Predictive modelling of post radiation therapy recurrence for gynecological cancer patients

Yujing Zou

Medical Physics Unit
McGill University
Montréal, Québec, Canada

December, 2022

A thesis presented for the degree of master of science in medical radiation physics.

Abstract

Outcome modeling can characterize the behavior of tissue response to a treatment based on patient-specific multi-omics data (e.g. clinical information, demographics, dosimetric profiles, radiomics, genomics, proteomics, and pathology) and predict probability of outcomes of the patient under changing therapy conditions.

Ionizing radiation kills cells or inhibits the cell cycle mainly by damaging DNA in the cell nucleus. Patient-specific radiation dose response may be influenced by inter-patient variation in cell/nucleus size, since target size impacts microscopic descriptors of energy deposition. The aim of this master's thesis was to investigate the correlation between microscopic influence of cell spacing and nuclei size, and gynecological cancer treatment outcomes (i.e. post radiation therapy recurrence) directly.

Therefore, **1)** a fully automated and generalizable method to extract patient-specific nuclei size and cell spacing distributions from cancerous and non-tumoral regions of diagnostic pre-treatment hematoxylin and eosin (H & E) stained digital histopathology whole slide image for gynecological cancer patients was developed; **2)** a model to predict the binary clinical endpoint of post radiation therapy recurrence of gynecological cancer patients using both aforementioned pre-treatment histopathology image features and clinical variables (ex. age at diagnosis, clinical & radiological stage at the time of radiation, p16 status, cancer type) was built upon data from 36 patients with a Tree-Based Pipeline Optimization Tool automated machine learning approach. An Area Under the Receiver Operating Characteristic Curve of Receiver Operating Characteristics score of 0.92 was achieved for our model and an empirical p-value of 0.036 was obtained from the permutation test ($n = 1000$). This implied first, real dependencies between our combined imaging and clinical features and outcomes were learned by a decision-tree-based ML classifier using an eXtreme gradient-boosting algorithm; and second, that the promising model performance was not by chance.

Résumé

La modélisation des résultats peut caractériser le comportement d'une réponse tissulaire à un traitement qui est basé sur des données multi-omiques spécifiques au patient (par ex. informations cliniques, démographie, profils dosimétriques, radiomique, génomique, protéomique et pathologie). Elle peut aussi prédire la probabilité des résultats du patient suite à une modification des conditions thérapeutiques.

Les rayonnements ionisants tuent les cellules ou inhibent le cycle cellulaire principalement en endommageant l'ADN dans le noyau cellulaire. La réponse à la dose de rayonnement spécifique au patient peut être influencée par la variabilité interpatients de la taille des cellules ou des noyaux, puisque la taille de la cible a un impact sur les descripteurs microscopiques du dépôt d'énergie. L'objectif de cette thèse de maîtrise est alors de directement examiner la corrélation entre l'influence microscopique de l'espacement entre les cellules et la taille des noyaux; et les résultats du traitement du cancer gynécologique (i.e. récurrence après la radiothérapie).

Par conséquent, nous avons développé **1)** une méthode entièrement automatisée et généralisable pour extraire : la taille des noyaux et les distributions de l'espacement entre les cellules, propres à chaque patiente. L'extraction des ces données est faite à partir d'histopathologies numériques sur lames entières scannées (Whole-Slide Images - whole slide image (WSI)), colorées à l'hématoxyline et à l'éosine (H & E), de régions cancéreuses et non tumorales, obtenues de patientes atteintes d'un cancer gynécologique avant le début du traitement; **2)** un modèle pour prédire le critère d'évaluation clinique binaire de la récidive après la radiothérapie (radiation therapy (RT)) des patientes atteintes d'un cancer gynécologique, en utilisant à la fois les caractéristiques des images histopathologiques susmentionnées et les variables cliniques (par ex. l'âge au moment du diagnostic, le stade clinique & radiologique au moment de la radiothérapie, statut p16, type de cancer). Ce modèle a été construit sur 36 patients suivant une approche basée sur le "Tree-Based Pipeline Optimization Tool (TPOT)" "Automated Machine Learning (autoML)". Un score Area Under the Receiver Operating Characteristic Curve (AUC) de Receiver Operating Characteristics (ROC) de 0,92 a été atteint pour notre modèle et une valeur p de 0,036 a été obtenue à partir du test de permutation (n = 1000). Cela implique, premièrement, de réelles dépendances entre nos caractéristiques d'images et nos caractéristiques cliniques combinées, et nos résultats appris par le classificateur. Et deuxièmement, que les performances prometteuses du modèle n'étaient pas par hasard.

Acknowledgements

First and foremost, I would like to extend my sincerest gratitude towards my supervisor Dr. Shirin Abbasinejad Enger, who has been an instrumental role model, mentor, supporter, and friend (and agent for my music). I have been extremely lucky to be working with her during my M.Sc. as I can always count on her to help flourish and guide our ideas while pushing me with just the right amount. I very much look forward to further exploring science and life during my Ph.D. studies with her.

During the course of this thesis, numerous collaborators have offered much appreciated generous guidance lending their domain expertise. I want to thank especially Dr. Magali Lecavalier-Barsoum for sharing knowledge as a gynecology radiation oncologist, Dr. Manuela Pelmus as a pathologist, Dr. Farhad Maleki in medical artificial intelligence, and Dr. Hamed Bekerat for being the first to show me a brachytherapy workflow in the treatment planning room at the Jewish General Hospital. I also want to thank Joseph DeCuhna for our fruitful discussions whose previous work set a basis for this thesis; and to Luca Weishaupt for co-founding McMedHacks with me, who I always had enormous fun spending countless hours working with, who motivates and inspires me as a friend and colleague, and who I can always rely on to turn our ideas into reality together with a simple mutual "let's do it!".

The McGill Medical Physics Unit (MPU) felt like a place where I grew up. Everyone I met here had something valuable to teach me and help me mature as a scientist. Especially, I want to genuinely thank Dr. Jan Seuntjens, Dr. Avishek Chatterjee, and Dr. André Diamant for leading me into the fascinating field of medical physics five years ago, Margery Knewstubb and Estrella Moyal for providing admin help anytime we needed, and Tatjana Nisic for research project management and always greeting me with a smile on any day.

To my dearest friends from the EngerLab: Hossein, Jingyi, Victor, Youstina, Liam, Joanna, Azin, Hailey, Jonathan, Jake, Sebastien, Fabian, Alana, Joshua, Joud, Maude, Melodie, Naim, and Behnaz, thank you so much for the most memorable good times at McCorlds, karaoke nights, fukTona coffee, too many to count, and getting me through the rougher times. The joy from making music with Dr. Jerry Battista (The Stochastics & Friends) and Jacob Sanz-Robinson can never be understated, thank you for exploring the interconnections of science and music with me.

Last but not the least, I would really like to thank my parents who packed their bags and immigrated to Canada with me almost 10 years ago for all their unconditional love, and endless encouragement. And to my dearest grandparents who I love so very deeply, who raised me tirelessly, showed me the meaning of love and carefree, taught me the essence of perseverance and drive, curiosity and constant learning. From the bottom of my heart, thank you for everything.

I would also like to acknowledge the funding sources of this project from the Canada Research Chair program and the Collaborative Health Research Project (CHRP grant #433418).

Contents

List of Figures

List of Tables

List of Acronyms

AAPM	American Association of Physicists in Medicine.
Adeno	Adenocarcinoma.
AI	Artificial Intelligence.
AUC	Area Under the Receiver Operating Characteristic Curve.
autoML	Automated Machine Learning.
CAMPEP	Commission on Accreditation of Medical Physics Education Programs.
CCR	continuously connected regions.
CI	confidence interval.
CMOS	complementary metal–oxide–semiconductor.
CNN	Convolutional Neural Network.
COBRA	Consortium for Brachytherapy Data Analysis.
CPU	Central Processing Unit.
cs	cell spacing.
CT	computed tomography.
CV	cross validation.
DL	Deep Learning.
EBRT	External Beam Radiation Therapy.
EHR	Electronic Health Record.
ESTRO	European Society for Radiotherapy & Oncology.

FCBF	Fast Correlation Based Filter.
FU	Follow Up.
GPU	Graphical Processing Unit.
GYN	gynecological.
H and E	Hematoxylin and Eosin.
HPV	human papillomavirus.
IHC	immunohistochemistry.
JGH	Jewish General Hospital.
KDE	kernel density estimation.
ML	machine learning.
MNIST	modified national institute of standards and technology.
MR	magnetic resonance.
MRI	magnetic resonance imaging.
MSE	mean squared error.
MUHC	McGill University Health Center.
NLP	Natural Language Processing.
NN	neural network.
ns	nuclei size.
NTCP	Tissue Complication Probability.
OHE	One Hot Encoding.
PC	Principal Component.
PCA	Principal Component Analysis.
PET	positron emission tomography.
RadOnc	Radiation Oncology.
RBF	Radial Basis Function Kernel.
ReLU	rectified linear unit.

ROC	Receiver Operating Characteristics.
ROI	Region of Interest.
RT	radiation therapy.
SCC	Squamous Cell Carcinoma.
SOTA	state of the art.
std	standard deviation.
TMA	Tissue Microarray.
TPOT	Tree-Based Pipeline Optimization Tool.
TPS	Treatment Planning Systems.
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.
US	ultrasound.
ViT	Vision Transformer.
WSI	whole slide image.
XGBoost	eXtreme gradient-boosting.

Chapter 1

Introduction

1.1 Outcome prediction modelling: from the micro to the macro

1.1.1 Motivation

Over 50% of all cancer patients receive radiotherapy during their cancer treatment [10]. Despite novel technology developments and improvement of treatment protocols in radiotherapy including brachytherapy and combination of treatments with surgery or chemotherapy, it remains clinically challenging to predict which patients will benefit from which treatment combination. This challenge can be approached by taking advantage of the enormous amounts of biological and digital information generated from each patient in the modern radiotherapy healthcare system [11–25]. The fundamental attraction in multimodal data integration for outcome prediction modelling lies in the augmentation of information drawn from multiple image modalities compared to any individual modality alone. This especially reveals promises when information contributed by each modality are orthogonal (or uncorrelated) and complementary with one another at multiple scales, such as radiological scans versus histopathology specimens to describe various tumor spatial characteristics [26, 27]. Therefore, this motivates the overarching goal of developing a patient-specific treatment outcome prediction model that can accurately predict optimal treatment combinations for gynecological (GYN) cancer patients using computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US) as well as cell spacing & nuclei size distribution data obtained from digital histopathology images (Fig. 1.1).

Ionizing radiation kills cells or inhibits the cell cycle mainly by damaging DNA in the cell nucleus. The dose received per nucleus correlates to dose response [28–30]. In microdosimetry, the specific energy z varies with a frequency distribution determined by the fluctuations in both the number of tracks passing the target (i.e. a cell or cell nucleus-sized volume), and the energy deposition per track [31]. The size of cells and nuclei depends on tissue type, cell cycle, and malignancy status, each of which differ between patients. The variance in specific energy imparted per cell (i.e. dose per cell) to its nuclei increases with decreasing cell size [28–30]. Thus, this varying target size distribution (i.e. cell spacing / nuclei size distributions) may result in differing specific energy deposited among tissues. Patient-specific radiation dose response may be influenced by inter-patient variation in cell/nucleus size, since target size impacts microscopic descriptors of energy deposition [32–34].

It is known that the magnitude of the microdosimetric spread may affect how the microdosimetric properties of radiation qualities influence the outcome analysis of biological assays, and it can be compared to uncertainties in macroscopic dose delivery in radiotherapy [31]. However, current research lacks an understanding of the bridge between microscopic level biological dose response to macroscopic level post radiation treatment clinical outcome due to the variation of patient-specific cell / nuclei target size among cancer patient population. This inspired our aim to evaluate the predictability of such pre-treatment microscopic cell / nuclei target size data for post radiation outcomes on retrospective data taken from patients diagnosed with gynecological cancers. Furthermore, DeCunha et.al. [32] developed an algorithm to extract cell spacing and nuclei size distributions from Hematoxylin and Eosin(H & E) histopathology Tissue Microarray (TMA) images, which formed a basis of this work.

Traditional machine learning (ML) approaches are wildly sought after in predictive modelling for clinical applications. Different machine learning algorithms can be developed once the Machine Learning (ML) problem has been defined with a clinical outcome of interest upon an understanding of the clinical context [35]. Once model performances on the medical dataset are compared using ML evaluation metrics, a best model is chosen for a specific problem and reported [15, 36–41]. However, the process of manual data pre-processing, feature engineering and selection, and hyperparameter fine tuning to achieve optimal model performances are labour-extensive as many algorithmic choices are available for each step. Furthermore, objectively selecting the best model is a delicate process as many ML algorithms could be potentially applicable, and the well-known "no free lunch" principle [42] suggests that there exists no "best" algorithm for all problems that exist. In contrast, Automated Machine Learning (autoML) uses ML itself to automatically construct candidates of data pre-processing pipelines, hyperparameter optimization, and model building, evaluation, and selection, for a specified problem. This returns the best pipeline, for a specified evaluation metric, considering a wide array of different combinations of ML pipeline possibilities [43–51]. This, at the very least, can provide a much more productive starting point for predictive modelling which can be adjusted further with specific domain knowledge. Additionally, it leads to a data-driven model building approach agnostic to dataset preprocessing and statistical model without making an *a priori* assumption about statistical properties of the data.

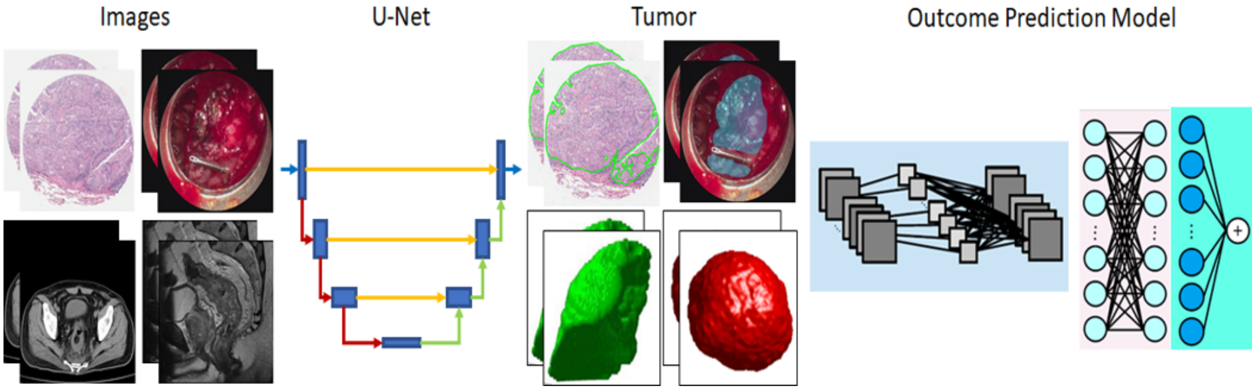


Figure 1.1: Multimodal treatment outcome prediction model we aim to build in the future. From both diagnostic images such at CT, magnetic resonance (MR), US, and digital histopathological images, whose Region of Interest (ROI) is delineated by an autosegmentation algorithm such as a U-Net [1, 2], extracted features can be fused and fed into a outcome prediction model.

1.1.2 Aim

In this work, we investigated how pre-treatment patient characteristics of various data types (i.e. clinical data and imaging features) influence treatment efficacy as measured by post-treatment response for gynecological patients. More specifically, with both statistical analyses and a Tree-Based Pipeline Optimization Tool (TPOT) autoML approach, we examined the microscopic influence of cell spacing (cs) and nuclei size (ns) data extracted from Hematoxylin and Eosin(H & E) stained digital histopathological whole slide image (WSI), as well as other pre-treatment clinical variables, on gynecological cancer treatment outcomes (i.e. post radiation therapy (RT) recurrence at the time of the Follow Up (FU)) in radiotherapy (external radiotherapy and brachytherapy). This work may be used in the future to bridge a gap between microscopic events during radiation, such as lineal energy deposition in patient-specific cell / nuclei targets with varying size, to the macroscopic patient clinical outcomes.

.

1.2 Preface and Contributions

The following are a list of oral / poster presentations based on the work of and within the duration of this master's thesis.

1. Yujing.Zou†*, M. Lecavalier-Barsoum, S.A. Enger, Treatment outcome Prediction for gynecological cancers patients with a machine learning model using pre/post diagnostic image modalities and digital histopathology images, CRUK RadNet Manchester AI for Optimizing Radiotherapy Outcomes Workshop, VIRTUAL. February 10, 2021. 1 min International Oral Presentation.

2. Yujing.Zou†*, M. Lecavalier-Barsoum, M. Pelmus, S.A. Enger, Deep Learning-Based Patient-Specific Multimodality Treatment Outcome Prediction for Gynecological Cancers using pre/post diagnostic image modalities and digital histopathology images, Friday Morning Talk, McGill Medical Physics Unit, October 1, 2021, 17 min Institutional Oral Presentation.

3. Yujing.Zou†*, Dr. Magali Lecavalier-Barsoum, Dr. Manuela Pelmus, Dr. Shirin A. Enger, Investigation of cell spacing, and nuclei size distribution extracted from H &E

histopathological Whole-Slide-Images for integration into a multimodality treatment outcome prediction model for gynecological cancer patients, AQPMC Annual Scientific Meeting, Dec 3rd, 2021, 15 min Provincial Oral Presentation.

4. Yujing.Zou†*, Magali Lecavalier-Barsoum, Manuela Pelmus, Farhad Maleki, Shirin A. Enger, Deep Learning-based Patient-Specific Multimodality Treatment Outcome Prediction for Gynecological Cancers, Lady Davis Institute Noon Seminar, Feb 11th, 2022, 20 min Oral Presentation.

5. Yujing.Zou†*, Luca L. Weishaupt, Shirin A. Enger, "McMedHacks: Deep learning for medical image analysis workshops and Hackathon in radiation oncology", proffered paper, European Society for Radiotherapy and Oncology (ESTRO) conference, May 7th, 2022, International 10 min Oral Presentation [52].

6. Yujing.Zou†*, Magali Lecavalier-Barsoum, Manuela Pelmus, Farhad Maleki, Shirin A. Enger, Predictive modeling of post radiation-therapy recurrence for gynecological cancer patients using clinical and histopathology imaging features, May 21st, 2022, Montreal, Canada, 10 min Oral Presentation, Friday Morning Talk (M.Sc. 2), Montreal, Canada, May 20th, 2022).

7. Yujing.Zou†*, Magali Lecavalier-Barsoum, Manuela Pelmus, Farhad Maleki, Shirin A. Enger, Predictive modeling of post radiation-therapy recurrence for gynecological cancer patients using clinical and histopathology imaging features, Curietherapies conference, Montreal, Canada, May 21st, 2022, 10 min Oral Presentation [53].

8. Yujing.Zou†*, Magali Lecavalier-Barsoum, Manuela Pelmus, Shirin A. Enger, Patient-specific nuclei size and cell spacing distribution extraction from histopathology whole slide images for treatment outcome prediction modeling, Celebration of Research and Training in Oncology (CORTO) conference, June 21st, 2022, poster presentation.

9. Yujing.Zou†*, Magali Lecavalier-Barsoum, Manuela Pelmus, Shirin A. Enger, Patient-specific nuclei size and cell spacing distribution extraction from histopathology whole slide images for treatment outcome prediction modeling, American Association of Physicists in Medicine (AAPM) conference, Washington, DC, USA, July 10th - 14th, 2022 , poster presentation [54].

1.3 Outline

Chapter 1 is an introduction of the motivation and hypothesis of this work. **Chapter 2** constitutes the background and literature reviews of the following concepts: 1) medical image analysis (particularly in digital histopathology), 2) gynecological cancer, 3) classical versus automated machine learning algorithms, and 4) the applications of these approaches in outcome prediction modelling in radiation oncology. **Chapter 3** includes two sub-sections that stem from three accepted conference abstracts that were presented at American Association of Physicists in Medicine (AAPM) [54], Curietherapies [53], and European Society for Radiotherapy & Oncology (ESTRO) [52] in 2022. **Chapter 3.1** describes a methodology [54] to extract patient-specific nuclei size and cell spacing distributions from pre-treatment gigapixel histopathology WSI for treatment outcome prediction modeling, which was extended from work of DeCunha et.al. [55]. Then we report results on outcome prediction modelling of post radiation therapy (RT) recurrence for thirty-six gynecological cancer patients, using both pre-treatment histopathology imaging features (i.e. patient-specific cell and nuclei size distributions) and clinical variables (ie., age at diagnosis, clinical & radiological stage at the time of radiation, p16 status, cancer type) [53], with a TPOT autoML [49] approach. **Chapter 4** discusses the merits and limitations of the current work and suggests future directions for each sub-chapter of chapter 3 such as implementation of a Deep Learning (DL)-based auto-segmentation algorithm to delineate cancerous regions of the histopathology WSI, and curation of more external patient data from multiple institutions. **Chapter 5** concludes this master's thesis.

1.4 Contributions of Authors

Authors: Yujing Zou[1], Magali Lecavalier-Barsoum[2], Manuela Pelmus[3], Farhad Maleki[4], Shirin Abbasinejad Enger[1,3,5,6]

[1] Medical Physics Unit, Department of Oncology, Faculty of Medicine, McGill University, Montreal, Quebec, Canada

[2]Department of Radiation Oncology, Jewish General Hospital, Montreal, QC, Canada

[3] Department of Pathology, Faculty of Medicine, McGill University, Montreal, Quebec, Canada

[4]Department of Radiology, The Research Institute of the McGill University Health Centre, McGill University, Montreal, QC, Canada

[5,6]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal,

Quebec, Canada

Contributions: Dr. Shirin Abbasinejad Enger proposed, investigated, and supervised this research. Yujing Zou investigated, introduced the methodology and analysed the results for all chapters. Dr. Magali Lecavalier-Barsoum selected the gynecological cancer patients included in the study, collected the clinical variables, and offered invaluable clinical knowledge. Dr. Manuela Pelmus retrieved the histopathology whole slide images, examined, selected the slide, and quality assured the cancerous region delineation. Dr. Farhad Maleki provided machine learning related discussions.

Chapter 2

Theory & Literature Review

2.1 Medical image analysis: digital histopathology WSI

2.1.1 H & E histopathology WSI vs. TMA

Histopathology image slides, which relies on visual interpretation of cellular biology captured in images, serves as the "gold standard" in diagnosis of cancer [4]. Routine cancer diagnostics adopts digital pathology as a mainstream option the WSI work-up, cutting, staining, scanning of histopathology WSI at the local pathology department. This process also preserves the underlying tissue architecture with disease characteristics which can be analyzed digitally at various resolutions [56]. It offers tremendous opportunities in computational pathology (CPath) image analysis research.

