REPUBLIC OF TURKEY

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY

INSTITUTE OF NATURAL AND APPLIED SCIENCES

AUTOMATIC SEMANTIC SEGMENTATION OF ORGANS-AT-RISK AND TARGET TUMOR VOLUME IN RADIOTHERAPY PLANNING CT IMAGES OF NASOPHARYNGEAL CANCER

Murat YÜCE

DEPARTMENT OF BIOMEDICAL ENGINEERING

SUPERVISOR

Lect. Phd. Seda Nilgün DUMLU

CO-SUPERVISOR

Assoc. Prof. Sinem Burcu ERDOĞAN

ISTANBUL - 2024

DECLARATION

I declare that the information reported in this thesis is the result of my own work, except where explicitly stated otherwise in the text. I am fully aware of all the ethical rules and there are not any violations of copyright and any kind of patent.

Murat YÜCE

ACKNOWLEDGEMENT

In the light of science, for a brighter future, I sincerely thank my respected professors, Lect. Phd. Seda Nilgün Dumlu and Associate Professor Sinem Burcu Erdoğan, who did not withhold their help and contributions by sparing their valuable time for this research which is conducted with determination and perseverance.

I express my gratitude to Professor Dr. Enis Özyar, who made the necessary conditions available for this research to be realized, and to his colleagues for their valuable efforts and time.

I thank my esteemed professors, Associate Professor Bükem Tanören and Professor Dr. Ata Akın, for providing all kinds of support to the students enrolled in our university's Master's programs and for their contributions to this study.

Lastly, I am grateful to my beloved family who raised me and my precious spouse Dr. Dilek Yüce for supporting me in my academic journey.

With the wish to enrich the legacy left to us by Mustafa Kemal ATATÜRK, who said, "My spiritual legacy is science and reason"...

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

MRI Magnetic Resonance Imaging

PET Positron Emission Tomography

ROI Regions of Interest

NPC Nasopharyngeal Carcinoma

OAR Organs-at-Risk

DFS 5-year Disease Free Survival

CT Computed Tomography

ASR Age-Standardized Rate

CNN Convolutional Neural Network

ASSD Average Symmetric Surface Distance

GTV Gross Tumor Volume

KNN K-Nearest Neighbors

SVM Support Vector Machine

XGBoost Extreme Gradient Boosting

SGD Stochastic Gradient Descent

GLCM Gray Level Co-Occurrence Matrix

HU Hounsfield Unit

DL Deep Learning

IMRT Intensity Modulated Radiation Therapy

ARC Arc Modulated Radiation Therapy

AUC Area Under the Curve

ROC Receiver Operating Characteristics

CTV Clinical Tumor Volume

IOV Interobserver Variability

NN Neural Network

NLP Natural Language Processing

ReLU Rectified Linear Unit

ViT Vision Transformer Network

DSC Dice Similarity Coefficient

SOM Self-Organizing Map

DBSCAN Density-Based Spatial Clustering Of Applications With Noise

SLIC Simple Linear Iterative Clustering

LoG Laplacian of Gaussian

MSRM Maximal Similarity Based Region Merging

LIST OF FIGURES

LIST OF TABLES

SUMMARY

Aim: The aim of this study is to develop automatic segmentation workflows for target tumor tissues in radiotherapy planning CT images of nasopharyngeal carcinoma (NPC) patients using U-Net, UNETR, and SwinUNETR models, and for organs-at-risk (OAR) using U-Net models. Furthermore, by employing feature extraction strategies (Radiomics and DINOv2), the study evaluates the predictive performances of extracted features both from manual and automated segmentations in combination with clinical features on 5-year Disease-Free Survival (DFS).

Materials and Methods: Enhanced and non-enhanced CT images from 120 NPC patients who received radiotherapy at Istanbul Acıbadem Maslak Hospital between 2009 and 2021collected. For OAR segmentation task all of the 120 CT series included and only models with 3D-UNET architecture developed, and for NPC segmentation task only the enhanced series (77 in total) included and UNETR, SwinUNETR vision transformer based models also developed alongside with 3D-UNET. CT series were resampled into same voxel dimensions and normalized before training. Models were trained using a 5-fold cross-validation method, with performance measured using Dice (DSC), Precision, Recall, Accuracy, and F1 values. Feature extraction utilized 2D slices from contrast-enhanced 75 CT volumes. A sum of 1059 distinct radiomic features using the "pyradiomics" Python library and 386 DINOv2 features using the DINOv2-small model extracted. Clinical features such as gender, age at diagnosis, and NPC stage were also included in the DFS prediction analysis. Lasso used for feature selection before training prediction models. Machine learning models trained with extracted featured from manual segmentations and predictions made both on features extracted from manual and automated segmentations. Prediction results compared to each other statistically.

Results: The models developed for the OAR task exhibited an average DSC score of 0.81 (std=0.11). The lowest DSC score was achieved in the optic chiasm (0.47), while the highest DSC score was obtained in the brain (0.99). Among the models developed for NPC segmentation, the highest DSC score of 0.64 was achieved with the SwinUNETR architecture. For DFS prediction, feature selection with Lasso resulted in the selection of 48 features from the combination of radiomics and clinical

feature set and 77 feature from the combination of DINOv2 and clinical feature set. mong the developed machine learning models, the best performance (F1=0.57) was achieved using the K-Nearest Neighbors model with combination of radiomics and clinical feature set, while the best performance (F1=0.58) was achieved using the Logistic Regression model with combination of DINOv2 and clinical feature set. There was no significant difference in the prediction performances of the trained models based on features extracted from manual and automatic segmentations ($p>0.05$).

Conclusion: Our study showcases the efficacy of cutting-edge deep learning architectures like U-Net, UNETR, and SwinUNETR in radiotherapy planning for nasopharyngeal cancer (NPC). It highlights the superiority of vision transformers in segmenting tumors using contrast-enhanced CT images, as evidenced by higher DSC scores. Despite having some limitations in the dataset, the study emphasizes the importance of data quality on the model performance in terms of accuracy and generalizability. Additionally, combining radiomics or DINOv2 features with clinical data to predict NPC outcomes suggests a promising avenue in precision oncology by enabling more tailored therapeutic approaches, improving prognosis estimation, and optimizing treatment protocols. As one of the novel and striking results of our study, this finding challenges the conventional assumption that manual segmentations invariably offer a superior basis for feature extraction, suggesting that features extracted from the segmentations of an advanced automatic segmentation technique can achieve comparable predictive performance. Overall, our research underscores the transformative potential of advanced deep learning and integration of radiomics and clinical features in refining radiotherapy planning and prognostic modeling for NPC, contributing to the broader goal of personalized medicine and potentially improving treatment outcomes and quality of life for patients.

ÖZET

Amaç: Bu çalışmanın amacı, nazofarengeal karsinom (NK) hastalarının radyoterapi planlaması CT görüntülerinde hedef tümör dokuları için 3D U-Net, UNETR ve SwinUNETR mimarileri, risk altındaki organlar (RAO) için ise 3D U-Net mimarisi kullanarak otomatik segmentasyon yöntemlerini uygulamak ve otomatize bir iş akışı geliştirmektir. Ayrıca, özellik çıkarma stratejileri (Radyomiks ve DINOv2) kullanarak, hem manuel hem de otomatik segmentasyonlardan çıkarılan özelliklerin klinik özelliklerle birleştirilerek 5 yıllık Hastalıksız Sağkalım (HS) üzerindeki tahmini performanslarını değerlendirmeyi amaçladık.

Materyal ve Metotlar: 2009 ve 2021 yılları arasında İstanbul Acıbadem Maslak Hastanesi'nde radyoterapi alan 120 nazofarenks kanseri hastasına ait kontrastlı ve kontrastsız bilgisayarlı tomografi (BT) görüntüleri toplandı. Risk Altındaki Organların segmentasyon görevi için 120 BT serisi dahil edildi ve yalnızca 3D-UNET mimarisiyle modeller geliştirildi; NK segmentasyon görevi için ise yalnızca kontrastlı seriler (toplamda 77) dahil edildi ve UNETR, SwinUNETR görüş dönüştürücüsü (vision transformer) tabanlı modeller ile 3D-UNET mimarisinde modeller geliştirildi. BT serileri, eğitimden önce aynı voksel boyutlarına örneklendi ve normalize edildi. Modeller, 5 kat çapraz doğrulama yöntemi kullanılarak eğitildi ve performans Dice (DSC), Hassasiyet, Duyarlılık, Doğruluk ve F1 değerleri kullanılarak ölçüldü. 5 yıllık hastalıksız sağ kalım (HS) tahmini için özellik çıkarma işlemi, kontrastlı 75 BT hacminden 2D dilimler kullanılarak gerçekleştirildi. "pyradiomics" Python kütüphanesi kullanılarak 1059 farklı radyomik özellik ve DINOv2-küçük modeli kullanılarak 386 DINOv2 özelliği çıkarıldı. Cinsiyet, tanı anındaki yaş ve NK evresi gibi klinik özellikler de HS tahmin analizine dahil edildi. Özellik seçimi için eğitimden önce Lasso kullanıldı. Manuel segmentasyonlardan çıkarılan özelliklerle eğitilen makine öğrenimi modelleri kullanılarak, hem manuel hem de otomatik segmentasyonlardan çıkarılan özellikler üzerinde tahminlerde bulunuldu. Tahmin sonuçları istatistiksel olarak birbirleriyle karşılaştırıldı.

Bulgular: RAO segmentasyon görevi için geliştirilen modellerde 0.81 (standart sapma=0.11) düzeyinde DSC skoru elde edildi. En düşük DSC skoru optik kiazmada (0.47) elde edilirken, en yüksek DSC skoru beyinde (0.99) elde edildi. NK

segmentasyonu için geliştirilen modeller arasında, en yüksek DSC skoru 0.64 ile SwinUNETR mimarisi ile elde edildi. DFS tahmini için, Lasso ile özellik seçimi yapıldı ve Radyomiks ile Klinik özellik kombinasyon setinden 48 özellik ve DINOv2 ile Klinik özellik kombinasyon setinden 77 özellik seçildi. Geliştirilen makine öğrenimi modelleri arasında, en iyi performanslar Radyomiks ile Klinik özellik kombinasyon setinde KNN modeli kullanılarak (F1=0.57), DINOv2 ile Klinik özellik kombinasyon seti ile lojistik regresyon modeli kullanılarak elde edilen en iyi performans (F1=0.58) ile elde edildi. Manuel ve otomatik segmentasyonlardan çıkarılan özelliklere dayanan eğitimli modellerin tahmin performansları arasında anlamlı bir fark bulunmadı (p>0.05).

Sonuç: Çalışmamız, nazofarengeal kanser (NK) için radyoterapi planlamasında U-Net, UNETR ve SwinUNETR gibi son teknoloji derin öğrenme mimarilerinin etkinliğini sergilemektedir. Kontrastlı BT görüntüleri kullanılarak tümörleri segmente etmede görü dönüştürücü yapılarının üstünlüğünü ve veri setindeki sınırlılıklara rağmen, çalışma model performansı üzerinde veri kalitesinin önemini geliştirilen modellerin doğruluğu ve genelleştirilebilirliği üzerinden vurgulamaktayız. Ayrıca, her ne kadar mevcut verilerle elde edilen sonuçlar doğrudan klinik uygulamalar için yeterli olmasada, radyomiks veya DINOv2 özelliklerinin klinik verilerle birleştirilerek NK hastalarının radyoterapi sonrası 5 yıllık sağ kalımlarının tahmin edilmesi hasta düzeyinde özelleştirilmiş onkolojik tedavi yöntemlerinin geliştirilebilmesi adına umut verici bir yol göstermektedir. Çalışmamızın yeni ve çarpıcı sonuçlarından biri olarak, manuel segmentasyonların her zaman daha üstün bir özellik çıkarma temeli sunduğu varsayımı sorgulanmış ve gelişmiş bir otomatik segmentasyon tekniği ile yapılan segmentasyonlardan çıkarılan özelliklerin benzer öngörü performansına ulaşabileceği gösterilmiştir. Genel olarak, araştırmamız, NK için radyoterapi planlamasını ve prognostik modellemeyi geliştirme konusunda derin öğrenme ve radyomiks özellikler ile klinik özelliklerin entegrasyonunun potansiyel uygulama alanlarını vurgulamakta ve kişiselleştirilmiş tıp yöntemlerine katkı sunmaktadır.

1. INTRODUCTION

Nasopharyngeal cancer (NPC) is an uncommon type of tumor, accounting for 87,000 new cases and making up 0.6% of all cancer types (1). It predominantly affects males, with a sex ratio of 2.3 males for every female (1). Nasopharyngeal carcinoma is a type of cancer that starts in the epithelial cells lining the nasopharynx. This tumor typically appears in the pharyngeal recess (also known as the fossa of Rosenmüller) within the nasopharynx. Although it arises from cell or tissue types similar to those of other epithelial head and neck cancers, nasopharyngeal carcinoma is notably different from these tumors (2). This cancer shows a significant geographic disparity in incidence rates, with the highest occurrence observed in South-Eastern Asia, where the age-standardized rate (ASR) is 6.4 for men and 2.4 for women (1). While it is a relatively rare cancer type, its mortality rates have decreased over the years due to the improvement in treatment outcomes. Radiotherapy is one of the first line preferred treatment strategy for NPC. Early stage (5-year) survival rates for NPC patients receiving radiotherapy could reach 80% (3). One of the important factors that determines the success of radiotherapy is accurate segmentation of target tissues by radiation oncologists, in order to precisely deliver radiation to tumor tissues while protecting surrounding organs. The manual segmentation process for radiotherapy in the head and neck region can take approximately 3 hours and represents a significant workload for radiation oncologists (4). In addition, the accuracy of tissue segmentation may vary according to the experience of the radiation oncologist (5). Therefore, there is a need for an automated, objective and accurate way to perform segmentations in NPC patients for radiotherapy planning.

Deep convolutional neural networks (CNNs) are currently widely used in medical image segmentation. The U-Net, a deep CNN developed by Ronneberger et al. in 2015, has enabled the development of automatic segmentation models with relatively few data in medical images (6). Studies in the literature have shown that successful models in the segmentation of tumor tissue and organs-at-risk for radiotherapy in brain, breast and rectal cancers can be developed with a range of acceptable dice (DSC) scores (7–10). Further with the continuous advancements in the field of computer vision, two novel visual transformer models, UNETR(11) and SwinUNETR(12) , have emerged as promising alternatives to the traditional U-Net architecture. UNETR, an extension of the transformer model originally designed for natural language processing tasks, introduces a self-attention mechanism to capture long-range dependencies in medical images. This innovation allows UNETR to effectively capture contextual information across the entire image, enabling more accurate and context-aware segmentation results compared to the conventional U-Net.

SwinUNETR, on the other hand, builds upon the success of Swin Transformer, a hierarchical vision transformer architecture introduced by Liu et al(13). The SwinUNETR model leverages the unique characteristics of the Swin Transformer, such as the shifted window partitioning strategy and hierarchical feature representation, to enhance the performance of medical image segmentation tasks. The shifted window approach reduces the computational complexity of self-attention mechanisms, making it more efficient for large medical images, while the hierarchical feature representation facilitates the extraction of multi-scale features, improving the model's ability to capture intricate details in medical images.

These innovations in UNETR and SwinUNETR address some of the limitations of the baseline U-Net model, particularly in handling contextual information and efficiently processing large-scale medical images. By incorporating transformer architectures into the medical image segmentation domain, these models showcase the potential for improved segmentation accuracy and generalization across different medical imaging tasks. As research in this area continues to progress, the exploration of transformer-based architectures opens new possibilities for advancing the state-of-the-art in medical image analysis and segmentation.

Feature extraction from medical images (especially using radiomics) and predicting relevant outcomes using extracted features in combination with clinical features (such as biochemical markers or demographic information) became a predominant focus of recent studies in the field of medical imaging. This trend underscores the increasing recognition of the synergistic value of integrating radiological and clinical data to enhance diagnostic accuracy, prognostic assessments, and therapeutic decision-making(14). By leveraging the complementary strengths of image-based biomarkers and clinical indicators, researchers and clinicians are better equipped to understand the complex interplay between anatomical, functional, and biological factors in disease processes. Consequently, this integrated approach facilitates the development of more personalized and precise medical interventions, aligning with the overarching goals of precision medicine.

Based on literature information, we aimed to develop an automatic segmentation infrastructure for target tumor tissues in radiotherapy planning CT images of NPC patients using U-Net, UNETR and SwinUNETR models, and by using feature extraction strategies, we evaluated the predictive performances of these segmentations, in combination with clinical features, on 5-year Disease Free Survival (DFS). Our objective is to enhance the accuracy and efficiency of the automated segmentation process for target tumor tissues in the radiotherapy planning CT images of NPC patients and to determine predicting performance of extracted features in 5-year DFS. In addition to the traditional U-Net model, we are incorporating the innovative UNETR and SwinUNETR models, both of which leverage transformer architectures, to explore the potential improvements in segmentation performance. To the best of our knowledge, there are no documented instances of applying UNETR and SwinUNETR models for segmenting CT images in patients with NPC. Also we tried to reproduce segmentation performance of U-Net architecture on organs-at-risk segmentation task.

In the following, nasopharyngeal cancer related literature is reviewed in chapter 2; especially segmentation and feature extraction related evidences are examined. In chapter 3, the segmentation and feature extraction methodology are explained in detail. In chapter 4, the findings from developed segmentation models and extracted features are given. In chapter 5, interpretation on the findings is discussed and future work is explained.

2. BACKGROUND AND LITERATURE REVIEW

2.1. Epidemiology of Nasopharyngeal Carcinoma

Nasopharyngeal cancer (NPC) represents a distinct type of head and neck cancer, characterized by its occurrence in the nasopharynx, the area located behind the nose and above the back of the throat. Although NPC accounts for approximately 0.6% of head and neck cancers globally, its incidence is notably higher in specific regions, particularly in South-Eastern Asia (**Figure 1**), where the annual incidence rates can reach up to 7 cases per 100,000 individuals (1). This geographical variance in incidence rates suggests the influence of genetic, environmental, and possibly viral factors (notably the Epstein-Barr virus) in the development of NPC (15).

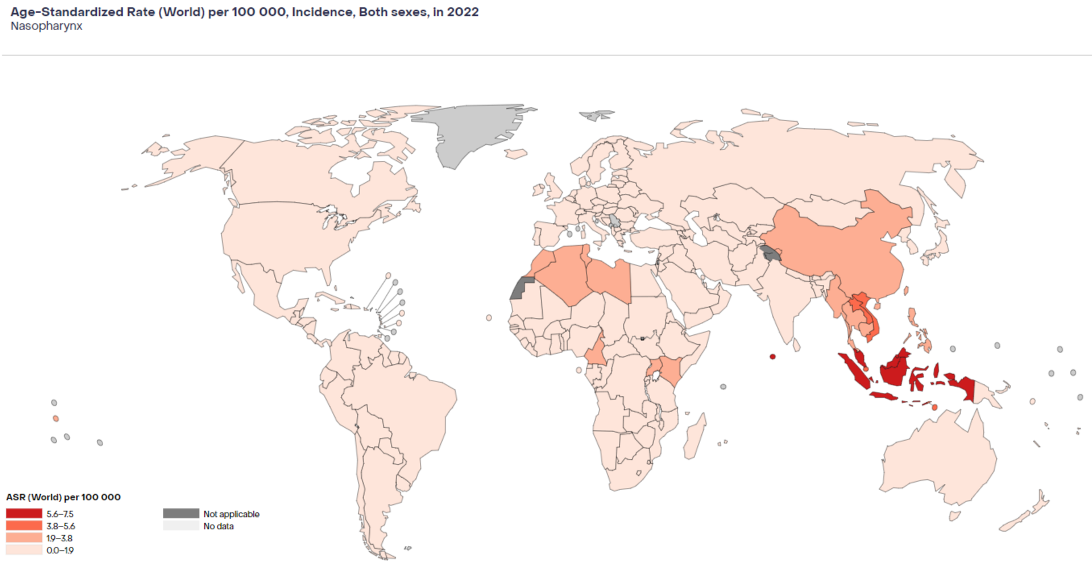

Figure 1. Age-Standardized incidence rate of NPC per 100 000 (World Graphic, 2022 data). Created with using https://gco.iarc.who.int/today web page of WHO International Agency for Research on Cancer.

Despite being considered relatively rare compared to other cancer types, the mortality rates associated with NPC have seen a decline over the years (16,17). This positive trend is largely attributed to advancements in treatment approaches, with radiotherapy emerging as a cornerstone in the management of NPC. The effectiveness of radiotherapy is underscored by the potential to achieve 5-year survival rates of up to 80% in patients diagnosed at early stages of the disease(18). Such outcomes highlight the significant strides made in improving the prognosis for individuals affected by NPC through advancements in therapeutic strategies and technologies.

The success of radiotherapy, however, is contingent upon the precise segmentation of target tissues. This critical process involves delineating the tumor and surrounding tissues to ensure the accurate delivery of radiation doses. The goal is to maximize the therapeutic effects on the tumor while minimizing exposure and potential damage to adjacent healthy organs and structures. The manual segmentation process, typically undertaken by radiation oncologists, is both time-consuming and subject to variability based on the clinician’s experience (19). Consequently, there is an ongoing need for more automated, objective, and accurate methods for tissue segmentation in the context of NPC treatment planning. This backdrop sets the stage for exploring innovative approaches and technologies that can enhance the accuracy and efficiency of radiotherapy for NPC, ultimately aiming to improve patient outcomes and quality of life.

2.2. Challenges in Radiotherapy for Nasopharyngeal Carcinoma

The treatment of nasopharyngeal cancer (NPC) with radiotherapy, while effective, is fraught with significant challenges that impact both the efficacy of the treatment and the quality of life for patients. These challenges stem from the anatomical complexity of the nasopharynx (**Figure 2**), the precision required in targeting tumor tissues, and the limitations of current manual segmentation processes. In a recent comprehensive review, Ng et. al. (2022) stated that strategies in accurately delineating tumor targets and organs at risk (OAR) are critical for achieving effective loco-regional control and minimizing radiation-associated complications (20). Authors also stressed the importance of high-resolution imaging and international consensus guidelines in refining target delineation.

