, G. (2019), Semantic segmentation with ex- tended deeplabv3 architecture, *in* '2019 27th Signal Processing and

Communications Applications Conference (SIU)', IEEE, pp. 1–4.

112. Yusiong, J. P. T. and Naval, P. C. (2019), Multi-scale autoencoders in autoencoder for semantic image segmentation, *in* 'Asian Conference on Intelligent Information andDatabase Systems', Springer, pp. 587–599.

113. Zeb, M. H., Al-Obeidat, F., Tubaishat, A., Qayum, F., Fazeel, A. and Amin, M. (2023), 'Denoising histopathology images for the detection of breast cancer', *Neural Comput-ing and Applications* pp. 1–15.

114. Zhang (2018), 'A better auto encoder for image: Convolutional autoencoder', http://users.cecs.anu.edu.au/Tom.Gedeon/conf/ABCs2018/paper/ ABCs2018_ paper_58.pdf

115. Zhang, A. (2022*a*), '14.9. semantic segmentation and the dataset', https://d2l.ai/chapter_computer-vision/semantic-segmentation-and-dataset.html.

116. Zhang, Q. (2022*b*), 'A novel resnet101 model based on dense dilated convolution for image classification', *SN Applied Sciences* **4**, 1–13.

117. Zhang, R., Du, L., Xiao, Q. and Liu, J. (2020), Comparison of backbones for seman- tic segmentation network, *in* 'Journal of Physics: Conference Series', Vol. 1544, IOP Publishing, p. 012196.

118. Zhang, Z., Li, J., Zhong, Z., Jiao, Z. and Gao, X. (2019), 'A sparse annotation strategybased on attention-guided active learning for 3d medical image segmentation', *arXiv preprint arXiv:1906.07367* .

119. Zheng, Y., Xie, J., Sain, A., Song, Y.-Z. and Ma, Z. (2023), 'Sketch-segformer: Transformer-based segmentation for figurative and creative sketches', *IEEE transac- tions on image processing* .

120. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. and Liang, J. (2018), Unet++: A nested u-net architecture for medical image segmentation, *in* 'Deep learning in medical image analysis and multimodal learning for clinical decision support', Springer, pp. 3–11.

121. Zhu, W., Zhu, L., Bao, Y., Zhong, X., Chen, Y. and Wu, Q. (2019), 'Clinical evaluation of neoadjuvant chemotherapy for osteosarcoma', *J BUON* **24**(3), 1181–1185.

Appendix A

Sample Images of Dataset

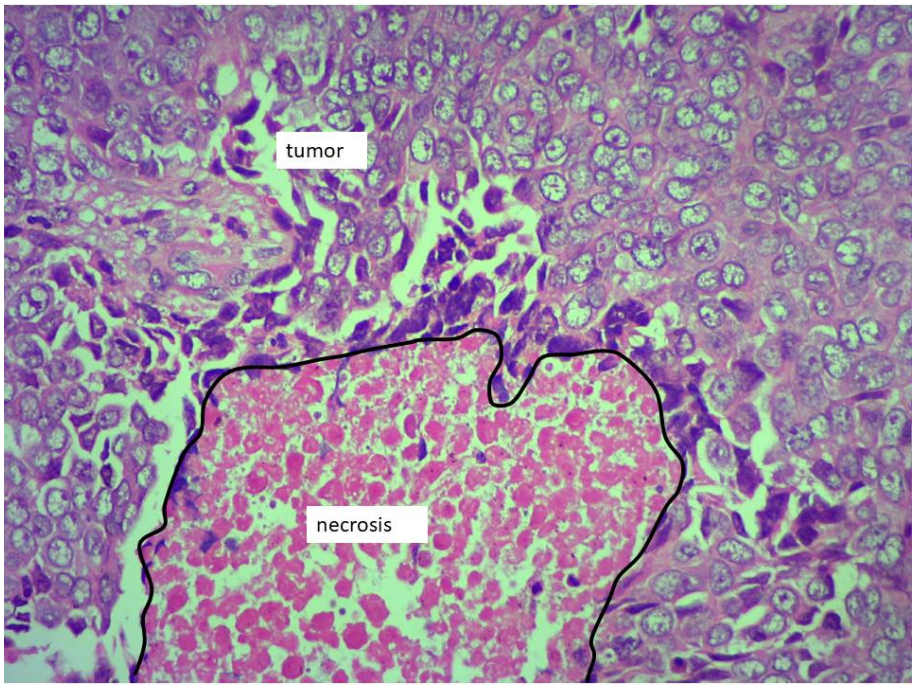


Fig. A.1 The Tumor tissues and Necrosis tissues are separately shown in this figure