During pathology tissue processing for the cancer diagnosis purposes, a fresh cancerous tissue specimen is placed in a liquid fixing agent such as formalin hardening to preserve the tissue. Following tissue specimen dehydration and clearing, it is infiltrated by paraffin wax and formed into a paraffin block. A microtome is used for section cutting of the paraffin block onto glass slides, each with thickness around 4 microns. The slides are then stained with Hematoxylin and Eosin (H & E) stain, the gold standard in cancer diagnosis with histopathology, which is the combination of two histological stains: hematoxylin and eosin. The Hematoxylin stains cell nuclei blue, and Eosin stains cytoplasm and connective tissue pink, combinations of these colors can be taken on by different shades, hues, and combinations of these colors. [4, 57–61].

On the other hand, Tissue Microarray TMAs are paraffin blocks produced by extracting cylindrical tissue cores from different paraffin donor blocks and re-embedding these into a single recipient (microarray) block at defined array coordinates. Using this technique, up to 1000 or more tissue samples can be arrayed into a single paraffin block [62]. It is a very effective approach for high-throughput molecular analysis of pathological tissues supporting the identification of new potential diagnostic and prognostic markers in human cancers. However, at the individual patient level, TMAs remain a research tool and are not used as a diagnostic/prognostic tool. Careful evaluation of the staining characteristics of the selected marker in whole tissue sections will be required to migrate promising predictive marker assays from TMAs into routine clinical use with development of precise quantification or analytical measurement approaches [63], [62]. Since a digitized TMA tissue sample is one of the 1000 tissue samples in a single paraffin block (Fig. 2.1), and a WSI is the entire tissue cut, there is a sizeable difference between the digitized image dimensions for the two modalities as presented in Fig. 2.2. For example, in the work of DeCunha et.al. [32] that developed an algorithm to extract cell spacing and nuclei size distributions from H & E histopathology TMA, the TMA image dimension was 7500 pixel by 7500 pixel. Whereas digital histopathology WSI used in this thesis were gigapixel with a dimension of 50,000 by 70,000 pixels, which were clinically used for diagnosis. However, this also foreshadows the computation expense digital histopathology WSI requires. This limitation opened opportunities for novel and innovative approaches to conquer this issue [64], as well as taking advantage of such large amount of personalized data to build DL-based segmentation algorithms [45, 65–75] and outcome prediction models [76–84].

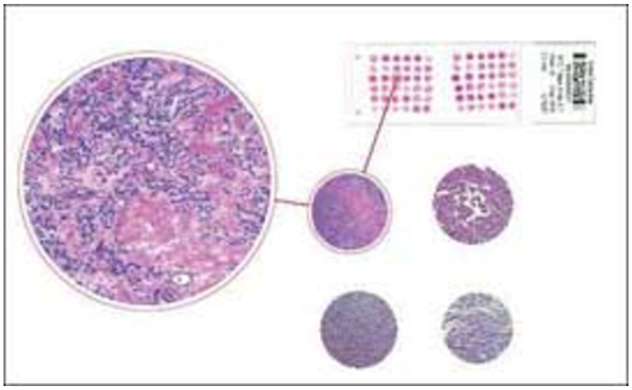


Figure 2.1: An example of tissue microarray, image extracted from [3].

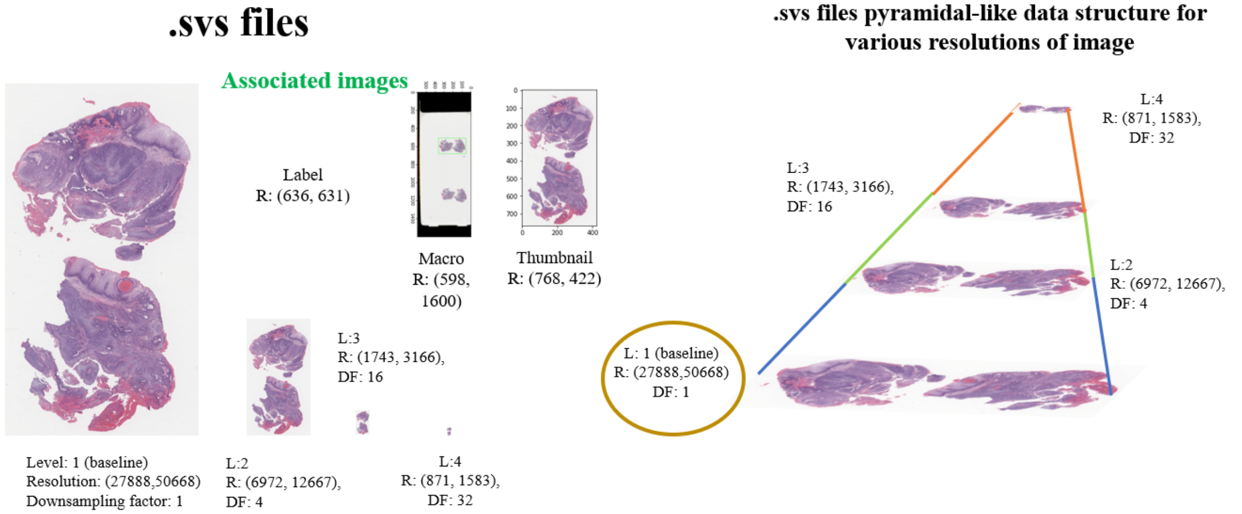


Figure 2.2: Pyramidal shaped data structure of histopathology whole slide images in the .svs file format. In this data format, there are usually four levels with various downsampling factor of the original highest magnification (40 x) in our case. In this thesis, the baseline (largest) resolution was used as circled in yellow. Meta files other than the digital pathology image itself are also associated such as an image of the macro view of the pathology glass slide, as well as the thumbnail of the slide.

2.1.2 Gynecologic cancer

Among women, nine percent of all female cancers worldwide were due to gynecologic malignancies [85]. Gynecologic malignant tumors come from reproductive organs such as ovaries, uterus, vagina vulva, and cervix in women [86, 87]. Its incidence rate is the highest in Asia, followed by Africa, Europe, Latin America, North America, reported by the World Health Organization [86]. While being among the most prevalent cancer occurring to women, they are heterogeneous diseases requiring various management paradigms [88] and are related to primary cancers from another site such as the breast, stomach, colon, and thyroid [89]. Cervical cancer incidence and mortality can be dramatically reduced via cancer screening human papillomavirus (HPV) vaccination, since with early detection and treatment, gynecologic cancers mediated by HPV are curable and preventable [88]. Very often, RT is used together with systemic therapy and surgery to improve locoregional control and extend overall survival for gynecologic cancer patients. However, RT itself can be curative for some cancers [88]. Especially, brachytherapy treatment techniques and new highly conformal External Beam Radiation Therapy (EBRT) have resulted in impressive recurrence and patient morbidity and mortality reduction [90]. Nonetheless, pelvic radiation delivered to gynecologic patients may lead to toxicity due to its unique anatomical locations, in addition to concurrent chemotherapy and/or surgery, and possible surgical interventions [90].

2.1.3 Immunohistochemistry and the p16 surrogate marker

Immunohistochemisty (IHC) links between immunology and pathology. Immunohistochemical staining can visualize epitopes in situ in histological tissue sections [91]. Grosset et.al. [92] recommended the combination of the the H & E-IHC protocol for improved diagnostic precision in the clinic as they were the first to show a potential in using hematoxylin counterstaining for immunohistochemistry interpretation.

In gynecologic pathology, p16 is an extensively studied biomarker [93], and immunohistochemistry (IHC) with p16 antibodies is used as a diagnostic tool [94]. In practice, p16 positive slides are used as positive controls for IHC staining using an anti-P16 antibody [95]. Furthermore, as an adjunct to morphologic assessment of cervical biopsies, to diagnose high-grade squamous intraepithelial lesion and its mimics, the p16 IHC has also been increasingly used. Yang et.al. [96] demonstrated the potential of better detection and upgrading of HSIL in patients who had a prior cytologic diagnosis with the use of p16 IHC. These evidence lead to application to our research that the cancerous regions of a H & E histopathology WSI is discerned with a reference of p16 positive IHC slides based on its p16 expression stained in brown. High-risk HPV's presence can be indicated by p16 diffuse positivity in the cervix as a surrogate marker [93]. This is because p16, or the p16 protein, has an important role in carcinogenesis [97]. Its loss is a critical event occurring early in tumor progression [98]. This is suggested by its its frequent mutations and deletions in human cancer cell lines [97] occuring early in many human cancers [98]. It is an inhibitor of cyclin-dependent kinases CDK4 and CDK6 as a cell cycle regulator [98], which gives rise to cell cycle arrest in epithelial cells undergoing cellular differentiation. The p16 is expressed excessively during the transforming HPV infection as a response to a transforming infection by high-risk HPV with oncogenic protein E7 overexpression [99]. This is why p16 diffuse positivity is a surrogate marker for high-risk HPV in gynecologic pathology [93, 99]. Therefore, in our work, p16 IHC stained slides were used as a guide to recognize corresponding cancerous regions on the H & E histopathology slide for the same slide.

The relationship between HPV status as expressed by the p16 surrogate marker on IHC slides with prognosis clinical outcome has been surveyed over the years. For example, enhanced prognosis was observed in p16-positive or HPV positive vaginal cancer patients as opposed to HPV negative or p16-negative patients [100]. Lee et.al. [101] investigated p16 status's prognostic impact on women with vulvar squamous cell carcinoma (SCC) treated with RT and found that those with a p16-positive status had a lower in-field recurrence (IFR) and extended survival than their p16-negative counterparts. Similarly, p16-positivity is related to fewer in-field relapses and improved progression-free survival in patients with vulvar cancer treated with adjuvant RT as shown by [102]. In addition, improved survival was also observed in p16-positive vaginal cancers across all tumor stages by [103]. Allo et.al. addressed the association between significantly worse prognosis in HPV independent vulvar squamous cell carcinoma compared to those that are HPV- or p16-positive. Favourable prognosis in cervical cancer patients may be predicted by p16 expression demonstrated by [104].

2.1.4 Nuclei size and cell spacing quantification and effects of radiation on cell nuclei targets

Ionizing radiation occurs when adequate kinetic energy is possessed by the radiation to eject one or more orbital electrons from the atom or molecule, which provokes physical and chemical processes. In directly ionizing radiation, individual particles with sufficient kinetic energy can directly disrupt the atomic structure of the absorbing medium through which they pass producing chemical and biological damage to molecules [105].

Radiation damages biologically important macromolecules such as the DNA. Free radicals and consequent molecular damage can be produced through direct and indirect actions from both electromagnetic and particulate radiations upon acting on cells [105, 106]. For indirect effects, water of an organism may absorb photons causing ionization and excitation in the water molecules. On the other hand for direct effects, the energy of ionizing radiation can be directly deposited in a macromolecule such as the DNA resulting in cell killing. There must be enough energy deposited in the DNA to cause sufficient number of double-strand breaks for a cell to be killed. It occurs by direct ionization of DNA moieties and reactions of hydroxyl radicals formed by water radiolysis [106–108]. In radiotherapy, DNA is the principal target for the biological effects of ionizing radiation, such as carcinogenesis, cell killing, and mutations. This is because DNA is present in only two copies, it therefore has very limited turnover. It is central to all cellular functions and is the largest molecule thus providing the biggest target. Furthermore, most DNA content is located in the cell nucleus. Since the intent of radiotherapy is to damage the DNA, the radiation dose to cell nuclei affects dose response [106–108].

Absorbed dose at a macroscopic level, is the average of the imparted energy per mass of the target volume [109]. The microscopic equivalent of absorbed dose is specific energy defined as the total energy imparted in a (small) volume of interest divided by the mass of that volume [110, 111]. In other words, the average of the stochastic quantity energy ε imparted per mass m at a point of interest defines the absorbed dose (D) [112]. The specific energy (z) is the microdosimetric equivalent to D, which is the direct quotient of ε divided by m [111]. However, molecular consequences of energy deposition of ionizing radiation encompassing stochastic events cannot be sufficiently described by an average effect to understand the cellular radiation dose response [113–115].

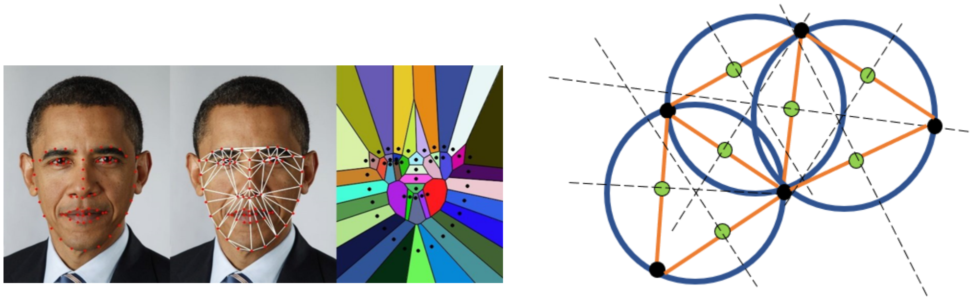
Individual cell nuclei of varying sizes do not necessarily receive a specific energy that is equal to absorbed dose D due to the presence of the microdosimetric spread. It is the standard deviation of dose dependent frequency distribution of specific energy for a group of cells or cell nuclei exposed to the macroscopic absorbed dose. The stochasticity of energy deposition leads to varying z, with a frequency distribution determined by the fluctuations in both the number of tracks passing the target, and the energy deposition per track, for a cell or cell nucleus-sized volume [116]. Its standard deviation (std) remarks a measure of the microdosimetric spread over a group of cells or cell nuclei. A microdosimetric spread in energy depositions for cell or cell nucleus-sized volumes can be induced by the stochastic nature of ionizing radiation interactions. When the target size decreases or when the energy of the radiation quality is reduced, the magnitude of the microdosimetric spread increases, as shown by Villegas et.al. [31]. Inter-patient variation in cell/nuclei target size variability leads to difference in energy deposition microscopic descriptors, which may result in patient-specific radiation dose response variation [28–30, 113].

Investigations on the direct impact of the microscopic level nuclei or cell target sizes to macroscopic post treatment outcomes for patients who underwent radiation therapy treatments is rare. Except the establishment that the morphology of tumor cell nuclei being significantly related with survival time of glioblastoma patients by Nafe et.al. [117], which supported the outlook of histomorphometry of tumor cell nuclei being a valueable prognostic marker for glioblastoma patients. However, sizes of tumor and outcome had been investigated by a few. Takahashi et.al. [118] found a correlation between the size of pancreatic cancer pathologic stage and overall survival. The continuous tumor-size–based end points as overall survival surrogates in randomized clinical trials were evaluated in metastatic colorectal cancer [119]. Erdogu et. al. also explored the correlation between outcomes of T4 non-small cell lung cancer patients and tumor size greater than 7 cm [120].

2.1.5 Graph-based image analysis

Delaunay triangulation [121] is a set of triangles made from a discrete set of vertices such that no vertex lies inside the circumcircle of any triangle in the set as shown in Fig. **??**. A Voronoi diagram partitions the 2D plane such that the boundary lines are equidistant from neighboring points giving rise to Voronoi cells [122] as illustrated in Fig. 2.3. Several studies have used Voronoi diagrams for their cell segmentation applications, nuclei are treated as a vertex of a Delaunay triangulation [4, 32, 123, 124] as shown in Fig. 2.5. For example, Sudbo et.al. [125] extracted the Voronoi features from the H & E digital histopathology slides and investigated the correlation between the Voronoi features in kidney cancer H & E stained slides and tumor growth parameters. Then, the authors also explored the prognostic value of such Voronoi features in slides obtained from prostate and cervical cancer patients and were able to distinguish good and poor outcomes [122].

Variables were scored based on their correlation with the discrete target variable with the



(a) Delaunay triangulation analogy in facial recognition applications

(b) Voronoi diagram analogy

Figure 2.3: Delaunay triangulation and Voronoi diagram analogy

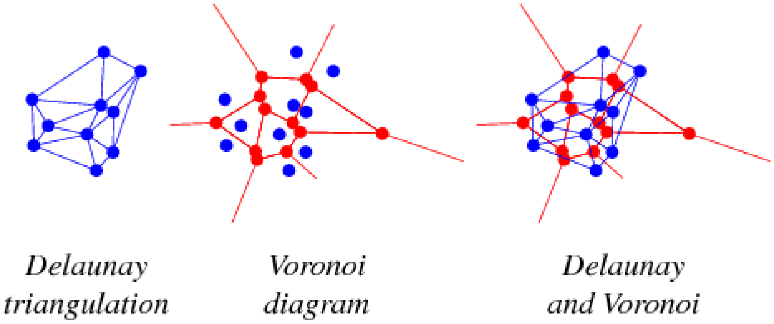


Figure 2.4: How Voronoi diagrams are constructed from Delaunay triangulation

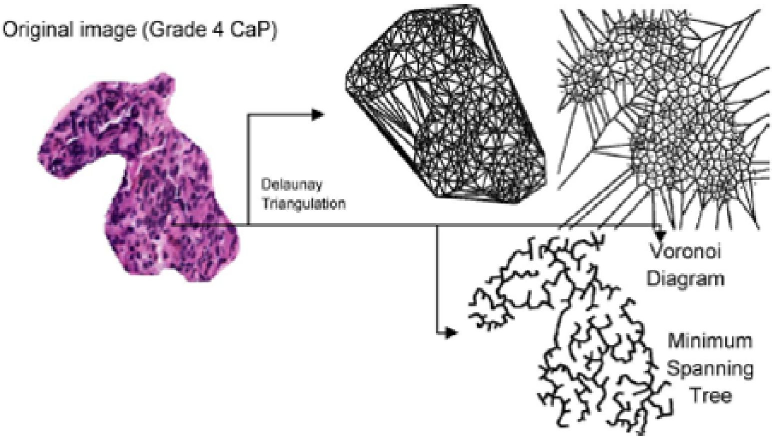


Figure 2.5: How Delaunay triangulations are constructed in histopathology images [4]

metrics of information gain (i.e.the expected amount of information (reduction of entropy)), gain ratio (i.e.,a ratio of the information gain and the attribute's intrinsic information, which reduces the bias towards multivalued features that occurs in information distribution gain), Gini (i.e. the inequality among values of a frequency), Chi-square (i.e. dependence between the feature and the class as measured by the chi-square statistic), ReliefF (i.e. the ability of an attribute to distinguish between classes on similar data instances), and Fast Correlation Based Filter (FCBF) (i.e. entropy-based measure, which also identifies redundancy due to pairwise correlations between features) [126] (Fig. 3.16).

2.2 autoML applications in outcome modelling

Automated Machine Learning (autoML) uses ML itself to automatically construct candidates of data pre-processing pipelines, hyperparameter optimization, and model building, evaluation, and selection, for a specified problem, which returns the best pipeline, for a specified evaluation metric, considering a wide array of different combinations of ML pipeline possibilities [43–51]. This, at the very least, can provide a much more productive starting point for predictive modelling which can be adjusted further with specific domain knowledge. Additionally, it leads to a data-driven model building approach agnostic to dataset preprocessing and statistical model without making an *a priori* assumption about statistical properties of the data. Instead of experimenting with many choices of data pre-processing, feature engineering, and hyperparamter tuning in classical ML, the TPOT autoML approach developed by Olson et.al. [48] does this automatically. With genetic programming, a number of operators such as preprocessing approaches, ML models, and their associated hyperparameters, are set to evolve iteratively untill the best pipeline with the highest user-defined evaluation metric is found.

Mustafa et.al. reviewed various autoML librarires such as autoSklearn, TPOT (used in this research), H2O, autoPrognosis as well as their pros and cons [127–129]. Dafflon et.

al. [51] evaluated a TPOT autoML approach to predict brain age from cortical anatomical measures. Meanwhile, autoML has also been used for the field of Radiomics. Su et. al. used autoML to perform radiomics feature selection to understand the feasibility of predicting H3 K27M mutation status.

Chapter 3

Predictive modelling of post RT recurrence for GYN cancer patients

3.1 Methods

To build a prediction model for post RT recurrence, we focused on obtaining patient-specific clinical variables at the time of diagnosis and RT described in section 3.1.1, and a method to extract patient-specific ns and cs distributions from pre-treatment H & E histopathology WSIs in section 3.1.

3.1.1 Patient cohort characteristics and data curation

Thirty-six gynecological (i.e., cervix, vaginal, and vulva) cancer patients (median age at diagnosis = 59.5 years) with a median follow-up time of 25.7 months were included into the current study. Nine of these patients (event rate of 25 %) experienced post RT recurrence at the time of the FU, the clinical outcome endpoint of interest. Two types of patient-specific data were computed and collected for each patient: H &E histopathology whole slide images and clinical information from the Jewish General Hospital ARIA database, respectively.

Once a list of diagnosed gynecological cancer patient cases were identified to be included in the study by our radiation oncologist, it was sent to the Pathology department for retrieval of the pre-treatment slides from the archives. A retrieval slip was filled and recorded for each case taken out of their archival box. Then, the slides were put in a new box with a divider separating each case. The pathologist analyzed the slides under a microscope and selected an H & E (hematoxylin and eosin stain) and p16 (immunohistochemistry stain) slide from each case for digital scanning (slice thickness around 4 microns). Since multiple sections can be cut for a cancerous tissue specimen paraffin wax resulting in various H & E WSIs belonging to the same patient, only the most representative (i.e., a slide with the most tissue coverage) was used for analysis upon visual inspection by our pathologist. Only slides expressing p16 (either positive or negative) were considered. All the selected slides were arranged on a tray and sent with a hard drive for digital scanning into .svs whole slide image files that were later viewed in the Aperio ImageScope ([130, 131]) software.

Furthermore, clinical variables were collected by our radiation oncologist for each patient. Clinical information retrieved from the Electronic Health Record (EHR) include age at diagnosis, time of diagnosis, total treatment time, time to recurrence (days), time to death (days), cancer type, follow up time length, clinical stage at time of radiation, radiological stage at time of radiation, histology (i.e., Squamous Cell Carcinoma (SCC) or Adenocarcinoma (Adeno)), and existence of post RT recurrence, among which whether the recurrence is locoregional or distant was distinctively collected as well. Lastly, the treatment modality (i.e., EBRT, brachytherapy, and concurrent chemotherapy) and the radiation dosimetry given by EBRT, and brachytherapy were also recorded. A summary of all clinical variables collected and their definitions can be found in Table 3.1.

3.1.1.1 Nuclei Size & Cell Spacing Distribution Extraction from pre-Treatment Histopathology Whole Slide Images (WSI)

Patient-specific nuclei size and cell spacing distributions were extracted from each selected digital histopathology WSI for downstream treatment outcome prediction modeling. Shown in Fig. 3.1, we depicted an example of a gigapixel digital H & E histopathological WSI with dimension of 50668 by 27888 pixels, retrieved in the .svs format from Aperio ImageScope [7], anonymized by being converted to .npy arrays using the OpenSlide [5, 6] library.