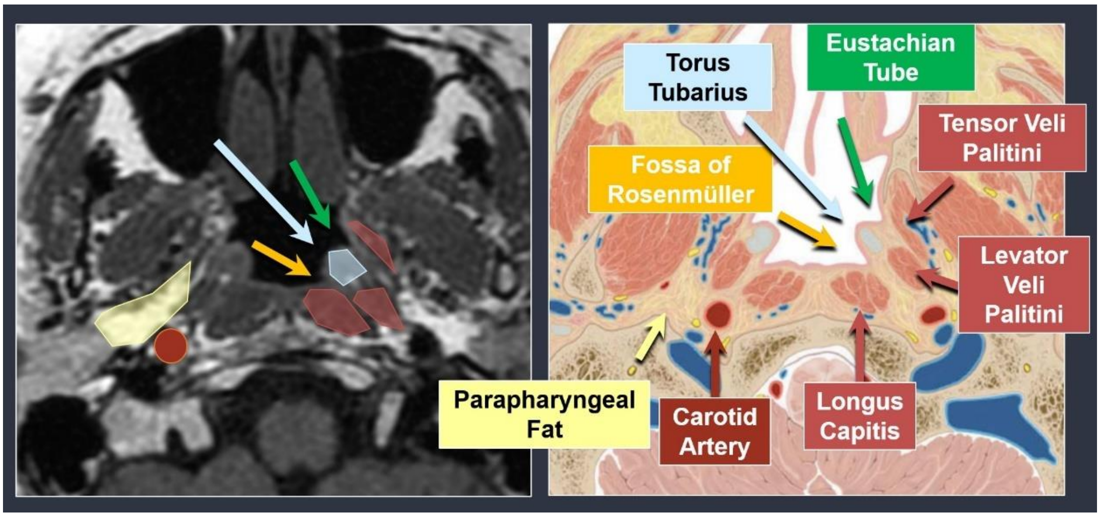

Figure 2. Cross sectional anatomy of nasopharynx. Left: T1 weighted MRI section, Right: Illustrative painting of same MRI section. From the "Head and Neuroanatomy" section of the "THIEME Atlas of Anatomy" authored by Scheuenke et. al. (21)

Manual segmentation is a time consuming and important step while planning radiotherapy. Susceptibility of this method to observer variability is another challenge for treatment planning and effectivity of the radiotherapy. Although it is considered as most accurate approach, these limitations underscore the necessity for auto-segmentation algorithms (22). Auto-segmentation technology evolved from basic intensity analysis and shape modeling to the more advanced machine learning approaches, including deep learning techniques like convolutional neural networks (CNNs) and specifically the U-Net architecture for medical image segmentation. Deep learning, particularly with CNNs and U-Net (**Figure 3**) architectures, represents the forefront of auto-segmentation technology, capable of learning complex patterns in medical images to accurately delineate target volumes and organs at risk with minimal human input. Despite the advances, challenges remain, including standardization of contouring protocols, adaptation to variations in imaging practices, and gaining clinicians' trust in automated systems (22).

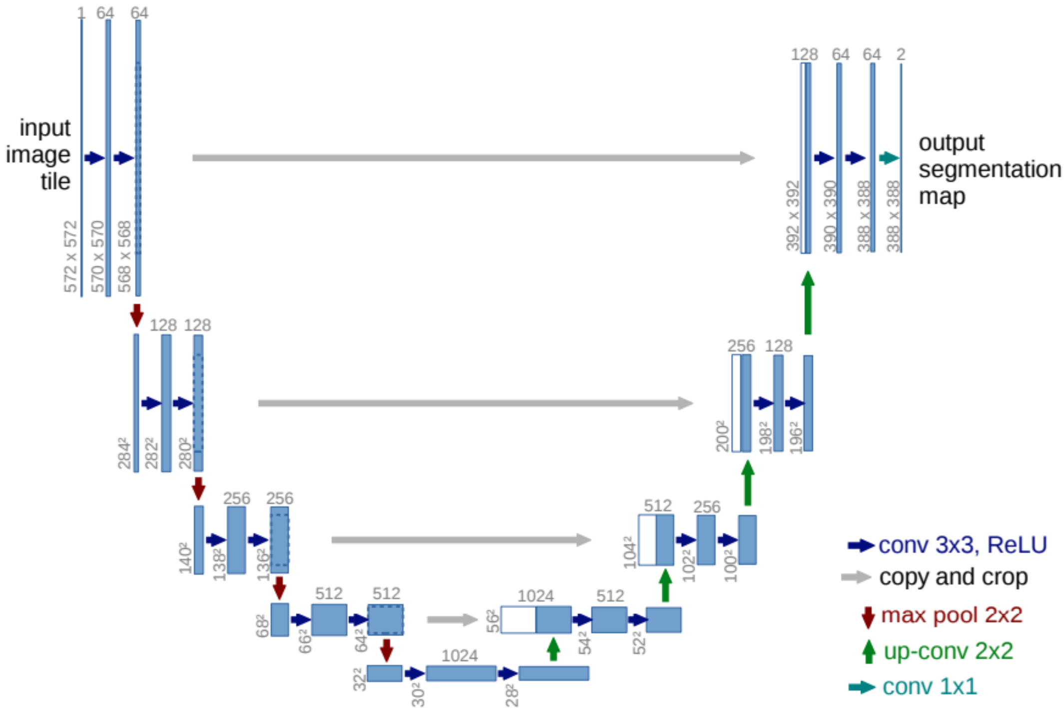

Figure 3. The U-net structure, illustrated with a minimum resolution of 32x32 pixels, features blue boxes that symbolize multi-channel feature maps, with the channel count specified above each box. The dimensions in the x and y directions are indicated at the bottom left corner of each box. White boxes are used to show feature maps that have been duplicated, and various operations are represented by arrows. From the article “U-Net: Convolutional Networks for Biomedical Image Segmentation” authored by Ronneberger et. al. (6).

2.3. Semantic Segmentation Techniques in Medical Imaging

Semantic segmentation in medical imaging is a critical process used for accurately delineating various structures within medical images, such as CT scans, MRI images, and X-rays. This technique involves classifying each pixel in an image into a specific category, making it invaluable for tasks like tumor detection, organ segmentation, and understanding anatomical structures (**Figure 4**). The methods for achieving semantic segmentation can be broadly classified into manual, semi-automated, and automated techniques, each with its own set of advantages, limitations, and applications. Before the widespread adoption of artificial intelligence applications in medical imaging, clinicians were largely dependent on manual segmentation techniques. Manual segmentation is performed using various medical image processing software and utilizes image processing techniques such as thresholding, histogram normalization, and edge detection during the application, these applications can also be considered as semi-automated segmentation (23). However, considering the volume of medical images, the careful preparation of a single case can consume a significant amount of the clinicians' already valuable time.

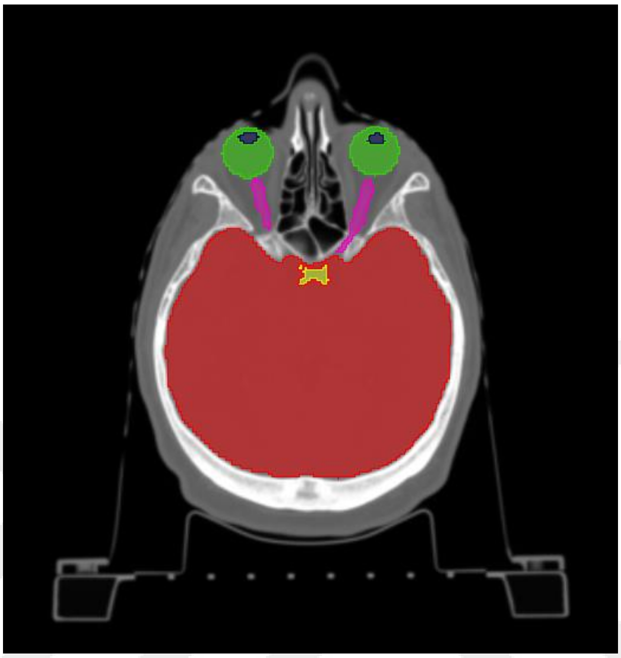

Figure 4. Semantic segmentation of a CT cross section. Red: Brain, Green: Eye Bulbs, Purple: Optic Nerves, Yellow: Optic Chiasm, Blue: Optic Lenses.

Herein, some of the key techniques and advancements are discussed in semantic segmentation specifically tailored for medical imaging.

2.3.1. Semi-Automated Segmentation Techniques

Semi-automated segmentation techniques are methods used in image processing and analysis, particularly in the context of medical imaging or any field where precise delineation of structures within an image is required. These techniques combine automated algorithms with human input to accurately segment, or partition, images into regions of interest. This approach helps to overcome the limitations of fully automated methods, which may struggle with variability in image quality, noise, and complex or ambiguous structures.

Semi-automated segmentation techniques are particularly valuable in applications where precision and reliability are crucial, and where variability across images makes fully automated segmentation challenging. These methods leverage the strengths of both humans and machines, offering a balance between automation efficiency and the nuanced judgment of human experts.

Thresholding with User Input:

Thresholding is a simple, yet effective, image processing technique used to separate objects from the background or to segment different regions of an image based on pixel intensity levels. It converts a grayscale image into a binary image, where each pixel is assigned to one of two categories (typically black or white) based on whether its intensity is higher or lower than a specified threshold value. This method is widely used in various applications, including image segmentation, object detection, and pre-processing steps in more complex image analysis workflows. There are several types of thresholding methods, including global (or simple) thresholding, adaptive thresholding and Otsu's method (24) (**Figure 5**). Even with different types of application methods, the choice of threshold value(s) is crucial for the success of the segmentation, and in many cases, this requires careful selection or the use of automatic thresholding methods that can adapt to the specific characteristics of the image.

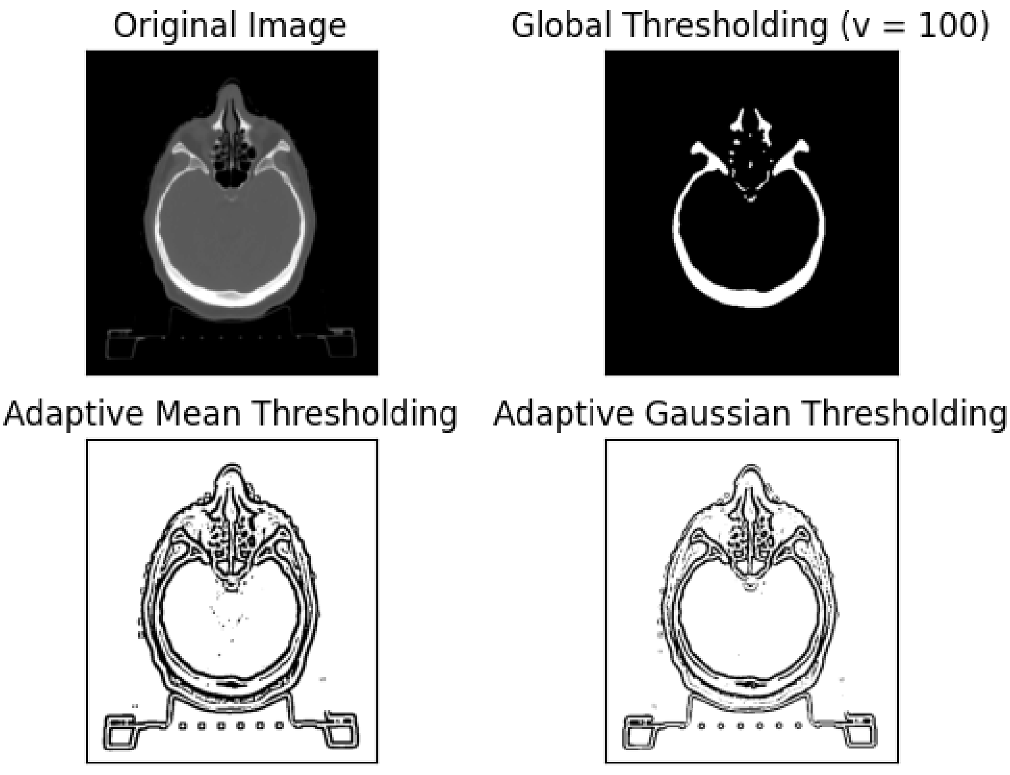

Figure 5. Example applications of thresholding methods on a sample CT image.

Chanapai and Ritthipravat (2009) explored an adaptive thresholding method for segmenting NPC images using Self-Organizing Maps (SOM) (25). The technique involves selecting initial tumor pixels manually and grouping pixels with similar threshold levels as tumor pixels. SOM is a neural network model which is used to cluster image pixels based on intensity values, facilitating the identification of tumor regions. The model adapts to variations in tumor appearance by learning from initial manually selected tumor pixels. This method aims to enhance the segmentation process by accurately distinguishing tumor pixels from non-tumor areas. The study demonstrated the effectiveness of this method in handling tumor heterogeneity in NPC images. Researchers proposed that this semi-automatic approach could potentially improve the accuracy and efficiency of medical image segmentation, particularly for NPC, by leveraging the SOM's capability to cluster image intensities into distinct groups for better threshold determination.

Region Growing and Seeding:

Region-based image segmentation begins by initializing seeds and then identifying pixels similar to these seeds, grouping these pixels into a class, and continually updating the seeds until a stable state is achieved (**Figure 6**). Key methods in region-based segmentation include region growing, region splitting and merging, and the use of superpixels (26–28).

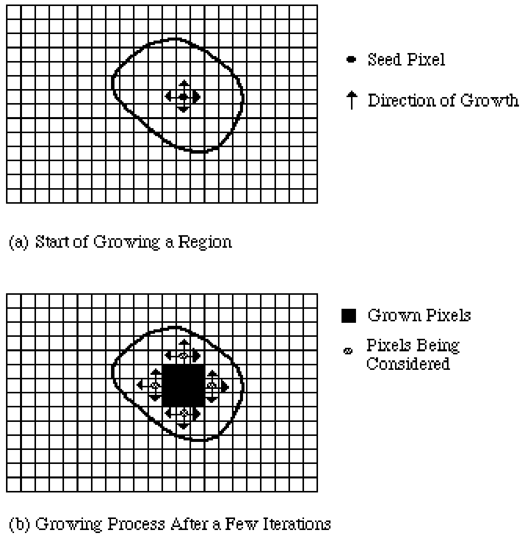

Figure 6. Visual example of a region growing algorithm. From Prof. Marshall A.D.'s (Cardiff University) lecture notes (URL: https://users.cs.cf.ac.uk/dave/Vision_lecture/node35.html)

Region growing focuses on expanding connected areas in an image by adhering to certain criteria, such as similarity in grayscale, edge definition, or texture, starting from one or more initial seed pixels and progressively incorporating adjacent pixels until no further merging is possible (29). Although region growing is typically part of a larger image analysis process and is useful for identifying small, simple areas like tumors, one limitation is the manual selection of seed points.

The concept behind region splitting and merging is to split a region into four equal parts if its features are inconsistent, and then merge any adjacent sub-regions that share consistent features, repeating these steps until further splitting and merging are not possible. This process, which may end when no further splitting can occur, also includes checking for and merging adjacent regions with similar features to complete the segmentation.

Region growing and region splitting and merging share a foundational principle and can enhance each other, with splitting potentially reducing an image to single pixels for re-merging based on specific criteria, which can be seen as an individual pixel-based region growing approach (29). Superpixel-based merging starts with segmenting the image into numerous small areas using superpixel algorithms, then merges these over-segmented areas to form the final segmentation, benefiting from the computational efficiency of superpixels in reducing pixel diversity for high-resolution images. This approach, however, relies on the success of the superpixel segmentation, with popular algorithms like SLIC (28), turbopixel (30), DBSCAN (31), LSC (31), DEL (32), GMM (33), and ISF (34) employing a local grid-based contour iterative optimization strategy for superpixel segmentation. These algorithms are advantageous for achieving precise local boundary delineation based on a predefined number of regions but may miss capturing the true contours of objects due to their focus on local rather than global information.

Interactive Edge Detection:

The edge-based segmentation technique operates on the premise that pixel values bridging the foreground and background exhibit significant differences, identifiable through changes in intensity (35). Such discontinuities are typically identified using derivative methods, either first or second order, like the gradient or Laplacian methods (36,37). Edge detection tools such as the Sobel, Roberts, and Prewitt detectors rely on the gradient method (36,37). They are straightforward to implement and can approximate the outline of objects, though they tend to be vulnerable to noise within the image. The Laplacian algorithm, which employs the second derivative for edge detection, is capable of pinpointing the edge direction but, like the gradient method, is noise-sensitive (36,37). To mitigate noise issues, the Laplace of Gaussian (LoG) technique was developed, applying a Gaussian filter to smooth the image before employing the Laplacian operator for edge detection. The Canny edge detector is noted for its superior performance in comparison to earlier methods, as it integrates filtering, enhancing, and detecting into a single operation. Despite the advancements in detecting object boundaries with these methods, they often result in the inclusion of numerous incorrect edges, necessitating further post-processing to refine the edge-based segmentation result.

Interactive edge detection is a technique in image processing and computer vision where the user interacts with the algorithm to refine or guide the detection of edges in images. This method typically involves the user specifying parameters or areas of interest, and the algorithm adjusts to these inputs to improve the accuracy of edge detection. It's particularly useful in applications where automated methods struggle with ambiguity or complex image features, enabling more precise segmentation or identification of structures within images.

Banu et. al. (2013) explored the significant role of semi-automated segmentation and edge detection in computer vision, particularly for medical images (38). Researchers emphasized that while unsupervised image segmentation without human intervention often falls short of accuracy, incorporating user input can greatly enhance outcomes. This approach, termed interactive image segmentation, leverages user-provided strokes or markers to delineate objects and backgrounds, employing a maximal similarity based region merging algorithm for segmentation. The research also introduces an autoadaptive edge detection algorithm, which contrasts with traditional methods like Sobel, Prewitt, Canny, and Laplacian for its robustness and efficiency in detecting clear and detailed edges. The segmentation process begins with an initial segmentation using the mean shift algorithm, followed by user interaction where markers indicate areas of interest. This input is then used in a maximal similarity based region merging (MSRM) algorithm to enhance the segmentation by merging similar regions as guided by the user's markers. For edge detection, the paper describes the autoadaptive algorithm's superiority in handling the intricacies of medical imaging, where noise and other factors can obscure edges, making traditional detection methods less effective. This research demonstrates the effectiveness of interactive methods in overcoming the limitations of fully automated systems, particularly in the nuanced field of medical image analysis. By allowing user guidance through simple inputs, it achieves more accurate segmentation and edge detection, crucial for detailed medical image analysis. The presented methodologies offer a promising direction for improving the precision of image-based diagnostics and research.

Graph Cut:

Graph-cut segmentation involves constructing a graph where vertices represent pixels and edges define pixel relationships. Segmentation is achieved through identifying a cut that minimizes an energy function, reflecting the optimal foreground-background partition (**Figure 7**). This method benefits from the formulation of an energy function combining regional terms, boundary terms, and, optionally, shape priors, balancing regional homogeneity and boundary smoothness.

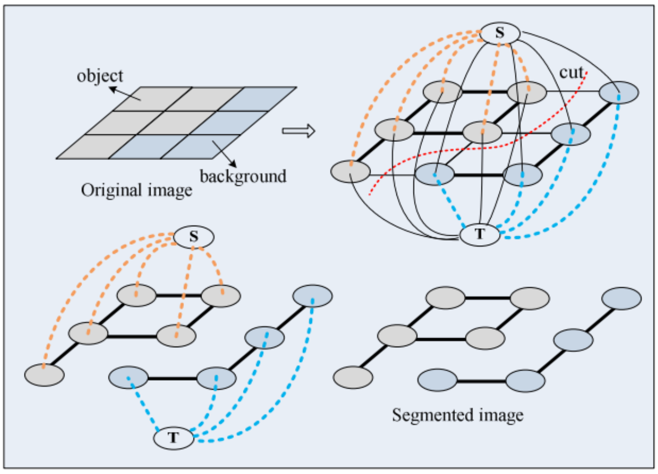

Figure 7. Illustration of graph cut for image segmentation. From the article "Image segmentation: A survey of graph-cut methods" authored by Yi and Moon (35).

Graph-cut methods have gained prominence due to their ability to use both boundary and regional information efficiently, offering globally optimal segmentation results. Originating from Boykov and Jolly (2001), graph-cut segmentation has evolved, showing effectiveness across various applications, from medical imaging to natural scene analysis (39). In their study, Yi and Moon (2012) surveyed graph-cut segmentation, categorizing the methods into speed-up-based, interactive-based, and shape prior-based graph cuts, facilitating a deeper understanding of this approach (35). In this study, graph-cut methods further categorized into three distinct approaches: Speed-up based graph-cut, interactive based graph-cut, and shape prior based graph-cut. Firstly, speed-up based graph-cut focuses on computational efficiency, often through pre-segmentation techniques like watershed algorithms or utilizing parallel computing to handle large graphs more rapidly. Secondly, Interactive based graph-cut incorporates user input to refine segmentation results. This approach is versatile, accommodating simple seed point selections or iterative interactions to progressively enhance segmentation accuracy. Lastly, shape prior based graph-cut integrates known shape information into the segmentation process, aiding in accurately delineating objects even in the presence of noise, diffuse edges, or occlusions. This method is particularly effective for images with pre-known object shapes.

Active Contour or Snakes:

Active contour segmentation, also known as snakes, is a method used in image processing and computer vision to detect and delineate the contours (edges) of objects within images. It is particularly useful for applications that require precise boundary detection, such as medical image analysis, object tracking, and shape recognition (40). The fundamental idea behind active contours is to evolve a curve, subject to constraints from the image data, until it closely approximates the contours of an object. This curve can move through the spatial domain of an image to minimize a specially designed energy function that represents the "fit" of the curve to the object's boundary.

The energy function typically comprises several terms (41), including:

1. Internal energy: Controls the smoothness and continuity of the curve. It is designed to prevent the curve from bending too much or stretching too far, ensuring a smooth and plausible shape.
2. External energy: Derived from the image data, this term guides the curve toward the object's edges. It can be based on gradient, intensity, or other image features that signify the presence of boundaries.
3. Constraint energy: Optionally, additional terms may be included to enforce specific shape or prior knowledge about the object's characteristics.

Active contour models are classified into two main types:

1. Parametric active contours (traditional snakes): These use an explicit parametric representation of the curve (e.g., a set of spline curves) (41). The curve is initialized close to the object of interest and iteratively adjusted to minimize the energy function, moving closer to the object's boundaries in each step.
2. Geometric active contours (level set methods): Instead of a parametric curve, these models use a level set representation where the evolving contour is implicitly represented as the zero level set of a higher-dimensional surface (42). This approach allows for more flexibility in handling changes in topology (e.g., merging and splitting of contours) and can more easily deal with complex shapes.

Active contour segmentation is widely appreciated for its flexibility and the level of control it offers over the segmentation process. By adjusting the energy function and the parameters governing the curve's behavior, it can be tailored to a wide range of applications and types of imagery. However, its success heavily relies on the appropriate initialization of the contour and the design of the energy function to adequately capture the characteristics of the object's boundary.

2.3.2. Automated Segmentation Techniques

Neural Networks and Deep Learning

Artificial Neural Networks (NN), also known as neural networks, are information processing systems designed inspired by biological neural networks in animals' brains (43). NNs are based on a collection of interconnected units or nodes known as artificial neurons, modeling neurons in the biological brain. Each connection, like synapses in the brain, can transmit signals to other neurons. An artificial neuron can receive a signal and process it before transmitting it to other connected neurons. The 'signal' in a connection is a real number, and the output of each neuron is processed by applying a set of nonlinear functions to the sum of its inputs. These connection lines are referred to as 'edges.' Neurons and edges typically have 'weights,' which are adjusted as learning progresses. Weights can decrease or increase the strength of the signal in a connection. Like their biological counterparts, neurons may have a threshold that only allows signal transmission when the combined signals exceed a certain threshold value. Neurons are typically grouped into layers. Different layers can apply various transformations to their inputs. As signals move from the first layer (input layer) to the last layer (output layer), they may undergo multiple transitions through different layers (**Figure 8**).

