SEPTEMBER 2023 M.Sc. in Electrical and Electronics Engineering AHMET SAİD DEDEOĞLU

REPUBLIC OF TÜRKİYE

GAZİANTEP UNIVERSITY

GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES

AUTOMATED SEGMENTATION OF MAXILLARY SINUS FROM COMPUTED TOMOGRAPHY IMAGES: A COMPARATIVE ANALYSIS OF U-NET AND ITS VARIANTS WITH PRE-TRAINED ENCODERS

M.Sc. THESIS

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

BY

AHMET SAİD DEDEOĞLU

SEPTEMBER 2023

AUTOMATED SEGMENTATION OF MAXILLARY SINUS FROM COMPUTED TOMOGRAPHY IMAGES: A COMPARATIVE ANALYSIS OF U-NET AND ITS VARIANTS WITH PRE-TRAINED ENCODERS

M.Sc. Thesis

in

Electrical and Electronics Engineering

Gaziantep University

Supervisor

Asst. Prof. Dr. Serkan ÖZBAY

Co-Supervisor

Assoc. Prof. Dr. Orhan TUNÇ

by

Ahmet Said DEDEOĞLU

September 2023

AUTOMATED SEGMENTATION OF MAXILLARY SINUS FROM COMPUTED TOMOGRAPHY IMAGES: A COMPARATIVE ANALYSIS OF U-NET AND ITS VARIANTS WITH PRE-TRAINED ENCODERS

submitted by **Ahmet Said DEDEOĞLU** in partial fulfillment of the requirements for the degree of Master of Science in **Electrical and Electronics Engineering, Gaziantep University** is approved by,

Prof. Dr. Çiğdem Aykaç
Director of the Graduate School of Natural and Applied Sciences ……….............

Prof. Dr. Ergun ERÇELEBİ
Head of the Department of Electrical and Electronics Engineering ....……..............

Asst. Prof. Dr. Serkan ÖZBAY
Supervisor, Electrical and Electronics Engineering
Gaziantep University ……….............

Assoc. Prof. Dr. Orhan TUNÇ
Co-Supervisor, Faculty of Medicine
Gaziantep University ……….............

Exam Date: 18 September 2023

Examining Committee Members:

Asst. Prof. Dr. Serkan ÖZBAY
Gaziantep University ……….............

Prof. Dr. Ergun ERÇELEBİ
Gaziantep University ……….............

Asst. Prof. Dr. Erkan KIYMIK
Hasan Kalyoncu University ……….............

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Ahmet Said DEDEOĞLU

ABSTRACT

AUTOMATED SEGMENTATION OF MAXILLARY SINUS FROM COMPUTED TOMOGRAPHY IMAGES: A COMPARATIVE ANALYSIS OF U-NET AND ITS VARIANTS WITH PRE-TRAINED ENCODERS

DEDEOĞLU, Ahmet Said
M.Sc. in Electrical and Electronics Engineering
Supervisor: Asst. Prof. Dr. Serkan ÖZBAY
Co- Supervisor: Assoc. Prof. Dr. Orhan TUNÇ
September 2023

63 pages

The maxillary sinus holds an important position in the respiratory system, and its accurate segmentation from computed tomography (CT) images remains challenging. Traditional manual and semi-automatic methods are time-consuming and inconsistent, making them less suitable for clinical applications. Automated segmentation has become possible thanks to the advancements in deep learning, but challenges still exist due to the complex sinus anatomy and inter-individual variations. This study presents the development and evaluation of four different deep learning models for maxillary sinus segmentation: a plain U-Net model, and three U-Net variations that utilize VGG16, ResNet50, and Inception-ResNet v2 as encoders. The main objectives are to evaluate these models for accurate segmentation, to further segment the air section and mucosal inflammation within the MS, to calculate the maxillary sinus volume and opacification ratios, and to demonstrate the potential of deep learning models in maxillofacial radiology. The Inception-ResNet v2 - UNet model demonstrated the best performance with a Dice score of 0.96671 and an Intersection over Union (IoU) score of 0.94004, while all models, including the plain U-Net, have shown noteworthy successes.