Tumor locations on the H & E stained WSI was confirmed by the p16 stained WSI. Cancer cells were identifiable from a large nucleus to cytoplasm ratio and a dark nucleus that contain multiple nucleoli. Another indication of tumour cells is lymphocytes located on the tumour site. They were shown under the microscope as small dark blue dots. There were three tumour grades: well-differentiated, moderately differentiated and poorly differentiated. Keratin pearls (pink, concentric spots) were found in well-differentiated and moderately differentiated tumours. H & E staining is a combination of both the hematoxylin stain and the eosin stain. Hematoxylin stains cell nuclei blue, while eosin stains cytoplasm and connective tissue pink. The expression of p16 in tumour cells was shown by their brown stain, while non-tumoral tissue with pale blue. Thus, tumour regions on the H & E WSI slide were identified using the p16 stained WSI slide, if positive, as a reference. Quality assured by our pathologist, cancerous regions were contoured on the Hematoxylin and Eosin (H and E) stained WSI in the Aperio ImageScope software [7]. The contours were saved as .xml files, which were then converted into a WSI binary mask.

Each gigapixel H & E WSI digitized at 40 x magnification (0.2482 microns/pixel) was divided into 5000 x 5000-pixel patches as shown in Fig. 3.1 b. Each 5000 by 5000 pixel patch from the H & E stained WSI and its binary mask with the same location indices on the WSI were then loaded into memory and processed iteratively. Within each patch, a Gaussian denoising filter was applied to the cyan channel of the RGB H & E patch, from which the blob detection algorithm found the nuclei center. A Delaunay triangulation from each continuously connected masked region of the patch were constructed (Fig. 3.3), giving rise to Voronoi diagrams (Fig. 3.4). A cell spacing radius was then approximated by the radius of a circle whose area was equivalent to a cell from the Voronoi diagram [4,32,132–134]. Similarly, a nucleus size radius was approximated as the radius of a circle whose area was the equivalent of a circle of combined stained area pixels dominated by hematoxylin content with an automatic thresholding algorithm [32] (Fig. 3.1 c).

Subsequently, cell spacing and nuclei size radii computed were stored into a vector for non-tumoral / cancerous regions for each patch as shown in Fig. 3.1 d. Then, non-tumoral / cancerous cell spacing and nuclei size distribution vectors at each patch level were concatenated into a WSI level non-tumoral / cancerous cell spacing and nuclei size distribution vector, as shown in Fig. 3.1 e.

With multiprocessing Central Processing Unit (CPU)s on a personal computer for each WSI per-patient, eight distribution feature types were calculated: the mean and standard deviation, of cell spacing and nuclei size distributions concatenated from all patches for cancerous and non-tumoral regions. A summary and definitions of these eight extracted imaging features can be found in Table 3.1.

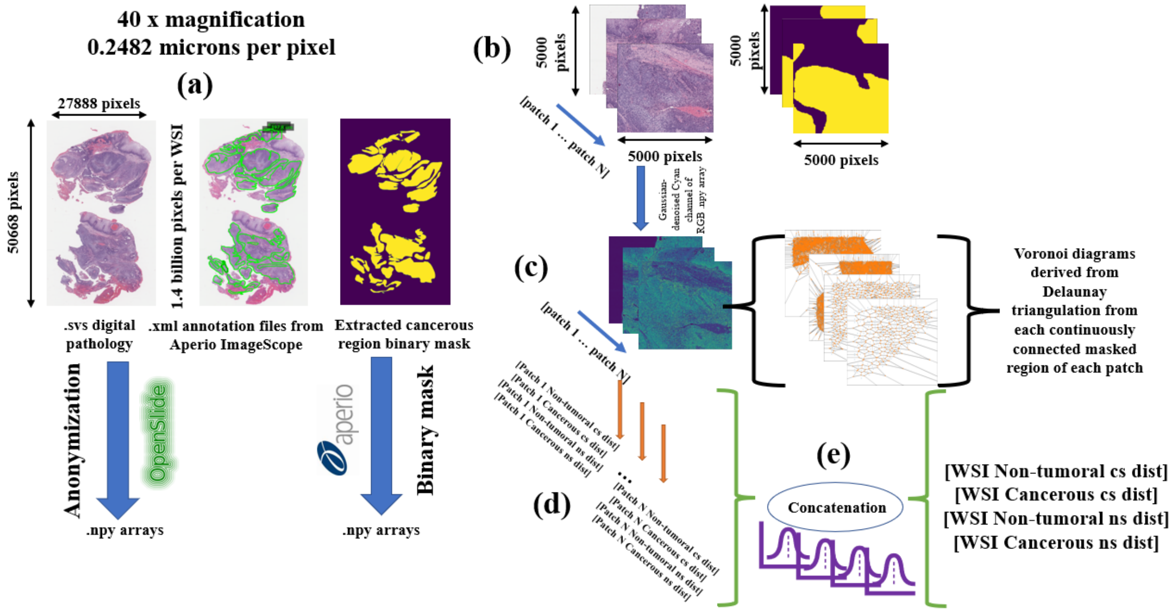


Figure 3.1: Methodology workflow: (a) an example of a gigapixel digital H & E histopathological WSI with dimension 50668 by 27888 pixels, retrieved in the .svs format, anonymized by being converted to .npy arrays using the OpenSlide [5, 6] library; contours of tumor regions were done on WSI in the Aperio ImageScope software [7], exported into .xml files, which is then converted into a WSI binary mask; (b) 5000 by 5000 pixel patch from the WSI and its binary mask were loaded into memory and processed iteratively; (c) A Gaussian denoising filter is applied to the cyan channel of the RGB H & E patch, from which the blob detection algorithm found the nuclei center and Delaunay triangulation from each continuously connected masked region of the patch were constructed, giving rise to Voronoi diagrams; (d) cell spacing and nuclei size radii were computed into a vector for non-tumoral / cancerous regions for each patch; (e) non-tumoral / cancerous cell spacing and nuclei size distribution patch-vectors were then concatenated into WSI non-tumoral / cancerous cell spacing and nuclei size distributions

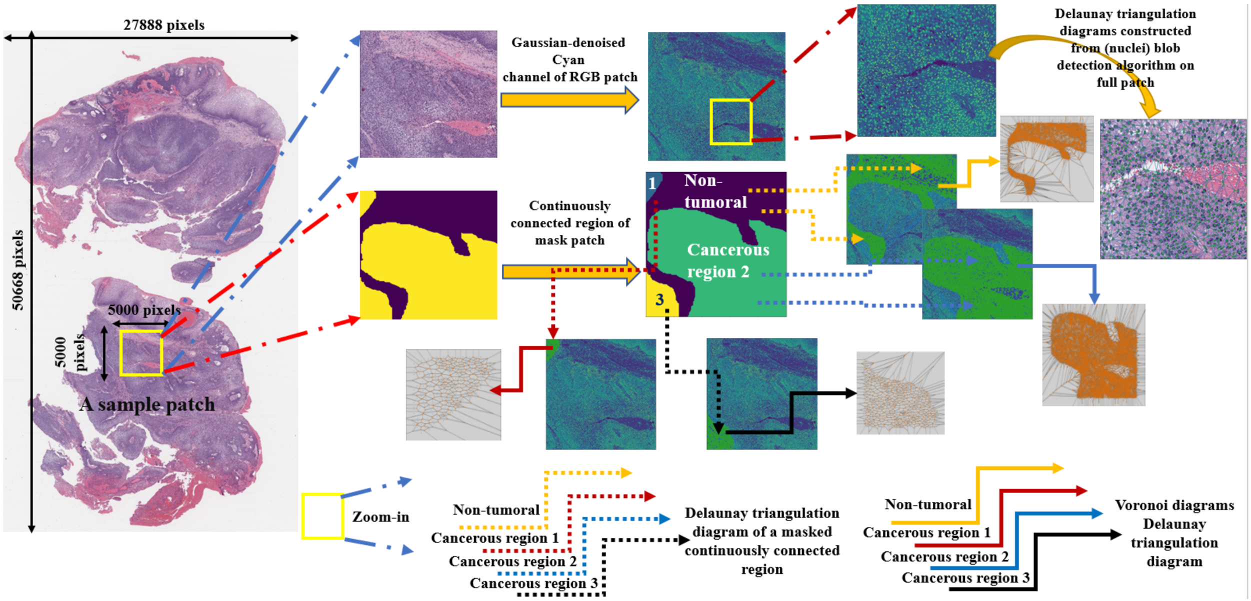


Figure 3.2: Zooming in on a single 5000 pixel by 5000 pixel patch showing the image processing details of Fig. 3.1 c. Each yellow box directs towards a zoomed-in view of the previous image. At the corresponding patch indices, both an H & E stained histopathology image patch, and its mask patch which was previously contoured from Aperio ImageScope [7] from the WSI level, were extracted together. In this example patch mask, the yellow region of the binary mask indicates a cancerous region of its H & E patch counterpart, therefore the purple background is the non-tumoral region of this patch. Three continuously connected regions (CCR) were identified within the masked cancerous region, and the rest of the patch is the non-tumoral region. A list of individual Delaunay triangulations from each CCR was constructed shown in dashed arrows colored by region name, from which its Voronoi diagram was then formed as shown in solid arrows.

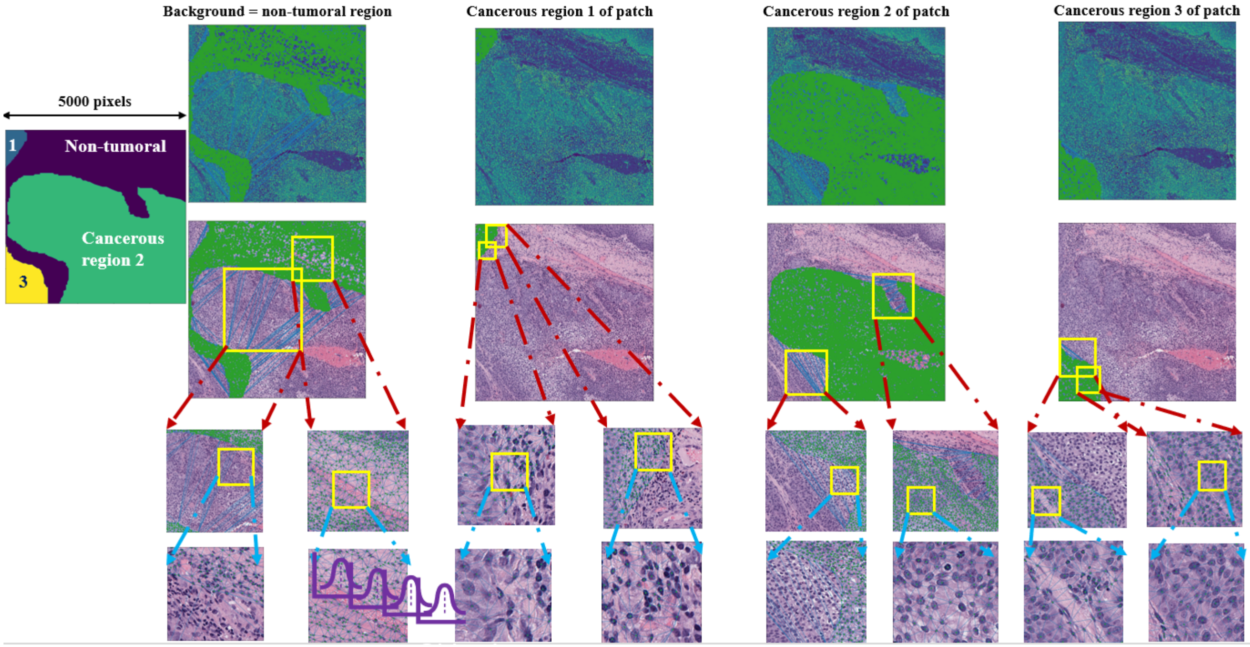


Figure 3.3: Zooming in on a single patch: Delaunay triangulation formed from a continuously connected region CCR as described in Fig.3.2. Here is a zoomed in display of how the Delaunay triangulation appears on the H & E patch upon nuclei center detection for the non-tumoral region and the three CCR. The same patch was shown in Fig. 3.2.

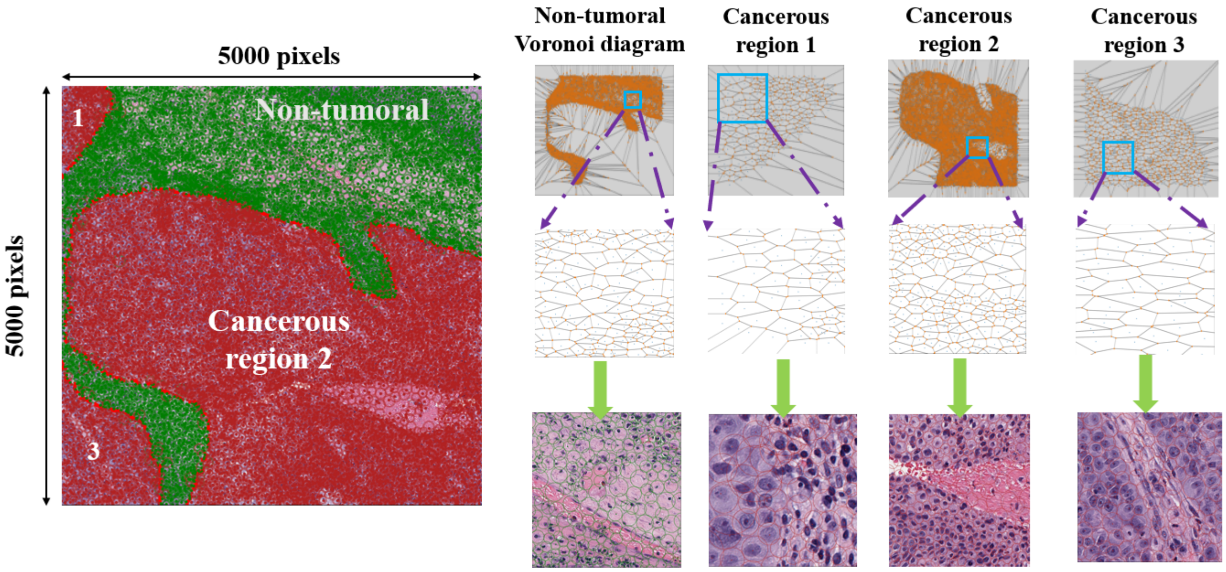


Figure 3.4: Zooming in on a single patch – Voronoi Diagram. Following the same patch shown in Fig. 3.2 and 3.4, here we show a detailed view of each Voronoi diagram constructed from each Delaunay triangulation from each cancerous region and non-tumoral region. The third row displays an approximate circle, with an equivalent area of that of the Voronoi cell, laying upon each H & E stained histopathology patch. Red indicates cancerous region and green nidicates non-tumoral region

3.1.2 Outcome Analysis

Prior to building a predictive model for the clinical endpoint of interest, post RT recurrence, it is crucial to understand the data itself. Therefore via visualization and statistical analyses, both imaging and clinical features summarized in Table 3.1 were explored independently in section 3.1.2.2 and in relation to the outcome of interest in section 3.1.2.3. Then we investigated a TPOT autoML approach to efficiently find the best ML pipeline to predict our outcome of interest in 3.1.2.5.

3.1.2.1 Data pre-processing

In the clinical data preprocessing process, ordinal encoding was used for categorical variables with a meaningful order, such as radiological or clinical stage at time of radiation, were encoded as its numerical value counterpart. However, subtypes within a specific stage were not distinguished. For example, a clinical stage of II was encoded as 2, whereas both IVA and IVB were encoded as 4. This is because stage subtype differentiation may result in more clinical feature variable types than there exists number of samples that the model could learn from. For a similar reason, the histology categorical variable was encoded as only Squamous Cell Carcinoma and Adenocarcinoma, without information regarding poorly / moderately/ well differentiated. For later modelling process, One Hot Encoding (OHE) was used for the categorical variables without an order whose input were simply processed as their categorical texts, such as the cancer type and histology.

3.1.2.2 Data exploration regardless of post RT recurrence outcome

To understand inter-patient variability for each of the eight derived imaging features (Table 3.1) notched box and whiskers plots were visualized for the mean and std of patient-specific cs and ns distributions (microns) across all thirty-six patients without making any assumptions of the underlying statistical distribution as presented in Fig. 3.6a (mean features) and Fig. 3.6b (std features). The "notch" indicates the 95 % confidence interval (CI) of the median. The median of each patient-population-wise distribution for each imaging feature (Table 3.1) were then computed in Table 3.2. To navigate correlation with imaging features, Pearson correlation coefficients between each of feature pairs (i.e. a cancerous region feature versus a non-tumoral region feature) were computed and visualized as scatter plots in Fig. 3.7a of cancerous vs. non-tumoral ns distribution mean, Fig. 3.7b of cancerous vs. non-tumoral ns distribution std, Fig. 3.8a of cancerous vs. non-tumoral cs distribution mean, Fig. 3.8b of cancerous vs. non-tumoral cs distribution std. Furthermore, while not assuming equal population variance, Welch's t-tests [135] were performed on each non-tumoral and cancerous imaging feature pairs as defined in (Table 3.1) and displayed in Table 3.3 obtaining t-statistics and its p-value.

3.1.2.3 Data exploration in relation to post RT recurrence outcome

Using the Orange [8] Python application, both the clinical and imaging features summarized in Table 3.1 (included features were highlighted in yellow) were investigated in relation to the binary classification target outcome as defined as post RT recurrence at the time of FU.

The outcome class imbalance can be visualized by a histogram as shown in Fig. 3.9. Stacked histogram were used to visually compare between categories (i.e., outcome) for each clinical feature (refer to Table 3.1) as presented in Fig. 3.9, 3.10, and imaging feature (refer to Table 3.1) as illustrated in Fig. 3.11 and Fig. 3.12. Values represented were proportional to the bars and were plotted vertically. To visually appreciate the trend of probability of post RT recurrence given a feature, the kernel density estimation (KDE) with a smoothing bandwidth of 19 were also displayed on top of each stacked histogram.

Each numerical feature was divided into two sub-groups based on the binary outcome of whether post RT recurrence had happened to this patient. Thus, a Student's t-test [136] was computed between these two subgroup values for each numerical feature returning the t-statistics and its p-value (Fig. 3.13, 3.15). For categorical values, a chi-squared test were performed (Fig. 3.14). Furthermore, each clinical and imaging feature was ranked in relation to the post RT recurrence outcome classification task with the Python Orange application [8].

Variables were scored based on their correlation with the discrete target variable with the metrics of information gain (i.e.,the expected amount of information (reduction of entropy)), gain ratio (i.e.,a ratio of the information gain and the attribute's intrinsic information, which reduces the bias towards multivalued features that occurs in information distribution gain), Gini (i.e., the inequality among values of a frequency), Chi-square (i.e., dependence between the feature and the class as measured by the chi-square statistic), ReliefF (i.e., the ability of an attribute to distinguish between classes on similar data instances), and FCBF (i.e., entropy-based measure, which also identifies redundancy due to pairwise correlations between features) [126] (Fig. 3.16).

3.1.2.4 Feature dimensionality reduction and visualization

To understand how all features (i.e., imaging and clinical) interact to classify an outcome (i.e., post RT recurrence), we explored two methods to reduce feature dimensionality and visualize them in a 2 dimensional space: Principal Component Analysis (PCA) and optimized multivariate linear projection using the FreeViz Orange application of Python [8, 9].

PCA is a well-known exploratory data analysis method to reduce dimensionality and interpret datasets while preserving "variability" of information as much as possible across disciplines. [137]. Since the number of Principal Component (PC) included in the analysis was set upon the cumulative proportion of variance of the data by hand, this makes PCA an adaptive data analysis technique [137]. It aims to find new uncorrelated variables, or PC, that were linear combinations (or projections) of the original dataset which successively maximize variance (i.e., the first PC accounts for the largest data variance, followed by the second and so on). Depending on an explained variance percentage accounting for the cumulative variance of the dataset, a number of PCs was then determined. Here with the Python Orange application [8], thirteen PCs were computed representing 99.5 percent of the original dataset cumulative variance (Fig. 3.17). The intelligent data visualization property of scatter plots in Orange [8, 9] can find the most informative projection pairs among the dataset's original features and the PCs to be visualized in a two-dimensional space. Orange [8] does so by finding 10 nearest neighbors in the projected 2D space on the combination of attribute pairs followed by checking how many of them belong to the same class for each data instance. The average number of same-classed neighbors determined the total score of the projection. The top five ranked projections were therefore visualized in Fig. 3.18.

FreeViz [9], also incorporated in the Orange Python application [8], uses optimization to find multivariate linear projections that produce scatter plots to best separate data instances from different class. It optimizes arbitrary linear transformations into lower dimensional spaces. A fascinating physics metaphor was analogized with this optimization procedure, where each data instance was imagined as a particle and those of the same class attract each other whereas those of different classes repel one another [9]. FreeViz [9] finds good class separation (i.e., optimization goal) by finding the projection with the minimal potential energy, which is minimized by gradient descent method. A global visualization (Fig. 3.19) of the classification task of distinguishing patients, given the clinical and imaging features, were built in a matter of a few seconds. In addition to its swift optimization, feature interactions and importance (based on the projection length), as well as intra-class similarities can also be easily visually interpreted.

3.1.2.5 Outcome modelling (post-RT recurrence) with imaging & clinical features using an Automated Machine Learning (AutoML) approach

Despite the small dataset and low event rate, the question of whether these features were "learnable" to predict an outcome (Fig. 3.5) was asked. Therefore, once the data itself was inspected and understood, a ML model was built to classify the outcome of interest (i.e., post RT recurrence) using both the imaging and clinical features as summarized in Table 3.1 (features included in the model were highlighted in yellow).

Uniquely, a TPOT autoML approach [48–51, 127, 128, 138–146] was adopted. TPOT [48] is a genetic programming-based autoML system that optimizes a number of available feature preprocessors and ML models with the objective to maximize a user-defined evaluation metric specific to the supervised classification problem. This was implemented to accelerate the finding of the best performing pipeline for this class-imbalanced and small dataset. Though a solid understanding and exploration of data at hand was still needed, expensive traditional manual processing of tasks such as missing value imputation, categorical values encoding, feature engineering such as normalization or scaling, iterative hyperparameter tuning, experimenting a series of ML algorithms, were all automatically optimized via TPOT [48] and autoML.

Here, all data from thirty six patients were fed into the TPOT algorithm as training data, with no specified test set. This is because further decreasing the number of already small dataset for training was not desirable and prone to overfitting. An internal cross validation (CV) score for each generation of the model that explored a different combination of feature preprocessors and ML models was produced. 100 generation and a population size of 50 were specified for the TPOT algorithm. A random state of one was set, while parallel processing of all CPU cores were utilized. A 10-fold stratified shuffle split was chosen to be the internal CV method, while the evaluation metric for the internal CV score was chosen to be the F1 score. Both methods were selected due to class imbalance and small size [147–149] of the dataset. This meant that for each TPOT model generation, an averaged interval CV F1 score was computed across all CV folds, and the "current best internal CV score" was reported at each generation in comparison to that of the previous generations.