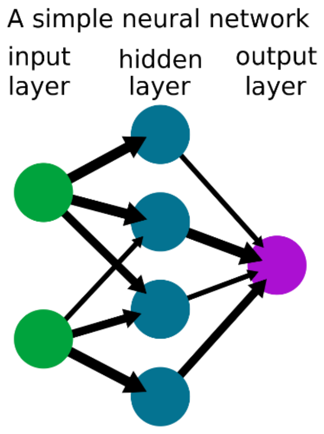

Figure 8. A simple neural network architecture.

Deep learning is achieved through models that utilize multiple layers of neurons between the network's inputs and outputs (44). With numerous layers, higher-level features can be progressively extracted from raw inputs. For example, in image processing, lower layers might identify edge features, while higher (or deeper) layers can recognize concepts more relatable to humans, such as digits, numbers, or human faces (45). Deep learning has significantly improved the performance of programs being developed in many different key research areas, such as computer vision, speech recognition, and image classification (46).

Convolutional neural network structures are often preferred in the layers used in deep learning. In a convolutional layer, each neuron accepts inputs only from a limited area of the previous layer, known as the neuron's receptive field. This approach significantly reduces the number of weighted connections between neurons, allowing for a hierarchical organization similar to the visual cortex in animals (47,48). The term 'deep' in deep learning refers to the presence of multiple layers in the neural network. While a linear perceptron structure cannot function as a universal classifier, a neural network with a single hidden layer, provided it has a non-polynomial activation function and is not limited in width, can achieve this. Deep learning is a modern variation that allows for practical and optimized applications under certain conditions, maintaining theoretical universality with an unlimited number of layers but with limited dimensions. Deep learning models can permit heterogeneity in layers and show significant differences from biologically analogized connection models. The goal here is to ensure efficiency, trainability, and comprehensibility. An example of a deep learning model is shown in (**Figure 9**).

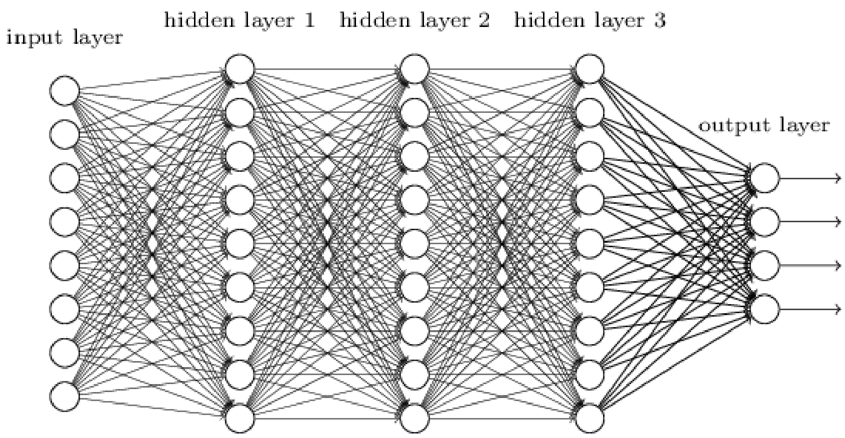

Figure 9. An example of a fully connected deep learning model. From Chapter 6 of "Neural Networks and Deep Learning" online book authored by Michael Nielsen (URL: https://neuralnetworksanddeeplearning.com/chap6.html).

U-Net Architecture and Automated Biomedical Image Segmentation

U-Net architecture has revolutionized the field of biomedical image segmentation, providing an efficient and robust means to segment complex images accurately. Developed Ronneberger et. al. (2015), U-Net is designed specifically for medical image processing, featuring a convolutional network architecture that excels in the precise segmentation of biomedical images (6).

U-Net extends the traditional convolutional neural network (CNN) architecture through a symmetric expanding path, which enables precise localization, a critical factor in medical image segmentation. The architecture is characterized by its U-shaped design, consisting of a contracting path to capture context and a symmetric expanding path that enables precise localization (6). The contracting path follows the typical architecture of a convolutional network, consisting of repeated application of convolutions, each followed by a rectified linear unit (ReLU) and a max-pooling operation. In contrast, the expanding path combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path. Additionally, Çiçek et al. (2016) advanced this framework by adapting the architecture to support 3D inputs, resulting in what is called 3D-Unet, for processing volumetric images (49). This enhancement in dimensionality enabled the models to capture spatial details more accurately, leading to an improvement in the precision and accuracy of the developed models.

U-Net has been successfully applied in a wide range of biomedical segmentation tasks, including but not limited to, the segmentation of neuronal structures in electron microscopic stacks, cell segmentation in light microscopy images, and lung lesion segmentation in CT images (49–51). Its ability to accurately segment complex anatomical structures and pathological features has significantly contributed to advancements in medical diagnostics, treatment planning, and biomedical research.

Currently, the majority of the highest-performing tumor segmentation models are built around a version of U-Net (52–55). The network's features and components can be extensively customized to enhance its effectiveness. Apart from the network design, critical factors such as preprocessing techniques, training strategies (for example, how data is split, choice of hyperparameters, loss functions, optimizers, and data augmentation methods), post-processing, and the use of ensembling strategies play a vital role in achieving optimal results. These choices are often made manually and can differ greatly depending on the dataset and the specific task at hand. nnU-Net has introduced an innovative approach by generating data fingerprints, which are designed to automatically tailor these preprocessing choices based on the data itself or on predetermined settings known for their effectiveness (56). While nnU-Net fundamentally relies on a U-Net structure, it modifies all other aspects to offer a universal solution for segmenting various types of medical imaging tasks. Without requiring any adjustments, nnU-Net has already achieved impressive results in multiple competitions. Nowadays, many entries in contemporary challenges are utilizing nnU-Net as a foundational model to compare it with the custom approaches.

Li et al. (2019) studied the application of a modified U-Net model, for automating the segmentation of tumor targets in nasopharyngeal carcinoma (NPC) patients (502 patients involved) undergoing radiotherapy (57). The study highlights the importance of accurate target segmentation in radiotherapy's outcome and aims to address the time-consuming and variable nature of manual segmentation dependent on the oncologist's experience. The U-Net model was trained using labeled computed tomography (CT) images and achieved an overall Dice Similarity Coefficient (DSC) of 65.86% for lymph nodes and 74.00% for the primary tumor, with respective Hausdorff distances of 32.10 mm and 12.85 mm. The study notes a decrease in segmentation accuracy with increasing cancer stage. The automatic segmentation process took approximately 2.6 hours per patient, compared to 3 hours using manual procedures, indicating that deep learning models can improve the accuracy, consistency, and efficiency of target segmentation in NPC, however, additional physician input may be necessary for accurate segmentation, especially in advanced stages.

Vision Transformer Networks

Vision Transformer Networks (ViTs) represent a paradigm shift in how deep learning models understand and process images (58). Building on the success of Transformer models in natural language processing (NLP), ViTs apply the Transformer architecture to visual data, offering a novel approach to image analysis that moves beyond the confines of convolutional neural networks (CNNs). Unlike traditional CNNs that process images through local filters and pooling layers, ViTs decompose images into a sequence of patches (**Figure 10**). These patches are then flattened and processed through a series of self-attention mechanisms, allowing the model to weigh the importance of different parts of the image relative to each other. This process facilitates a more global understanding of the image content, enabling the model to capture complex relationships and dependencies across the entire image. Furthermore, the self-attention mechanism allows ViTs to focus on the most relevant parts of an image, making them particularly effective for tasks requiring detailed analysis and interpretation, such as fine-grained classification, object detection, and semantic segmentation (58).

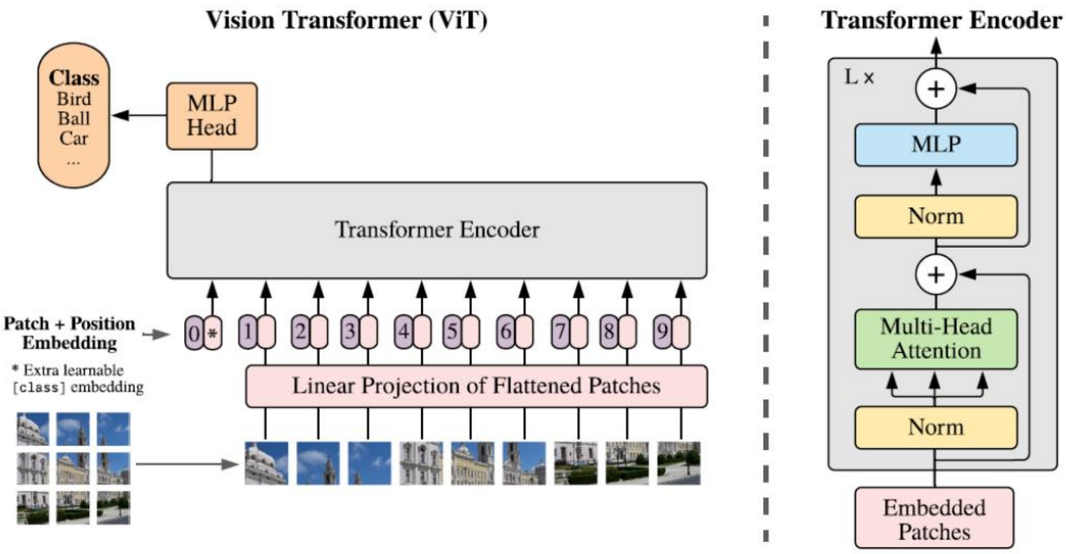

Figure 10. Framework of a visual transformer architecture. From the article "A Survey on Visual Transformer" authored by Liu et. al. (58).

In 2022, Hatamizadeh et al. proposed a novel architecture called UNETR, which utilizes a transformer as the encoder within a U-Net framework for the segmentation of volumetric medical images (11). This innovation addressed the limitation of Fully Convolutional Neural Networks (FCNNs) in capturing long-range spatial dependencies by reformulating 3D medical image segmentation as a sequence-to-sequence prediction problem. The UNETR model incorporates a transformer encoder to process 3D patches of the input volume, capturing global multi-scale information effectively (**Figure 11**). This is a significant departure from traditional CNNs that utilize localized receptive fields, which may not efficiently handle multi-scale information. The encoder is connected to a CNN-based decoder through skip connections at different resolutions, ensuring that the model benefits from both global contextual representations and localized spatial information. Researchers evaluated UNETR's performance on two public datasets: the Multi-Atlas Labeling Beyond the Cranial Vault (BTCV) dataset for multi-organ segmentation and the Medical Segmentation Decathlon (MSD) dataset for brain tumor and spleen segmentation tasks. UNETR achieved state-of-the-art performance on the BTCV leaderboard, demonstrating its effectiveness across various segmentation tasks in CT and MRI modalities (**Table 1**).

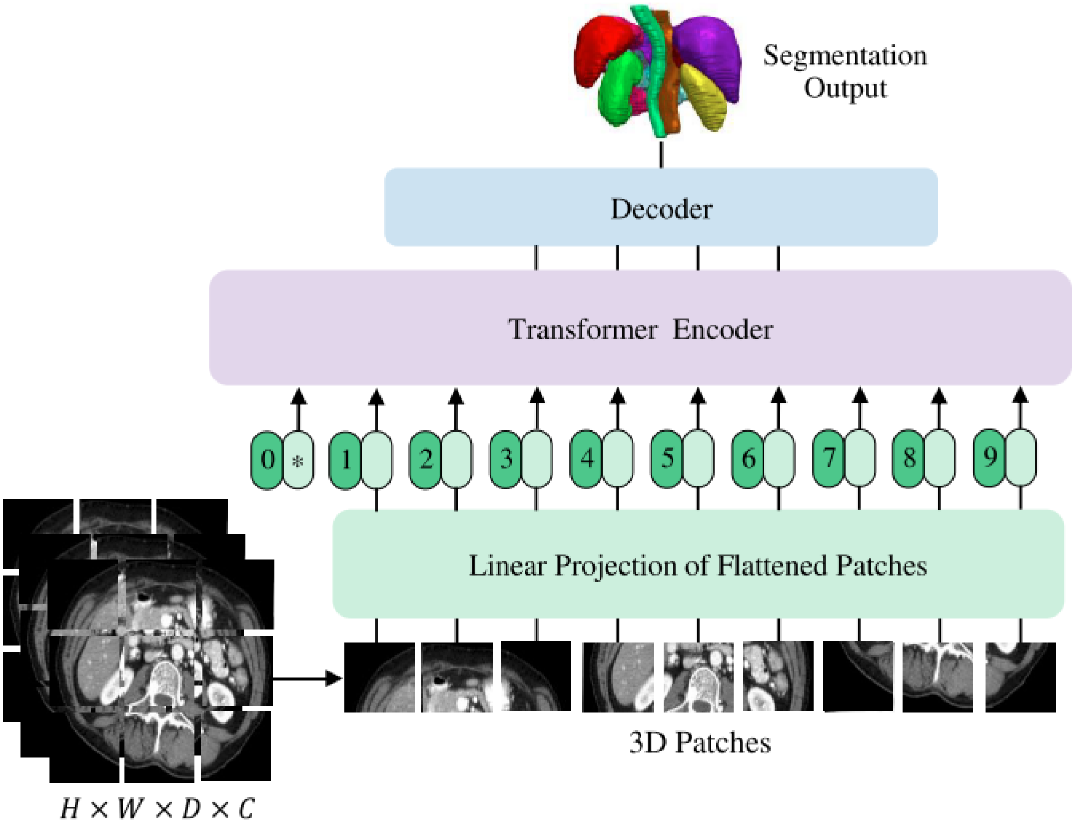

Figure 11. Framework of UNETR architecture. From the article “UNETR: Transformers for 3D Medical Image Segmentation” authored by Hatamizadeh et. al. (11).

Table 1. Quantitative comparisons of segmentation performances of different models in BTCV test set.

Methods	**Spl**	**RKid**	**LKid**	**Gall**	**Eso**	**Liv**	**Sto**	**Aor**	**IVC**	**Veins**	**Pan**	**AG**	**Avg.**
SETR NUP	0,93	0,89	0,9	0,65	0,76	0,95	0,81	0,87	0,75	0,72	0,72	0,62	0,8
SETR PUP	0,93	0,89	0,89	0,65	0,76	0,95	0,82	0,87	0,74	0,72	0,71	0,62	0,8
SETR MLA	0,93	0,89	0,89	0,65	0,76	0,95	0,82	0,87	0,74	0,72	0,72	0,61	0,8
nnUNet	0,94	0,89	0,91	0,7	0,72	0,95	0,82	0,88	0,78	0,72	0,68	0,62	0,8
ASPP	0,94	0,89	0,91	0,69	0,76	0,95	0,81	0,92	0,81	0,7	0,72	0,63	0,81
TransUNet	0,95	0,93	0,93	0,66	0,76	0,97	0,89	0,92	0,83	0,79	0,78	0,64	0,84
CoTr w/o CNN encoder	0,94	0,89	0,91	0,71	0,72	0,95	0,82	0,88	0,78	0,72	0,67	0,62	0,8
CoTr	0,96	0,92	0,94	0,7	0,76	0,96	0,85	0,92	0,84	0,79	0,78	0,69	0,84
UNETR	**0,97**	**0,92**	**0,94**	**0,75**	**0,77**	**0,97**	**0,91**	**0,89**	**0,85**	**0,79**	**0,77**	**0,74**	**0,86**

Note: Spl: spleen, RKid: right kidney, LKid: left kidney, Gall: gallbladder, Eso: esophagus, Liv: liver, Sto: stomach, Aor: aorta IVC: inferior vena cava, Veins: portal and splenic veins, Pan: pancreas, AG: adrenal gland. All results obtained from BTCV leaderboard. From the article “UNETR: Transformers for 3D Medical Image Segmentation” authored by Hatamizadeh et. al. (11).

Even with aforementioned accomplishments, UNETR faces two main challenges. Firstly, it is constrained to processing patches at a single scale, limiting its effectiveness in handling features that vary in size. Secondly, its self-attention mechanism's computational cost increases quadratically with the size of the image, leading to inefficiencies, especially with high-resolution images. To overcome these issues, within the same year as UNETR paper published, Hatamizadeh et al. (2022) introduced the Swin UNet TRansformer (Swin UNETR) (**Figure 12**), utilizing the innovative Swin Vision Transformer (Swin ViT) as its core (13,59). The Swin ViT distinguishes itself by having a computational complexity that scales linearly with the size of the image and by accommodating patches of varying scales. Through the adoption of the advanced Swin ViT and a self-supervised pre-training strategy applied to a substantial set of CT images, Swin UNETR outstripped its predecessor, UNETR, in terms of performance on the BTCV dataset (**Table 2**) (60). Swin UNETR also features a U-shaped design, connecting the Swin Transformer encoder to an FCNN-based decoder at various resolutions via skip connections. This design ensures the integration of deep semantic information extracted by the encoder with spatial details necessary for precise segmentation, recovered in the decoding phase.

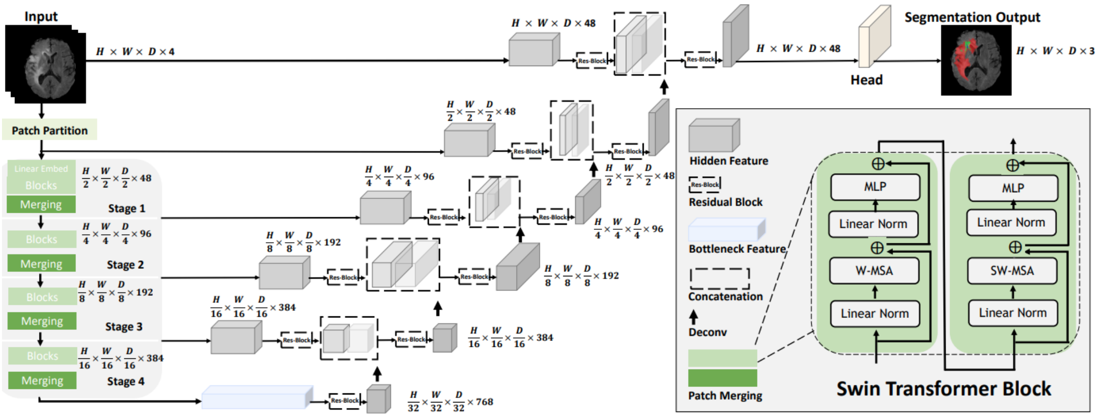

Figure 12. Overview of the Swin UNETR architecture. The Swin UNETR creates non-overlapping patches of the input data and uses a patch partition layer to create windows with a desired size for computing the self-attention. The encoded feature representations in the Swin transformer are fed to a CNN-decoder via skip connection at multiple resolutions. From the article "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images" authored by Hatamizadeh et. al. (59).

Table 2. Five-fold cross-validation benchmarks of models at BraTS 2021 dataset in terms of mean DSC score values. ET, WT and TC denote Enhancing Tumor, Whole Tumor and Tumor Core respectively.

	Swin UNETR				nnU-Net				SegResNet				TransBTS			
DSC Score	ET	WT	TC	Avg.	ET	WT	TC	Avg.	ET	WT	TC	Avg.	ET	WT	TC	Avg.
Fold 1	**0,88**	**0,93**	**0,91**	**0,91**	0,87	0,92	0,90	0,90	0,87	0,92	0,91	0,90	0,86	0,91	0,90	0,88
Fold 2	**0,91**	**0,94**	**0,92**	**0,92**	0,90	0,93	**0,92**	0,92	0,90	0,93	0,92	0,92	0,89	0,92	0,90	0,90
Fold 3	**0,89**	**0,93**	**0,92**	**0,91**	0,89	0,93	0,91	0,91	0,88	0,93	0,92	0,91	0,87	0,90	0,90	0,89
Fold 4	**0,89**	**0,94**	**0,92**	**0,92**	0,89	0,93	0,91	0,91	0,89	0,92	0,92	0,91	0,87	0,91	0,90	0,89
Fold 5	**0,89**	**0,93**	**0,92**	**0,91**	0,88	0,93	**0,92**	0,91	0,88	0,93	0,91	0,91	0,87	0,92	0,89	0,89
Avg.	**0,89**	**0,93**	**0,92**	**0,91**	0,88	0,93	0,91	0,91	0,88	0,93	0,91	0,91	0,87	0,91	0,90	0,89

From the article “Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images” authored by Hatamizadeh et. al. (59).

2.3.3. Current Developments in Nasopharyngeal Carcinoma Segmentation

The segmentation of gross tumor volume (GTV) for radiotherapy planning in nasopharyngeal carcinoma (NPC) is a critical step that significantly influences treatment effectiveness and patient prognosis. Traditional methods, which rely heavily on manual segmentation by radiation oncologists, are not only time-consuming but also suffer from considerable interobserver variability (IOV), leading to inconsistencies in treatment outcomes (61–63). The emergence of deep learning-based automated segmentation approaches, particularly utilizing multi-modality imaging data, has shown promise in enhancing the precision, efficiency, and consistency of GTV segmentation (64). Recent advancements in automated GTV segmentation for head and neck cancer, including NPC, have focused on developing deep learning models that can efficiently process complex imaging data to accurately identify tumor boundaries. A notable development in this domain is the application of 3D convolutional neural networks (CNNs) that leverage multi-modality imaging inputs—such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET)—to improve the accuracy and reliability of tumor segmentation (64).

Bollen et. al. (2023) showed that the accuracy of automated GTV segmentation has been significantly enhanced through the use of multi-modality CNNs (64). In this study, researchers demonstrated that the mean Dice Similarity Coefficient (DSC) between automated and manual segmentations was 69% for primary tumors (GTVp) and 79% for pathologic lymph nodes (GTVn) when utilizing CT and PET imaging data. When leveraging CT and MRI data, the mean DSC values were slightly lower, at 59% for GTVp and 71% for GTVn. However, when automated segmentations were corrected by observers, the mean DSC values increased to 81% and 89% for CT and PET, and 69% and 77% for CT and MRI, for GTVp and GTVn respectively, indicating substantial agreement between automated and expert-corrected segmentations. Also, before the introduction of automated segmentations, interoperator variability (IOV) measured by DSC between observers was reported at approximately 76% for GTVp and 86% for GTVn. Following the correction of automated segmentations, the IOV significantly decreased, with DSC values improving to 95% and 96% for GTVp and GTVn respectively. This reduction in variability underscores the potential of automated segmentation to standardize the segmentation process and improve the consistency of radiotherapy planning.