Key Words: Maxillary Sinus, U-Net, Transfer Learning, Medical Image Segmentation

ÖZET

BİLGİSAYARLI TOMOGRAFİ GÖRÜNTÜLERİNDEN MAKSİLLER SİNÜSÜN OTOMATİK BÖLÜMLENDİRİLMESİ: U-NET VE ÖNCEDEN EĞİTİMLİ KODLAYICILAR İLE KARŞILAŞTIRILMALI BİR ANALİZ

DEDEOĞLU, Ahmet Said
Yüksek Lisans Tezi, Elektrik ve Elektronik Mühendisliği
Tez Yöneticisi: Dr. Öğretim Üyesi Serkan ÖZBAY
Tez Yönetici Yardımcısı: Doç. Dr. Orhan TUNÇ
September 2023
63 pages

Maksiller sinüs, solunum sisteminde önemli bir yere sahiptir ve bilgisayarlı tomografi görüntülerinden doğru bir şekilde ayırt edilmesi hâlâ zordur. Geleneksel manüel ve yarı otomatik yöntemler çok zaman alıcı ve tutarsızdır, dolayısıyla klinik uygulamalar için pek uygun değillerdir. Derin öğrenmenin gelişmesi sayesinde otomatik segmentasyon mümkün hale gelmiştir, ancak karmaşık sinüs anatomisi ve bireyler arası varyasyonlar gibi nedenlerle hâlâ zorluklar yaşanmaktadır. Bu çalışma, maksiller sinüs segmentasyonu için dört farklı derin öğrenme modelinin geliştirilmesi ve değerlendirilmesini sunmaktadır: düz bir U-Net modeli ve VGG16, ResNet50 ve Inception-ResNet v2'yi kodlayıcı olarak kullanan üç U-Net varyasyonu. Ana hedefler, bu modellerin doğru segmentasyon için değerlendirilmesi, MS içindeki hava bölümü ve mukoza iltihabının ayrıca segmente edilmesi, maksiller sinüs hacminin ve opasifikasyon oranlarının hesaplanması ve derin öğrenme modellerinin maksillofasiyal radyolojideki potansiyelini göstermektir. Inception-ResNet v2 - UNet modeli 0.96671 Dice ve 0.94004 intersection over union (IoU) skorları ile en iyi performansı gösterirken, düz U-Net dahil tüm modeller dikkate değer başarılar göstermiştir.

Anahtar Kelimeler: Maksiller sinüs, U-Net, Öğrenme aktarması, Tıbbi görüntü bölütleme

"Dedicated to my family"

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Asst. Prof. Dr. Serkan ÖZBAY and Assoc. Prof. Dr. Orhan TUNÇ for their excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research.

I would also like to thank my parents. They were always supporting me and encouraging me with their best wishes.

This work was supported by Gaziantep University Scientific Research Projects Coordination Unit. Project Number: MF.YLT.23.07. I would also like to thank Gaziantep University Scientific Research Projects Coordination Unit. for providing financial support throughout my graduate study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CBCT	Cone-Beam Computed Tomography
CNN	Convolutional Neural Network
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DSC	Dice Similarity Coefficient
FESS	Functional Endoscopic Sinus Surgery
FN	False Negative
FP	False Positive
MRI	Magnetic Resonance Imaging
MRCs	Mucous Retention Cysts
MS	Maxillary Sinus
MT	Mucosal Thickening
ReLU	Rectified Linear Unit
ResNet	Residual Network
ROI	Region of Interest
TP	True Positive

CHAPTER 1
INTRODUCTION

This chapter presents an overview of the maxillary sinus (MS), its definition, functions, and the diseases and operations related to it. It further introduces the concept of MS segmentation and its various types, and its importance in medical imaging. Then discusses the problem statement: the need for more efficient and precise segmentation techniques. Subsequently, explains the motivation for the study and its objectives. Lastly, the outline of the thesis is presented.

1.1 Overview of Maxillary Sinus

The maxillary sinuses are pyramid-shaped, air-filled cavities located within the maxilla, the upper jawbone. They are part of the paranasal sinus system and are situated on either side of the nose, underneath the eyes. Their health and structure are very important for functions such as the humidification and warming of inspired air, and voice resonance. Moreover, they are often implicated in various disease states, ranging from sinusitis and allergic reactions to more severe conditions such as sinus tumors [1], [2].

Sinusitis is one of the most common diseases affecting the MS. It is an inflammation of the sinus lining, usually caused by a viral or bacterial infection. Symptoms typically include nasal congestion, facial pain or pressure, and a reduced sense of smell. In CT scans, sinusitis may manifest as an increased opacification within the sinus cavity [3], [4]. Sinus polyps are soft, non-cancerous growths on the lining of the sinuses or nasal passages, often linked with chronic inflammation due to conditions like sinusitis or allergies. They may cause similar symptoms to sinusitis and can sometimes block the sinus passages entirely [5], [6]. Maxillary sinus tumors are relatively rare but serious conditions. Both benign and malignant tumors can occur in the maxillary sinus. Depending on the tumor type and stage, symptoms can vary from sinusitis-like symptoms to more severe signs like facial numbness or swelling [7], [8]. Maxillary sinus mucocoele is a benign, cyst-like lesion resulting from the obstruction of the sinus ostium, the opening from the sinus to the nasal cavity. A mucocoele can cause sinusitis-like symptoms and may lead to bone erosion if not treated [9], [10]. Odontogenic diseases are conditions that originate from the teeth or related tissues and can extend into the maxillary sinus. Such conditions include periodontal disease, dental abscesses, or periapical pathology [11].