Once 100 generations of TPOT algorithm were run, the model pipeline yielding the generation that had the best internal averaged CV F1 score was created. A Python script entailing the chosen hyperparamters, feature preprocessors, and ML algorithms giving rise to the best pipeline used were exported. To replicate the resulting best pipeline, a ML model was manually re-trained using exactly the optimized hyperparameters, feature processing steps, and ML algorithms suggested by the TPOT approach on the training data with the 36 patients. This was to further describe and understand the "best-pipeline" model with other evaluation metrics such as accuracy, balanced accuracy, precision, recall, and Receiver Operating Characteristics (ROC) Area Under the Receiver Operating Characteristic Curve (AUC). Finally, A permutation test (n=1000) was performed to validate the significance of the classification scores. During this process, the "best-pipeline" model accuracy were computed each time the features and outcome of the original dataset were permuatated for 1000 times. More specifically, randomized data of the original features and their permuted targets were generated. The model was applied to each permutation with its accuracy score calculated. The percentage of all 1000 scores obtained from randomized permuted data that were higher than that of the original data was represented as an empirical p value. Such p value less than 0.05 would indicate real dependencies between the features and the outcome of interest.

3.2 Results

A summary of all clinical features collected, imaging features derived from H & E histopathology WSI, their definitions were presented in Table 3.1. The results section here is organized in the same order as the outcome analysis method section in 3.1.2.

3.2.1 Statistical analysis and visualization regardless of outcome

Inter-patient variability in each of the eight imaging features (as defined in Table 3.1) extracted from H & E histopathology WSIs, as described in the method section 3.1.1.1, can

Clinical Variables	Variable Name	Definition
	age_at_diagnosis (year)	biopsy date - date of birth
	time_of_diagnosis	biopsy date
	total_treatment_time (days)	RT end - RT beginning
	time to recurrence (days)	post-RT recurrence date - RT beginning date
	time to death (days)	death date - RT beginning date
	cancer type class	vulva, vagina, cervix
	Follow_up_length	last FU (Follow-up) date - date of diagnosis
	Clinical stage at time of radiation class	I, IB1, IB2, II, IIA2, IIB, IIIB, IIIC, IIIC1, IIIC2, IVA, IVB Ordinal encoding was used in analysis such that: I, IB1, IB2 were encoded as 1 II, IIA2, IIB were encoded as 2 IIIB, IIIC, IIIC1, IIIC2, were encoded as 3 IVA, IVB were encoded as 4
	Radiological stage at time of radiation class	I, IB1, IB2, II, IIA, IIB, IIIB, IIIC, IIIC1, IIIC2, IVA, IVB Ordinal encoding was used in analysis such that: I, IB1, IB2 were encoded as 1 II, IIA, IIB were encoded as 2 IIIB, IIIC, IIIC1, IIIC2 were encoded as 3 IVA, IVB were encoded as 4
	EBRT class	1 if the patient received External Beam Radiation Therapy (EBRT), 0 if the patient did not receive EBRT.
	concurrent chemo class	1 if the patient received concurrent chemotherapy,0 if the patient did not receive concurrent chemotherapy.
	brachy class	1 if the patient received brachytherapy, 0 if the patient did not receive brachytherapy.
	p16 class	1 if positive, 0 if negative
Imaging Features		
	mean_radius_corrected_ns_masked_all	cancerous region nuclei size distribution mean
	mean_radius_corrected_ns_background_all	non-tumoral region nuclei size distribution mean
	std_radius_corrected_ns_masked_all	cancerous region nuclei size distribution standard deviation
	std_radius_corrected_ns_background_all	non-tumoral region cell spacing distribution standard deviation
	mean_radius_corrected_cs_masked_all	cancerous region cell spacing distribution mean
	mean_radius_corrected_cs_background_all	non-tumoral region cell spacing distribution mean
	std_radius_corrected_cs_masked_all	cancerous region cell spacing distribution standard deviation
	std_radius_corrected_cs_bakground_all	non-tumoral region cell spacing distribution standard deviation
Target Outcome		
	post radiation therapy recurrence	1 if post RT recurrence occurred 0 if post RT recurrence did not occur
	locoregional recurrence	For those who experienced post RT recurrence: 1 if there is local recurrence, 0 if there is no local recurrence
	distant recurrence	For those who experienced post RT recurrence: 1 if there is distant recurrence, 0 if there is no distant recurrence

Table 3.1: Patient-specific clinical, histopathology WSI extracted imaging features and target clinical endpoint outcome definitions. Only features and outcome highlighted in yellow were used in our analysis. Time-to-event variables were not included in the binary outcome prediction classification task. Treatment modality type clinical variables were also not included since most patients had the same treatment combination (i.e., all had EBRT, all but two had brachytherapy and concurrent chemotherapy). Since the number of patients and event rate is relatively low, locoregional and distance recurrence outcomes were not investigated at current stage of the study.

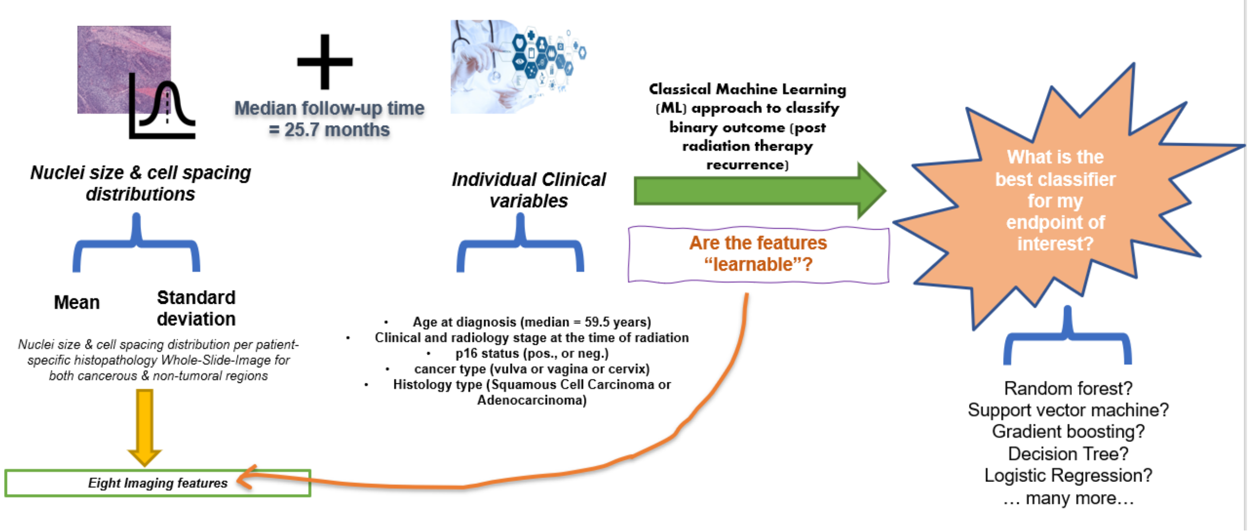


Figure 3.5: Workflow of combining nuclei size and cell spacing distributions imaging features and individual clinical variables such as the age at diagnosis, clinical and radiological stage at the time of radiation, p16 status, cancer type, and histology type for post radiation therapy recurrence outcome prediction

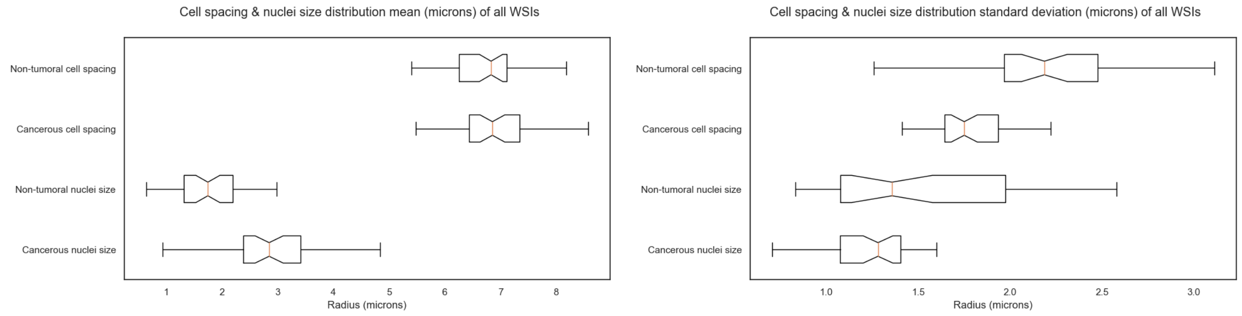
be clearly observed in the notched whisker and box plots. In particular, for cs and ns distribution mean (microns) of non-tumoral and cancerous regions for all thirty six WSIs as shown in Fig. 3.6a, and for cs and ns distribution std (microns) of non-tumoral and cancerous regions for all thirty six WSIs as depicted in Fig. 3.6b. Among the four cs and ns distribution mean (microns) features across all patients (Fig. 3.6a), the cancerous ns mean feature had the largest patient variability; and among the four cs and ns distribution std (microns) feature across all patients (Fig. 3.6b), the largest patient-population-wise variance was exhibited with the non-tumoral cs std feature.

To quantify observations from Fig. 3.6, the median of each patient-population-wise distribution for each imaging feature can be found in Table 3.2. Furthermore, the results of the Welch's t-test [135], or an unequal variances t-test, performed on each cancerous and non-tumoral feature pair was shown in Table 3.3. Note that each imaging feature was computed at a patient-specific level from their individual histopathology WSI. The following results, however, were described at a patient-population level in reference to Fig. 3.6. At the $\alpha = 0.05$ level of significance, there was sufficient evidence to remark a feature difference between the cancerous (median = 2.837 microns) and non-tumoral (median =

Imaging features	Median (microns) across all patients
mean_radius_corrected_ns_masked_all	2.837
mean_radius_corrected_ns_background_all	1.729
std_radius_corrected_ns_masked_all	1.282
std_radius_corrected_ns_background_all	1.301
mean_radius_corrected_cs_masked_all	6.938
mean_radius_corrected_cs_background_all	6.796
std_radius_corrected_cs_masked_all	1.788
std_radius_corrected_cs_bakground_all	1.36

Table 3.2: The median of notched boxplots of the population-wise cell spacing and nuclei size distribution mean (microns) and standard deviation (microns) as shown in Fig. 3.6 in a table. Definition of imaging features referred in this table can be found in Table 3.1.

1.729 microns) ns distribution mean (p = 7.136 e-08), the cancerous (median = 1.282 microns) and non-tumoral (median = 1.301 microns) ns distribution std (p = 0.027), and the cancerous (median = 1.788 microns) and non-tumoral (median = 1.36 microns) cs distribution std (p = 3.142 e-05). However, there was no adequate evidence to support a significant difference between the cancerous (median = 6.938 microns) and non-tumoral (median = 6.796 microns) cs distribution mean (p = 0.167).



(a) distribution mean radius (microns)

(b) distribution std

Figure 3.6: Notched boxplots of the cell spacing and nuclei size distribution mean (microns) in (a), and cell spacing and nuclei size distribution standard deviation (microns) in (b) for all patients

To further understand imaging feature correlations at a patient-population level, results of Pearson correlations and its p value (for testing non-correlation) between each cancerous

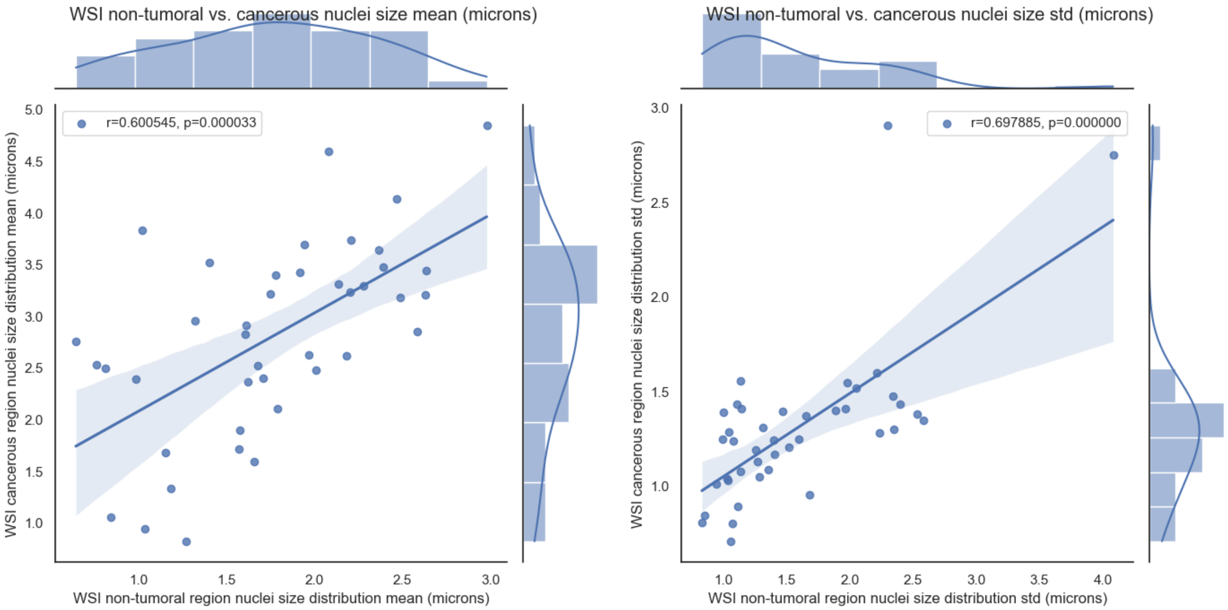
two-sided t test performed on	t-statistic	p-value
mean__radius__corrected__ns__masked__all **mean__radius__corrected__ns__background__all**	6.049	7.136 e-08
mean__radius__corrected__cs__masked__all **mean__radius__corrected__cs__background__all**	1.393	0.167
std__radius__corrected__ns__masked__all **std__radius__corrected__ns__background__all**	-2.265	0.027
std__radius__corrected__cs__masked__all **std__radius__corrected__cs__bakground__all**	-4.425	3.142 e-05

Table 3.3: Independent Welch's t test results (t-statistics and p-value) between each ns and cs imaging feature "pair" (i.e., a cancerous region quantity against a non-tumoral region quantity) across all thirty-six patients. Definition of imaging features referred in this table can be found in Table 3.1.

and non-tumoral feature pair across patients were also computed (shown in method section 3.1.2.2) and displayed as joint (scatter) plots made by the Seaborn Python library shown in Fig. 3.7. A strong Pearson correlation between the cancerous and non-tumoral regions of ns distribution mean ($r = 0.600$, $p < 0.001$) shown in Fig. 3.7a and std ($r = 0.700$, $p < 0.001$) shown in Fig. 3.7b were observed. Meanwhile, we discerned a weak Pearson correlation between the cancerous and non-tumoral regions of cs mean ($r = 0.252$, $p = 0.112$) shown in Fig. 3.8a and std ($r = 0.116$, $p = 0.471$) shown in Fig. 3.8b.

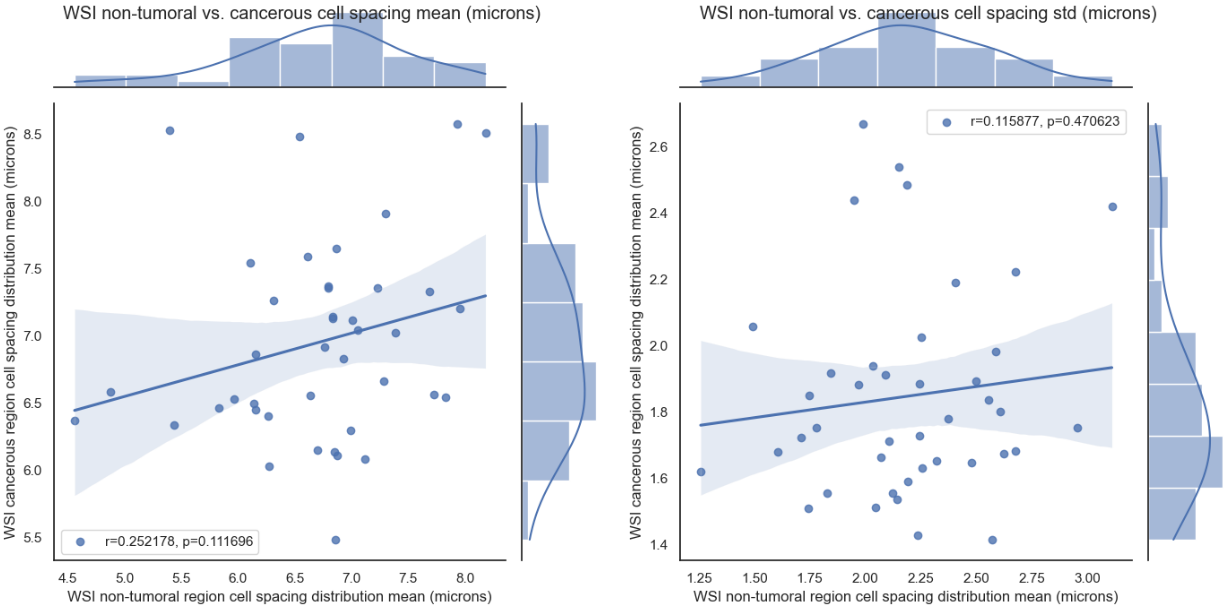
3.2.2 Statistical analysis and visualization in relation to outcome

Once the imaging features statistics had been analyzed independent of the outcome of interest (i.e., post RT recurrence at the time of follow up) at a whole patient-population-level, we scrutinized the behaviors of both imaging and clinical features between subgroups of patients with the opposite outcome (i.e., those who experienced post RT recurrence, or a positive class, versus those who did not, or a negative class). For all visualizations from this point on, a red color indicates a positive class and blue means a negative class. The size of a shape,



(a) (b)

Figure 3.7: Scatter plot of cancerous vs. non-tumoral regions nuclei size distribution mean (micron) in (a) and standard deviation in (b) Pearson correlation and p-value for non-correlation null hypothesis testing



(a) **(b)**

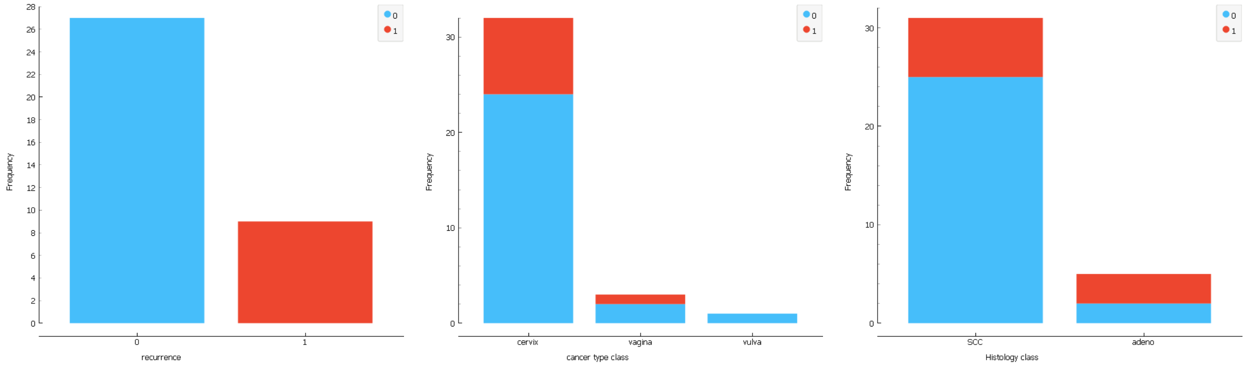
Figure 3.8: Scatter plot of cancerous vs. non-tumoral regions cell spacing distribution mean (micron) in (a) and standard deviation in (b) Pearson correlation and p-value for non-correlation null hypothesis testing

particularly in Figs. 3.18 and 3.19 stipulates the radiological stage at the time of radiation (i.e., the larger the shape, the more advanced the radiological stage).

Firstly, an imbalance in the outcome class (nine out of thirty six patients experienced post RT recurrence at the time of follow up, event rate = 25 %) can be recognized in Fig. 3.9a, and similarly in the stacked histograms of each cancer type (i.e., vulva, vaginal, and cervix) in Fig. 3.9b and histology type (SCC, Adeno) in Fig. 3.9c. Under our retrospective patient cohort, the probability of post RT recurrence given an age at diagnosis (Fig. 3.10a), a clinical stage at time of radiation, and a radiological stage at time of radiation were visualized in Fig. 3.10. Interestingly, a sharp increase in the probability of post RT recurrence was witnessed starting at the radiological stage of III at time of radiation as illustrated in Fig. 3.10c. In contrast, a relatively more flattened probability trajectory of post RT recurrence given a clinical stage at time of radiation was seen across all clinical stages as shown in Fig. 3.10b.

Intriguing pair-wise (i.e., a cancerous vs. non-tumoral imaging features) remarks can be provoked from patient-specific ns distributions (Fig. 3.11) and cs distributions (Fig. 3.12). Specifically, a monotonic decrease in probability of post RT recurrence was observed for cancerous ns distribution mean (in range 1 to 5 microns shown in Fig. 3.11a) and non-tumoral ns distribution mean (in range 0.5 to 2.5 microns shown in Fig. 3.11b). The same probability behavior was seen for non-tumoral cs distribution mean (in range 4 to 10 microns in Fig. 3.12b), however not in cancerous cs distribution mean (in range 5 to 8 microns shown in Fig. 3.12a). In terms of std features, opposite tendencies for probability of post RT recurrence were displayed for both pairs of cancerous ns distribution std (Fig. 3.11c) and non-tumoral ns distribution std (Fig. 3.11d), and cancerous cs distribution std (Fig. 3.12c) and non-tumoral cs distribution std (Fig. 3.12d). Notably, a monotonic increase in the probability of post RT recurrence was observed in cancerous cs distribution std shown in Fig. 3.12c.

At the $\alpha = 0.05$ level of significance, the radiological stage of two subgroups of patients who ended up experiencing post RT recurrence or not was revealed to be a statistically significant clinical feature analyzed by a Student's t-test in Fig. 3.13 (t-statistic = 3.794, $p = 0.001$). However, not enough evidence could suggest a significant difference in age at diagnosis and clinical stage at time of radiation and p-16 status (Fig. 3.15) for patient subgroups with opposite outcomes with the Student's t-test. Similarly, no significant insight can be drawn from Chi-square tests conducted between the opposite class subgroups for the cancer type ($p = 0.801$) feature as well as the histology feature ($p = 0.164$) (Fig. 3.14). Furthermore, a combination of metrics (i.e., Info gain, gain ratio, Gini, Chi-square, ReliefF, and FCBF) ranked the following features as the top five for their importance to the classification of post RT recurrence based on their correlation with the target outcome using the Orange Python application [8] in descending order: radiological stage at time of radiation, non-tumoral cs distribution std, non-tumoral ns distribution mean, non-tumoral cs distribution mean, and histology class.