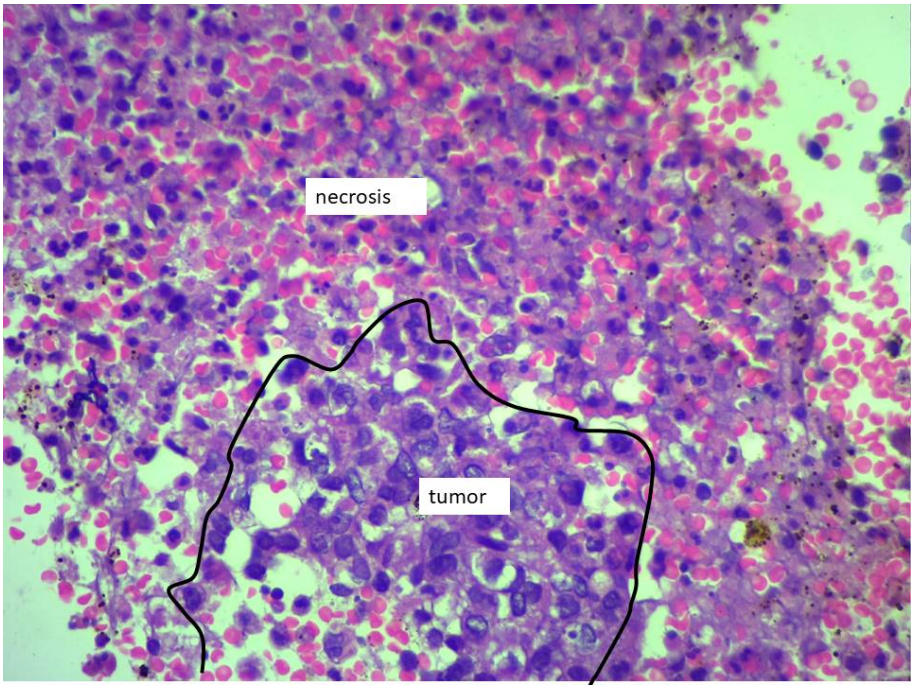


Fig. A.2 The Tumor tissues and Necrosis tissues are separately shown in this figure

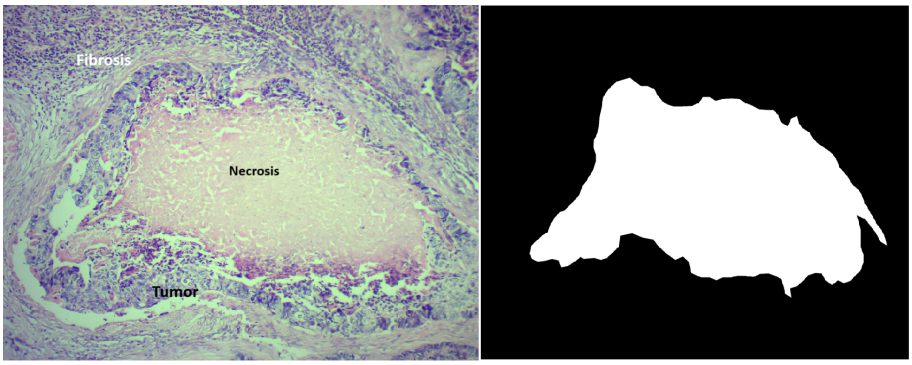


Fig. A.3 The figure include Tumor tissues, Fibrosis tissues and Necrosis

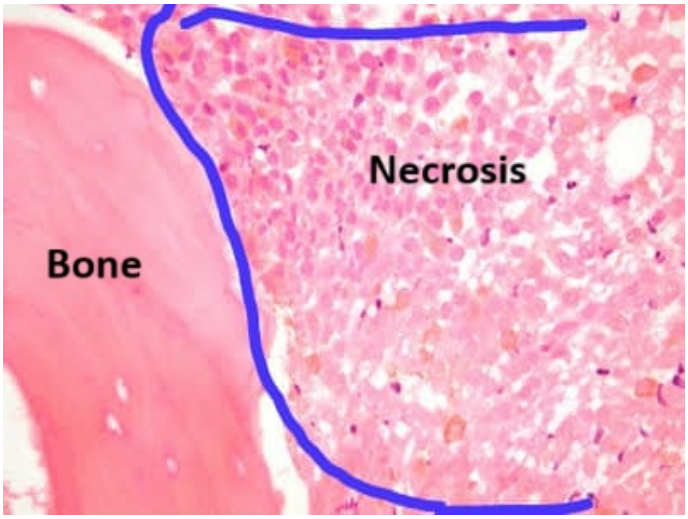


Fig. A.4 The Bone tissues and Necrosis tissues are separately shown in this figure