Due to their excellent spatial resolution in imaging soft tissue, MR images are commonly used in diagnosing nasopharyngeal cancer. They play a crucial role in evaluating radiation and monitoring the treatment of nasopharyngeal carcinoma (NPC) because pinpointing NPC lesions with MR images offers significant reference value. At present, radiologists are required to conduct manual examinations to detect and verify the presence of nasopharyngeal cancer lesions in MR images. One of the notable architectural advancements in NPC segmentation in MRI images is the Dilated Convolution Transformer Residual (DCTR) U-Net model, developed by Zeng et. al. (2023) (65). This innovative architecture integrates dilated convolution, transformer, and residual modules into the traditional U-Net framework, aiming to tackle the limitations of restricted convolutional fields of perception and achieve multi-scale feature fusion both globally and locally. This combination allows for superior segmentation performance, particularly in segmenting the target region of NPC from medical images. The DCTR U-Net model validated through 10-fold cross-validation on a dataset comprising 300 MRI series from NPC patients. The evaluation metrics used were the Dice Similarity Coefficient (DSC) and the Average Symmetric Surface Distance (ASSD), where the DCTR U-Net achieved DSC and ASSD values of 0.852

and 0.544 mm, respectively. These results signify a significant improvement in segmentation accuracy, underscoring the effectiveness of integrating dilated convolution, transformer, and residual modules. When compared with other mainstream segmentation models like U-Net (0.772 DSC, 0.823 mm ASSD), MultiResUNet (0.795 DSC, 0.749 mm ASSD), TransUNet (0.807 DSC, 0.685 mm ASSD), Swin-Unet (0.819 DSC, 0.636 mm ASSD), and UNETR (0.837 DSC, 0.597 mm ASSD), DCTR U-Net (0.852 DSC, 0.544 mm ASSD) demonstrated superior performance. Specifically, it outperformed these models in both DSC and ASSD metrics, confirming its effectiveness in NPC segmentation tasks. Such comparative analysis highlights the advancements made in deep learning architectures for medical image segmentation, showcasing the potential of custom architectures introducing both convolution and transformer structures in clinical applications.

2.3.4. Organs-At-Risk Segmentation in Nasopharyngeal Carcinoma

The accurate delineation of organs at risk (OARs) in computed tomography (CT) images is a pivotal step in the planning of radiation therapy for the treatment of nasopharyngeal carcinoma (NPC). This task, traditionally performed manually by radiation oncologists or radiologists, is time-consuming, labor-intensive, and subject to inter-operator variability, directly affecting the quality of the treatment plan and, consequently, the dose distribution for OARs. Thus, the development of robust, accurate, and automatic algorithms for OAR segmentation in NPC from CT images is of significant clinical interest.

Recent advancements in deep learning have shown promising results in the detection, classification, and segmentation of images, making them the state-of-the-art approach for these tasks. In this context, the study by Liang et al. (2019) introduces a fully automated deep-learning-based method termed the organs-at-risk detection and segmentation network (ODS net) (66). The ODS net is designed to improve the efficiency of radiation therapy planning by automatically detecting and segmenting OARs in CT images of patients with NPC (**Figure 13**). The ODS net architecture comprises two CNNs: the first proposes organ bounding boxes with associated scores, while the second uses these bounding boxes to predict segmentation masks for each organ. This approach focuses the segmentation process on the detected area, reducing the influence of similar or neighboring structures on the results. The study included 185 subjects, with sensitivity, specificity, and DSC assessments performed to evaluate the detection performance of the ODS net. The ODS net demonstrated high accuracy in the automated detection of OARs, with sensitivity ranging from 0.997 to 1 for most organs and specificity from 0.983 to 0.999. Additionally, the segmentation results from the ODS net showed strong correlation with manual segmentation, achieving a DSC of more than 0.85 for most organs.

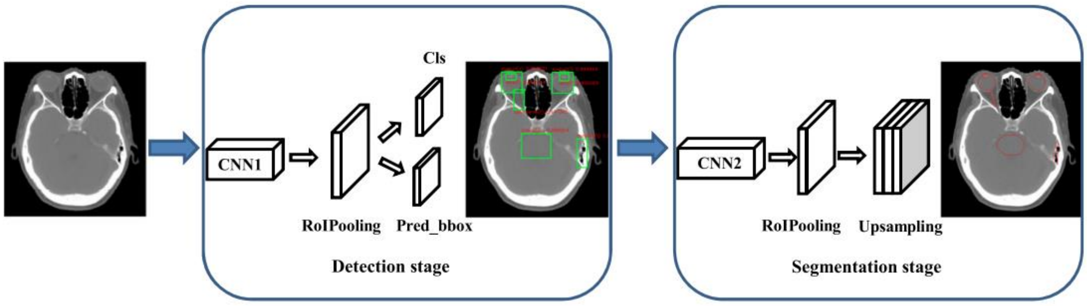

Figure 13. Diagram of the ODS net assembly process. This two-phase network is comprised of detection and segmentation components. During detection, a first convolutional neural network (CNN1) suggests potential bounding boxes for organs, accompanied by corresponding confidence scores. In the segmentation phase, another convolutional neural network (CNN2) takes these suggested bounding boxes and proceeds to generate predictive segmentation masks for the organs at risk (OARs). CNN: convolutional neural network, Cls: classification, Pred_bbox: predicted bounding box. From the article “Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning” authored by Liang et. al. (66).

The SegRap2023 challenge, organized in conjunction with MICCAI 2023 (URL: https://conferences.miccai.org/2023/en/), provided a platform to benchmark the latest advancements in deep learning for the segmentation of OARs in NPC (67). This challenge offered a unique dataset comprising 400 computed tomography (CT) scans from 200 NPC patients, each with a pair of pre-aligned non-contrast and contrast-enhanced CT scans. The primary goal was to segment 45 OARs and 2 gross tumor volumes (GTVs) from these scans, highlighting the current state-of-the-art in medical image segmentation. A total of 387 teams registered for the challenge, with 12 teams submitting containerized algorithms for OAR segmentation. The methodologies employed by participating teams varied but generally followed two main approaches: leveraging existing segmentation frameworks like nnUNet (56) with custom modifications and employing multi-step processes that involved both coarse localization and fine segmentation of OARs. Key innovations introduced by the participants included structure-specific label generation, boundary refinement techniques, and the use of ensemble models to improve segmentation accuracy.

During the challenge, following significant architectural developments had been proposed:

- **Two-Step Approach by Y. Zhong et al. (1st place):** This method introduced a novel two-step segmentation process comprising structure-specific label generation followed by boundary refinement. Initially, the model segmented OARs by dividing them into distinct classes, considering anatomical symmetries and overlapping structures. The segmentation was powered by an adaptation of the nnUNetV2 framework, fine-tuned to utilize both non-contrast and contrast-enhanced CT images. The subsequent step involved boundary refinement, where regions of interest (ROIs) were extracted for further precision enhancement, leveraging multiple output layers for diverse organs.
- **UniSeg-Based Method by Y. Ye et al. (2nd place)** (68)**:** Leveraging the pre-trained UniSeg model, this approach focused on fine-tuning the network to NPC OAR segmentation tasks through bespoke pre-processing and an ensemble strategy. The model was trained on processed 3D patches from CT scans, utilizing an ensemble of predictions to derive the final segmentation. This approach highlighted the efficacy of transfer learning and ensemble models in achieving high segmentation accuracy.
- **Vanilla nnUNet Approach by Y. Su et al. (3rd place):** This model employed the nnUNet framework with an increased patch size and without mirror data augmentation to account for the anatomical symmetry in the head and neck region. The model focused on leveraging extensive data augmentation techniques to improve generalization and robustness.

These approaches aimed to address the common challenges in NPC OAR segmentation, such as dealing with the high variability in organ shapes and sizes, managing the presence of overlapping structures, and ensuring the accurate differentiation between contrast-enhanced and non-contrast scans. The results of the challenge demonstrated significant progress in automatic OAR segmentation for NPC. The top-performing teams achieved average Dice similarity coefficient (DSC) scores ranging from 76.68% to 86.70% for OARs, indicating the effectiveness of deep learning approaches in this domain. Organs such as the brain, brainstem, and eyes showed high segmentation accuracy across most models. This is likely due to their distinct anatomical boundaries and larger size compared to other OARs, making them easier to segment. Smaller or thin-structure OARs, such as the optic nerves, cochleas, and optic-chiasm, presented greater challenges, with models showing lower accuracy. These organs' proximity to other structures and their small size contribute to the difficulty in achieving precise segmentation. The challenge underscored the need for further improvements, especially in the segmentation of smaller or thin-structure OARs and in enhancing the generalizability of the models to new, unseen data.

2.4. Feature Extraction in Medical Imaging

2.4.1. Radiomics Feature Extraction

Radiomics plays a pivotal role in the analysis of medical images by extracting a large number of quantitative features that are not always visually perceptible to the human eye (69). These features, encompassing shape, intensity, texture, and wavelet-based attributes, provide comprehensive information about tumor phenotype and its surrounding environment. The process starts with the accurate segmentation of the tumor and regions of interest (ROI), following which standardized feature extraction protocols, such as those defined by the Image Biomarker Standardization Initiative (IBSI), are applied to ensure reproducibility and comparability across studies (**Figure 14**) (70).

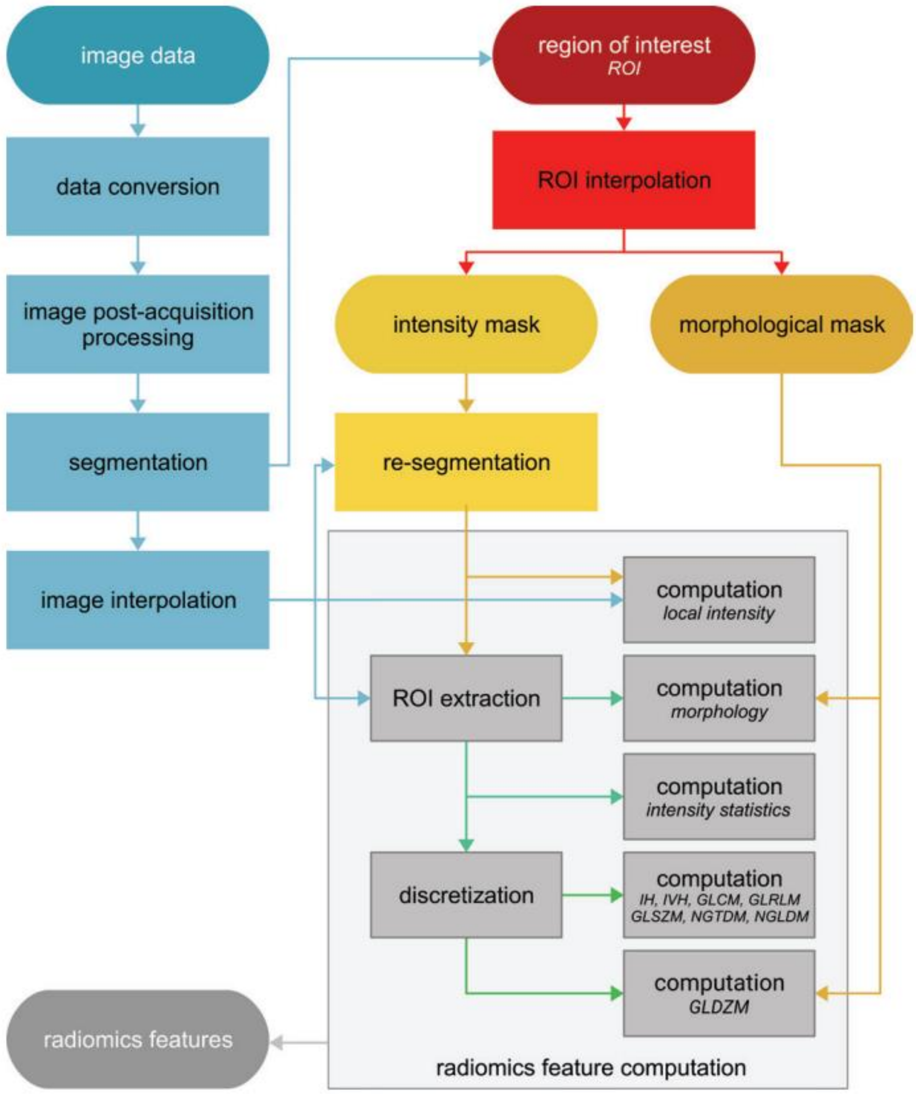

Figure 14. The flowchart outlines the process for calculating radiomics features from images, beginning with the reconstructed images. These images undergo several optional processes, such as data conversion to standardized values, post-acquisition modifications like denoising, and interpolation. A region of interest (ROI) is either automatically generated during segmentation or selected from an existing one. This ROI undergoes interpolation, and both intensity and morphological masks are produced as duplicates. The intensity mask can be further segmented based on intensity levels to ensure consistency in intensity ranges across studies. Radiomics features are calculated using the masked image around the ROI (for local intensity features) or from the ROI itself (for all other features). Before calculating features from the intensity histogram and other analyses (such as intensity-volume histogram, gray-level co-occurrence matrix, and several others), image intensities are discretized. From the article of Zwanenburg et. al. “The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping” (70).

Radiomics successfully applied in the management of different types of oncologic imaging tasks such as lung, esophageal, breast, rectal, and prostate cancers (71–74). These applications have demonstrated the potential of radiomic analyses to significantly impact cancer care across multiple fronts:

- Diagnosis: Radiomics can enhance diagnostic accuracy by identifying subtle imaging biomarkers indicative of NPC, potentially allowing for earlier detection and intervention.
- Treatment Personalization: By correlating radiomic features with treatment outcomes, clinicians can tailor therapies to individual patients, optimizing therapeutic efficacy while minimizing adverse effects.
- Prognostication: The prognostic value of radiomic signatures can provide insights into disease progression, recurrence risks, and overall survival, aiding in patient counseling and management decisions.
- Response Monitoring: Radiomics can offer a sensitive and early indicator of tumor response to radiotherapy, enabling timely modifications to treatment plans if necessary. Outcome and Toxicity
- Prediction: Predictive models incorporating radiomic features can forecast potential treatment-related toxicities, guiding the choice of therapeutic modalities and intensities.

For NPC patients, radiomics offers the potential to unearth biomarkers that are predictive of treatment outcomes, such as Disease-Free Survival (DFS). For example, texture analysis on segmented NPC tumors from CT or MRI scans can reveal heterogeneity within the tumor, which has been associated with different biological behaviors and treatment responses. Zhang et al. (2022) provided a comprehensive overview of the application of radiomics in the management of NPC, emphasizing its potential in diagnosis, treatment planning, and prognostication (75). Studies included in this review have shown that radiomics models, can classify NPC survival subgroups with higher accuracy than the TNM staging system, potentially differentiate recurrency and radiation resistance, predict treatment response to radiotherapy and chemotherapy, and can show possible DFS and PFS (progression free survival) suggesting its potential for individualized diagnosis and treatment planning (75–77).

Radiotherapy serves as the cornerstone of Nasopharyngeal Carcinoma (NPC) management, particularly in the disease's early stages, with the addition of chemotherapy for advanced stages (78). However, intensity-modulated radiotherapy, the modern standard for precision, encounters challenges such as significant dose inaccuracies due to factors like patient weight loss and tumor regression during treatment (79). Adaptive radiotherapy, which revisits and adjusts the treatment plan mid-therapy, emerges as a solution, albeit at a cost of increased economic burden, time, and labor. Identifying patients who could benefit from adaptive radiotherapy upfront could substantially enhance treatment outcomes. In this context, Yu et al. (2019) leveraged pre-treatment MRI scans to analyze tumor marker features, utilizing enhanced T1 and T2 images to generate Area Under the Curve (AUC) scores of 0.852, 0.750, and 0.930 for different model verification groups, suggesting a robust method to pinpoint patients needing adaptive therapy (79).

For predicting responses to induction chemotherapy in NPC patients showing advanced local progression, Zhao et al. (2020) crafted a nomogram merging multi-sequence MRI features and clinical parameters (80). Their model demonstrated superior predictive capabilities (with C-indexes of 0.952 for the training group and 0.863 for the validation group) over models relying solely on clinical parameters (C-indexes of 0.708 and 0.549 for training and validation groups, respectively). Similarly, Youngfeng et al.'s (2021) work resulted in the highest AUC of 0.905 for a combined model, focusing on specific MRI sequence imaging features, signifying the potential for tailored treatment schemes and adjustments based on neoadjuvant chemotherapy sensitivity (81).

In the realm of prognosis, the application of MRI-based radiomics has illuminated its efficacy in evaluating critical outcomes like progression-free survival (PFS), disease-free survival (DFS), and overall survival among NPC patients. This is significantly enriched when combined with clinical information, such as lymph node status, Epstein-Barr virus presence, and tumor staging, thus fostering personalized treatment approaches (82–85). Ouyang et al.'s (2017) analysis of the Radscore, a radiomic biomarker, affirmed its capability to predict PFS (82). Furthermore, Shen et al. (2020) developed a comprehensive model that incorporated radiomics, staging, and

Epstein-Barr virus data, achieving high C-index scores for PFS prediction (0.805 for the training group and 0.874 for the validation group) (83). Notably, Yang et al.'s (2019) nomogram, which integrated lymph node data, Dose Volume Histogram signatures, and TNM stages, presented an impressive C-index of 0.811 for PFS prediction, showcasing enhanced performance compared to the TNM staging system alone (C-index: 0.613) (85).

The predictive power of radiomics extends beyond treatment effects to anticipate recurrence and side effects. For instance, Zhang et al.'s (2019) models based on MRI radiomics provided AUC scores of 0.827 and 0.792 for training and validation groups, respectively, effectively stratifying patients into risk groups for metastasis and aiding in strategic treatment planning (86). The potential for radiomics to forecast radiotherapy-induced complications, like acute xerostomia, was exemplified in a study with an accuracy of 0.9220 and a sensitivity of 100% for early prediction post-radiotherapy (87).

In summary, radiomics in NPC treatment not only offers a granular view of tumor heterogeneity but also holds the key to refining diagnosis, optimizing treatment protocols, and improving prognostication. Its integration into clinical workflows promises a leap toward personalized and precision oncology, underscoring the critical role of advanced imaging and machine learning in revolutionizing NPC management.

2.4.2. Deep Learning Based Feature Extraction

The advent of deep learning has revolutionized the field of medical imaging, offering transformative approaches in diagnostics, treatment planning, and patient monitoring. Central to this revolution is the concept of feature extraction, a process by which deep learning algorithms identify and utilize intricate patterns within imaging data to inform clinical decisions. Feature extraction in the context of deep learning involves automatically discovering the representations needed for classification or prediction from raw data. This contrasts sharply with traditional machine learning techniques, which typically rely on handcrafted features that require domain expertise and extensive preprocessing. Deep learning models, particularly convolutional neural networks (CNNs), are adept at learning hierarchies of features directly from data.

These hierarchies range from simple edge detectors at lower levels to complex structures pertinent to the specific problem domain at higher levels.

The application of deep learning-based feature extraction in medical imaging is of paramount importance due to several factors:

- **High Dimensionality and Complexity:** Medical images, such as MRI, CT scans, and X-rays, are inherently high-dimensional and contain complex structures. Deep learning models are capable of navigating this complexity to extract relevant features without explicit human guidance, facilitating more accurate and reliable analyses.
- **Automated Analysis:** With the ability to automatically extract and utilize features, deep learning models can significantly reduce the time and labor associated with manual image analysis. This automation holds the potential to accelerate diagnostics and improve the efficiency of healthcare delivery.
- **Enhanced Diagnostic Accuracy:** By leveraging learned features that may not be visible to the human eye, deep learning models can enhance diagnostic accuracy. This is particularly crucial in early disease detection, where subtle imaging markers can be indicative of disease progression.
- **Personalized Treatment Planning:** Deep learning-based feature extraction can uncover patient-specific characteristics from medical images, supporting personalized treatment planning. This tailored approach can lead to improved patient outcomes and optimized resource utilization.

Various deep learning models have been employed for feature extraction in medical imaging, with CNNs being the most prominent due to their ability to process grid-like data (e.g., image pixels). Shen et. al. (2017) introduced the fundamentals of deep learning methods in medical imaging, highlighting their success in various applications such as image registration, detection of anatomical and cellular structures, tissue segmentation, and computer-aided disease diagnosis and prognosis, emphasizing the shift from handcrafted features to data-driven hierarchical feature representations (88). Additionally, Apostolopoulos et al. (2022) evaluated deep learning (DL) methods for feature extraction in medical imaging, emphasizing the challenge of explainability in DL, which affects its reliable use in clinical practice. The study found that DL can reveal potential biomarkers, especially when models are trained from scratch in domains where low-level features are insufficient (89).

In 2024, Meta AI researchers introduced the DINOv2 feature extractor, marking a significant leap in self-supervised learning for computer vision (90). This model aims to create a versatile foundation for generating general-purpose visual features by learning task-agnostic visual representations, similar to the success of pretrained models in natural language processing. Utilizing a large, diverse dataset and a Vision Transformer (ViT) with 1 billion parameters, DINOv2 excels in generating high-quality features without needing fine-tuning for different tasks (**Figure 15**). Its innovation stems from a unique pretraining process that combines existing techniques with new improvements, focusing on detailed visual information at both the image and patch levels. The model's use of a specially curated dataset, LVD-142M, ensures high-quality, diverse training data, leading to robust feature development. DINOv2 not only sets new benchmarks in self-supervised learning but also excels across various vision tasks, demonstrating the potential of self-supervised pretraining with a large, sophisticated model on a carefully curated dataset to match or exceed supervised learning methods.

Figure 15. Visualisation of first 3 PCA components calculated from DINOv2 model feature predictions. Each component is matched to a one of RGB color channels and background is removed by thresholding the first PCA component. From the article "DINOv2: Learning Robust Visual Features without Supervision" authored by Oquab et. al. in affiliation with Meta AI (90).

Huang et. al. (2024) evaluated the performance of ImageNet pre-trained models (VGG16, ResNet50, DenseNet121) and DINOv2 variants (DINOv2-s, DINOv2-b, DINOv2-l) in medical imaging classification, using accuracy, precision, recall, and F1-score as metrics (91). The research provides a detailed comparison across different datasets, revealing insights into each model's effectiveness under various conditions, including the impact of a freezing mechanism during transfer learning. In the classification of chest X-ray images, DINOv2-l without the freezing mechanism achieved the highest F1-score of 87.73%, demonstrating superior performance over other models, including a matching accuracy of 87.73%. For the iChallenge-AMD dataset, focusing on retinal fundus images for age-related macular degeneration detection, DINOv2-l with freezing outperformed with an F1-score of 86.73%. In this domain, DINOv2 models generally surpassed ImageNet-based models, with DINOv2-l reaching an accuracy of 87.16% without freezing. The study further examined the HAM10000 dataset, used for classifying skin cancer from dermoscopic images, where DINOv2 models, especially the larger variants, significantly outperformed ImageNet pre-trained models. The highest F1-score and accuracy were recorded by DINOv2-l without freezing, at 78.78% and 78.71% respectively.

Features extracted with deep learning methods are also studied for NPC research, including diagnosis and prognostic prediction. Jing et. al. (2020) explored the application of deep learning for predicting the risk of disease progression in patients with nasopharyngeal carcinoma (NPC), utilizing multi-parametric MRI data (92). The study introduced a novel end-to-end multi-modality deep survival network (MDSN), extending from a 3D dense net framework, designed to extract deep representations from multi-parametric MRIs, including T1-weighted, T2-weighted, and contrast-enhanced T1-weighted images. The incorporation of clinical stages into the model further refined the overall risk score (ORS) prediction for individual NPC patients. The research included a training and validation cohort of 1,417 patients, alongside a retrospective cohort of 429 patients for testing. The study's findings are particularly notable for demonstrating that the deep learning model, when integrated with clinical stages, achieves superior performance compared to existing methods and clinical staging alone. Specifically, the C-index for the MDSN model, both with and without clinical stage incorporation, was reported at 0.672 and 0.651, respectively, on the test set. These results surpass the performance of the clinical stage grouping alone, which had a C-index of 0.610. By effectively leveraging deep learning techniques and multi-parametric MRI data, the MDSN model presents an accurate and robust tool for predicting NPC patient outcomes.