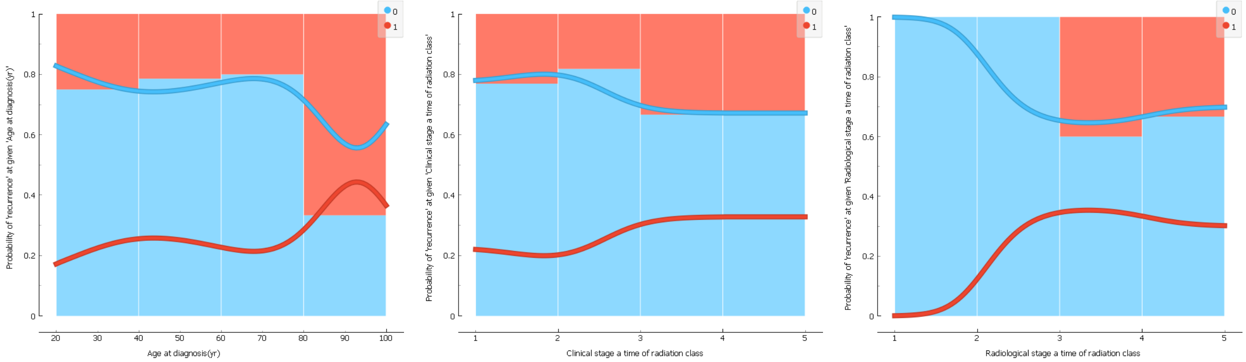


(a) recurrene **(b)** cancer type **(c)** histology class

Figure 3.9: Post RT recurrence frequency. Fig. 3.9a: outcome class imbalance was shown; Fig. 3.9b: stacked histogram of each cancer type (vulva, vaginal, cervix); Fig. 3.9c: stacked histogram of each histogram of each histology type (SCC and Adeno). Red means a positive class of post RT recurrence, and blue is a negative class of no post RT recurrence.

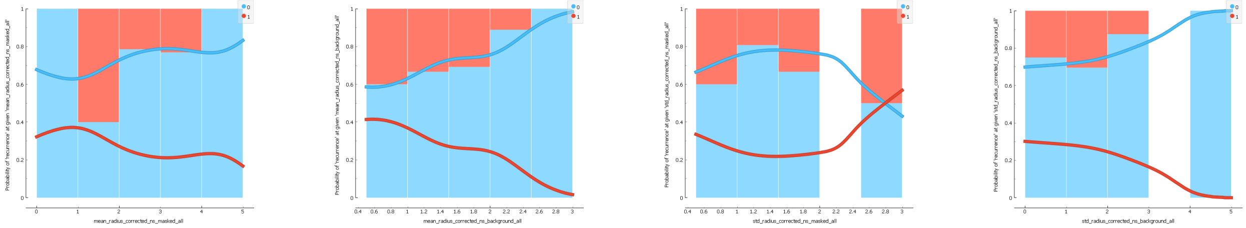
3.2.2.1 Feature dimensionality reduction and visualization

From the PCA, thirteen PCs were chosen to account for 99.5 % cumulative explained variance of the imaging and clinical features data set. From Fig. **??**, we observed that the first and the second PC account for 22.4 % and 17.3 %, respectively, of the total data set variation. This variance accountability (red curve) decreases for further PC included on the x-axis. The cumulative variance increases and reaches exponential plateau (in yellow curve) to 99.5 % at the thirteenth PC, which itself only accounted for 0.8 % of the total variance. With the Intelligent Data Visualization property of scatter plotting from the Orange [8] application, we showed the top five most informative 2 dimensional projections representing our features



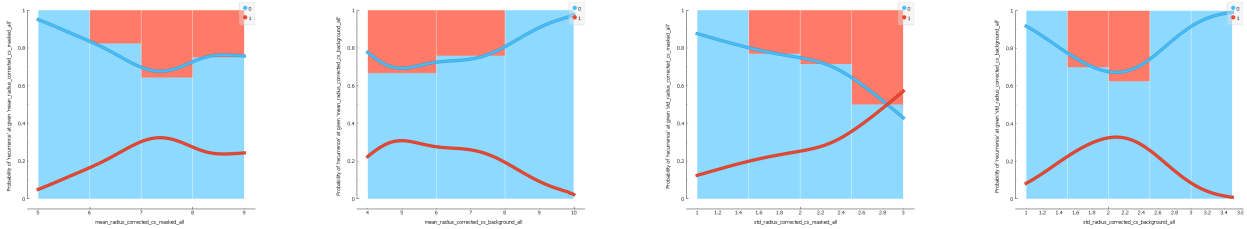
(a) age at diagnosis **(b)** clinical stage at time of radiation **(c)** radiological stage at time of radiation

Figure 3.10: Probability of post RT recurrence given 3.10a age at diagnosis, 3.10b clinical stage at time of radiation, 3.10c radiological stage at time of radiation, plotted as stacked histogram of the two-classed outcomes (Red means a positive class of post RT recurrence, and blue is a negative class of no post RT recurrence). A KDE with a bandwidth of 19 were laid on top of the stacked histogram.



(a) cancerous ns dist. mean **(b)** non-tumoral ns dist. mean **(c)** cancerous ns dist. std **(d)** non-tumoral ns dist. std

Figure 3.11: Probability of post RT recurrence given a ns imaging feature across patients. Each subfigure is displayed as a combination of a stacked histogram (red: post RT recurrence, blue: no post RT recurrence) and its corresponding KDE with smoothing bandwidth of 19 for each feature.



(a) cancerous cs dist. mean **(b)** non-tumoral cs dist. mean **(c)** cancerous cs dist. std **(d)** non-tumoral cs dist. std

Figure 3.12: Probability of post RT recurrence given a cs feature across patients. Each subfigure is displayed as a combination of a stacked histogram (red: post RT recurrence, blue: no post RT recurrence) and its corresponding KDE with smoothing bandwidth of 19 for each feature.

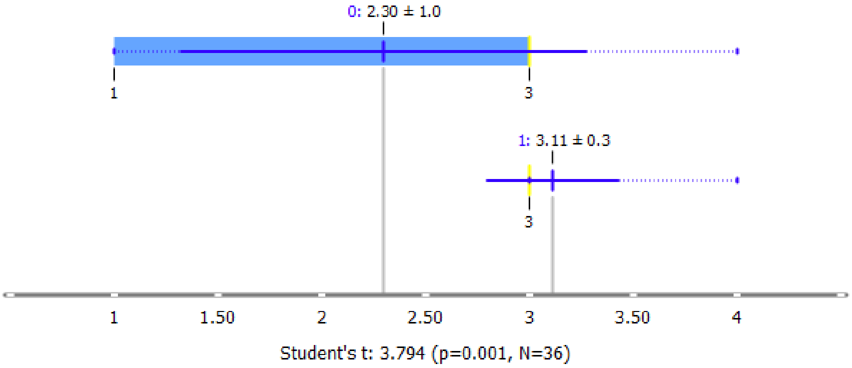
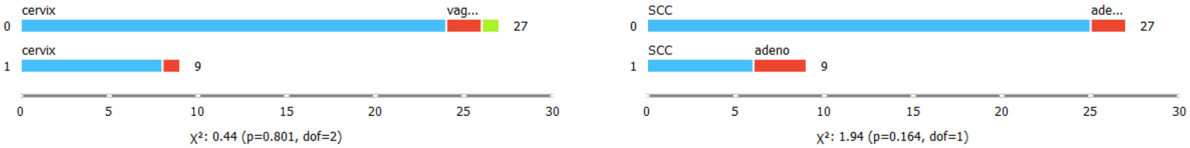
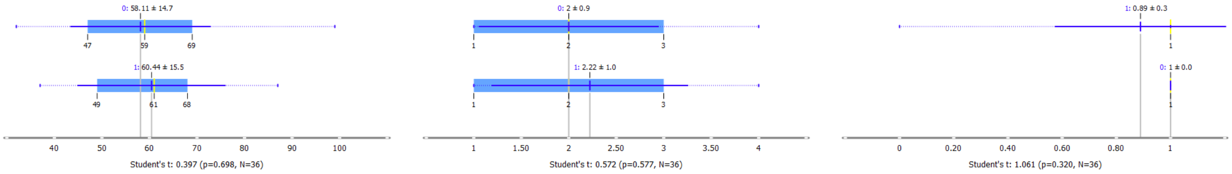


Figure 3.13: Student's t-test between subgroups (i.e., experienced post RT recurrence or not) for numerical feature of radiological stage at time of radiation.



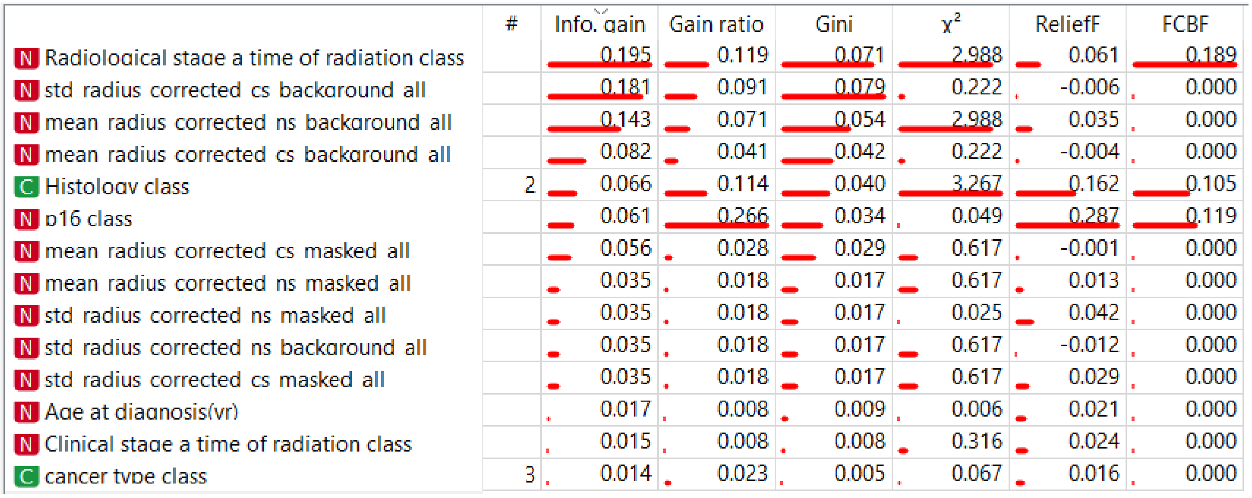
(a) cancer type class **(b)** histology class

Figure 3.14: Chi-square test between subgroups (i.e., experienced post RT recurrence or not) for categorical features of cancer type (i.e., cervix, vaginal, cervix) and histology type (i.e., SCC and Adeno).



(a) age at diagnosis **(b)** clinical stage at time of radiation **(c)** p16 status

Figure 3.15: Student's t-test between subgroups (i.e., experienced post RT recurrence or not) for numerical features of age at diagnosis, clinical stage at time of radiation, and p16 status.



	#	Info. gain	Gain ratio	Gini	χ²	ReliefF	FCBF
N Radiological stage a time of radiation class		0.195	0.119	0.071	2.988	0.061	0.189
N std radius corrected cs background all		0.181	0.091	0.079	0.222	-0.006	0.000
N mean radius corrected ns background all		0.143	0.071	0.054	2.988	0.035	0.000
N mean radius corrected cs background all		0.082	0.041	0.042	0.222	-0.004	0.000
C Histology class	2	0.066	0.114	0.040	3.267	0.162	0.105
N p16 class		0.061	0.266	0.034	0.049	0.287	0.119
N mean radius corrected cs masked all		0.056	0.028	0.029	0.617	-0.001	0.000
N mean radius corrected ns masked all		0.035	0.018	0.017	0.617	0.013	0.000
N std radius corrected ns masked all		0.035	0.018	0.017	0.025	0.042	0.000
N std radius corrected ns background all		0.035	0.018	0.017	0.617	-0.012	0.000
N std radius corrected cs masked all		0.035	0.018	0.017	0.617	0.029	0.000
N Age at diagnosis(yr)		0.017	0.008	0.009	0.006	0.021	0.000
N Clinical stage a time of radiation class		0.015	0.008	0.008	0.316	0.024	0.000
C cancer type class	3	0.014	0.023	0.005	0.067	0.016	0.000

Figure 3.16: Feature Rank from the Orange Python application [8, 8] based on Info gain, gain ratio, gini, chi-square, ReliefF, FCBF

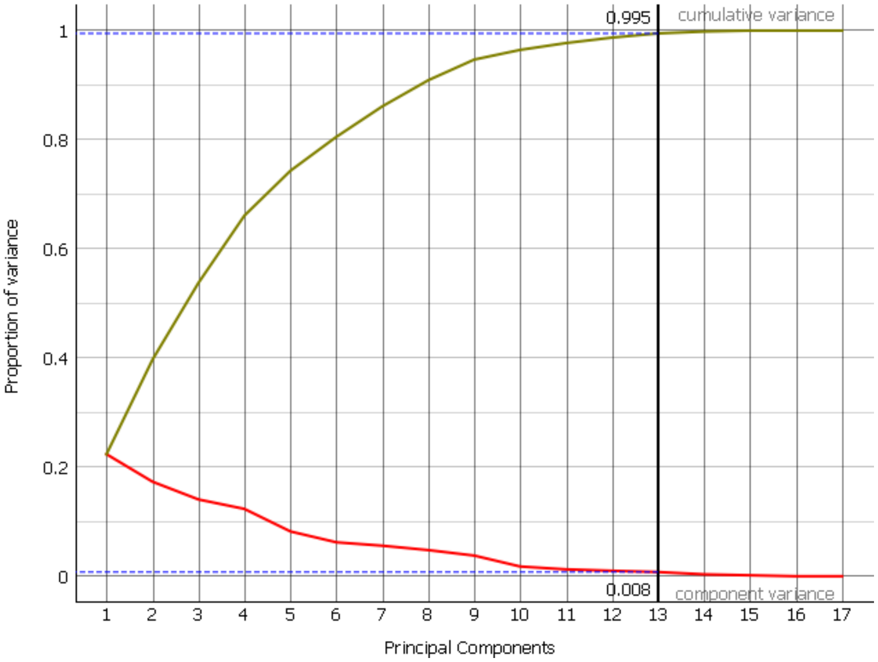
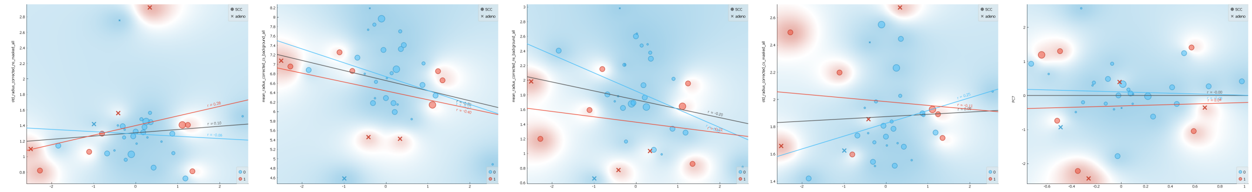


Figure 3.17: PCA of the imaging and clinical feature dataset. It shows the proportion of variance as a function of PC included to represent the original data. The yellow exponential plateau towards 1 (i.e., a full representation of the original data variance) shows the cumulative variance the PCs represent as the number of PC increases. The decaying red curve illustrates the variance each PC accounts for itself. For example, the first PC accounted for 22.4 % of the original data variance, and the second PC 17.3 %. The first and the second PC together accounts for a cumulative variance of 39.7 % of the original data. In our analysis, 13 PCs were chosen to altogether account of 99.5 % of the total data variance. The green curve indicates the cumulative variance, and the red curve shows the component variance.

where the color of each instance indicates the outcome (red: had post RT recurrence, blue: no post RT recurrence) (Fig.3.18). Though promising neighbor instances "clustering" can be observed (for example at the left bottom corner of Fig. 3.18a), no visually interpretable linear separation between the two outcome classes can be easily observed from all PCA scatter plots in Fig. 3.18. However interestingly for patient who experienced post RT recurrence (shown as red point), Fig. 3.18a shows a Pearson correlation of 0.28 between the cancerous ns std and PC7, and -0.4 between non-tumoral cs mean and PC 7 in Fig. 3.18b.



(a) Rank 1 **(b)** Rank 2 **(c)** Rank 3 **(d)** Rank 4 **(e)** Rank 5

Figure 3.18: The top 5 most informative 2D projections where each projection pair as selected the original clinical and imaging features and from thirteen PCs determined from a PCA accounting for a 99.5% explained variance with the Intelligent Data Visualization property of scatter plots in the Orange Python application [8]. Red color indicates an instance that experienced post RT recurrence, and blue means otherwise. The red line in each subplot is a best fitted linear regression of the data instances with post RT recurrence and the blue line is the same but for those without post RT recurrence with a computed Pearson correlation coefficient. A shape of a circle is the SCC histology type, and an x is the Adeno type. The size of each shape represents the radiological stage at time of radiation. For example, a larger red circle means a more advanced radiological stage for a patient who experienced post RT recurrence with SCC. Fig. 3.18a shows cancerous ns std on the y axis and PC 7 on the x axis; Fig. 3.18b: non-tumoral cs mean (y axis) and PC 7 (x axis); Fig. 3.18c: non-tumoral ns mean (y axis) and PC 7 (x axis); Fig. 3.18d: cancerous cs std (y axis) and PC 7 (x axis); and Fig. 3.18e: PC 7 (x axis) and PC 12 (y axis).

On the contrary, a clearly visible class separation can be observed in Fig. 3.19 from the multivariate linear projection from the FreeViz Orange application of Python [8, 9]. Since instances on the FreeViz 2D projection figure was grouped by proximity where the more connected the base vectors (features) were, the closer they locate on the 2D plane. Features more important to the classification had longer projection length and pointed to a more orthogonal direction to the line or curve separating the binary class cluster. A feature is more prominent at making the classification decision if it sits at a region mostly populated by instances of a specific class. [9]. Therefore for the positive class of post RT recurrence, radiological stage at time of radiation, histology class, patient-specific cancerous cs distribution mean, and age at diagnosis were of the most importance, due to its base vectors' orthogonal direction to the curve that separates the two classes and the normalized length of the base vector. The clinical stage at time of radiation has the least significance out of all clinical features because of its parallel direction of its base vector to the separation curve. For the negative class of no post RT recurrence, among all imaging features, the patient-specific non-tumoral ns distribution std and the non-tumoral ns distribution mean were shown to be the most promising due to its base vector length and the cluster of data instances surrounding these two vectors.

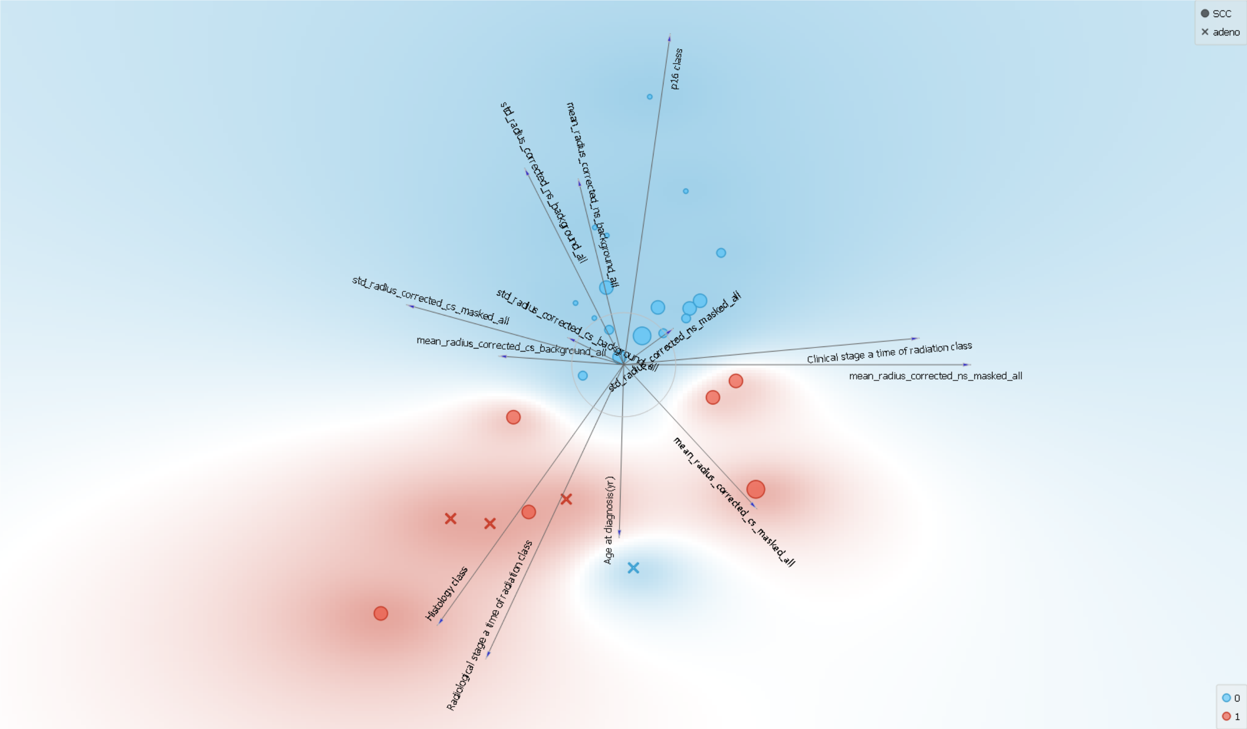


Figure 3.19: Multivariate linear projection implemented from FreeViz of the Python Orange package [8, 9] successfully visually separated the binary outcome classes (i.e., post RT recurrence or not) by optimizing arbitrary linear transformations into lower dimensional spaces.

3.2.3 TPOT autoML approach to predict clinical endpoint: post RT recurrence

Trained on Google Colab, on CPU in a few hours, the best pipeline yielded by TPOT autoML achieved a best averaged internal CV F1 score of 0.83 after 100 generations.

In this best pipeline automatically optimized by TPOT, feature preprocessors were selected, and each of the following steps included hyperparameter values that were automatically fine tuned. A Radial Basis Function Kernel (RBF) sampler (gamma = 0.25) was applied to combined imaging and clinical input variables for training. The resulting features were fed into an eXtreme gradient-boosting (XGBoost) classifier (`learning_rate` = 0.1, `max_depth` = 1, `min_child_weight` = 15, `n_estimators` = 100, `n_jobs` = 1, `subsample` = 0.45, `verbosity` = 0). One hot encoding was then applied. Its outputs were propagated as "synthetic features" followed by polynomial feature transforms. All raw and transformed features were trained with a decision tree classification algorithm(`criterion` = "entropy", `max_depth = 9`, `min_samples_leaf = 12`, `min_samples_split` = 7).

Now that the optimized feature preprocessing steps and fine-tuned hyperparameters were obtained, we reproduced this result by retraining the entire dataset where a 10-fold stratified shuffle split CV was implemented whose evaluation metrics (i.e., ROC AUC, accuracy, balanced accuracy, precision, recall, and F1 score) were averaged from each CV fold. Fig. 3.20) shows the ROC AUC scores at each fold of the CV. Note that the same CV method was chosen here to recreate that used in the TPOT autoML approach.