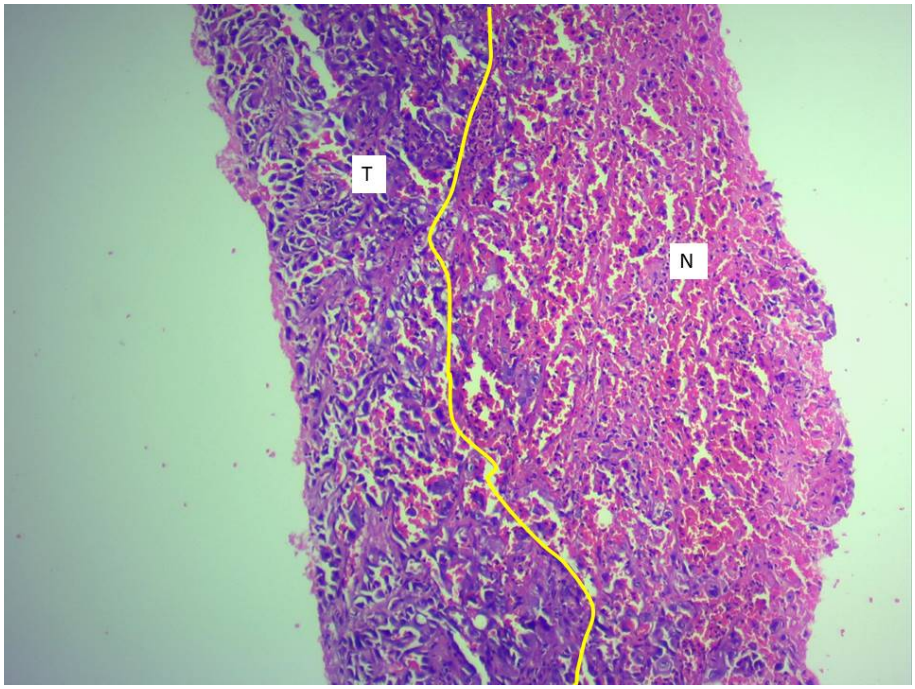


Fig. A.5 T denotes Tumor and N denotes Necrosis

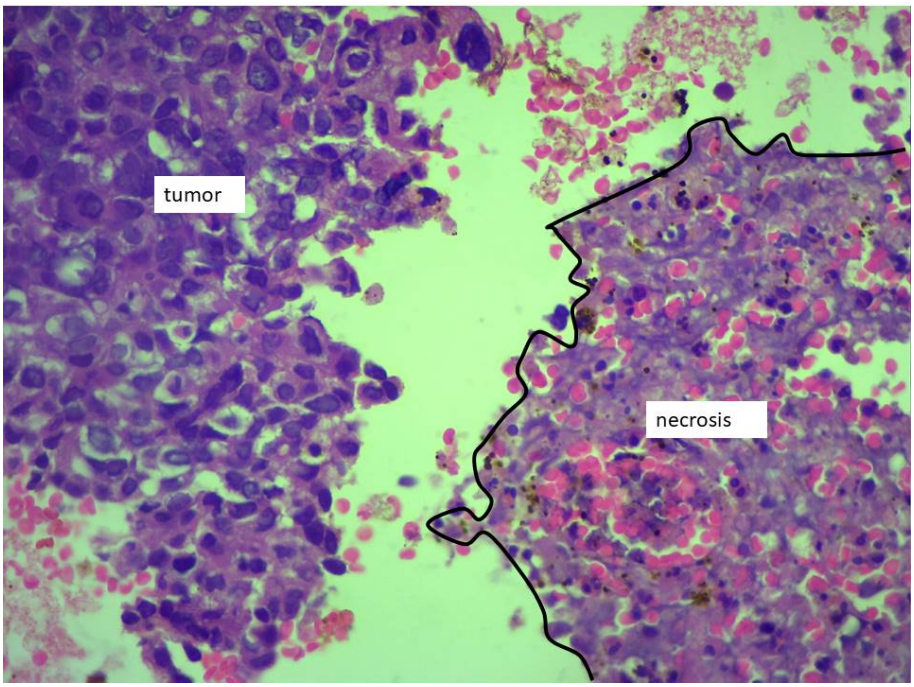


Fig. A.6 The Tumor tissues and Necrosis tissues are separately shown in this figure