3. MATERIALS AND METHODS

Data Collection:

CT images of 120 non-metastatic nasopharyngeal cancer patients who received radiotherapy treatment for a diagnosis of nasopharyngeal cancer at the Istanbul Acıbadem Maslak Hospital Radiation Oncology unit between 2009 and 2021, along with tumor and organs-at-risk contour data marked by an experienced radiation oncologist were included in the study. The CT images were reconstructed with a matrix size of 512x512 and a slice thickness of 3 mm. For OAR task, all available CT series included, and for NPC segmentation task (using GTV tumor contour data), only enhanced series with intravenous contrast media were used. The image and related contour data were randomly splitted into a training and a test dataset in a ratio of %80 and %20, respectively. Five fold cross validation strategy including stratification was used by incorporating scikit-learn python library.

Initial Data Preprocessing:

To feed the 3D pixel matrices ("arrays") obtained from images and their corresponding segmentation masks to the models in an organized and accurate manner, the CT series and DICOM contour data were converted to the NIfTI format using Platipy (Processing Library and Analysis Toolkit for Medical Imaging in Python), an open source python medical image processing library.

For NPC segmentation task, to overcome overflowing contours into neighboring bone and air containing pixels, areas of the contour that match spatially with the underlying image voxels with Hounsfield Unit (HU) values below -50 and above 180 were excluded.

U-Net Based NPC and OAR Segmentation Model Structure (nnUNet):

Isensee et al. (2018) provided an adaptive and generalizable framework for training U-Net segmentation models in biomedical domain (56). In addition, the authors shared the open source codes for this study at the URL 'github.com/MIC-DKFZ/nnUNet'. Using the nnUNet (v1.7) framework 3D full resolution U-Net models were trained. In the nnUNet framework, two convolution layers and one pooling layer were used in the encoder blocks, while a transpose convolution layer was used in the decoder blocks. The activation function "leaky ReLU" was preferred. "Instance normalization" was used for normalization in the network layers.

A 5-fold cross-validation was performed during model training. The cross-entropy loss function, defined in the nnUNet framework, was preferred as the loss function. The SGD (Stochastic Gradient Descent) function was used for optimization and the learning rate chosen at different levels according to the model training process and the function with the highest success rate preferred. The model trainings were carried out using an NVIDIA GTX 3090 24GB GPU. Models were trained for total number of 250 epochs at each fold of cross validation.

During training of the models within the nnUNet framework, online data augmentation was performed using random rotation, random scaling, random elastic deformation, gamma correction. A custom trainer class was generated using nnUNet's "insane data augmentation" pipeline combining with Cross-Entropy loss function. Because of the potential decrease in the performance of models, mirroring was not used in both NPC and OAR tasks.

UNETR and SwinUNETR NPC Segmentation Model Structure

These models were trained only for NPC tumor segmentation. For creating UNETR and SwinUNETR, we used the same architectures as defined in the article "UNETR: Transformers for 3D Medical Image Segmentation" and "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images", both of which are authored by Hatamizadeh et. al. (2022) Different than nnUNet platform (automatized), CT series was resampled into same voxel dimentions (1.5x1.5x2 mm) before training these models. 96x96x96 array was used as an input patch shape (11,59). To overcome class imbalance, only minor amount of background patches were randomly sampled. Additionally, we again used data augmentation strategies such as random rotation of 90, 180 and 270 degrees, random flip in axial, sagittal and coronal views and random scale and shift intensity for training these models.

A 5-fold cross-validation was performed during model training. Combined cross-entropy and DSC loss function was preferred as the main loss function. The

Adam (Adaptive Moment Estimation) function was used for optimization and the learning rate was chosen at different levels according to the model training process and the function with the highest success rate was preferred. The model trainings were carried out using an NVIDIA GTX 3090 24GB GPU. Models were trained for total number of 20.000 steps (around 300 epochs) at each fold of cross validation. For the convention, same exact cross validation splits that were created to train 3D Unet algorithm were used for model training.

Model Metrics and Statistical Evaluation:

For measuring the performance of the models, 'DSC', 'Precision', 'Recall', and 'Accuracy', ‘F1’ values were obtained by comparing the model outputs with the test data set marked by the radiation oncologist. Also, different model structures were compared with each other. All calculations and coding in the model development process were done using the Python 3 programming language.

Feature Extraction and Disease Free Survival Prediction:

In this phase of our study, we engaged in a comprehensive feature extraction and disease-free survival (DFS) prediction process. Only NPC segmentation maps from enhanced CT series were used for feature extraction.

Initially, all contrast-enhanced (CE) computed tomography (CT) volumes along with their corresponding segmentation masks were resampled to align within the same spatial domain, ensuring uniform voxel dimensions across all samples. This step is crucial for maintaining consistency in feature extraction and subsequent analysis.

Following resampling, we performed case-wise normalization of the volumes. This process aims to reduce variability between the imaging data of different patients, enhancing the reliability of subsequent feature extraction and analysis steps.

Features were extracted using 2D slices to increase total amount of data to open some room to dealing with class imbalance between recurred and disease-free survived patients and to conduct further statistical analysis.

Initially, feature extraction had been carried out on manual segmentations (from contour data) to ascertain the most indicative features and features also were extracted from automatic segmentations for comparison.

Radiomics Features: Within the radiomics features first order statistics, shape descriptors, gray level co-occurrence matrix (GLCM) features, and other gray level (GL) attributes, along with their filtered and logarithmic transformations were included. A sum of 1059 distinct radiomics features were extracted. For extracting these features "pyradiomics" python library was used with following settings:

Table 3. Pyradiomics feature extraction settings.

enableAllImageTypes	True
enableAllFeatures	True
normalize	True
normalizeScale	100
binWidth	5
resampledPixelSpacing	[2,2,2]
preCrop	True
sigma	[1,2,3,4,5]
start_level	0
distances	[1,2]
wavelet	haar
force2D	True
force2Ddimension	0
minimumROISize	16

In addition to radiomic features, we also considered DINOv2 (a self-supervised learning framework for visual representation) and clinical data to enrich our feature set, aiming to encapsulate a comprehensive array of predictive factors for DFS.

Clinical Features:

Gender, pathology type, NPC stage, T stage, N stage, number of adjuvant and neoadjuvant treatment courses, treatment protocol (only radiotherapy or additional chemotherapy duration), presence of cranial nerve involvement, gross tumor volume and clinical tumor volume, given radiotherapy dose, and radiotherapy technique (IMRT, 3ARC, 4ARC, Other ARC) were encoded categorically and included in the disease-free survival prediction features for patients included in the study from hospital electronic records. Additionally, age at diagnosis and time elapsed from diagnosis to radiotherapy were included in the feature set as continuous variables (Table 4).

Table 4. Encoding of clinical features as a preprocessing step.

Clinical Feature	**Encoding**
Gender	M, F
Pathology Type	Indifferentiated, Non-Keratinized Differentiated, Keratinized
NPC Stage	I,II,III,IV
T Stage	I, II, III, IV
N Stage	0, I, II, III
Protocol	RT, CRT 1-3W, CRT 3-6W, CRT 6-7W, CRT >7W
Diagnosis to RT	Continuous (Days)
Age at Diagnosis	Continuous (Years)
Gross Tumor Volume RT Dose	>=6900, <6900
Clinical Tumor Volume RT Dose	Equal to 6000, not equal to 6000
RT Technique	IMRT, 3 ARC, 4 ARC, Other ARC
DFS	Recurred or Non recurred within 5 years after RTx

DINOv2 Features: For the part of feature extraction from NPC regions with deep learning we preferred a novel pretrained image feature extractor named as DINOv2 (90). The architecture and detailed information about this model are available at the article "DINOv2: Learning Robust Visual Features without Supervision" published in affiliation with Meta AI (URL: https://dinov2.metademolab.com/). DINOv2-s model will be used for extracting 386 different features. These features are extracted by transferred frozen weights of the model and defined as "general purpose visual features" in the article.

Disease Free Survival Prediction

Using the 5-fold cross validated same split strategy in training segmentation models, Random Forest, Linear Kernel Support Vector Machine, Logistic Regression, XGboost, Decision Tree and K-Nearest Neighbors classifier models were trained to predict disease free survival within 5 years. In combination with clinical features, models were trained with either radiomics or DINOv2 features that extracted from manually segmented NPC regions. For the feature selection, Lasso regression with alpha of 0.01 was applied. Minority class upsampled by using SMOTENC algorithm in trials and Z-score normalization was applied before training. All the baseline models initially were trained with pycaret (v3.0.0) python library with predefined hyperparameters and without finetuning. ROC-AUC, sensitivity, specificity and F1 scores calculated. Best performing model for each feature set was determined. After selecting best performing models, optuna python library with criterion to optimize as F1 score and with 100 iterations was used for fine tuning the ML model.

Feature correlations were calculated and model performances were compared to each other for ML models trained with radiomics and DINOv2 features in combination with clinical features. As stated before, ML models for DFS predictions trained with using features from manually segmented tumor regions. So, features were also extracted from predictions of best performing automatic NPC segmentation model to compare performance of DFS ML models on extracted features from automatic segmentations. For this approach, both automatic segmentation predictions and DFS predictions using extracted features from these predicted segmentations were made on same unseen validation data. No data leakage allowed between both segmentation and ML models (**Figure 16**).

Ethical Approval

This study was conducted in strict accordance with the guidelines and regulations set forth by the Acıbadem University's Ethics Committee. Prior to commencement, the research protocol was thoroughly reviewed and granted full approval (Approval No. 2023-05/163).

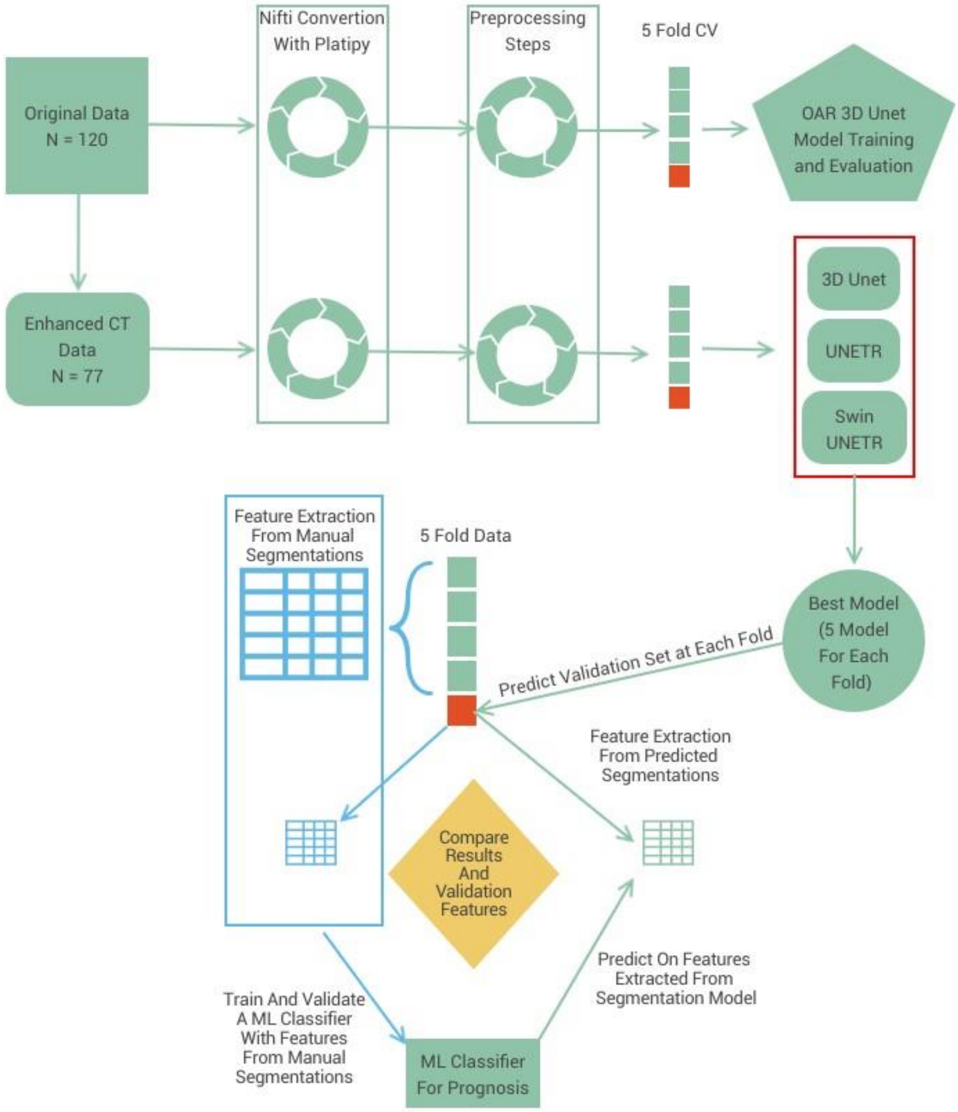

Figure 16. Flowchart of proposed methodology in the study.

RESULTS

The median age of the patients included in our study is 46 (n=120, std=14.94) (Figure 17).

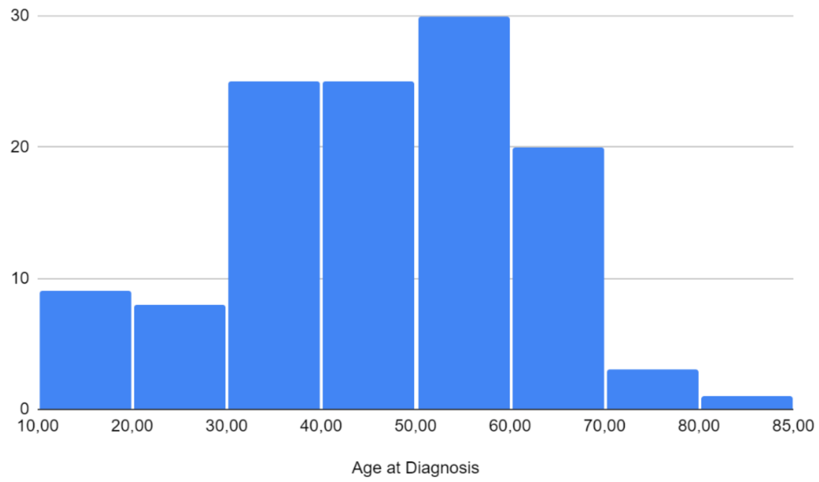

Figure 17. Distribution of age at diagnosis of nasopharyngeal carcinoma.

Using CT series from 120 patients and segmentation data of organs at risk contoured by an experienced radiation oncologist, training and test sets were created in an 80/20 ratio with 5-fold cross-validation. 3D UNET algorithms for automatic segmentation of relevant organs were trained and tested using the nnUNet framework. Among the included 120 patients, contrast-enhanced CT images were available for 77. With this dataset, 3D UNET models with nnUNet, UNETR, and SwinUNETR models were trained again with 5-fold cross-validation, for gross tumor volume (GTV) segmentation. Two patients with rare pathologies known to be nasopharyngeal cancer but not classifiable with any of the differentiated, non-keratinized differentiated, or keratinized pathologies were not included in the training of the DFS prediction ML models which will be developed later.

The distribution of clinical data for the 75 patients included in the DFS prediction is shown in **Figure 18** and **Figure 19**.

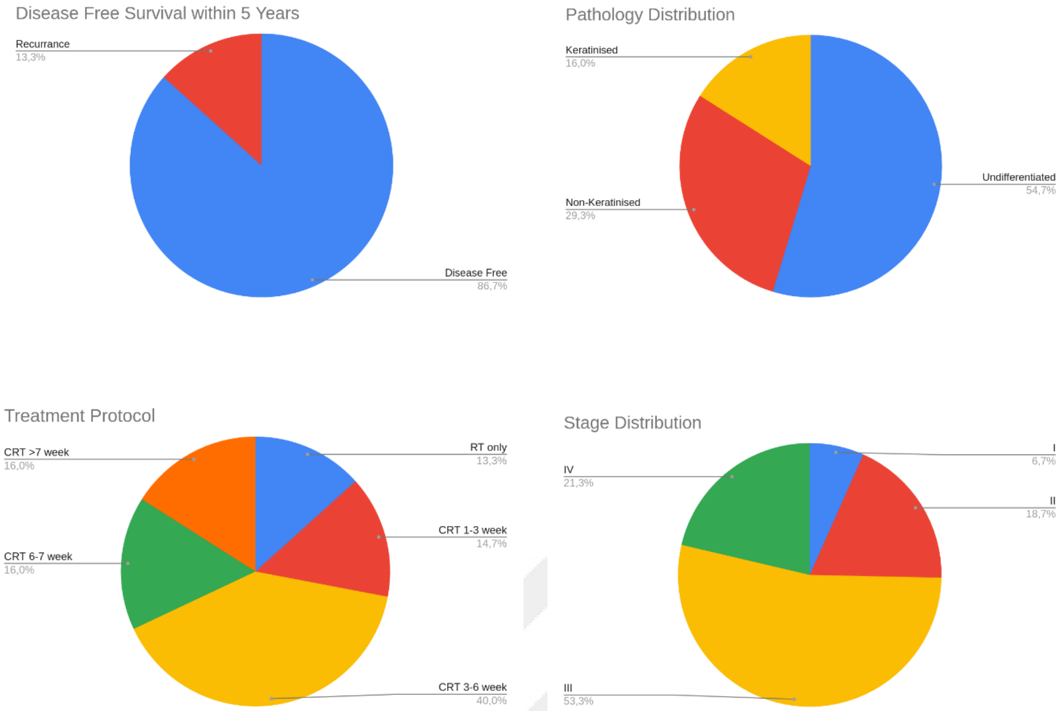

Figure 18. Distributions of different cathegorically encoded clinical features.

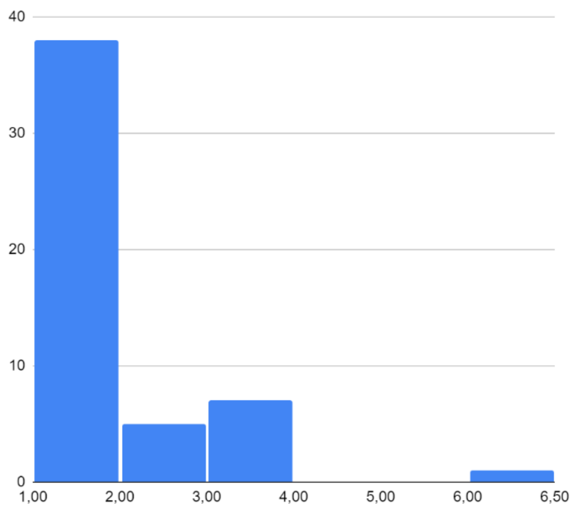

Figure 19. Histogram of Waiting Time to RT after Diagnosis (Days).

Among the patients included in the DFS prediction, 54 were male and 21 were female. Cranial nerve involvement was present in 8 patients at diagnosis. Among them, 38 patients were at T1 stage, 10 patients at T2 stage, 14 patients at T3 stage, and 13 patients were at T4 stage, while 16 patients were at N1 stage, 42 patients were at N2 stage, and 4 patients were at N3 stage. Lymph node metastasis was absent in 13 patients.

One patient received 1 course, 8 patients received 3 courses, 2 patients received 4 courses, 1 patient received 6 courses, and 1 patient received 18 courses of neoadjuvant chemotherapy, while 62 patients did not receive neoadjuvant chemotherapy. Additionally, 1 patient received 1 course, 3 patients received 2 courses, 12 patients received 3 courses, 1 patient received 4 courses, and 1 patient received 6 courses of adjuvant chemotherapy, whereas 57 patients did not receive adjuvant chemotherapy.

Patients included in the DFS prediction received radiotherapy doses of 70 Gy for 71 patients, 66 Gy for 3 patients, and 62.5 Gy for 1 patient in the GTV region. Similarly, among these patients, radiotherapy doses of 60 Gy were administered for 69 patients, 66 Gy for 3 patients, 57.6 Gy for 1 patient, 58 Gy for 1 patient, and 59.4 Gy for 1 patient in the CTV region. Regarding the applied radiotherapy technique, the distribution is as follows: 30 patients received 4ARC, 22 patients received 3ARC, 13 patients received IMRT, and 9 patients received Other-ARC techniques.

For OAR segmentation task, using 5-fold cross-validation, nnUNet-based 3D UNET algorithms were developed with training sets consisting of 96 CT series and validated with validation sets consisting of 24 CT series. In the developed algorithms, DSC scores for different classes varied between 0.99 and 0.47 (**Figure 20**). In the conducted ANOVA and post-hoc analysis, there was no significant difference in validation DSC scores between folds except for the optic chiasm class ($p>0.05$). However, in the optic chiasm class, DSC scores were noticeably lower compared to other classes, and there was a significant difference in validation DSC scores between folds ($p<0.05$) (**Figure 21**). Mean DSC score values for each class for 5-fold cross-validation are presented in **Table 5**.

Table 5. 3D UNET model 5-fold cross validation mean DSC scores for each organ at risk.