The model achieved a 10-fold stratified shuffle split cross-validation scores of 0.875 for mean accuracy, 0.917 for mean balanced accuracy, 0.783 for precision, 1 for recall, 0.85 for F1 score, and 0.917 for Area Under the Curve of Receiver Operating Characteristics Curve (Table 3.4), to predict the patient cohort's post RT recurrence binary outcome. A p-value of 0.036 was obtained from the permutation test (Fig. 3.21). This implies real dependencies between the combined imaging and clinical features and outcomes which were learned by the classifier, and the promising model performance was not by chance.

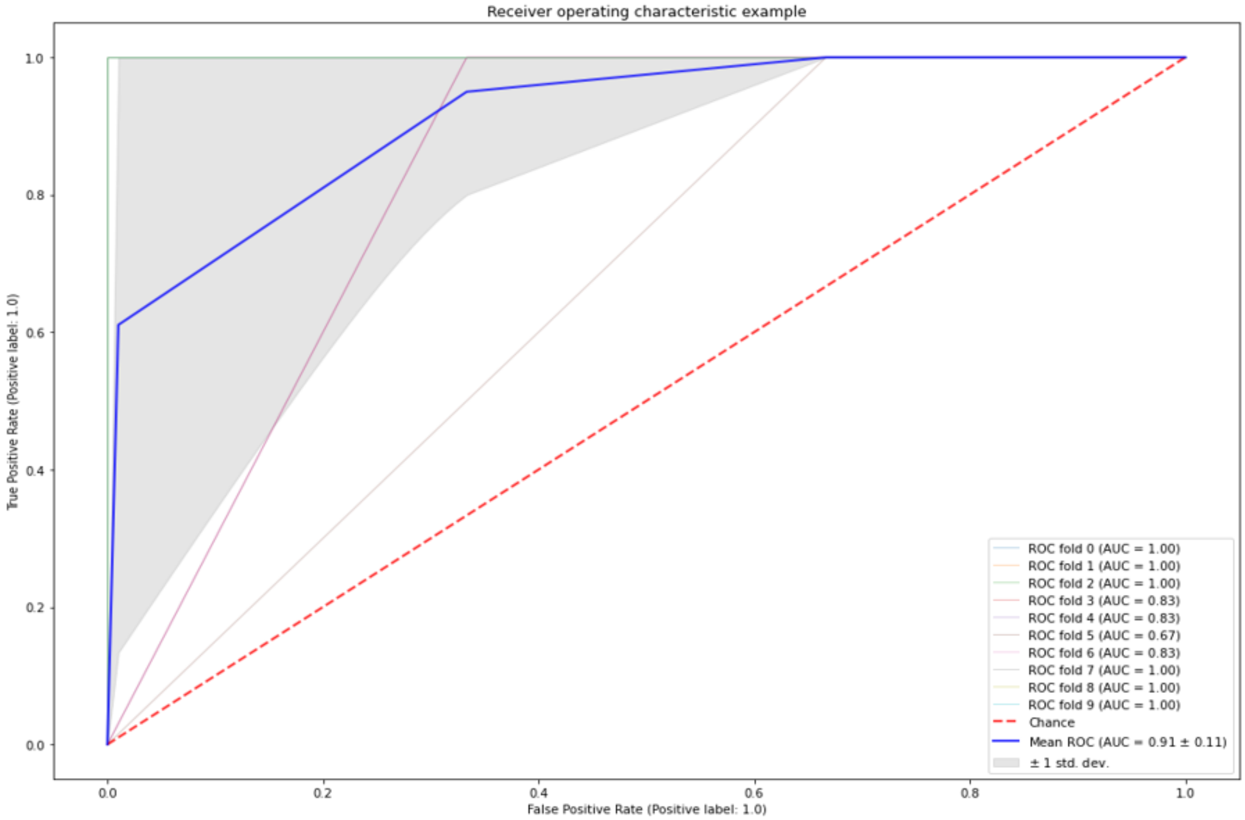


Figure 3.20: ROC AUC of 10 stratified shuffle split CV for the best pipeline selected by TPOT autoML

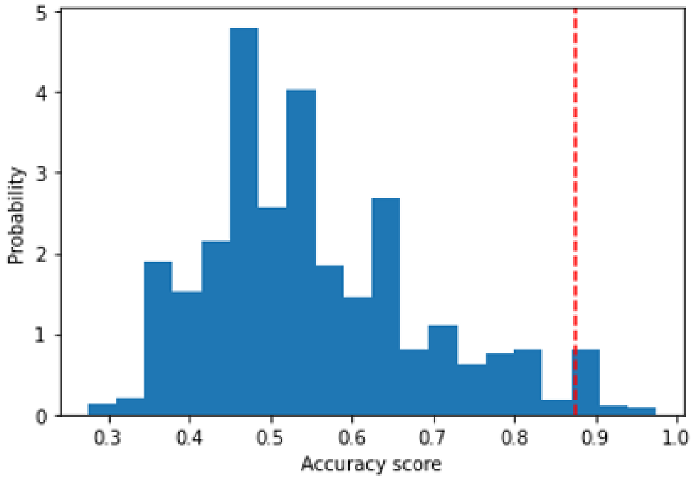


Figure 3.21: Permutation test (n = 1000) performed the best model pipeline produced from TPOT autoML based on accuracy, p-value = 0.036. The permutation test encompasses the process of permuting the order of the features and their corresponding labels or target or outcome (here, for 1000 times), and apply the same best pipeline model on the permuted dataset. A model accuracy score was produced at each permutation contributing to the probability density function shown. The red vertical dash line shows the accuracy score of the model on the original dataset which is 0.88. An empirical p-value was computed to be 0.036, meaning that only 3.6 % of the classification accuracy scores from the permuted dataset was higher than that of the accuracy score of the model applied to the original dataset. This indicates that the chance of an accuracy score from the permuted dataset being higher than that of the original dataset was by chance, therefore concluding real dependency between the features and the target outcome in the classification model.

Model performance evaluation metrics	Scores
Accuracy	0.875
Balanced accuracy	0.917
Precision	0.783
Recall	1
F1	0.85
ROC AUC	0.917

Table 3.4: Model performance evaluation metrics from TPOT autoML derived best pipeline

Chapter 4

Discussion

4.1 Extraction of cell and nucleus size distribution from H & E stained histopathology WSI

A method to automatically extract patient-specific cell and nuclei size distributions from a gigapixel H & E stained histopathology WSI was developed for gynecological patients. This approach is generalizable for histopathology WSI of any cancer type as long as it was stained by hematoxylin and eosin. The basis of this approach was built upon work of DeCunha et.al. [32]. However, a major difference lies in the fact that H & E histopathology TMA was used in the previous work, which had a pixel dimension roughly one tenth for that of a gigapixel WSI used in this study. Though a 5000 pixel by 5000 pixel patch-based method was implemented to process our WSI iteratively to not exceed available CPU memory as described in section 3.1.1.1, the heavy computation expense cannot be understated. Another limitation of this current method is attributed to a lack of accountability for the cell membrane extent in the cell size approximation with cell segmentation using Voronoi tessellation diagrams [32, 150], since the cell spacing radius was essentially the distance (in microns) between two nuclei. Moreover, a unanimous threshold parameter determining pixels dominated by hematoxylin over eosin content was set to compute the nucleus radii (see section 3.1.1.1). However, a common issue in digital pathology image analysis rose from slides coming from various scanners. This might especially affect our algorithm's generalizability when the patient cohort in a future study extends to external hospitals with various pathology scanners. The fourth source of inefficiency in this approach was by virtue of manual delineation of cancerous regions of the H & E stained histopathology WSI using the p16 positive slides as a reference, a DL-based automatic segmentation algorithm would be much preferred due to its speed and reproducibility of segmentation results.

Therefore, a number of improvements can be made upon this current method. First of all, since predicting post RT recurrence outcome at the time of FU is the downstream goal subsequent to extracting patient-specific cell and nuclei size distributions for cancerous and non-tumoral regions, a decision for the distribution extraction algorithm to only "pay attention" to WSI regions that are most relevant to the outcome may be computationally economical. Major advances on Vision Transformer (ViT) has been made to the computer vision domain [151]. The Convolutional Neural Network (CNN) is popularly considered as a de-facto approach in the medical computer vision community from achieving state-of-the-art results for tasks such as detection, image segmentation, synthesis, registration, reconstruction, etc. for a wide range of image modalities such as radiography, CT, mammography, ultrasound, MRI, positron emission tomography (PET) [152]. The mathematical operation called convolution is how CNNs extract discriminative features in its local receptive field [153, 154]. This locality attribute allows it to model local visual feature of images. However, this very operator also inevitably limits CNN's ability to learn from long-range pixel relationships within the global context of an image [152]. Vision transformers (ViT) [155], on the other hand, bears the capability of encoding long-range dependencies of the input image to learn its full representation. This is achieved by breaking down an original image into patches which are then fed into a standard transformer encoded as used in Natural Language Processing (NLP) [152, 155, 156]. Many state of the art (SOTA) ViT results are indicating the potential full replacement of convolution operators by transformer modules. Especially in medical image segmentation, [157] showed that due to a superior ability to encode the global context of an image, nearly all transformer-based autosegmentation models such as TransUNet [158] performed better than CNN-based models. [159] introduced TransMIL, transformer-multiple-instance-learning method, where the WSI histopathology classification algorithm was not only weakly supervised, but also interpretable due to attention scores visualization with heatmaps. Interpretability is crucial for WSI histopathology in an outcome prediction context as well. Lu et.al. [160] developed a data efficient weakly supervised DL method to automatically identify sub-regions of the histopathology whole slide image with high diagnostic value. Therefore, an application of the ViT architecture guiding its attention to pixels most relevant to the outcome of interest will not only decrease input data size, but also shorten time spent on extracting cell and nuclei size distributions from unimportant regions. In a sense, we would ideally want to mimic how a pathologist diagnoses cancer in real life under the microscope, which is scrutinizing the slide at a lower resolution on the digital pathology slide, identifying where to look further, then magnify the pixels in just those parts. This might also offer competitive advantage for automatically segmenting cancerous regions. Because frankly speaking, data itself is not lacking when training a DL model for a task when it is in gigapixel digital pathology. However, what matters more is the useful supervisory signal the data gives to the downstream task, instead of a large amount of repetitive data inadequate of distribution variety. Furthermore, self-supervised learning algorithms is causing a paradigm shift in medical Artificial Intelligence (AI) [**?**, 22, 161–169] and is particularly suitable for histology image pixel classification (i.e. cancerous vs. non-tumoral), which alleviates labelling intensity and generalizability of domain specific data via contrastive learning. Other weakly supervised learning for DL autosegmentation on histopathology WSI has also been extensively surveyed in the recent years which can be further experimented on [83, 169–171]. In addition, there exists non-tumoral cells in the mixes of cancerous cells such as: stromal, immune, and vascular cells. Non-tumoral cells are also crucial in tumor growth and disease progression [32]. This alludes to the notion that instead of simply classifying a cell or nuclei as cancerous or non-tumoral, it may be worthwhile to [172–174] classify cell types within the cancerous or non-tumoral region to investigate how their cell/nuclei size distributions differ and impact the post treatment outcome.

4.2 Image and clinical feature analyzed with statistical tests regardless of and in relation to the clinical outcome of interest, post RT recurrence at the time of FU

To understand how patient-specific pre-treatment cell/nuclei size distributions might infer post RT recurrence at the time of FU, thorough statistical analyses and visualization were conducted with the clinical and imaging features (detailed in Table 3.1) by themselves and in relation to the outcome of interest.

It is worthwhile to note that time-to-event variables were not included since the event rate (25 %) was too low to conduct robust survival analysis at the current stage of the study. Furthermore, the categorical variable of treatment modality (i.e. EBRT, brachytherapy, and concurrent chemotherapy) combinations will be considered in the future when a larger patient cohort is curated. It was not included at the current stage of the analysis due to lack of variation of these variables among different patients, since all patients had EBRT, and all but two underwent concurrent chemotherapy and brachytherapy, out of all thirty six patients.

Similar to DeCunha et.al. [32] showed, the median of a cancerous image feature is larger than that of a non-tumoral image feature among the entire patient population. And at the $\alpha = 0.05$ level of significance, there was significant difference between cancerous and non-tumoral ns distribution mean, the cancerous and non-tumoral ns distribution std, and the cancerous and non-tumoral cs distribution std.

Imaging features of patient-specific cell/nuclei target size distributions was at the heart of this investigations. However in reality, a post radiation treatment outcome of recurrence cannot be ingenuously quantified in relation to these imaging features directly without the consideration of other clinical variables that were pivotal to an outcome. The analyses of probability of post RT recurrence at a given numerical range (or determined bin size) of a certain feature, however, is not a violation towards ignoring confounding variables to the outcome.

The radiological stage at time of radiation is the most prominent variable among all clinical features, as demonstrated by a number of results. For example, Student's t-test between patient subgroups based on their outcome ($p = 0.001$) in Fig. 3.13, a sharp increment in the probability of post RT recurrence beginning at the radiological stage of III at time of radiation. Similarly, in the FreeViz [9] result from Fig. 3.19, we saw the base vector indicating radiological stage at time of radiation to extend the farthest (i.e. most important projection to the classification decision) in the red region (i.e. post RT recurrence). However, it is equally intriguing to recognize that despite both being a measure of staging, the clinical stage at time of radiation was not nearly shown as significant, as illustrated in the rather uniform probability distribution of probability of recurrence given a clinical stage at time of radiation in Fig. 3.10b, as well as the insignificant difference portrayed by a Student's t test in Fig. 3.15b. In turn, this probes the question of the extent of a role radiological stage at time of radiation played in the "best pipeline" model built with the TPOT autoML approach described in section 3.1.2.5. Such that had this feature been eliminated, would the model still perform almost as well, as we would like to primarily understand how the patient-specific cell/nuclei distribution may impact the post RT recurrence outcome.

An interesting observation was that the cancerous cs mean (Fig. 3.12a) and non-tumoral cs mean (Fig. 3.12b) exhibited opposite trend for probability of post RT recurrence in the 5 to 7 microns range for cs distribution mean. Same for patient-specific cancerous cs distribution std in Fig. 3.12c and non-tumoral cs distribution std in Fig. 3.12d in the 2 to 3.6 miconrs range for cs distribution std.

Instances on the FreeViz 2D projection figure was grouped by proximity where the more connected the base vectors (features) are, the closer they locate on the 2D plane. Features more important to the classification had longer projection length and pointed to a more orthogonal direction to the line or curve separating the binary class cluster. A feature is more prominent at making the classification decision if it sits at a region mostly populated by instances of a specific class. [9]. Among imaging feature space, the patient-specific cancerous cs distribution mean plays the largest role in post RT recurrence in the red region of Fig. 3.19. The non-tumoral ns std base vector is approximately orthogonal to the class separation line and a cluster of SCC patients who did not experience post RT recurrence.

Moreover, target size influences the energy deposition spatial distribution. As the radiation sensitive volume decreases for any radiation quality, the no-hit peak becomes the dominant factor in the energy deposition spatial distribution, or microdosimetric spread [175]. This knowledge gives a potential explanation for the increased probability of recurrence of post RT recurrence at a smaller cancerous ns distribution mean, as shown noticeably in Fig. 3.11a. Similarly, Fig. 3.12c shows a monotonically increasing probability as the cancerous cs distribution std increases. A larger cancerous cs distribution std stipulates an increasing cell size variation for the cancerous region cell population. It can then also be theorized that the no-hit peak factor domination of the microdosimetric spread occurring to a volume of cells with a larger dispersion of cell size.

The microdosimetric spread magnitude enlarges with decreasing the target size, thus the variance of specific energy imparted per cell increases with decreasing cell size [31, 175]. This might offer insights to why in Fig. 3.12a between 5 to 7 microns range, a decreasing probability of post RT recurrence was shown as the cancerous cs distribution mean decreases. It might be due to an increased variance of specific energy with lowered cancerous cell size [114, 116]. In contrast, smaller non-tumoral cs distribution mean (4 - 10 microns) led to a higher probability of post **??** recurrence. This could potentially be explained by the same idea that the specific energy deposition increases with smaller target cell size. However, when the target belongs to a non-tumoral region, healthy tissues were therefore injured by the increase in microdosimetric spread whose biological response may have led to an increase in post RT recurrence.

It is especially curious that some non-tumoral imaging features were ranked and quantified as being more imperative to classifying the outcome of post RT recurrence. More specifically, patient-specific non-tumoral ns distribution std and the non-tumoral ns distribution mean were the most promising at prediction for the negative class of no post RT recurrence shown in FreeViz Fig. 3.19. This observation invites further investigation in cell subtype classification within the non-tumoral region as prognostic subtypes potential has been shown with,for example, the presence of tumor-infiltrating immune cells, cancer-associated fibroblasts, and vascular invasion, in relation to tumor microenvironment [176]. Concepts in relation to Normal Tissue Complication Probability (NTCP) Tissue Complication Probability (NTCP) might also be drawn depending on how exactly these non-tumoral features behave. Furthermore, as dimensionality reduction and visualization approaches, the FreeViz [8, 9] Orange application attested superior class separation visualization ability than PCA.

4.3 A TPOT autoML approach towards predictive modelling

Automated machine learning has not yet been commonly explored in the medical physics and radiation oncology domain. In this work, we explored a TPOT autoML approach to predict the post RT recurrence outcome given pre-treatment clinical and imaging (i.e. patient-specific cell / nuclei distribution data) features, and achieved reproducible results with excellent model performance (see evaluation metrics for best pipeline produced by TPOT autoML and replicated in Table 3.4). This validated its purpose to accelerate the ML pipeline construction and selection process. More critically, it emphasized the data-driven model building nature that is agnostic to any inherent bias towards the data itself, which also struck a good balance between agnostic modelling and adding domain expertise into the problem to optimize the model [51]. Most definitively, it provided at least a head start to the data pre-processing, feature engineering, ML model selection, and hyperparameter tuning process, which would be especially beneficial to researchers and clinicians who are not necessarily experienced with ML knowledge yet.

With our experimentation, TPOT autoML worked very well and relatively fast in under a few hours to deduce a best pipeline after 100 generations. However, based on the TPOT creators [48], since the number of generation (i.e. 100 in this analysis) was set by hand, a different number chosen could yield different "best pipelines" had we let it run longer. This is because genetic programming is used to allow TPOT to search through different operators (i.e. data preprocessors, ML algorithms, and hyperparameter choices). Therefore, a deeper and thorough comparison of TPOT best pipeline results yielded by various number of generations should be conducted to increase the confidence in trusting the "best pipeline" results, as well as investigating the top ranked (which were not the best) pipelines from each set generation number. A noticeable limitation is that the results of the model used in this thes were averaged internal CV scores over ten folds of stratified shuffle split cross validation on the entire patient cohort of thirty-six patients. In other words, all available data were used for training and there were no model performance evaluation on any test set. Though good averaged internal CV scores were shown in our results whose permutation test 3.21 appraised real dependency between the features and the outcome, the model must be tested in an unseen dataset in the future, preferable from an external institution, while following the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) criteria [177, 178].

Chapter 5

Conclusion

5.1 Conclusion

In this thesis a fully-automated software was developed to extract microscopic nuclei size cell spacing distribution from clinically used histopathology whole slide images. Given the extracted pre-treatment patient-specific cell and nuclei data as well as clinical variables, the outcome of post radiation therapy recurrence at the time of follow up was investigated. Radiological stage at time of radiation was shown to be the most important clinical feature to predict the post radiation recurrence outcome. Cell and nuclei target size (image) features from non-tumoral regions were of crucial importance to our clinical endpoint of interest. We also showed that predictive modeling with an automated machine learning system, Tree-Based Pipeline Optimization Tool, was able to utilize combined (cell spacing nuclei size) imaging and clinical features to predict post-radiation therapy recurrence outcomes for gynecological cancer patients, despite the small dataset and low event rate.

Bibliography

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015. Number: arXiv:1505.04597 arXiv:1505.04597 [cs].

[2] O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins, and L. Soler, “U-Net Transformer: Self and Cross Attention for Medical Image Segmentation,” in *MICCAI workshop MLMI*, (Strasbourg (virtuel), France), Sept. 2021.

[3] “Tissue Microarray (TMA) - Cancer and Normal Arrays | OriGene.”

[4] M. N. Gurcan, L. E. Boucheron, A. Can, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, “Histopathological Image Analysis: A Review,” *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, Oct. 2009. MAG ID: 2103243046.

[5] A. Goode, B. Gilbert, J. Harkes, D. Jukic, and M. Satyanarayanan, “OpenSlide: A vendor-neutral software foundation for digital pathology,” *Journal of Pathology Informatics*, vol. 4, p. 27, Sept. 2013.

[6] “openslide-python: Python interface to OpenSlide.”

[7] “Aperio format.”

[8] J. Demšar, T. Curk, A. Erjavec, Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar,

M. Žitnik, and B. Zupan, “Orange: Data Mining Toolbox in Python,” *Journal of Machine Learning Research*, vol. 14, no. 35, pp. 2349–2353, 2013.

[9] J. Demšar, G. Leban, and B. Zupan, “FreeViz—An intelligent multivariate visualization approach to explorative analysis of biomedical data,” *Journal of Biomedical Informatics*, vol. 40, pp. 661–671, Dec. 2007.

[10] R. Atun, D. A. Jaffray, M. B. Barton, F. Bray, M. Baumann, B. Vikram, T. P. Hanna, F. M. Knaul, Y. Lievens, T. Y. M. Lui, M. Milosevic, B. O’Sullivan, D. L. Rodin, E. Rosenblatt, J. Van Dyk, M. L. Yap, E. Zubizarreta, and M. Gospodarowicz, “Expanding global access to radiotherapy,” *The Lancet Oncology*, vol. 16, pp. 1153–1186, Sept. 2015.

[11] I. El Naqa and M. J. Murphy, eds., *Machine and Deep Learning in Oncology, Medical Physics and Radiology.* Cham: Springer International Publishing, 2022.

[12] I. El Naqa and S. Das, “The role of machine and deep learning in modern medical physics,” *Medical Physics*, vol. 47, no. 5, pp. e125–e126, 2020. _eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.14088.

[13] O. Morin, M. Vallières, S. Braunstein, J. B. Ginart, T. Upadhaya, H. C. Woodruff, A. Zwanenburg, A. Chatterjee, J. E. Villanueva-Meyer, G. Valdes, W. Chen, J. C. Hong, S. S. Yom, T. D. Solberg, S. Löck, J. Seuntjens, C. Park, and P. Lambin, “An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication,” *Nature Cancer*, vol. 2, pp. 709–722, July 2021. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Breast cancer;Cancer models;Cancer therapy;Lung cancer Subject_term_id: breast-cancer;cancer-models;cancer-therapy;lung-cancer.

[14] Y. Zhou, S.-C. Huang, J. A. Fries, A. Youssef, T. J. Amrhein, M. Chang, I. Banerjee, D. Rubin, L. Xing, N. Shah, and M. P. Lungren, "RadFusion: Benchmarking Performance and Fairness for Multimodal Pulmonary Embolism Detection from CT and EHR," *arXiv:2111.11665 [cs, eess]*, Nov. 2021. arXiv: 2111.11665 version: 1.