There are several factors that can influence the shape and size of the MS. Dental anatomy is one of the primary factors that affect the morphology of the MS, specifically the upper posterior teeth, which are closely linked to the sinus. The roots of these teeth can encroach on the MS cavity and cause convolutions in the sinus floor [12], [13]. Another important factor that can significantly affect maxillary sinus morphology is sinus septa. These are thin bony walls within the sinus that can vary significantly in size, location, and orientation. They can divide the sinus into separate compartments, significantly affecting its morphology [1], [14], [15]. Age is another important factor for maxillary sinus morphology. The morphology of the MS change with age. In infants and children, the sinuses are small, but they grow rapidly during the first few years of life and continue to enlarge throughout adolescence. In adulthood, the sinuses have usually reached their full size [16], [17]. Diseases such as chronic sinusitis, benign and malignant tumors, cysts, and odontogenic infections can alter the morphology of the maxillary sinus [18]–[21].

There are various maxillary sinus related operations. Sinus Lift (or Sinus Augmentation) operation is often performed by oral and maxillofacial surgeons or periodontists. It's mainly done when the upper jaw doesn't have enough bone, or the sinuses are too near the jaw, to put in dental implants. The procedure involves lifting the sinus membrane and grafting bone into the sinus, thus providing enough bone to securely place the implant [13], [22]. Functional Endoscopic Sinus Surgery (FESS) is another procedure which is usually performed by otolaryngologists to treat chronic sinusitis that has not responded to medication [23], [24]. These surgeries underscore the importance of accurate preoperative maxillary sinus segmentation. This segmentation can provide critical information about the sinus's structure, shape, size, and any present abnormalities, supporting surgeons in surgical planning and execution [25], [26].

1.2 Maxillary Sinus Segmentation

Maxillary sinus segmentation, the process of identifying and delineating the boundaries of the MS in medical images, plays a pivotal role in a variety of clinical applications as explained in the previous section. Despite its importance, the segmentation of the MS is challenging due to its complex anatomical structure, variations among individuals, and the presence of neighboring structures with similar radiographic characteristics. There are manual, semi-automatic, and automatic segmentation methods in the literature [27].

1.2.1 Manual Maxillary Sinus Segmentation

Manual segmentation of the MS usually involves a radiologist or a trained operator going through the slices of a CT or MRI scan and delineating the sinus region of interest slice by slice by using the special purpose softwares for biomedical applications [28]–[31]. While manual segmentation can be more accurate because it allows for the radiologist's expertise and knowledge of the structure of the MS to be used, it is also laborious and time-consuming. Furthermore, it can suffer from variability between different raters, making it less consistent and reliable. Therefore, the need for semi-automated and automated segmentation methods is stated in various studies [32]–[34].

1.2.2 Semi-automatic Maxillary Sinus Segmentation

Semi-automatic segmentation methods combine both computer algorithms and human intervention to achieve optimal results. They reduce the time and effort required for segmentation, while still incorporating some level of expert knowledge and control. These methods often require an initial seed point or contour provided by the user, after which the segmentation is completed automatically based on predefined criteria such as intensity thresholds or region growing algorithms.

Shi et al. [35] used an intensity-based algorithm for three-dimensional (3D) segmentation of the MS based on cone-beam computed tomography (CBCT) datasets. The segmentation procedure is a methodical, step-by-step process that selects a reference point in each sinus, examines neighboring voxels, and classifies points based on intensity values and geometric features. Cho et al. [36] carried out a retrorespective case-control study to identify key factors for volumetric changes in the adult maxillary sinus. The MS was segmented using a semi-automatic process that began with the conversion of the original CT image into a binary format. This helped to filter out soft tissue densities and clearly outline the bony boundary of the sinus. The binary image underwent a morphological closing operation to maintain a continuous sinus boundary. The MS was further segmented using a region-growing algorithm. Farias Gomes et al. [37] developed a formula through maxillary sinus measurements using CBCT scans for sex estimation. The maxillary sinus segmentation was performed using ITK-SNAP 3.0 software in a semi-automatic process. Initially, the boundaries of the maxillary sinus were manually defined on different planes to set the region of interest (ROI). Following this, a specific threshold range was set for voxel selection. 'Seeds' were then placed within the ROI to initialize the segmentation. Hung et al. [38] used 3D-Slicer software platform to determine the presence, location, and volume of mucous retention cysts (MRCs) in the MS and to correlate these data with factors such as age and sex. To measure MRCs, a region-growing algorithm was used involving steps such as manual marking of the MRCs, outlining the boundary of the maxillary sinus, and automatic segmentation of MRCs and the remaining air-filled cavity of the sinus, using the Grow From Seeds tool in the software for volume and surface measurement.