Segmented Class	**DSC Score (Mean of 5 Fold CV)**	**Segmented Class**	**DSC Score (Mean of 5 Fold CV)**
Body	0,98	**Brain**	0,99
Lens (Right)	0,77	**Larynx**	0,77
Lens (Left)	0,75	**Spinal Cord**	0,82
Cochlea (Right)	0,71	**Eusophagus**	0,74
Cochlea (Left)	0,73	**Lungs**	0,98
Oral Cavity	0,84	**Hypophysis**	0,74
Parotid gland (Left)	0,83	**Brainstem**	0,86
Parotid gland (Right)	0,81	**Eye Bulb (Left)**	0,88
Mandible	0,88	**Eye Bulb (Right)**	0,88
Submandibular Gland (Left)	0,84	**Optic Nerve (Left)**	0,75
Submandibular Gland (Right)	0,84	**Optic Nerve (Right)**	0,74
Optic Chiasm	0,47	**Mean**	0,81 (std = 0,11)

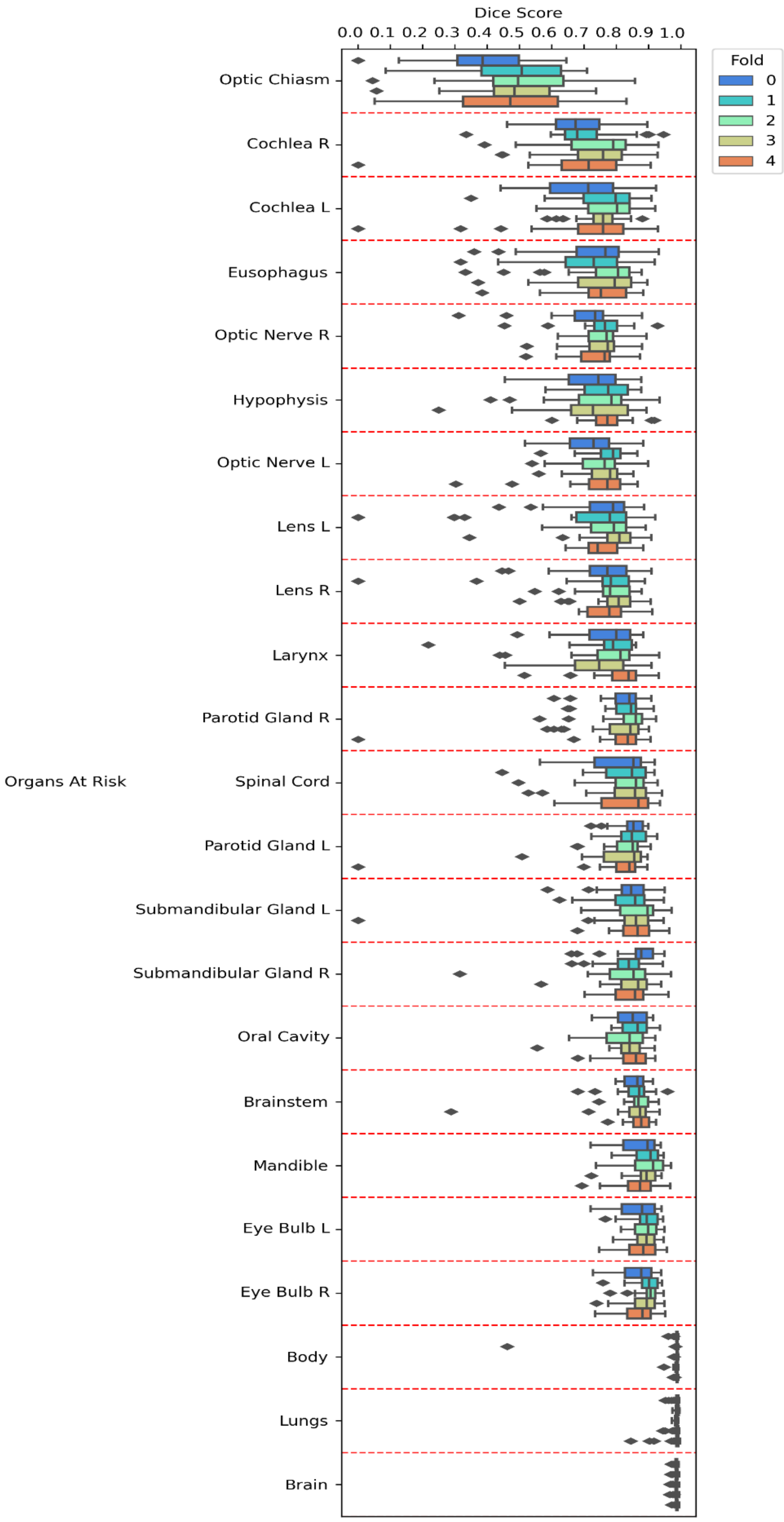

Figure 20. Boxplots of DSC scores at each cross-validation fold for organs at risk segmentation with 3D UNET.

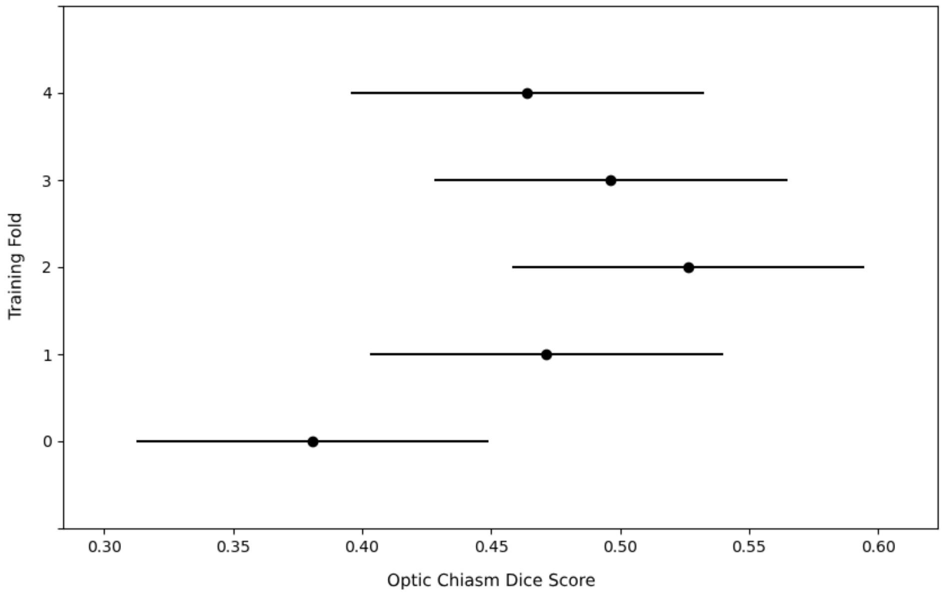

Figure 21. Distribution of DSC scores among validation sets for Optic Chiasm segmentation.

In nasopharyngeal GTV segmentation, three different architectures (3D UNET, UNETR, and SwinUNETR) of automatic segmentation algorithms were trained using a 5-fold cross-validation method with an 80%/20% training/validation ratio. Examples of loss curves generated during the training of the models are shown in **Figure 22**, **Figure 23**, and **Figure 24**.

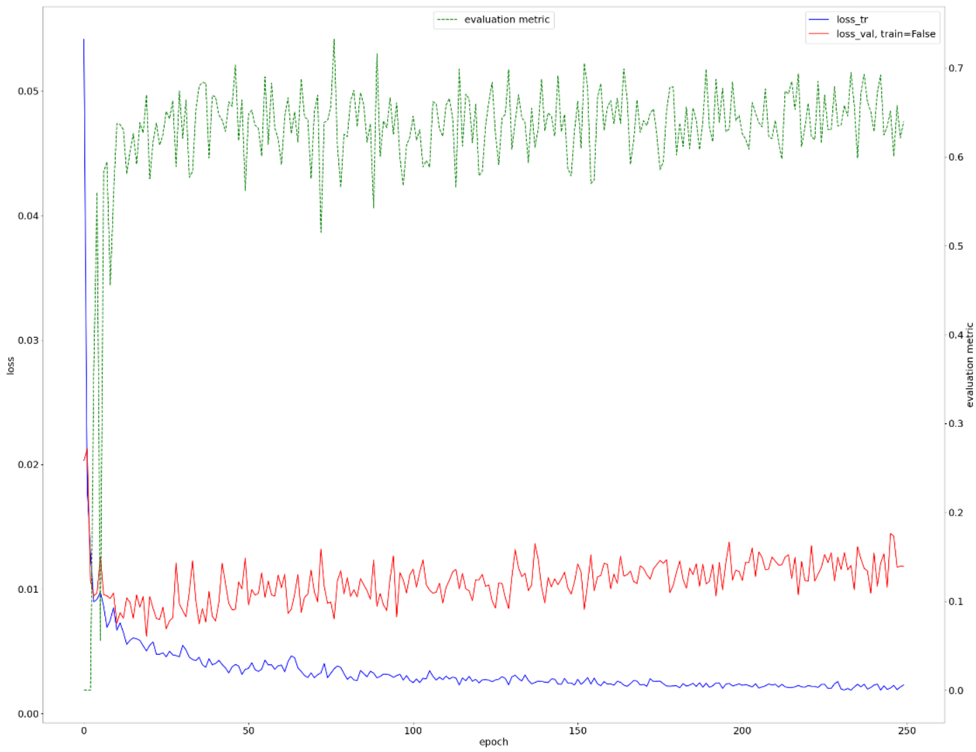

Figure 22. Train and validation loss curve example of 3D UNET algorithm (created by using nnUNet framework).

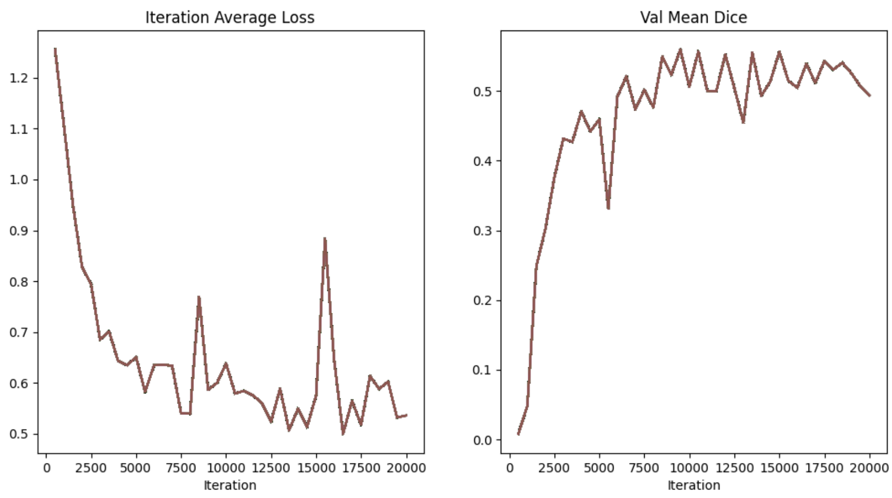

Figure 23. Example of validation loss and mean DSC curve for UNETR algorithm.

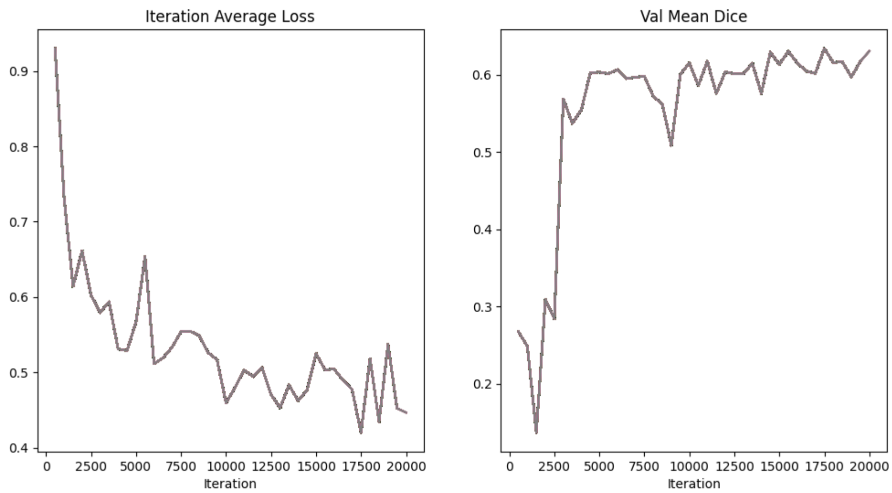

Figure 24. Example of validation loss and mean DSC curve for SwinUNETR algorithm.

The DSC scores for nasopharyngeal GTV for each fold are shown in Table 6. Accordingly, the average DSC score was highest in the SwinUNETR model with 0.637 and lowest in the UNETR model with 0.559.

Table 6. 5-fold cross validation DSC scores of different models for nasopharyngeal GTV.

Training Fold	**3D UNET**	**UNETR**	**SwinUNETR**
Fold 0	0,58	0,591	**0,618**
Fold 1	0,625	0,602	**0,634**
Fold 2	**0,614**	0,537	0,591
Fold 3	0,672	0,636	**0,689**
Fold 4	0,622	0,62	**0,653**
Mean	0,622	0,559	**0,637**

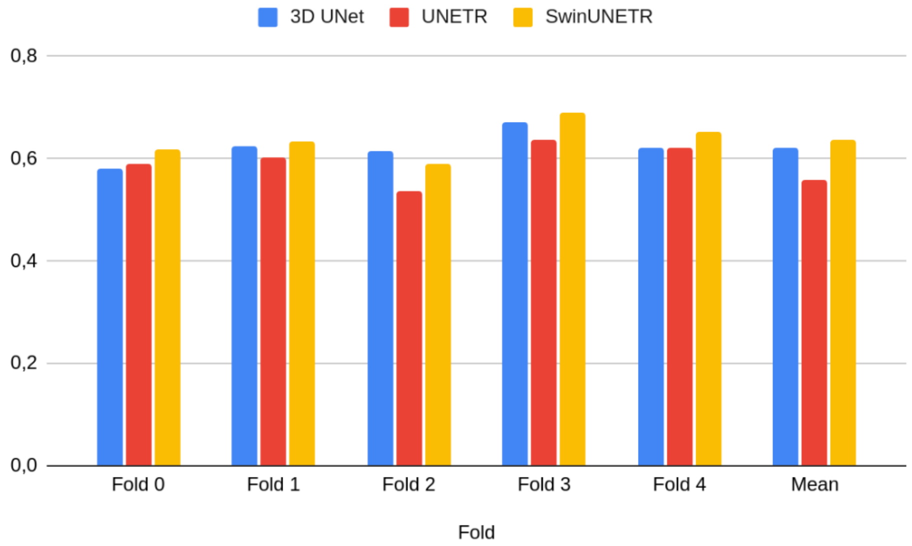

Figure 25. Bar graph of nasopharyngeal GTV segmentation model DSC scores.

For the 5-year disease-free survival (DFS) prediction, features were extracted from 2D slices of manually and automatically segmented NPC (Nasopharyngeal Carcinoma) tumor tissues from 75 CT volumes using radiomics, resulting in 1059 features, and using the DINOv2 model, resulting in 386 features. Fourteen clinical features, were encoded using OneHotEncoding (only cathegorical features), resulting in a total of 21 clinical features.

Regarding the relationship with DFS, the clinical features showed the following correlations: "Waiting Time To RT After Diagnosis" had a moderate positive correlation (r=0.48), while "4 courses of Adjuvant Chemotherapy", "Positive Cranial Nerve Involvement", "T stage 4", and "18 courses of Neoadjuvant Chemotherapy" showed a low positive correlation ($0.4 > r \geq 0.2$). On the other hand, "4ARC RT Technique" and "60Gy CTV" showed a low negative correlation ($-0.2 \geq r > -0.4$). Other clinical features showed very low correlation ($-0.2 < r < 0.2$) (Figure 26).

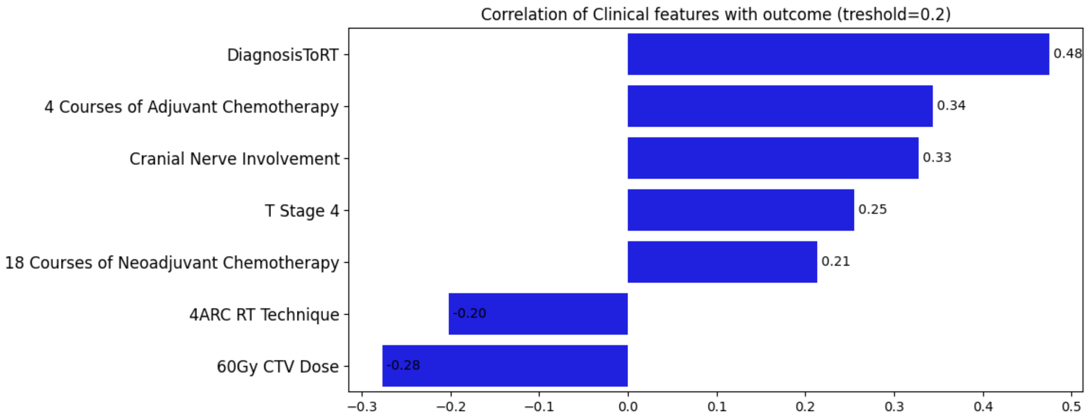

Figure 26. Correlation of clinical features with 5-year disease free survival of nasopharyngeal carcinoma patients treated with radiotherapy.

All DINOv2 features extracted from both manual and SwinUNETR model-based automatic segmentations showed very low correlation with DFS. From the radiomics features extracted from manual segmentations, only one feature (logarithm_gldm_DependenceNonUniformity, r=0.2) showed a low correlation with DFS, while from the radiomics features extracted from automatic segmentations, 24 features showed a low correlation with DFS ($0.2 \geq |r| > 0.4$) (**Figure 27**).

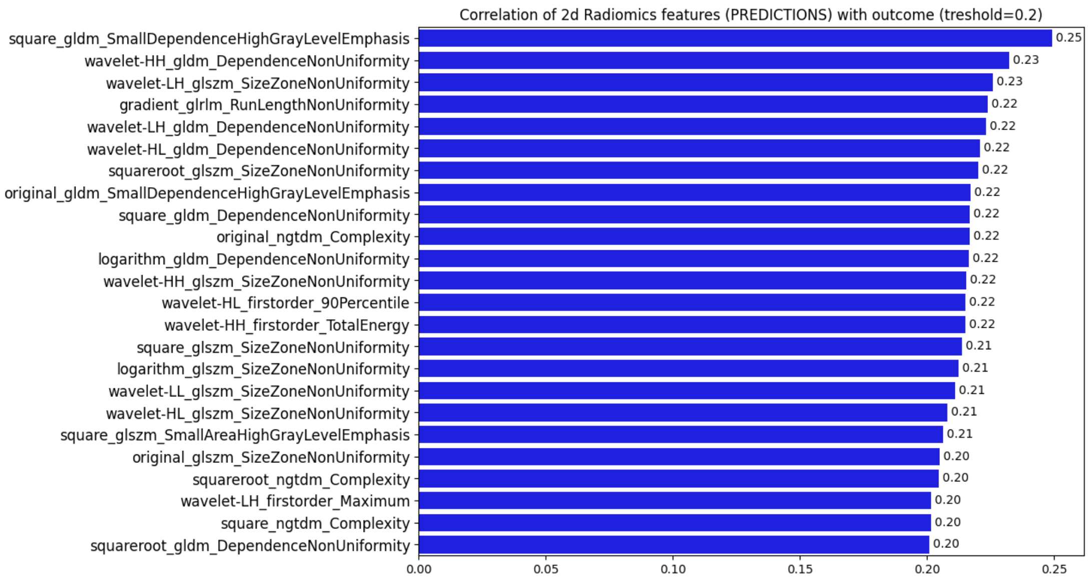

Figure 27. Correlation of 2D radiomics features extracted from predicted segmentations with 5-year disease free survival of nasopharyngeal carcinoma patients treated with radiotherapy.

A Mann-Whitney U analysis was conducted to test whether features extracted from manual and automatic segmentations showed statistically significant differences. According to the analysis, out of the 386 extracted DINOv2 features, 112 (29%) did not show significant differences between manual and automatic segmentations, while out of the 1059 radiomics features, 160 (15%) did not show significant differences ($p>0.05$) (Figure 28 and Figure 29).

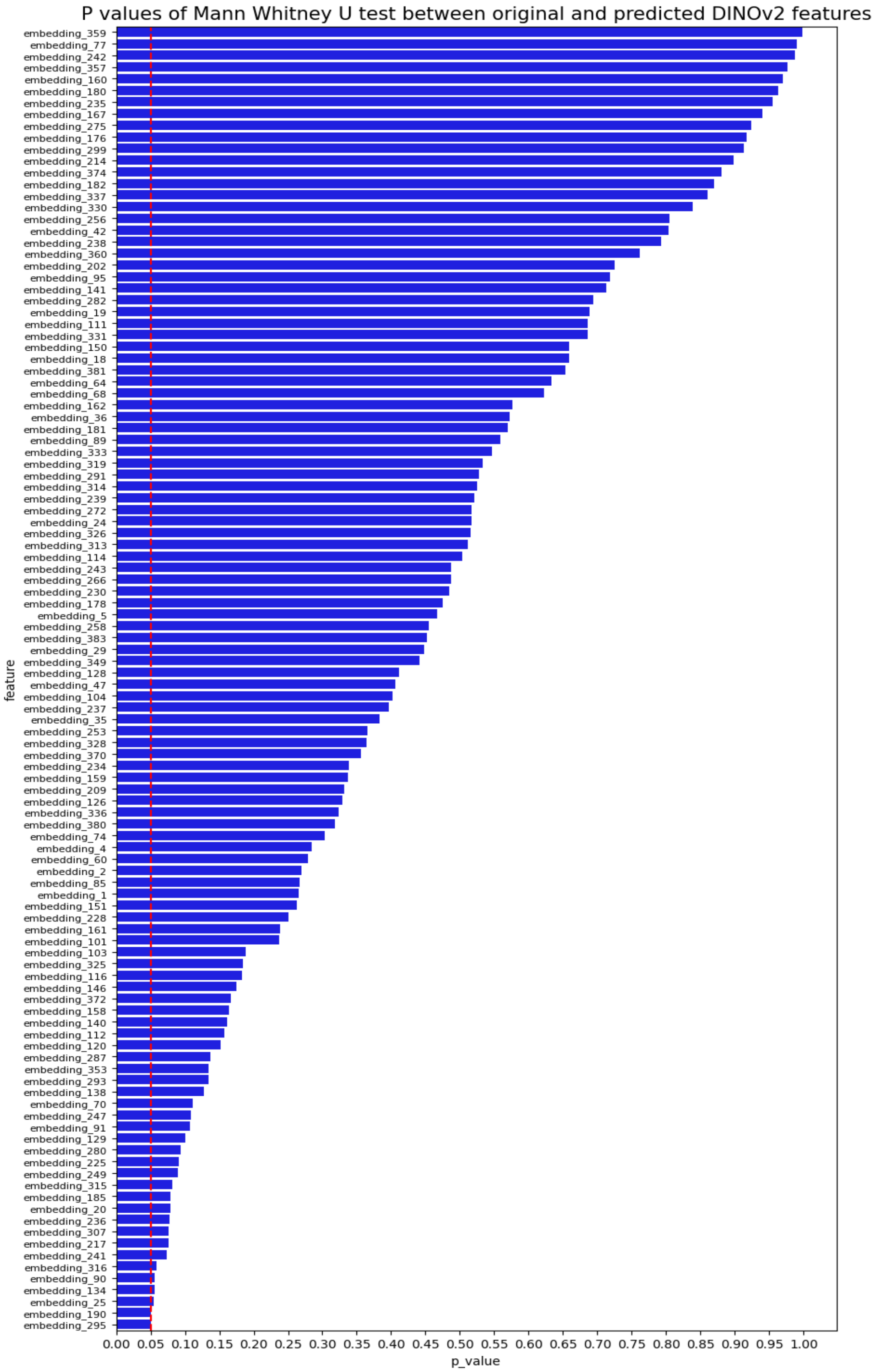

Figure 28. P values of DINOv2 features from MannWhitneyU analysis conducted for testing the difference between features extracted from manual and automatic segmentations.