[15] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. Velázquez Vega, D. J. Brat, and L. A. D. Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences*, vol. 115, pp. E2970–E2979, Mar. 2018.

[16] R. J. Chen, M. Y. Lu, J. Wang, D. F. K. Williamson, S. J. Rodig, N. I. Lindeman, and F. Mahmood, "Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis," *arXiv:1912.08937 [cs, q-bio]*, Sept. 2020. arXiv: 1912.08937.

[17] R. J. Chen, M. Y. Lu, D. F. K. Williamson, T. Y. Chen, J. Lipkova, M. Shaban, M. Shady, M. Williams, B. Joo, Z. Noor, and F. Mahmood, "Pan-Cancer Integrative Histology-Genomic Analysis via Interpretable Multimodal Deep Learning," *arXiv:2108.02278 [cs, q-bio]*, Aug. 2021. arXiv: 2108.02278 version: 1.

[18] S.-J. Sammut, M. Crispin-Ortuzar, S.-F. Chin, E. Provenzano, H. A. Bardwell, W. Ma, W. Cope, A. Dariush, S.-J. Dawson, J. E. Abraham, J. Dunn, L. Hiller, J. Thomas, D. A. Cameron, J. M. S. Bartlett, L. Hayward, P. D. Pharoah, F. Markowetz, O. M. Rueda, H. M. Earl, and C. Caldas, "Multi-omic machine learning predictor of breast cancer therapy response," *Nature*, Dec. 2021.

[19] M. Crispin-Ortuzar, R. Woitek, E. Moore, M. Reinius, L. Beer, V. Bura, L. Rundo, C. McCague, S. Ursprung, L. E. Sanchez, P. Martin-Gonzalez, F. Mouliere, D. Chandrananda, J. Morris, T. Goranova, A. M. Piskorz, N. Singh, A. Sahdev, R. Pintican, M. Zerunian, H. Addley, M. Jimenez-Linan, F. Markowetz, E. Sala,

and J. D. Brenton, "Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer," preprint, Oncology, July 2021.

[20] K. M. Boehm, P. Khosravi, R. Vanguri, J. Gao, and S. P. Shah, "Harnessing multimodal data integration to advance precision oncology," *Nature Reviews Cancer*, vol. 22, pp. 114–126, Feb. 2022.

[21] L. Zhuang, J. Lipkova, R. Chen, and F. Mahmood, "Deep learning-based integration of histology, radiology, and genomics for improved survival prediction in glioma patients," in *Medical Imaging 2022: Digital and Computational Pathology*, vol. 12039, p. 120390Z, SPIE, Apr. 2022.

[22] X. Wang and B.-b. Li, "Deep Learning in Head and Neck Tumor Multiomics Diagnosis and Analysis: Review of the Literature," *Frontiers in Genetics*, vol. 12, p. 42, 2021.

[23] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pancancer prognosis prediction," *Bioinformatics*, vol. 35, pp. i446–i454, July 2019.

[24] H. Hermessi, O. Mourali, and E. Zagrouba, "Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain," *Neural Computing and Applications*, vol. 30, pp. 2029–2045, Oct. 2018.

[25] M. Kazmierski, M. Welch, S. Kim, C. McIntosh, P. M. Head, N. C. Group, K. Rey-McIntyre, S. H. Huang, T. Patel, T. Tadic, M. Milosevic, F.-F. Liu, A. Hope, S. Bratman, and B. Haibe-Kains, "A Machine Learning Challenge for Prognostic Modelling in Head and Neck Cancer Using Multi-modal Data," *arXiv:2101.11935 [cs, eess]*, Jan. 2021. arXiv: 2101.11935.

[26] K. M. Boehm, P. Khosravi, R. Vanguri, J. Gao, and S. P. Shah, "Harnessing multimodal data integration to advance precision oncology," *Nature Reviews Cancer*, vol. 22, pp. 114–126, Feb. 2022.

[27] S. K. Zhou, H. N. Le, K. Luu, H. V. Nguyen, and N. Ayache, "Deep reinforcement learning in medical imaging: A literature review," Mar. 2021. Number: arXiv:2103.05115 arXiv:2103.05115 [cs, eess].

[28] S. A. Enger, A. Ahnesjö, F. Verhaegen, and L. Beaulieu, "Dose to tissue medium or water cavities as surrogate for the dose to cell nuclei at brachytherapy photon energies," *Physics in Medicine and Biology*, vol. 57, pp. 4489–4500, June 2012. Publisher: IOP Publishing.

[29] N. Truong, M. Evans, N. Ybarra, and S. Enger, "Nuclei Size Distribution as a Predictor for Radiosensitivity with 192Ir Brachytherapy," *Brachytherapy*, vol. 16, p. S83, May 2017.

[30] B. Behmand, M. Evans, Y. Kamio, and S. Abbasinejad Enger, "OC-0048 Correlation between radiation-induced foci from 192Ir brachytherapy and tumor nuclei Size," *Radiotherapy and Oncology*, vol. 158, pp. S37–S38, May 2021.

[31] F. Villegas, N. Tilly, and A. Ahnesjö, "Monte Carlo calculated microdosimetric spread for cell nucleus-sized targets exposed to brachytherapy ^{125} I and ^{192} Ir sources and ^{60} Co cell irradiation," *Physics in Medicine and Biology*, vol. 58, pp. 6149–6162, Sept. 2013.

[32] J. M. DeCunha, C. M. Poole, M. Vallières, J. Torres, S. Camilleri-Broët, R. F. Rayes, J. D. Spicer, and S. A. Enger, "Development of patient-specific 3D models from histopathological samples for applications in radiation therapy," *Physica Medica*, vol. 81, pp. 162–169, Jan. 2021.

[33] J. M. DeCunha, F. Villegas, M. Vallières, J. Torres, S. Camilleri-Broët, and S. A. Enger, "Patient-specific microdosimetry: a proof of concept," *Physics in Medicine & Biology*, vol. 66, p. 185011, Sept. 2021.

[34] J. DeCunha, *Microdosimetric evaluation of photon emitting brachytherapy sources in tissue-specific models*. PhD thesis, McGill University, 2021. Publisher: McGill University.

[35] F. Shamout, T. Zhu, and D. A. Clifton, "Machine Learning for Clinical Outcome Prediction," *IEEE reviews in biomedical engineering*, vol. 14, pp. 116–126, 2021.

[36] I. Chamseddine, Y. Kim, B. De, I. El Naqa, D. G. Duda, J. Wolfgang, J. Pursley, H. Paganetti, J. Wo, T. Hong, E. J. Koay, and C. Grassberger, "Predictive Modeling of Survival and Toxicity in Patients With Hepatocellular Carcinoma After Radiotherapy," *JCO Clinical Cancer Informatics*, p. e2100169, Feb. 2022.

[37] Y. Cho, Y. Kim, I. Chamseddine, W. H. Lee, H. R. Kim, I. J. Lee, M. H. Hong, B. C. Cho, C. G. Lee, S. Cho, J. S. Kim, H. I. Yoon, and C. Grassberger, "Lymphocyte dynamics during and after chemo-radiation correlate to dose and outcome in stage III NSCLC patients undergoing maintenance immunotherapy," *Radiotherapy and Oncology*, vol. 168, pp. 1–7, Mar. 2022.

[38] T. M. Deist, F. J. W. M. Dankers, G. Valdes, R. Wijsman, I.-C. Hsu, C. Oberije, T. Lustberg, J. van Soest, F. Hoebers, A. Jochems, I. El Naqa, L. Wee, O. Morin, D. R. Raleigh, W. Bots, J. H. Kaanders, J. Belderbos, M. Kwint, T. Solberg, R. Monshouwer, J. Bussink, A. Dekker, and P. Lambin, "Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers," *Medical Physics*, vol. 45, pp. 3449–3459, July 2018.

[39] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, Jan. 2015.

[40] K. Wakabayashi, Y. Koide, T. Aoyama, H. Shimizu, R. Miyauchi, H. Tanaka, H. Tachibana, K. Nakamura, and T. Kodaira, “A predictive model for pain response following radiotherapy for treatment of spinal metastases,” *Scientific Reports*, vol. 11, p. 12908, Dec. 2021.

[41] H. Naseri, S. Skamene, M. Tolba, M. D. Faye, P. Ramia, J. Khriguian, H. Patrick, A. X. Andrade, M. David, and J. Kildea, “A radiomics-based machine learning pipeline to distinguish between metastatic and healthy bone using lesion-center-based geometric regions of interest,” preprint, In Review, Mar. 2022.

[42] T. F. Sterkenburg and P. D. Grünwald, “The no-free-lunch theorems of supervised learning,” *Synthese*, vol. 199, pp. 9979–10015, Dec. 2021.

[43] X. He, K. Zhao, and X. Chu, “AutoML: A survey of the state-of-the-art,” *Knowledge-Based Systems*, vol. 212, p. 106622, Jan. 2021.

[44] S. K. K. Santu, M. M. Hassan, M. J. Smith, L. Xu, C. Zhai, and K. Veeramachaneni, “AutoML to Date and Beyond: Challenges and Opportunities,” May 2021. Number: arXiv:2010.10777 arXiv:2010.10777 [cs].

[45] L. Hou, K. Singh, D. Samaras, T. M. Kurc, Y. Gao, R. J. Seidman, and J. H. Saltz, “Automatic histopathology image analysis with CNNs,” in *2016 New York Scientific Data Summit (NYSDS)*, pp. 1–6, Aug. 2016.

[46] H. Eldeeb, M. Maher, O. Matsuk, A. Aldallal, R. Elshawi, and S. Sakr, “AutoMLBench: A Comprehensive Experimental Evaluation of Automated Machine Learning Frameworks,” Apr. 2022. Number: arXiv:2204.08358 arXiv:2204.08358 [cs].

[47] T. Perk, T. Bradshaw, S. Chen, H.-J. Im, S. Cho, S. Perlman, G. Liu, and R. Jeraj, “Automated classification of benign and malignant lesions in 18F-NaF PET/CT images

using machine learning," *Physics in Medicine and Biology*, vol. 63, p. 225019, Nov. 2018.

[48] R. S. Olson and J. H. Moore, "TPOT: A Tree-based Pipeline Optimization Tool for Automating Machine Learning," in *Proceedings of the Workshop on Automatic Machine Learning*, pp. 66–74, PMLR, Dec. 2016. ISSN: 1938-7228.

[49] R. S. Olson and J. H. Moore, "TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning," in *Automated Machine Learning* (F. Hutter, L. Kotthoff, and J. Vanschoren, eds.), pp. 151–160, Cham: Springer International Publishing, 2019. Series Title: The Springer Series on Challenges in Machine Learning.

[50] J. D. Romano, T. T. Le, W. Fu, and J. H. Moore, "TPOT-NN: augmenting tree-based automated machine learning with neural network estimators," *Genetic Programming and Evolvable Machines*, vol. 22, pp. 207–227, June 2021.

[51] J. Dafflon, W. H. L. Pinaya, F. Turkheimer, J. H. Cole, R. Leech, M. A. Harris, S. R. Cox, H. C. Whalley, A. M. McIntosh, and P. J. Hellyer, "An automated machine learning approach to predict brain age from cortical anatomical measures," *Human Brain Mapping*, vol. 41, pp. 3555–3566, Sept. 2020.

[52] Y. Zou, L. Weishaupt, and S. Enger, "SP-0014 McMedHacks: Deep learning for medical image analysis workshops and Hackathon in radiation oncology," *Radiotherapy and Oncology*, vol. 170, pp. S4–S5, May 2022.

[53] Y. Zou, M. Lecavalier-Barsoum, M. Pelmus, F. Maleki, and S. Enger, "Predictive modeling of post radiation-therapy recurrence for gynecological cancer patients using clinical and histopathology imaging features," (Montreal, QC, Canada), May 2022.

[54] Y. Zou, M. Lecavalier-barsoum, M. Pelmus, and S. Enger, "Patient-Specific Nuclei Size and Cell Spacing Distribution Extraction From Histopathology Whole Slide Images

for Treatment Outcome Prediction Modelling," in *MEDICAL PHYSICS*, vol. 49, pp. E266–E266, WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2022. Issue: 6.

[55] J. M. DeCunha, C. M. Poole, M. Vallières, J. Torres, S. Camilleri-Broët, R. F. Rayes, J. D. Spicer, and S. A. Enger, "Development of patient-specific 3D models from histopathological samples for applications in radiation therapy," *Physica Medica*, vol. 81, pp. 162–169, Jan. 2021.

[56] S. W. Jahn, M. Plass, and F. Moinfar, "Digital Pathology: Advantages, Limitations and Emerging Perspectives," *Journal of Clinical Medicine*, vol. 9, p. 3697, Nov. 2020.

[57] "Steps to Tissue Processing for Histopathology."

[58] Vanessa, "Sample and slide preparation - Histology," July 2019.

[59] D. K. Tucker, J. F. Foley, S. A. Hayes-Bouknight, and S. E. Fenton, "Preparation of high quality hematoxylin and eosin-stained sections from rodent mammary gland whole mounts for histopathologic review," *Toxicologic pathology*, vol. 44, pp. 1059–1064, Oct. 2016.

[60] J. P. Hinton, K. Dvorak, E. Roberts, W. J. French, J. C. Grubbs, A. E. Cress, H. A. Tiwari, and R. B. Nagle, "A Method to Reuse Archived H&E Stained Histology Slides for a Multiplex Protein Biomarker Analysis," *Methods and Protocols*, vol. 2, p. 86, Nov. 2019.

[61] F. Aeffner, M. D. Zarella, N. Buchbinder, M. M. Bui, M. R. Goodman, D. J. Hartman, G. M. Lujan, M. A. Molani, A. V. Parwani, K. Lillard, O. C. Turner, V. N. Vemuri, A. G. Yuil-Valdes, and D. Bowman, "Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association," *Journal of Pathology Informatics*, vol. 10, p. 9, Jan. 2019.

[62] N. M. Jawhar, “Tissue Microarray: A rapidly evolving diagnostic and research tool,” *Annals of Saudi Medicine*, vol. 29, no. 2, pp. 123–127, 2009.

[63] C. Le Page, A.-M. Mes-Masson, and A. M. Magliocco, “Tissue Microarrays in Studying Gynecological Cancers,” in *Cancer Genomics*, pp. 65–76, Elsevier, 2014.

[64] S. Kalra, H. Tizhoosh, C. Choi, S. Shah, P. Diamandis, C. J. Campbell, and L. Pantanowitz, “Yottixel – An Image Search Engine for Large Archives of Histopathology Whole Slide Images,” *Medical Image Analysis*, vol. 65, p. 101757, Oct. 2020.

[65] A. Serag, A. Ion-Margineanu, H. Qureshi, R. McMillan, M.-J. Saint Martin, J. Diamond, P. O’Reilly, and P. Hamilton, “Translational AI and Deep Learning in Diagnostic Pathology,” *Frontiers in Medicine*, vol. 6, p. 185, 2019.

[66] C. Srinidhi, “Self-Supervised Driven Consistency Training for Annotation Efficient Histopathology Image Analysis,” Oct. 2021. original-date: 2021-02-04T20:14:05Z.

[67] A. Baardwijk, G. Bosmans, L. Boersma, J. Buijsen, R. Wanders, M. Hochstenbag, R.-J. Suylen, A. Dekker, C. Oberije, R. Houben, S. Bentzen, M. Kroonenburgh, P. Lambin, and D. Ruysscher, “PET-CT–Based Auto-Contouring in Non–Small-Cell Lung Cancer Correlates With Pathology and Reduces Interobserver Variability in the Delineation of the Primary Tumor and Involved Nodal Volumes,” *International journal of radiation oncology, biology, physics*, vol. 68, pp. 771–8, July 2007.

[68] A. G. Berman, W. R. Orchard, M. Gehrung, and F. Markowetz, “PathML: A unified framework for whole-slide image analysis with deep learning,” preprint, Pathology, July 2021.

[69] L. L. Weishaupt, J. Torres, S. Camilleri-Broët, R. F. Rayes, J. D. Spicer, S. C. Maldonado, and S. A. Enger, “Deep learning-based tumor segmentation on digital

images of histopathology slides for microdosimetry applications," preprint, In Review, Feb. 2021.

[70] M. Aubreville, C. A. Bertram, T. A. Donovan, C. Marzahl, A. Maier, and R. Klopfleisch, "A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research," *Scientific Data*, vol. 7, p. 417, Dec. 2020.

[71] G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Kone, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, and P. Aguiar, "BACH: Grand challenge on breast cancer histology images," *Medical Image Analysis*, vol. 56, pp. 122–139, Aug. 2019.

[72] S. Sornapudi, R. Addanki, R. J. Stanley, W. V. Stoecker, R. Long, R. Zuna, S. R. Frazier, and S. Antani, "Cervical Whole Slide Histology Image Analysis Toolbox," preprint, Pathology, July 2020.

[73] S. Hemati, S. Kalra, C. Meaney, M. Babaie, and H. R. Tizhoosh, "CNN and Deep Sets for End-to-End Whole Slide Image Representation Learning," p. 11.

[74] L. Hou, R. Gupta, J. S. Van Arnam, Y. Zhang, K. Sivalenka, D. Samaras, T. M. Kurc, and J. H. Saltz, "Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of ten cancer types," *Scientific Data*, vol. 7, p. 185, June 2020. Bandiera_abtest: a Cc_license_type: cc_publicdomain Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer imaging;Data publication and archiving;Image processing;Microscopy Subject_term_id: cancer-imaging;data-publication-and-archiving;image-processing;microscopy.

[75] A. Pedersen, M. Valla, A. M. Bofin, J. P. de Frutos, I. Reinertsen, and E. Smistad, "FastPathology: An open-source platform for deep learning-based research and decision support in digital pathology," *arXiv:2011.06033 [cs, eess]*, Nov. 2020. arXiv: 2011.06033.

[76] S. Kuntz, E. Krieghoff-Henning, J. N. Kather, T. Jutzi, J. Höhn, L. Kiehl, A. Hekler, E. Alwers, C. von Kalle, S. Fröhling, J. S. Utikal, H. Brenner, M. Hoffmeister, and T. J. Brinker, "Gastrointestinal cancer classification and prognostication from histology using deep learning: Systematic review," *European Journal of Cancer*, vol. 155, pp. 200–215, Sept. 2021.

[77] J. A. Diao, J. K. Wang, W. F. Chui, V. Mountain, S. C. Gullapally, R. Srinivasan, R. N. Mitchell, B. Glass, S. Hoffman, S. K. Rao, C. Maheshwari, A. Lahiri, A. Prakash, R. McLoughlin, J. K. Kerner, M. B. Resnick, M. C. Montalto, A. Khosla, I. N. Wapinski, A. H. Beck, H. L. Elliott, and A. Taylor-Weiner, "Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes," *Nature Communications*, vol. 12, p. 1613, Dec. 2021.

[78] B.-I. Song, "A machine learning-based radiomics model for the prediction of axillary lymph-node metastasis in breast cancer," *Breast Cancer*, vol. 28, no. 3, pp. 664–671, 2021. MAG ID: 3126105087 S2ID: 140fac7ac2deeeb89256b9a273849ef84ddf78e9.

[79] Z. Song, S. Zou, W. Zhou, Y. Huang, L. Shao, J. Yuan, X. Gou, W. Jin, Z. Wang, X. Chen, X. Ding, J. Liu, C. Yu, C. Ku, C. Liu, Z. Sun, G. Xu, Y. Wang, X. Zhang, D. Wang, S. Wang, W. Xu, R. C. Davis, and H. Shi, "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nature Communications*, vol. 11, p. 4294, Aug. 2020. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1

Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Gastric cancer;Machine learning Subject_term_id: gastric-cancer;machine-learning.

[80] R. Gupta, T. Kurc, A. Sharma, J. S. Almeida, and J. Saltz, "The Emergence of Pathomics," *Current Pathobiology Reports*, vol. 7, pp. 73–84, Sept. 2019.

[81] J. H. Oh, W. Choi, E. Ko, M. Kang, A. Tannenbaum, and J. O. Deasy, "PathCNN: interpretable convolutional neural networks for survival prediction and pathway analysis applied to glioblastoma," *Bioinformatics*, vol. 37, pp. i443–i450, July 2021.

[82] J. Yang, R. Shi, and B. Ni, "MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 191–195, Apr. 2021. ISSN: 1945-8452.

[83] F. Yang, "Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer," Nov. 2021. original-date: 2020-02-23T12:59:33Z.

[84] R. Cao, F. Yang, S.-C. Ma, L. Liu, Y. Zhao, Y. Li, D.-H. Wu, T. Wang, W.-J. Lu, W.-J. Cai, H.-B. Zhu, X.-J. Guo, Y.-W. Lu, J.-J. Kuang, W.-J. Huan, W.-M. Tang, K. Huang, J. Huang, J. Yao, and Z.-Y. Dong, "Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer," *Theranostics*, vol. 10, no. 24, pp. 11080–11091, 2020.

[85] O. Ortashi and M. Al Kalbani, "Gynecological Cancer Services in Arab Countries: Present Scenario, Problems and Suggested Solutions," *Asian Pacific Journal of Cancer Prevention*, vol. 14, no. 3, pp. 2147–2150, 2013. Publisher: Asian Pacific Journal of Cancer Prevention.

[86] "Global burden of gynecological cancers | LIAQUAT MEDICAL RESEARCH JOURNAL," Dec. 2021.

[87] B. P. Whitcomb, "Gynecologic malignancies," *The Surgical Clinics of North America*, vol. 88, pp. 301–317, vi, Apr. 2008.

[88] G. Suneja and A. Viswanathan, "Gynecologic Malignancies," *Hematology/Oncology Clinics of North America*, vol. 34, pp. 71–89, Feb. 2020.

[89] E.-J. Yang, J.-H. Lee, A.-J. Lee, N.-R. Kim, Y.-T. Ouh, M.-K. Kim, S.-H. Shim, S.-J. Lee, T.-J. Kim, and K.-A. So, "Multiple Primary Malignancies in Patients with Gynecologic Cancer," *Journal of Clinical Medicine*, vol. 11, p. 115, Dec. 2021.

[90] A. N. Viswanathan, L. J. Lee, J. R. Eswara, N. S. Horowitz, P. A. Konstantinopoulos, K. L. Mirabeau-Beale, B. S. Rose, A. G. von Keudell, and J. Y. Wo, "Complications of pelvic radiation in patients treated for gynecologic malignancies," *Cancer*, vol. 120, no. 24, pp. 3870–3883, 2014. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cncr.28849.

[91] D. M. Haines and K. H. West, "Immunohistochemistry: Forging the links between immunology and pathology," *Veterinary Immunology and Immunopathology*, vol. 108, pp. 151–156, Oct. 2005.

[92] A.-A. Grosset, K. Loayza-Vega, Adam-Granger, M. Birlea, B. Gilks, B. Nguyen, G. Soucy, D. Tran-Thanh, R. Albadine, and D. Trudel, "Hematoxylin and Eosin Counterstaining Protocol for Immunohistochemistry Interpretation and Diagnosis," *Applied Immunohistochemistry & Molecular Morphology*, vol. 27, pp. 558–563, Aug. 2019.