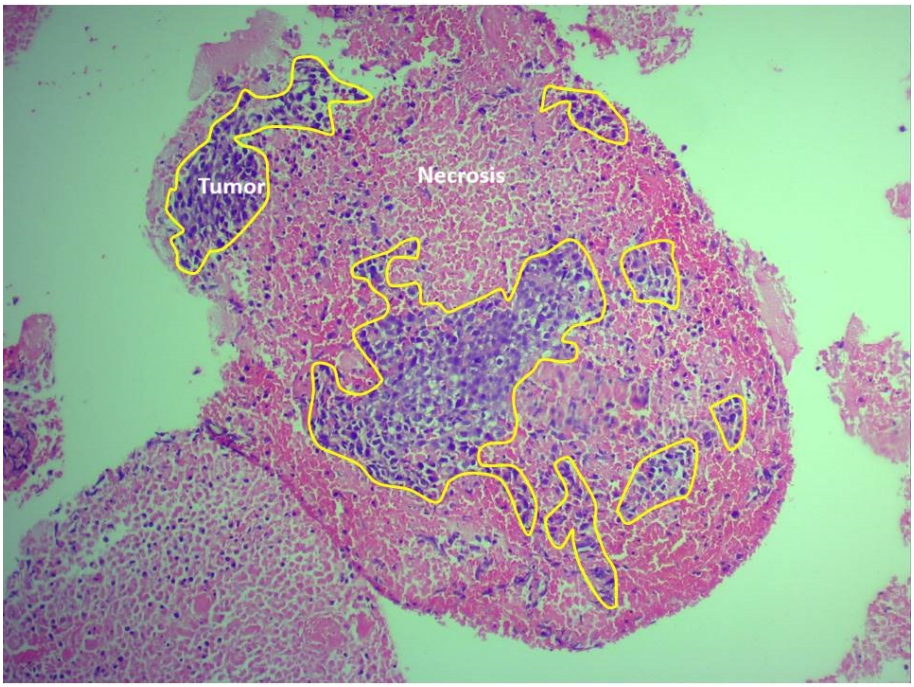


Fig. A.7 The Tumor tissues and Necrosis tissues are separately shown in this figure

LIST OF PUBLICATIONS

Journals

1. Saleena, T. S., Ilyas, P. M., Sajna, V. M. K., and Haque, A. K. M. B. (2023). Deep Learning Techniques for Quantification of Tumour Necrosis in Post-neoadjuvant Chemotherapy Osteosarcoma Resection Specimens for Effective Treatment Planning. Acta Informatica Pragensia, 12(1), https://doi.org/10.18267/j.aip.207 (Scopus Indexed)

2. T. S. Saleena, M. I. P and S. V. M. Kutty, "Synthetic Dataset Generation From Histopathology Images For Quantizing Necrosis In Post Neo-Adjuvant Chemotherapy Resection Specimen," 2023 2nd International Conference on Computational Systems and Communication (ICCSC), Thiruvananthapuram, India, 2023, pp. 1-5, doi: 10.1109/ICCSC56913.2023.10143027(Scopus Indexed).

3. Saleena, T. S., Ilyas, P. M., and Kutty, S. V. (2024, October). SegFormer-Based Approach for Semantic Segmentation of Necrotic Tissues from Histopathology Images. In 2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER) (pp. 383-387)(Scopus Indexed).

4. Saleena, T. S., Muhamed Ilyas P. (February, 2021) (18)F-FDG PET/CT- An Effective Diagnostic Tool for Non-Small Cell Lung Cancer: A Review Journal of the Maharaja Sayajirao University of Baroda ISSN : 0025-0422 – UGC CARE GROUP 1 journal(print only) ISSN: 0025-0422

5. Saleena, T. S., Ilyas, P. M., and Kareem, K. S. (2023). Gabor filter as feature extractor in anomaly detection from radiology images. In Intelligent Systems and Applications in Computer Vision (pp. 179-187). Taylor and Francis Group, CRC Press.

6. Saleena Thorayanpilackal Sulaiman, Muhamed Ilyas Poovankavil, Abdul Jabbar Perumbalath (2023, August). Classification Tool to Predict the Presence of Colon Cancer Using Histopathology Images - Pp. 33-46 (14), Deep Learning: Theory, Architectures, and Applications in Speech, Image, and Language Processing, DOI: 10.2174/9789815079210123010001

Conference Presentation

1. ”SegFormer-Based Approach for Semantic Segmentation of Necrotic Tissues from Histopathology Images” in 2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)

2. “Synthetic Dataset Generation from Histopathology Images for Quantizing Necrosis in Post Neo-Adjuvant Chemotherapy Resection Specimen” in IEEE conference at LBS Institute of Technology for Women, Poojappura, Thiruvananthapuram

3. “Semantic Segmentation of Necrotic Tissues in Post Neo-adjuvant Chemotherapy Resection Specimens using Auto-Encoder Network” in International Conference on FOSS Approaches towards Computational Intelligence and Language Technology (FOSS-CIL T24) organized by ICFOSS, Thiruvananthapuram

4. “AI-Powered Semantic Segmentation and Fluid Volume Calculation of Lung CT images in Covid-19 Patients” International Conference on “Emergent Converging Technologies and Biomedical Systems, (ETBS-2021)” 28th -29th April 2021 Organized by University Institute of Engineering and Technology Kurukshetra University, Kurukshetra