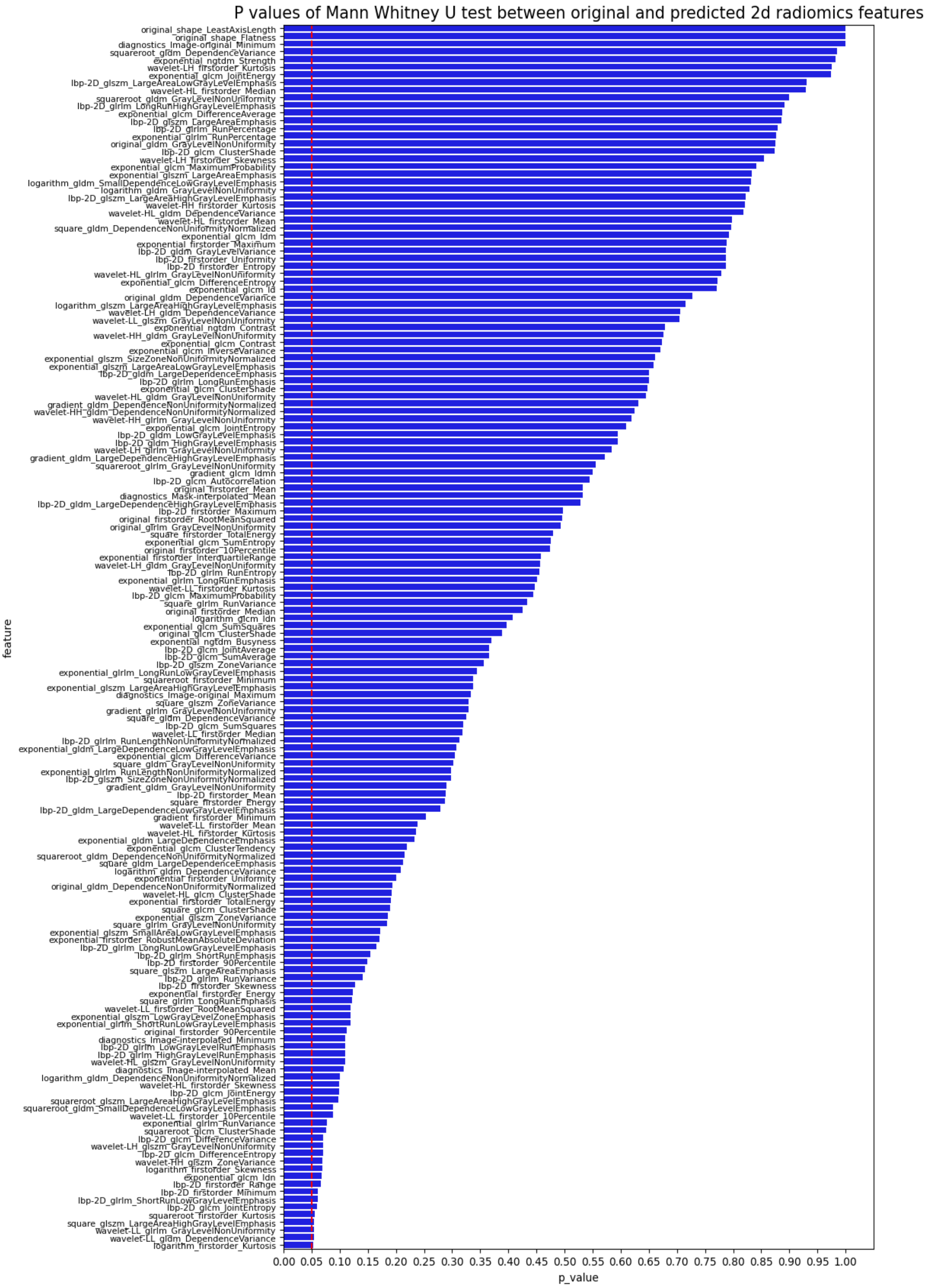

Figure 29. P values of radiomics features from MannWhitneyU analysis conducted for testing the difference between features extracted from manual and automatic segmentations.

Using radiomics + clinical features and DINOv2 + clinical features, DFS prediction models were trained with 5-fold cross-validation (80% training, 20% validation). Since features were obtained from 2D slices for each patient, the slices were grouped based on originated CT volumes, and GroupKFold cross-validation was preferred to prevent data leakage in the separation of training and validation sets. While the slice groups distributed to cross-validation folds remained the same for slices obtained from manual and automatic segmentations, the number of slices allowing the feature extraction varied within the groups. Accordingly, the distribution of the number of 2D slices obtained from manual and automatic segmentations for DINOv2 and Radiomics feature extraction in each validation set is shown in **Table 7** and **Table 8**.

Table 7. Number of 2D slices that features can be extracted using DINOv2 model for each validation set in 5-fold cross validation.

Number Of Slices At Each Validation Set	**Fold 0**	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Total Number Of Slices**
2D Slices From Manual Segmentations	496	451	442	423	407	2219
2D Slices From Automated Segmentations	469	466	464	463	461	2323

Table 8. Number of 2D slices that features can be extracted using radiomics model for each validation set in 5-fold cross validation.

Number Of Slices At Each Validation Set	**Fold 0**	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Total Number Of Slices**
2D Slices From Manual Segmentations	437	394	391	380	370	1972
2D Slices From Automated Segmentations	450	444	442	436	434	2206

During preprocessing, initially z-normalization, feature selection with Lasso model (alpha=0.01), and upsampling with SMOTENC were applied. Since the upsampling has negatively effected model performances in the trials, it is excluded at further steps. As a result, in the model created using Radiomics and Clinical features, a total of 48 features were selected. Among these features, 30 were radiomics features and 18 were clinical features (**Table 9**). In the model created using DINOv2 + Clinical features, a total of 77 features were selected. Among these features, 57 were DINOv2 features and 20 were clinical features (**Table 10**). Same clinical features except for gender and pathological type of tumor was selected consistently in both feature sets.

Table 9. Selected features for Radiomics + Clinical Feature models.

Selected Clinical Features	2 Courses of Adjuvant Chemotherapy	Selected Radiomics Features	exponential_glrlm_GrayLevelNonUniformity
	3 Courses of Adjuvant Chemotherapy		exponential_glszm_SmallAreaLowGrayLevelEmphasis
	6 Courses of Adjuvant Chemotherapy		gradient_firstorder_10Percentile
	60 Gy CTV		gradient_glrlm_RunVariance
	N Stage 1		gradient_glszm_GrayLevelNonUniformity
	N Stage 3		gradient_glszm_LowGrayLevelZoneEmphasis
	T Stage 2		gradient_glszm_SmallAreaLowGrayLevelEmphasis
	Stage 2		gradient_glszm_ZoneVariance
	Positive Cranial Nerve Involvement		lbp-2D_firstorder_Median
	18 Courses of Neoadjuvant Chemotherapy		lbp-2D_gldm_SmallDependenceLowGrayLevelEmphasis
	3 Courses of Neoadjuvant Chemotherapy		logarithm_firstorder_TotalEnergy
	4 Courses of Neoadjuvant Chemotherapy		logarithm_glszm_ZonePercentage
	6 Courses of Neoadjuvant Chemotherapy		original_glcm_ClusterProminence
	70 Gy GTV		original_glcm_InverseVariance
	CRT 6w-7w		original_ngtdm_Complexity
	Diagnosis To RT		original_shape2D_Sphericity
	IMRT Radiotherapy Technique		original_shape_SurfaceVolumeRatio
	RT only Treatment Protocol		square_gldm_LargeDependenceEmphasis
			square_glrlm_LongRunLowGrayLevelEmphasis
			square_glrlm_ShortRunEmphasis
			square_ngtdm_Complexity
			squareroot_glcm_InverseVariance
			wavelet-HH_glcm_InverseVariance
			wavelet-HL_firstorder_Mean
			wavelet-HL_firstorder_Median
			wavelet-HL_glcm_Correlation
			wavelet-LH_glcm_ClusterShade
			wavelet-LL_glcm_Autocorrelation
			wavelet-LL_glcm_MCC
			wavelet-LL_ngtdm_Busyness

Table 10. Selected features for DINOv2 + Clinical Feature models.

Selected Clinical Features	2 Courses of Adjuvant Chemotherapy
	3 Courses of Adjuvant Chemotherapy
	6 Courses of Adjuvant Chemotherapy
	Female Sex
	60 Gy CTV
	N Stage 1
	N Stage 3
	T Stage 2
	Stage 2
	Positive Cranial Nerve Involvement
	18 Courses of Neoadjuvant Chemotherapy
	3 Courses of Neoadjuvant Chemotherapy
	4 Courses of Neoadjuvant Chemotherapy
	6 Courses of Neoadjuvant Chemotherapy
	70 Gy GTV
	Indifferentiated Pathology
	CRT 6w-7w
	DiagnosisToRT
	IMRT Radiotherapy Technique
*57 of selected DINOv2 Features not noted down in this table since the features only named by numbers and can't be directly interpreted because of the black-box nature of deep learning models.	

Initially, separate KNN, SVM, Logistic Regression, Random Forest, XGBoost, and Decision Tree models were trained for each feature set. The models that exhibited the best performance during pre-training were selected, and then fine-tuned.

The performance metrics of the preliminary models trained using Radiomics+Clinical Features are shown in **Table 11**.

Table 11. Preliminary models trained using Radiomics+Clinical Features.

Model	Accuracy	AUC	Recall	Precision	F1
K Neighbors Classifier	0,78	**0,66**	0,51	**0,31**	**0,38**
Logistic Regression	0,73	0,64	0,57	0,29	0,36
SVM - Linear Kernel	0,74	0,00	0,58	0,27	0,34
Random Forest Classifier	0,79	0,66	0,32	0,18	0,23
Extreme Gradient Boosting	0,74	0,58	0,32	0,15	0,20
Decision Tree Classifier	0,72	0,56	0,32	0,14	0,19

The KNN algorithm, which showed the best performance with an F1 score of 0.38, has been fine-tuned. The results of 5-fold cross-validation for this algorithm are shown in **Table 12**.

Table 12. Performance metrics of trained and fine-tuned K-Nearest Neighbors model with Radiomics+Clinical features.

Fold Number	F1 Score	ROC AUC	Precision	Recall
0	0,44	0,47	0,41	0,47
1	0,45	0,50	0,41	0,50
2	0,65	0,64	0,66	0,64
3	0,41	0,42	0,40	0,42
4	0,91	0,88	0,93	0,89
Mean	0,57	0,58	0,56	0,58
Std.	0,21	0,19	0,23	0,19

The performance metrics of the K-Nearest Neighbors model with Radiomics+Clinical features on features extracted from predicted segmentations are shown in **Table 13**.

Table 13. Performance metrics of the K-Nearest Neighbors model predictions on features extracted from automatic segmentations.

Fold Number	F1 Score	ROC AUC	Precision	Recall
0	0,44	0,48	0,40	0,48
1	0,44	0,50	0,39	0,50
2	0,66	0,65	0,68	0,65
3	0,45	0,46	0,44	0,46
4	0,74	0,70	0,81	0,70
Mean	0,55	0,56	0,54	0,56
Std.	0,14	0,11	0,19	0,11

In the t-test analysis, there was no statistically significant difference between the K-Nearest Neighbors (KNN) model predictions on features extracted from manual and automatic segmentations ($p>0.05$). **Figure 30** demonstrates confusion matrices of cumulative predictions from 5-fold cross validation both for manual segmentations and automatically predicted segmentation.

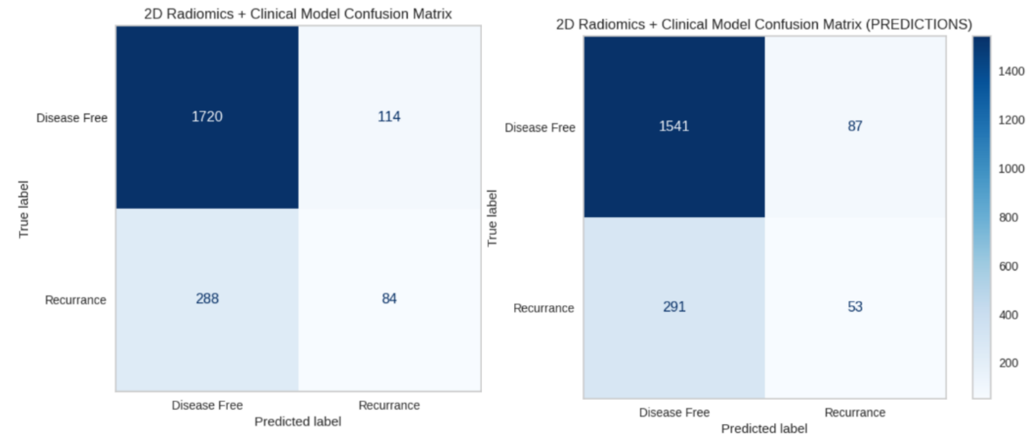

Figure 30. Confusion matrices of KNN model predictions. Left: Predictions on features from manual segmentations, Right: Predictions on features from automatic segmentations.

The performance metrics of the preliminary models trained using DINOv2+Clinical Features are shown in

Table 14.

Model	Accuracy	AUC	Recall	Precision	F1
Logistic Regression	0,76	0,71	0,51	0,32	0,37
SVM -Linear Kernel	0,74	0,00	0,48	0,26	0,32
K Neighbors Classifier	0,79	0,64	0,37	0,35	0,32
Random Forest Classifier	0,83	0,63	0,29	0,29	0,29
Extreme Gradient Boosting	0,76	0,56	0,31	0,19	0,22
Decision Tree Classifier	0,78	0,57	0,25	0,18	0,20

Table 14. Preliminary models trained using DINOv2+Clinical Features.

The Logistic Regression (LR) algorithm, which showed the best performance with an F1 score of 0.37, has been fine-tuned. The results of 5-fold cross-validation for this algorithm are shown in **Table 15**.

Table 15. Performance metrics of trained and fine-tuned Logistic Regression model with DINOv2+Clinical features.

Fold Number	**F1 Score**	**ROC AUC**	**Precision**	**Recall**
0	0,43	0,45	0,41	0,45
1	0,46	0,50	0,42	0,50
2	0,66	0,73	0,65	0,73
3	0,60	0,70	0,62	0,70

4	0,74	0,70	0,83	0,70
Mean	0,58	0,62	0,59	0,62
Std.	0,13	0,13	0,18	0,13

The performance metrics of the LR model with DINOv2+Clinical features on features extracted from predicted segmentations are shown in **Table 16**.

Table 16. Performance metrics of the Logistic Regression model predictions on features extracted from automatic segmentations.

Fold Number	**F1 Score**	**ROC AUC**	**Precision**	**Recall**
0	0,43	0,47	0,40	0,47
1	0,46	0,50	0,42	0,50
2	0,72	0,77	0,70	0,77
3	0,61	0,76	0,63	0,76
4	0,60	0,58	0,83	0,58
Mean	0,56	0,62	0,60	0,62
Std.	0,12	0,14	0,18	0,14

In the t-test analysis, there was no statistically significant difference between the LR model predictions on features extracted from manual and automatic segmentations ($p>0.05$). Figure 31 demonstrates confusion matrices of cumulative predictions from 5-fold cross validation both for manual segmentations and automatically predicted segmentation.

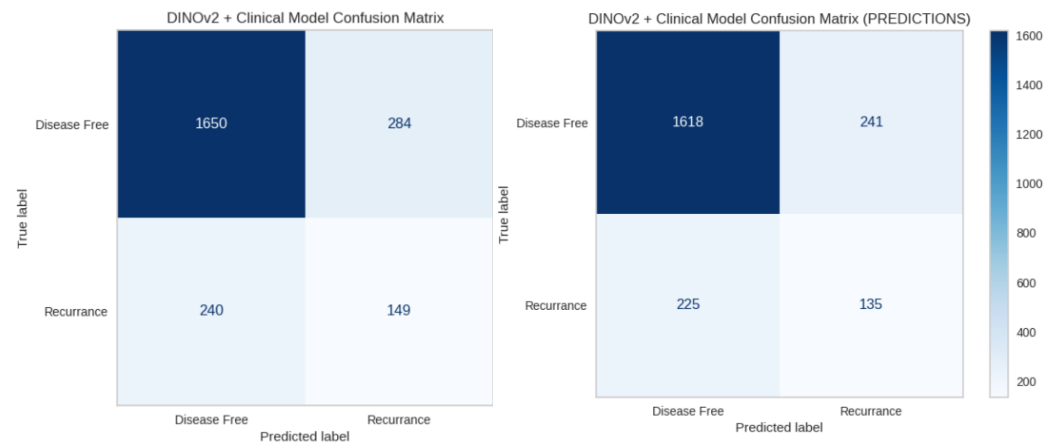

Figure 31. Confusion matrices of LR model predictions. Left: Predictions on features from manual segmentations, Right: Predictions on features from automatic segmentations.

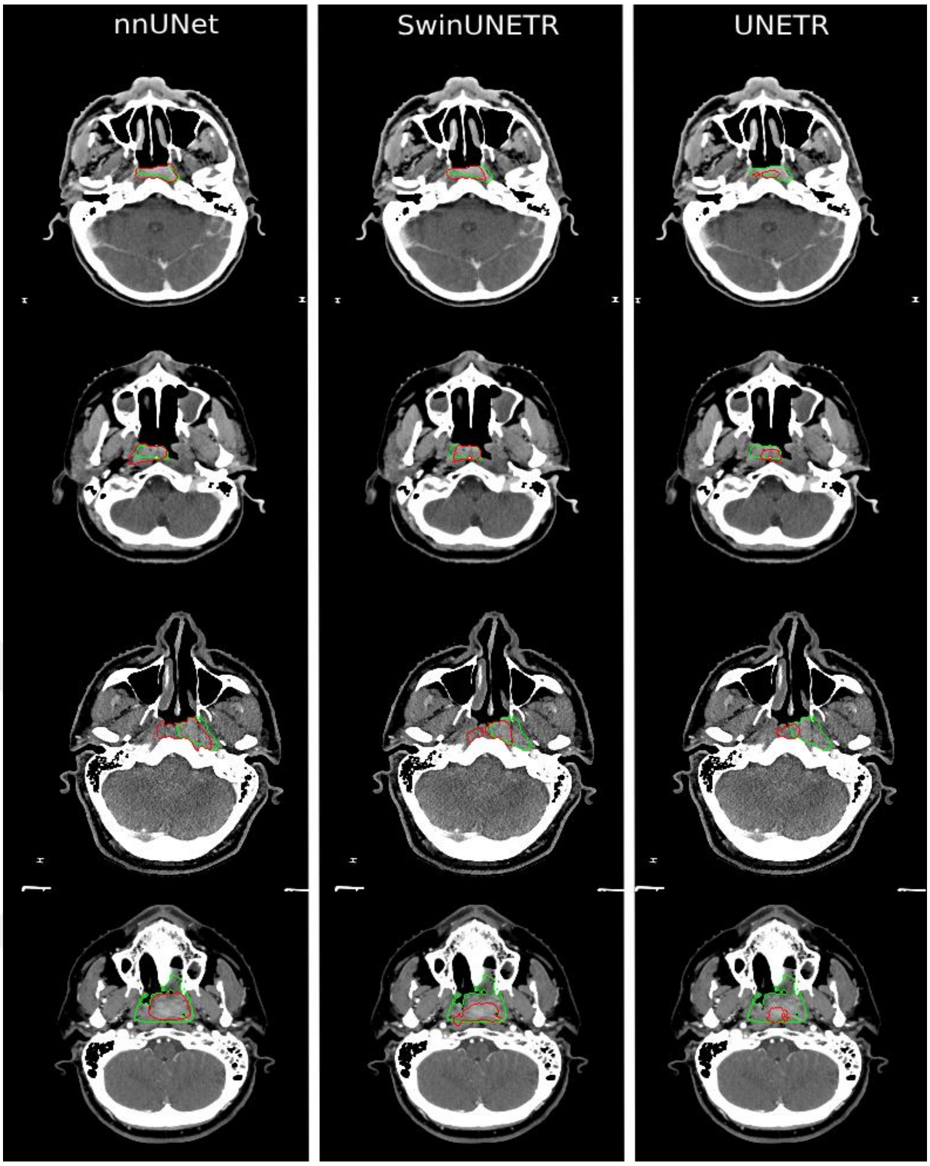

Figure 32. Sample ground truth and predicted NPC segmentations of 3D-UNET (nnUNet), SwinUNETR and UNETR models. Green: Ground truth, Red: Prediction.

DISCUSSION

This research marks an advancement in the automatic semantic segmentation of target tumor volumes in radiotherapy planning for nasopharyngeal cancer (NPC), utilizing cutting-edge deep learning architectures such as U-Net, UNETR, and SwinUNETR, alongside with contributing the cumulative knowledge in automated segmentation of organs-at-risk by deep learning architectures. These models, especially the vision transformer-based architectures, have demonstrated promising segmentation performances, highlighting their potential to enhance the precision and efficiency of radiotherapy planning.

In the use of contextual information, models in the U-Net architecture developed by Ronneberger et. al. (2015) have been replaced by models in the vision transformer structure for automatic biomedical image segmentation today (6,11,12,58–60). In the literature, these models can demonstrate higher performance than architectures with convolutional structures in different modalities and organ segmentation tasks by taking the advantage of their capability to capture long range dependencies. Based on this information, in our study, the SwinUNETR model, whose architecture was developed by Hatamizadeh et. al. (2022) (59), has shown higher performance in the automatic segmentation of nasopharyngeal cancer using contrast-enhanced CT images, compared to 3D-UNET and UNETR architectures. The higher performance of SwinUNETR model can be attributed to ability of extracting features at different resolutions and shifted windowing mechanism, as stated in the study of Hatamizadeh et. al. (2022) (59). To our knowledge, there is no other study in the literature that proposes the SwinUNETR and UNETR model for nasopharyngeal cancer segmentation in CT images. However, in their article published for NPC segmentation in MRI images, Zeng et. al. (2023) have demonstrated the performance of various U-Net architectures along with Swin-UNET (0.819 DSC score) and UNETR (0.837 DSC score) models, and stated that their own developed transformer-based DCTR U-Net (0.852 DSC score) model performed superiorly to these models (65). The architectures based on convolutional U-Net developed in this study have shown lower performance compared to these models (<0.81 DSC score).

In the literature, the DSC scores of automatic segmentation algorithms based on deep learning developed for NPC GTV and CTV segmentation in radiotherapy planning CT images vary between 0.57 and 0.84 (93–97). The highest average DSC score we could achieve in our study with the SwinUNETR model is approximately 0.64. A meta-analysis observed that in studies with a similar methodology, the DSC scores often fell below 0.6 in studies with a sample size of less than 100 (96). We attribute our result, which is relatively lower compared to studies in the literature, to the small number of data (only 77 contrast-enhanced CT images were used for NPC segmentation in our study) and the data being in an uncurated state directly from clinical practice without being specifically prepared for research.

Studies have shown that models developed for the segmentation of organs at risk in nasopharyngeal cancer radiotherapy planning CT images have average DSC scores ranging from 0.76 to 0.86 (66,67,98). When examined on an organ basis, it has been found in studies that the DSC scores for visually very small structures such as the optic chiasm and optic nerves, as well as for organs like the esophagus and Eustachian tube, which are both narrow in cross-sectional area and can appear in complex shapes, are significantly lower compared to other organs (66,67,98). Using the nnUNet framework, we also obtained similar results in line with the literature with our models in 3D UNET architecture trained by using contrast-enhanced and non-contrast CT images of 120 patients. In our developed models, the lowest average DSC score was obtained in the segmentation of the optic chiasm (0.47), and the highest DSC score (0.99) was in brain segmentation. As previously mentioned, this situation may depend on the size of the organ to be segmented, as can be seen in the case of the optic chiasm, as well as the clarity of the boundaries of the organ in CT imaging, the cross-sectional area, and the contrast difference created by the surrounding structures. In clinical practice, radiation oncologists also benefit from the fusion of MR and PET examinations with the relevant CT images for more accurate manual segmentation of such structures in a 3D plane. Therefore, it can be said that techniques incorporating the automated segmentation with multimodal input might be more successful in the segmentation of small and complex organs at risk. There is a clear need for further research in this area.