[93] E. Kostopoulou, M. Samara, P. Kollia, K. Zacharouli, I. Mademtzis, A. Daponte, I. E. Messinis, and G. Koukoulis, "Different patterns of p16 immunoreactivity in cervical biopsies: correlation to lesion grade and HPV detection, with a review of the

literature," *European Journal of Gynaecological Oncology*, vol. 32, no. 1, pp. 54–61, 2011.

[94] C. J. O'Neill and W. G. McCluggage, "p16 Expression in the Female Genital Tract and Its Value in Diagnosis," *Advances in Anatomic Pathology*, vol. 13, pp. 8–15, Jan. 2006.

[95] S. da Mata, J. Ferreira, I. Nicolás, S. Esteves, G. Esteves, S. Lérias, F. Silva, A. Saco, D. Cochicho, M. Cunha, M. del Pino, J. Ordi, and A. Félix, "P16 and HPV Genotype Significance in HPV-Associated Cervical Cancer—A Large Cohort of Two Tertiary Referral Centers," *International Journal of Molecular Sciences*, vol. 22, p. 2294, Jan. 2021. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[96] J. Yang, A. Elliott, A. L. Hoffa, N. Herring, and P. M. Houser, "Potential influence of p16 immunohistochemical staining on the diagnosis of squamous cell lesions in cervical biopsy specimens: observation from cytologic-histologic correlation," *Cancer Cytopathology*, vol. 126, no. 12, pp. 1003–1010, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cncy.22063.

[97] W. H. Liggett and D. Sidransky, "Role of the p16 tumor suppressor gene in cancer.," *Journal of Clinical Oncology*, vol. 16, pp. 1197–1206, Mar. 1998. Publisher: Wolters Kluwer.

[98] D. M. O'Connor, "The Use of p16 as a Marker for Cervical Cancer Precursor Lesions," *Journal of Lower Genital Tract Disease*, vol. 11, pp. 67–68, Apr. 2007.

[99] R. H. Tambouret, "Use of immunohistochemical staining for p16 in gynecological cytology," *Cancer Cytopathology*, vol. 124, no. 9, pp. 611–612, 2016. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cncy.21760.

[100] C. L. Rasmussen, H. K. Bertoli, F. L. Sand, A. K. Kjær, L. T. Thomsen, and S. K. Kjær, “The prognostic significance of HPV, p16, and p53 protein expression in vaginal cancer: A systematic review,” *Acta Obstetricia et Gynecologica Scandinavica*, vol. 100, no. 12, pp. 2144–2156, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/aogs.14260.

[101] L. J. Lee, B. E. Howitt, E. Oliva, H. Zhang, P. J. Catalano, C. Crum, P. Bu, N. Cimbak, R. Demaria, R. Murphy, N. S. Horowitz, U. A. Matulonis, and A. L. Russo, “Prognostic importance of p16 status for women with vulvar squamous cell carcinoma (SCC) treated with radiotherapy.,” *Journal of Clinical Oncology*, vol. 35, pp. 5599–5599, May 2017. Publisher: Wolters Kluwer.

[102] J. K. Schwarz, J. S. Lewis, J. Pfeifer, P. Huettner, and P. Grigsby, “Prognostic Significance of p16 Expression in Advanced Cervical Cancer Treated With Definitive Radiotherapy,” *International Journal of Radiation Oncology, Biology, Physics*, vol. 84, pp. 153–157, Sept. 2012. Publisher: Elsevier.

[103] V. M. Feldbaum, L. C. Flowers, and G. M. Oprea-Ilies, “Improved Survival in p16-Positive Vaginal Cancers Across All Tumor Stages but No Correlation With MIB-1,” *American Journal of Clinical Pathology*, vol. 142, pp. 664–669, Nov. 2014.

[104] M. J. Dohopolski, Z. D. Horne, D. Pradhan, R. Bhargava, R. P. Edwards, J. L. Kelley, J. T. Comerci, A. B. Olawaiye, M. Courtney-Brooks, J. L. Berger, P. Sukumvanich, and S. Beriwal, “The Prognostic Significance of p16 Status in Patients With Vulvar Cancer Treated With Vulvectomy and Adjuvant Radiation,” *International Journal of Radiation Oncology, Biology, Physics*, vol. 103, pp. 152–160, Jan. 2019. Publisher: Elsevier.

[105] E. B. Podgoršak, "Kinetics of Radioactive Decay," in *Radiation Physics for Medical Physicists* (E. B. Podgorsak, ed.), Biological and Medical Physics, Biomedical Engineering, pp. 451–473, Berlin, Heidelberg: Springer, 2010.

[106] J. Cadet, T. Douki, and J.-L. Ravanat, "Oxidatively Generated Damage to the Guanine Moiety of DNA: Mechanistic Aspects and Formation in Cells," *Accounts of Chemical Research*, vol. 41, pp. 1075–1083, Aug. 2008. Publisher: American Chemical Society.

[107] G. Scholes, J. F. Ward, and J. Weiss, "Mechanism of the radiation-induced degradation of nucleic acids," *Journal of Molecular Biology*, vol. 2, pp. 379–391, Dec. 1960.

[108] E. B. Podgoršak, "Interactions of Neutrons with Matter," in *Radiation Physics for Medical Physicists* (E. B. Podgorsak, ed.), Biological and Medical Physics, Biomedical Engineering, pp. 429–449, Berlin, Heidelberg: Springer, 2010.

[109] E. B. Podgorsak, *Radiation Physics for Medical Physicists.* Biological and Medical Physics, Biomedical Engineering, Berlin, Heidelberg: Springer, 2010.

[110] A. M. Kellerer and D. Chmelevsky, "Concepts of microdosimetry II. Probability distributions of the microdosimetric variables," *Radiation and Environmental Biophysics*, vol. 12, pp. 205–216, Oct. 1975.

[111] A. Kellerer, "A survey of microdosimetric quantities and concepts," *Ultramicroscopy*, vol. 14, pp. 169–174, Jan. 1984.

[112] A. A. Jansen, "Icru Report 33: Radiation Quantities and Units," *Journal of Nuclear Medicine Technology*, vol. 8, pp. 241–241, Dec. 1980. Publisher: Society of Nuclear Medicine Section: JNMT Bookshelf.

[113] G. Gruel, C. Villagrasa, P. Voisin, I. Clairand, M. Benderitter, J.-F. Bottollier-Depois, and J. F. Barquinero, "Cell to Cell Variability of Radiation-Induced Foci: Relation

between Observed Damage and Energy Deposition," *PLoS ONE*, vol. 11, p. e0145786, Jan. 2016.

[114] G. A. Santa Cruz, "Microdosimetry: Principles and applications," *Reports of Practical Oncology & Radiotherapy*, vol. 21, pp. 135–139, Mar. 2016. Publisher: VM Media SP. zo.o VM Group SK.

[115] E. Kouhkan, N. Chegeni, and A. Hussain, "The Effect of Nucleus Size on the Cell Dose in Targeted Radionuclide Therapy – A Monte Carlo Study," *Journal of Medical Signals and Sensors*, vol. 10, pp. 113–118, Apr. 2020.

[116] H. Griffiths, "Microdosimetry. ICRU Report no. 36," *Radiology*, vol. 154, pp. 528–528, Feb. 1985. Publisher: Radiological Society of North America.

[117] R. Nafe, K. Franz, W. Schlote, and B. Schneider, "Morphology of tumor cell nuclei is significantly related with survival time of patients with glioblastomas.," *Clinical Cancer Research*, vol. 11, pp. 2141–2148, Mar. 2005. MAG ID: 2130819084 S2ID: 9540d726628c8c45f7f240555ffdfb006ec7b927.

[118] C. Takahashi, R. Shridhar, J. Huston, and K. Meredith, "Correlation of tumor size and survival in pancreatic cancer," *Journal of Gastrointestinal Oncology*, vol. 9, pp. 910–921, Oct. 2018.

[119] T. Burzykowski, E. Coart, E. D. Saad, Q. Shi, D. W. Sommeijer, C. Bokemeyer, E. Díaz-Rubio, J.-Y. Douillard, A. Falcone, C. S. Fuchs, R. M. Goldberg, J. R. Hecht, P. M. Hoff, H. Hurwitz, F. F. Kabbinavar, M. Koopman, T. S. Maughan, C. J. A. Punt, L. Saltz, H.-J. Schmoll, M. T. Seymour, N. C. Tebbutt, C. Tournigand, E. Van Cutsem, A. de Gramont, J. R. Zalcberg, and M. Buyse, "Evaluation of Continuous Tumor-Size–Based End Points as Surrogates for Overall Survival in Randomized Clinical Trials

in Metastatic Colorectal Cancer," *JAMA Network Open*, vol. 2, p. e1911750, Sept. 2019.

[120] V. Erdoğu, N. Çitak, C. B. Sezen, L. Cansever, C. Aker, S. Onay, M. V. Doğru, Saydam, M. A. Bedirhan, and M. Metin, "Correlation between outcomes and tumor size in >7cm T4 non-small cell lung cancer patients: A tumor size-based comparison study," *Asian Cardiovascular and Thoracic Annals*, vol. 29, pp. 784–791, Oct. 2021. Publisher: SAGE Publications Ltd STM.

[121] "PII: 0012-365X(72)90093-3 | Elsevier Enhanced Reader." ISSN: 0012-365X.

[122] J. Sudbø, R. Marcelpoil, and A. Reith, "New algorithms based on the Voronoi Diagram applied in a pilot study on normal mucosa and carcinomas," *Analytical Cellular Pathology: The Journal of the European Society for Analytical Cellular Pathology*, vol. 21, no. 2, pp. 71–86, 2000.

[123] C. Poole, A. Ahnesjö, and S. Enger, "Determination of subcellular compartment sizes for estimating dose variations in radiotherapy," *Radiation protection dosimetry*, vol. 166, May 2015.

[124] B. Boots, A. Okabe, and K. Sugihara, "Spatial tessellations," vol. 1, pp. 503–526, 1999.

[125] J. Sudbø, A. Bankfalvi, M. Bryne, R. Marcelpoil, M. Boysen, J. Piffko, J. Hemmer, K. Kraft, and A. Reith, "Prognostic Value of Graph Theory-Based Tissue Architecture Analysis in Carcinomas of the Tongue," *Laboratory Investigation*, vol. 80, pp. 1881–1889, Dec. 2000.

[126] "Rank — Orange Visual Programming 3 documentation."

[127] A. Mustafa, Mostafa Rahimi Azghadi, Mostafa Rahimi Azghadi, and M. R. Azghadi, "Automated machine learning for healthcare and clinical notes analysis,"

The first computers, vol. 10, no. 2, p. 24, 2021. MAG ID: 3130376580 S2ID: 8de9635bc852df3da4e5b0407606976a5d2da78b.

[128] N. Erickson, J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, and A. Smola, "AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data," Mar. 2020. Number: arXiv:2003.06505 arXiv:2003.06505 [cs, stat].

[129] D. Laadan, R. Vainshtein, Y. Curiel, G. Katz, and L. Rokach, "MetaTPOT: Enhancing A Tree-based Pipeline Optimization Tool Using Meta-Learning," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, (Virtual Event Ireland), pp. 2097–2100, ACM, Oct. 2020.

[130] "ImageScope Annotations - Tutorial."

[131] CHOOSEHAPPY, "Working with Aperio SVS files in Matlab – Converting Annotations to Binary Masks," Jan. 2015. Section: Digital Histology.

[132] "Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, 2nd Edition | Wiley."

[133] A. Saribudak, Y. Dong, S. Gundry, J. Hsieh, and M. Uyar, "Mathematical models of tumor growth using Voronoi tessellations in pathology slides of kidney cancer," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4454–4457, Aug. 2015. ISSN: 1558-4615.

[134] M. Bock, A. K. Tyagi, J.-U. Kreft, and W. Alt, "Generalized Voronoi Tessellation as a Model of Two-dimensional Cell Tissue Dynamics," *Bulletin of Mathematical Biology*, vol. 72, pp. 1696–1731, Oct. 2010.

[135] Z. Lu and K.-H. Yuan, "Welch's t test," pp. 1620–1623, Jan. 2010.

[136] J. de Winter, "Using the Student's t-test with extremely small sample sizes," *Practical Assessment, Research & Evaluation*, vol. 18, Aug. 2013.

[137] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, vol. 374, p. 20150202, Apr. 2016.

[138] L. Parmentier, O. Nicol, L. Jourdan, and M.-E. Kessaci, "TPOT-SH: A Faster Optimization Algorithm to Solve the AutoML Problem on Large Datasets," *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 471–478, Nov. 2019. MAG ID: 3006528928 S2ID: 62fe8d0955ca639c56e6de22553116d1566d765b.

[139] X. Su, N. Chen, H. Sun, Y. Liu, X. Yang, W. Wang, S. Zhang, Q. Tan, J. Su, Q. Gong, and Q. Yue, "Automated machine learning based on radiomics features predicts H3 K27M mutation in midline gliomas of the brain.," *Neuro-oncology*, vol. 22, pp. 393–401, Sept. 2019. MAG ID: 2975172371 S2ID: a168be21ffee2143e7e4931e3361fb2a45cfeddd.

[140] K. Mohan, "Predicting Heart Disease Using Machine Learning Classification Algorithms And Along With TPOT (AUTOML)," 2020. S2ID: 6008d08596a6753677cca36b0fd44c655f4e06f8.

[141] Phani Siginamsetty and V. K. Reddy, "Machine Learning Classifiers and Along with TPOT Classifier (Automl) to Predict the Readmission Patterns of Diabetic Patients," *International journal of recent technology and engineering*. MAG ID: 4240912358 S2ID: 6d8ac650c9b09b30e9c3c70969e7b43241f457b7.

[142] P. Wanlin, Tian-Jing Zhang, K. Shi, Hai-Xia Li, Ying Li, S. He, C. Li, Dong Xia, C. Xia, and Z.-L. Li, "Automatic machine learning based on native T1 mapping can identify myocardial fibrosis in patients with hypertrophic

cardiomyopathy," *European Radiology*, pp. 1–10, Sept. 2021. MAG ID: 3196452975 S2ID: a7fd9ae86d706c1ffc8c99beaa47133d92af390d.

[143] Tammo P.A. Beishuizen, J. Vanschoren, Joaquin Vanschoren, P. A. J. Hilbers, and D. Bošnački, "Automated Feature Selection and Classification for High-Dimensional Biomedical Data," Oct. 2021. MAG ID: 3206803568 S2ID: 5f6cf017edf3a0d4b1f3ebabf60eb6ff4cb5157a.

[144] N. S. M. Haniff, M. K. B. A. Karim, N. S. Ali, M. A. A. Rahman, N. H. Osman, and M. I. Saripan, "Magnetic Resonance Imaging Radiomics Analysis for Predicting Hepatocellular Carcinoma," in *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, pp. 1–5, July 2021.

[145] A. Orlenko, D. Kofink, L.-P. Lyytikäinen, K. Nikus, P. Mishra, P. Kuukasjärvi, P. J. Karhunen, M. Kähönen, J. O. Laurikka, T. Lehtimäki, F. W. Asselbergs, and J. H. Moore, "Model selection for metabolomics: predicting diagnosis of coronary artery disease using automated machine learning," *Bioinformatics*, vol. 36, pp. 1772–1778, Mar. 2020.

[146] P. Yang, H. Zhang, X. Lai, K. Wang, Q. Yang, and D. Yu, "Accelerating the Selection of Covalent Organic Frameworks with Automated Machine Learning," *ACS Omega*, vol. 6, pp. 17149–17161, July 2021.

[147] F. Maleki, K. Ovens, R. Gupta, C. Reinhold, A. Spatz, and R. Forghani, "Generalizability of Machine Learning Models: Quantitative Evaluation of Three Methodological Pitfalls," p. 13.

[148] F. Maleki, N. Muthukrishnan, K. Ovens, C. Reinhold, and R. Forghani, "Machine Learning Algorithm Validation," *Neuroimaging Clinics of North America*, vol. 30, pp. 433–445, Nov. 2020.

[149] F. Maleki, K. Ovens, K. Najafian, B. Forghani, C. Reinhold, and R. Forghani, "Overview of Machine Learning Part 1," *Neuroimaging Clinics of North America*, vol. 30, pp. e17–e32, Nov. 2020.

[150] S. Kaliman, C. Jayachandran, F. Rehfeldt, and A.-S. Smith, "Limits of Applicability of the Voronoi Tessellation Determined by Centers of Cell Nuclei to Epithelium Morphology," *Frontiers in Physiology*, vol. 7, 2016.

[151] A. Zhang, L. Xing, J. Zou, and J. C. Wu, "Shifting machine learning for healthcare from development to deployment and from models to data," *Nature Biomedical Engineering*, pp. 1–16, July 2022. Publisher: Nature Publishing Group.

[152] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in Medical Imaging: A Survey," Jan. 2022. Number: arXiv:2201.09873 arXiv:2201.09873 [cs, eess].

[153] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.*

[154] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into Deep Learning," p. 1018.

[155] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," June 2021. Number: arXiv:2010.11929 arXiv:2010.11929 [cs].

[156] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *arXiv:1706.03762 [cs]*, Dec. 2017. arXiv: 1706.03762.

[157] C. Nguyen, Z. Asad, and Y. Huo, "Evaluating Transformer-based Semantic Segmentation Networks for Pathological Image Segmentation," Sept. 2021. Number: arXiv:2108.11993 arXiv:2108.11993 [cs, eess].

[158] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," Feb. 2021. Number: arXiv:2102.04306 arXiv:2102.04306 [cs].

[159] Z. Shao, H. Bian, Y. Chen, Y. Wang, J. Zhang, X. Ji, and Y. Zhang, "TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification," p. 15.

[160] M. Y. Lu, D. F. K. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature biomedical engineering*, vol. 5, pp. 555–570, June 2021.

[161] R. Krishnan, P. Rajpurkar, and E. J. Topol, "Self-supervised learning in medicine and healthcare," *Nature Biomedical Engineering*, pp. 1–7, Aug. 2022. Publisher: Nature Publishing Group.

[162] Q. Chen, D. R. Cherry, V. Nalawade, E. M. Qiao, A. Kumar, Abhishek Kumar, Abhishek Kumar, A. M. Lowy, D. R. Simpson, and J. D. Murphy, "Clinical Data Prediction Model to Identify Patients With Early-Stage Pancreatic Cancer.," *JCO clinical cancer informatics*, vol. 5, no. 5, pp. 279–287, 2021. MAG ID: 3136362669 S2ID: d611b69ddde00db95d9658ec8e4c7c6d4a9bfe1e.

[163] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Medical Image Analysis*, vol. 67, p. 101813, Jan. 2021. arXiv: 1912.12378.

[164] F. C. Ghesu, B. Georgescu, A. Mansoor, Y. Yoo, D. Neumann, P. Patel, R. S. Vishwanath, J. M. Balter, Y. Cao, S. Grbic, and D. Comaniciu, "Self-supervised Learning from 100 Million Medical Images," Jan. 2022. Number: arXiv:2201.01283 arXiv:2201.01283 [cs].

[165] N. A. Koohbanani, B. Unnikrishnan, S. A. Khurram, P. Krishnaswamy, and N. Rajpoot, “Self-Path: Self-Supervision for Classification of Pathology Images With Limited Annotations,” *IEEE Transactions on Medical Imaging*, vol. 40, pp. 2845–2856, Oct. 2021. Conference Name: IEEE Transactions on Medical Imaging.

[166] O. Ciga, T. Xu, and A. Martel, “Self supervised contrastive learning for digital histopathology,” *Machine Learning with Applications*, vol. 7, p. 100198, Nov. 2021.

[167] M. Assran, M. Caron, I. Misra, P. Bojanowski, F. Bordes, P. Vincent, A. Joulin, M. Rabbat, and N. Ballas, “Masked Siamese Networks for Label-Efficient Learning,” Apr. 2022. Number: arXiv:2204.07141 arXiv:2204.07141 [cs, eess].

[168] T. Truong, S. Mohammadi, and M. Lenga, “How Transferable are Self-supervised Features in Medical Image Classification Tasks?,” in *Proceedings of Machine Learning for Health*, pp. 54–74, PMLR, Nov. 2021. ISSN: 2640-3498.

[169] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen, V. Natarajan, and M. Norouzi, “Big Self-Supervised Models Advance Medical Image Classification,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, (Montreal, QC, Canada), pp. 3458–3468, IEEE, Oct. 2021.

[170] P. Stanitsas, A. Cherian, V. Morellas, R. Tejpaul, N. Papanikolopoulos, and A. Truskinovsky, “Image Descriptors for Weakly Annotated Histopathological Breast Cancer Data,” *Frontiers in Digital Health*, vol. 2, p. 37, 2020.

[171] G. Xu, Z. Song, Z. Sun, C. Ku, Z. Yang, C. Liu, S. Wang, J. Ma, and W. Xu, “CAMEL: A Weakly Supervised Learning Framework for Histopathology Image Segmentation,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, (Seoul, Korea (South)), pp. 10681–10690, IEEE, Oct. 2019.

[172] J. Pocock, S. Graham, Q. D. Vu, M. Jahanifar, S. Deshpande, G. Hadjigeorghiou, A. Shephard, R. M. Saad Bashir, M. Bilal, W. Lu, D. Epstein, F. Minhas, N. M. Rajpoot, and S. E. Ahmed Raza, "TIAToolbox: An End-to-End Toolbox for Advanced Tissue Image Analytics," preprint, Bioengineering, Dec. 2021.

[173] C. Lin, L. Zhang, L. Mao, M. Wu, and D. Hu, "Separable-HoverNet and Instance-YOLO for Colon Nuclei Identification and Counting," Mar. 2022. Number: arXiv:2203.00262 arXiv:2203.00262 [cs, eess].

[174] Chu-Fang Lin, Liukun Zhang, Lijian Mao, Min Wu, and Dong Hu, "Separable-HoverNet and Instance-YOLO for Colon Nuclei Identification and Counting," *ArXiv*, 2022. ARXIV_ID: 2203.00262 S2ID: 12c8b6d51d831a19a5d6cfaec3c5d0e212842855.

[175] F. Villegas, N. Tilly, and A. Ahnesjö, "Target Size Variation in Microdosimetric Distributions and its Impact on the Linear-Quadratic Parameterization of Cell Survival," *Radiation Research*, vol. 190, Aug. 2018.

[176] Y. Yuan, "Spatial Heterogeneity in the Tumor Microenvironment," *Cold Spring Harbor Perspectives in Medicine*, vol. 6, p. a026583, Aug. 2016.

[177] "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement | The EQUATOR Network."

[178] G. S. Collins, P. Dhiman, C. L. Andaur Navarro, J. Ma, L. Hooft, J. B. Reitsma, P. Logullo, A. L. Beam, L. Peng, B. Van Calster, M. van Smeden, R. D. Riley, and K. G. Moons, "Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence," *BMJ Open*, vol. 11, p. e048008, July 2021.