Studies on the estimation of patient outcomes and disease-free survival through the combination of features extracted from CT or MR images with a set of clinical characteristics of patients appear in the literature as novel approaches (83,84,99,100). Building upon the innovative strategies identified in the literature for enhancing the estimation of patient outcomes and disease-free survival through imaging and clinical data integration, this study further explores the potential of combining Radiomics or DINOv2 feature sets with clinical characteristics to predict nasopharyngeal carcinoma outcomes. Notably, the model leveraging Radiomics and Clinical features selected 48 distinct features, comprising 30 radiomics and 18 clinical features, whereas the DINOv2 and Clinical features model incorporated a broader set of 77 features, with a notable increase in DINOv2 features to 57 and a consistent selection of 20 clinical features, excluding gender and the pathological type of tumor.

The variance in feature selection between the two models underscores the unique contributions of the DINOv2 architecture in capturing complex patterns not represented by radiomics features alone. This observation is particularly relevant when considering the findings from the literature (83,84,99,100), which emphasize the robustness of MRI-based radiomics signatures in predicting progression-free survival (PFS) for patients with nonmetastatic nasopharyngeal carcinoma. The integration of DINOv2 features suggests an expanded analytical scope, capturing subtle nuances within the imaging data that may be pivotal for prognosis.

The comparative analysis of machine learning algorithms revealed that the K-Nearest Neighbors (KNN) algorithm, fine-tuned on the Radiomics + Clinical features set, achieved an F1 score of 0.57, with corresponding performance metrics indicating balanced precision and recall. This outcome is juxtaposed with the performance of the Logistic Regression (LR) model, fine-tuned on the DINOv2 + Clinical features set, which demonstrated a superior F1 score of 0.58, alongside higher AUC, precision, and recall metrics.

Remarkably, the t-test analysis revealed no statistically significant difference in the predictive accuracy of model predictions on features extracted from manual versus automatic segmentations. As one of the novel and striking results of our study, this finding challenges the conventional assumption that manual segmentations invariably offer a superior basis for feature extraction, suggesting that features extracted from the segmentations of an advanced automatic segmentation technique can achieve comparable predictive performance.

In light of these findings, this study not only corroborates the pivotal role of imaging and clinical data fusion in enhancing outcome prediction, as documented in the literature, but also emphasizes the promising potential of leveraging advanced deep learning features, such as those derived from DINOv2. The nuanced interplay between these feature sets, alongside the consistent selection of core clinical characteristics, heralds a sophisticated approach to prognostic modeling in nasopharyngeal carcinoma, aligning with the evolving landscape of precision oncology.

CONCLUSION

This study represents a significant contribution to the field of radiotherapy planning for nasopharyngeal cancer (NPC), showcasing the efficacy of cutting-edge deep learning architectures such as U-Net, UNETR, and SwinUNETR in the automatic semantic segmentation of target tumor volumes. It also sheds light on the potential of automated segmentation of organs-at-risk using these deep learning architectures. The introduction of vision transformer-based architectures, particularly SwinUNETR, marks a pivotal advancement, demonstrating superior segmentation performance that could revolutionize the precision and efficiency of radiotherapy planning.

Our findings reveal that the SwinUNETR model outperforms traditional 3D-UNET and UNETR architectures in segmenting nasopharyngeal cancer using contrast-enhanced CT images. This observation is bolstered by a comparative analysis with the literature, where transformer-based models have shown superior performance over conventional convolutional U-Net architectures in various segmentation tasks. Despite achieving a lower average DSC score compared to some literature-reported scores for NPC segmentation, this study's outcome is influenced by the limited and uncurated dataset, highlighting the impact of data quality and quantity on model performance. The study's exploration into the combination of Radiomics or DINOv2 feature sets with clinical characteristics for predicting NPC outcomes further underscores the evolving landscape of precision oncology. The nuanced integration of advanced imaging features with clinical data through sophisticated machine learning algorithms holds promise for enhancing patient outcome prediction.

In conclusion, this research underscores the transformative potential of advanced deep learning architectures and the integration of radiomics and clinical features in refining radiotherapy planning and prognostic modeling for nasopharyngeal carcinoma. It highlights the need for further research into optimizing data quality, exploring multimodal input techniques, and expanding the analytical scope beyond traditional imaging features. Through this, the study contributes to the broader goal of personalized medicine, where treatment plans are not only precise but are also tailored to the individual characteristics of both the tumor and the patient, thereby potentially improving treatment outcomes and quality of life for patients with NPC.

As a future work, we aim to develop new models by creating a hybrid dataset with the data offered as open source in the SegRap2023 competition to overcome the problem of data scarcity and to more accurately evaluate model performances in tumor volume segmentation.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer [Internet]. 2015 Mar 1;136(5):E359–86. Available from: https://doi.org/10.1002/ijc.29210

2. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. The Lancet. 2019 Jul;394(10192):64–80.

3. Hua YJ, Han F, Lu LX, Mai HQ, Guo X, Hong MH, et al. Long-term treatment outcome of recurrent nasopharyngeal carcinoma treated with salvage intensity modulated radiotherapy. Eur J Cancer. 2012 Dec;48(18):3422–8.

4. Harari PM, Song S, Tomé WA. Emphasizing Conformal Avoidance Versus Target Definition for IMRT Planning in Head-and-Neck Cancer. International Journal of Radiation Oncology*Biology*Physics. 2010 Jul;77(3):950–8.

5. Breen SL, Publicover J, De Silva S, Pond G, Brock K, O'Sullivan B, et al. Intraobserver and Interobserver Variability in GTV Delineation on FDG-PET-CT Images of Head and Neck Cancers. International Journal of Radiation Oncology*Biology*Physics. 2007 Jul;68(3):763–70.

6. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In 2015. p. 234–41.

7. Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning <scp>CT</scp> for rectal cancer using deep dilated convolutional neural networks. Med Phys. 2017 Dec 28;44(12):6377–89.

8. Kamnitsas K, Ferrante E, Parisot S, Ledig C, Nori A V., Criminisi A, et al. DeepMedic for Brain Tumor Segmentation. In 2016. p. 138–49.

9. Dong H, Yang G, Liu F, Mo Y, Guo Y. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. In 2017. p. 506–17.

10. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. Physica Medica. 2018 Jun;50:13–9.

11. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, et al. UNETR: Transformers for 3D Medical Image Segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2022. p. 574–84.

12. Cao Hu and Wang Y and CJ and JD and ZX and TQ and WM. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky Leonid and Michaeli T and NK, editor. Computer Vision – ECCV 2022 Workshops. Cham: Springer Nature Switzerland; 2023. p. 205–18.

13. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021. p. 10012–22.

14. Pinto dos Santos D, Dietzel M, Baessler B. A decade of radiomics research: are images really data or just patterns in the noise? Eur Radiol [Internet]. 2021;31(1):1–4. Available from: https://doi.org/10.1007/s00330-020-07108-w

15. Wong KCW, Hui EP, Lo KW, Lam WKJ, Johnson D, Li L, et al. Nasopharyngeal carcinoma: an evolving paradigm. Nat Rev Clin Oncol [Internet]. 2021;18(11):679–95. Available from: https://doi.org/10.1038/s41571-021-00524-x

16. Carioli G, Negri E, Kawakita D, Garavello W, La Vecchia C, Malvezzi M. Global trends in nasopharyngeal cancer mortality since 1970 and predictions for 2020: Focus on low-risk areas. Int J Cancer [Internet]. 2017 May 15;140(10):2256–64. Available from: https://doi.org/10.1002/ijc.30660

17. Gamal AH. Epidemiology and outcomes of nasopharyngeal carcinoma. In: Pharynx-Diagnosis and Treatment. 2021. p. 43–57.

18. Sun X, Su S, Chen C, Han F, Zhao C, Xiao W, et al. Long-term outcomes of intensity-modulated radiotherapy for 868 patients with nasopharyngeal

carcinoma: An analysis of survival and treatment toxicities. Radiotherapy and Oncology [Internet]. 2014;110(3):398–403. Available from: https://www.sciencedirect.com/science/article/pii/S016781401300529X

19. Baroudi H, Brock KK, Cao W, Chen X, Chung C, Court LE, et al. Automated Contouring and Planning in Radiation Therapy: What Is 'Clinically Acceptable'? Vol. 13, Diagnostics. 2023.

20. Ng WT, Chow JCH, Beitler JJ, Corry J, Mendenhall W, Lee AWM, et al. Current Radiotherapy Considerations for Nasopharyngeal Carcinoma †. Vol. 14, Cancers. 2022.

21. Kehr P. Michael Schuenke, Erik Schulte, Udo Schumacher, Laurence M. Ross, Edward D. Lamperti, Ethan Taub: THIEME atlas of anatomy head and neuranatomy. European Journal of Orthopaedic Surgery & Traumatology. 2010;20(1).

22. Harrison K, Pullen H, Welsh C, Oktay O, Alvarez-Valle J, Jena R. Machine Learning for Auto-Segmentation in Radiotherapy Planning. Clin Oncol. 2022;34(2).

23. Dar AS, Padha D. Medical Image Segmentation A Review of Recent Techniques, Advancements and a Comprehensive Comparison. International Journal of Computer Sciences and Engineering. 2019;7(7).

24. Otsu N. THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS. IEEE Trans Syst Man Cybern. 1979;SMC-9(1).

25. Chanapai W, Ritthipravat P. Adaptive thresholding based on SOM technique for semi-automatic NPC image segmentation. In: 8th International Conference on Machine Learning and Applications, ICMLA 2009. 2009.

26. Oghli MG, Fallahi A, Pooyan M. Automatic region growing method using gsmap and spatial information on ultrasound images. In: Proceedings - 2010 18th Iranian Conference on Electrical Engineering, ICEE 2010. 2010.

27. Deng W, Xiao W, Deng H, Liu J. MRI brain tumor segmentation with region growing method based on the gradients and variances along and inside of the

boundary curve. In: Proceedings - 2010 3rd International Conference on Biomedical Engineering and Informatics, BMEI 2010. 2010.

28. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell. 2012;34(11).

29. Shan X, Gong X, Nandi AK. Active Contour Model Based on Local Intensity Fitting Energy for Image Segmentation and Bias Estimation. IEEE Access. 2018;6.

30. Levinshtein A, Stere A, Kutulakos KN, Fleet DJ, Dickinson SJ, Siddiqi K. TurboPixels: Fast superpixels using geometric flows. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009.

31. Ester M, Kriegel HP, Sander J, Xu X. A Density-Based Algorithm for Discovering Clusters A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proceedings - 2nd International Conference on Knowledge Discovery and Data Mining, KDD 1996. 1996.

32. Liu Y, Jiang PT, Petrosyan V, Li SJ, Bian J, Zhang L, et al. DEL: Deep embedding learning for efficient image segmentation. In: IJCAI International Joint Conference on Artificial Intelligence. 2018.

33. Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings - International Conference on Pattern Recognition. 2004.

34. Falcão AX, Stolfi J, De Alencar Lotufo R. The Image Foresting Transform: Theory, Algorithms, and Applications. IEEE Trans Pattern Anal Mach Intell. 2004;26(1).

35. Yi F, Moon I. Image segmentation: A survey of graph-cut methods. In: 2012 International Conference on Systems and Informatics, ICSAI 2012. 2012.

36. Luo W. Comparison for Edge Detection of Colony Images. Journal of Computer Science. 2006;6(9).

37. Musoromy Z, Ramalingam S, Bekooy N. Edge detection comparison for license plate detection. In: 11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010. 2010.

38. Banu S, Apparao G, Syed AS. Interactive Image Segmentation and Edge Detection of Medical Images. International Journal of Advanced Computer Research. 2013;3(4):211–211.

39. Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. Int J Comput Vis. 2006;70(2).

40. Bong CW, Liew CC, Lam HY. Chapter 5 - Ground-glass opacity nodules detection and segmentation using the snake model. In: Yang XS, Papa JP, editors. Bio-Inspired Computation and Applications in Image Processing [Internet]. Academic Press; 2016. p. 87–104. Available from: https://www.sciencedirect.com/science/article/pii/B9780128045367000053

41. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. Int J Comput Vis. 1988;1(4).

42. Caselles V, Kimmel R, Sapiro G. Geodesic Active Contours. Int J Comput Vis. 1997;22(1).

43. Comprehensive Biomedical Physics. Comprehensive Biomedical Physics. 2014.

44. Deng L, Yu D. Deep learning: Methods and applications. Vol. 7, Foundations and Trends in Signal Processing. 2013.

45. Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2012.

46. Habibi Aghdam H, Jahani Heravi E. Guide to Convolutional Neural Networks. Guide to Convolutional Neural Networks. 2017.

47. Fukushima K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. Neural Networks. 1988;1(2).

48. Schmidhuber J. Deep Learning in neural networks: An overview. Vol. 61, Neural Networks. 2015.

49. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2016.

50. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods. 2019;16(1).

51. Kaur A, Kaur L, Singh A. GA-UNet: UNet-based framework for segmentation of 2D and 3D medical images applicable on heterogeneous datasets. Neural Comput Appl. 2021;33(21).

52. Iantsen A, Visvikis D, Hatt M. Squeeze-and-Excitation Normalization for Automated Delineation of Head and Neck Primary Tumors in Combined PET and CT Images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2021.

53. Michael E, Ma H, Li H, Kulwa F, Li J. Breast Cancer Segmentation Methods: Current Status and Future Potentials. Vol. 2021, BioMed Research International. 2021.

54. Samarasinghe G, Jameson M, Vinod S, Field M, Dowling J, Sowmya A, et al. Deep learning for segmentation in radiation therapy planning: a review. Vol. 65, Journal of Medical Imaging and Radiation Oncology. 2021.

55. Liu X, Li KW, Yang R, Geng LS. Review of Deep Learning Based Automatic Segmentation for Lung Cancer Radiotherapy. Vol. 11, Frontiers in Oncology. 2021.

56. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, et al. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. 2018 Sep 27; Available from: http://arxiv.org/abs/1809.10486

57. Li S, Xiao J, He L, Peng X, Yuan X. The Tumor Target Segmentation of Nasopharyngeal Cancer in CT Images Based on Deep Learning Methods. Technol Cancer Res Treat. 2019;18.

58. Liu Y, Zhang Y, Wang Y, Hou F, Yuan J, Tian J, et al. A Survey of Visual Transformers. IEEE Trans Neural Netw Learn Syst. 2023;

59. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2022.

60. Tang Y, Yang D, Li W, Roth HR, Landman B, Xu D, et al. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2022.

61. Boero IJ, Paravati AJ, Xu B, Cohen EEW, Mell LK, Le QT, et al. Importance of radiation oncologist experience Among patients with head-and-neck cancer treated with intensity-modulated radiation therapy. Journal of Clinical Oncology. 2016;34(7).

62. Riegel AC, Berson AM, Destian S, Ng T, Tena LB, Mitnick RJ, et al. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. Int J Radiat Oncol Biol Phys. 2006;65(3).

63. Hong TS, Tomé WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. Radiotherapy and Oncology. 2012;103(1).

64. Bollen H, Willems S, Wegge M, Maes F, Nuyts S. Benefits of automated gross tumor volume segmentation in head and neck cancer using multi-modality information. Radiotherapy and Oncology. 2023;182.

65. Zeng Y, Zeng PH, Shen SD, Liang W, Li J, Zhao Z, et al. DCTR U-Net: automatic segmentation algorithm for medical images of nasopharyngeal cancer in the context of deep learning. Front Oncol. 2023;13.

66. Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. Eur Radiol. 2019;29(4).

67. Luo X, Fu J, Zhong Y, Liu S, Han B, Astaraki M, et al. SegRap2023: A Benchmark of Organs-at-Risk and Gross Tumor Volume Segmentation for Radiotherapy Planning of Nasopharyngeal Carcinoma. arXiv preprint arXiv:231209576. 2023;

68. Ye Y, Xie Y, Zhang J, Chen Z, Xia Y. UniSeg: A Prompt-Driven Universal Segmentation Model as Well as A Strong Representation Learner. In: Greenspan H, Madabhushi A, Mousavi P, Salcudean S, Duncan J, Syeda-Mahmood T, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Cham: Springer Nature Switzerland; 2023. p. 508–18.

69. Yip SSF, Liu Y, Parmar C, Li Q, Liu S, Qu F, et al. Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer. Sci Rep. 2017;7(1).

70. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295(2).

71. Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and deep learning in lung cancer. Vol. 196, Strahlentherapie und Onkologie. 2020.

72. Sah BR, Owczarczyk K, Siddique M, Cook GJR, Goh V. Radiomics in esophageal and gastric cancer. Vol. 44, Abdominal Radiology. 2019.

73. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. Vol. 169, Breast Cancer Research and Treatment. 2018.

74. Sun Y, Reynolds HM, Parameswaran B, Wraith D, Finnegan ME, Williams S, et al. Multiparametric MRI and radiomics in prostate cancer: a review. Vol. 42, Australasian Physical and Engineering Sciences in Medicine. 2019.

75. Zhang YM, Gong GZ, Qiu QT, Han YW, Lu HM, Yin Y. Radiomics for Diagnosis and Radiotherapy of Nasopharyngeal Carcinoma. Vol. 11, Frontiers in Oncology. 2022.

76. Zhuo EH, Zhang WJ, Li HJ, Zhang GY, Jing BZ, Zhou J, et al. Radiomics on multi-modalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups. Eur Radiol. 2019;29(10).

77. Akram F, Koh PE, Wang F, Zhou S, Tan SH, Paknezhad M, et al. Exploring MRI based radiomics analysis of intratumoral spatial heterogeneity in locally advanced nasopharyngeal carcinoma treated with intensity modulated radiotherapy. PLoS One. 2020;15(10 October).

78. Nayak U, Satya Prasad R, Sekhar S. Clinical Radiology of Head and Neck Tumors. Clinical Radiology of Head and Neck Tumors. 2018.

79. Yu TT, Lam SK, To LH, Tse KY, Cheng NY, Fan YN, et al. Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients. Front Oncol. 2019;9.

80. Zhao L, Gong J, Xi Y, Xu M, Li C, Kang X, et al. MRI-based radiomics nomogram may predict the response to induction chemotherapy and survival in locally advanced nasopharyngeal carcinoma. Eur Radiol. 2020;30(1).

81. Yongfeng P, Chuner J, Lei W, Fengqin Y, Zhimin Y, Zhenfu F, et al. The usefulness of pretreatment mr-based radiomics on early response of neoadjuvant chemotherapy in patients with locally advanced nasopharyngeal carcinoma. Oncol Res. 2021;28(6).

82. Ouyang FS, Guo BL, Zhang B, Dong YH, Zhang L, Mo XK, et al. Exploration and validation of radiomics signature as an independent prognostic biomarker in stage III-IVb nasopharyngeal carcinoma. Oncotarget. 2017;8(43).

83. Shen H, Wang Y, Liu D, Lv R, Huang Y, Peng C, et al. Predicting Progression-Free Survival Using MRI-Based Radiomics for Patients With Nonmetastatic Nasopharyngeal Carcinoma. Front Oncol. 2020;10.

84. Ming X, Oei RW, Zhai R, Kong F, Du C, Hu C, et al. MRI-based radiomics signature is a quantitative prognostic biomarker for nasopharyngeal carcinoma. Sci Rep. 2019;9(1).

85. Yang K, Tian J, Zhang B, Li M, Xie W, Zou Y, et al. A multidimensional nomogram combining overall stage, dose volume histogram parameters and radiomics to predict progression-free survival in patients with locoregionally advanced nasopharyngeal carcinoma. Oral Oncol. 2019;98.

86. Zhang L, Dong D, Li H, Tian J, Ouyang F, Mo X, et al. Development and validation of a magnetic resonance imaging-based model for the prediction of distant metastasis before initial treatment of nasopharyngeal carcinoma: A retrospective cohort study. EBioMedicine. 2019;40.

87. Liu Y, Shi H, Huang S, Chen X, Zhou H, Chang H, et al. Early prediction of acute xerostomia during radiation therapy for nasopharyngeal cancer based on delta radiomics from CT images. Quant Imaging Med Surg. 2019;9(7).

88. Shen D, Wu G, Suk H Il. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 2017;19.

89. Apostolopoulos ID, Papathanasiou ND, Papandrianos NI, Papageorgiou EI, Panayiotakis GS. Deep Learning Assessment for Mining Important Medical Image Features of Various Modalities. Diagnostics. 2022;12(10).

90. Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, et al. DINOv2: Learning Robust Visual Features without Supervision. 2024.

91. Huang Y, Zou J, Meng L, Yue X, Zhao Q, Li J, et al. Comparative Analysis of ImageNet Pre-Trained Deep Learning Models and DINOv2 in Medical Imaging Classification. 2024.

92. Jing B, Deng Y, Zhang T, Hou D, Li B, Qiang M, et al. Deep learning for risk prediction in patients with nasopharyngeal carcinoma using multi-parametric MRIs. Comput Methods Programs Biomed. 2020;197.

93. Bai X, Hu Y, Gong G, Yin Y, Xia Y. A deep learning approach to segmentation of nasopharyngeal carcinoma using computed tomography. Biomed Signal Process Control. 2021;64.

94. Wang X, Yang G, Zhang Y, Zhu L, Xue X, Zhang B, et al. Automated delineation of nasopharynx gross tumor volume for nasopharyngeal carcinoma by plain CT combining contrast-enhanced CT using deep learning. J Radiat Res Appl Sci. 2020;13(1).

95. Domoguen JKL, Manuel JJA, Cañal JPA, Naval PC. Automatic segmentation of nasopharyngeal carcinoma on CT images using efficient UNet-2.5D ensemble with semi-supervised pretext task pretraining. Front Oncol. 2022;12.

96. Zamanian M, Abedi I. Convolutional neural networks in auto-segmentation of nasopharyngeal carcinoma tumor — a systematic review and meta-analysis. Oncology in Clinical Practice. 2024 Feb 21;20(1):27–39.

97. Xue X, Qin N, Hao X, Shi J, Wu A, An H, et al. Sequential and Iterative Auto-Segmentation of High-Risk Clinical Target Volume for Radiotherapy of Nasopharyngeal Carcinoma in Planning CT Images. Front Oncol. 2020;10.

98. Zhong T, Huang X, Tang F, Liang S, Deng X, Zhang Y. Boosting-based cascaded convolutional neural networks for the segmentation of CT organs-at-risk in nasopharyngeal carcinoma. Med Phys. 2019;46(12).

99. Yan C, Shen DS, Chen XB, Su DK, Liang ZG, Chen KH, et al. CT-based radiomics nomogram for prediction of progression-free survival in locoregionally advanced nasopharyngeal carcinoma. Cancer Manag Res. 2021;13.

100. Sun MX, Zhao MJ, Zhao LH, Jiang HR, Duan YX, Li G. A nomogram model based on pre-treatment and post-treatment MR imaging radiomics signatures: application to predict progression-free survival for nasopharyngeal carcinoma. Radiation Oncology. 2023;18(1).