These methods significantly reduce the workload of manual segmentation while maintaining a reasonable level of accuracy. They, however, still require some level of manual intervention, hence they may not be ideal for large-scale or real-time applications. This has led to the development of fully automatic methods, such as deep learning-based models, which can provide high-quality segmentation results with no need for manual intervention.

1.2.3 Automatic Maxillary Sinus Segmentation

Automatic segmentation methods use algorithms to partition an image into multiple segments, without user intervention. In recent years, a number of innovative automatic segmentation methods have been developed. Giacomini et al. [39] created an automated algorithm for the volumetric assessment of the MS based on CT images. The automated segmentation process involves reading the CT image, applying a threshold to remove soft tissue and highlight bone regions, using morphological image processing operators for fine-tuning, utilizing the watershed technique for segmentation, employing a rule-based system to compare major areas, and finally classifying areas of air and opacification, repeated until the entire CT scan is analyzed and a volumetric region of interest is derived. Therefore, the algorithm automates the process of identifying and measuring the volume of the MS, thereby reducing the time required for manual segmentation. However, as stated in the study, the methodology was only tested using one protocol that relied on a CT slice thickness of 0.5 mm and volume quantification errors can increase with thicker slices.

Xu et al. [40] proposed an automatic segmentation method based on combination of the VGG classification network and a V-Net architecture for maxillary sinus in CT images. The VGG classification network determines whether each input slice contains the MS region. Once the slices containing the MS region have been identified, the V-Net performs a detailed segmentation of the MS. An additional feature of this V-Net is the application of edge supervision, which helps to reduce region misjudgments in the segmentation process. The VGG network achieved a classification accuracy of 97.04 ± 2.03% in identifying CT slices with and without the MS. In terms of segmentation, the improved V-Net method achieved a Dice coefficient of 94.40 ± 2.07%, an intersection over union (IoU) score of 90.05 ± 3.26% .

Kuo et al. [41] introduced a technique for segmentation of the inferior turbinate and MS for volume calculation. They first found the area they wanted to look at (ROI) in CT images using something called a back propagation neural network. This network uses special matching techniques and similarities in small areas as inputs. Next, they used a level set method to draw the shape of the inferior turbinate and MS. For creating a 3D model and visualization, they used the marching cubes algorithm. The automatic recognition accuracy and sensitivity for the inferior turbinate and MS were found to be 96.3% and 95.1%, respectively.

Deng et al. [42] proposed a neural network, BE-FNet, for 3D segmentation of MS. The network uses a symmetrical encoder-decoder architecture for bounding box estimation and in-region 3D segmentation tasks with overestimation strategy, residual dense blocks, an attention excitation mechanism, and a pyramid network with multilevel feature fusion. The model displayed a good performance with an average Dice of 0.947 ± 0.031.

It was presented a method for the level set segmentation of maxillary sinus using a convolutional neural network (CNN) model in [23]. This method uses CNN to identify points on the zero level set contour, adjusting their speed based on proximity to lesions, which enables more accurate and adaptive MS segmentation, particularly in challenging cases with heterogeneous lesions.

In [43], an active learning framework was utilized for segmentation of the MS and lesions using a CNN model on CBCT images. The dice similarity coefficients showed an incremental improvement in air segmentation as opposed to lesion segmentation across the steps of the framework.

Zhang et al. [44] developed a network built upon the U-Net backbone, for segmentation of the MS and MS septum in CBCT images. The experimental results indicated that this method significantly improved the average Dice similarity coefficient and Hausdorff distance, outperforming networks without boundary Cross-fusion and boundary attention components.

Jung et al. [43] recently developed a network based on a 3D U-Net architecture and applied to CBCT images for maxillary sinus segmentation. It was reported in the study, the network yielded a high Dice Similarity Coefficient (DSC) of 98.4%.

Another CNN model developed in [45] for segmentation and separation of MS to air, mucosal thickening (MT), and MRCs, using low and full-dose CBCT images. The model was a three-step CNN algorithm, which leaned on V-Net with support vector regression. The median Dice scores for air, MT, and MRCs parts were recorded at 0.972, 0.729, and 0.678 for low-dose, and 0.968, 0.663, and 0.787 for full-dose scans, respectively.

Another deep learning-based method employing the U-Net architecture was developed for segmentation of MS from CBCT images, achieving a DSC of 0.9090 ± 0.1921 [46]. To refine and correct prediction errors from U-Net segmentation, post-processing techniques were applied. After post-processing, the DSC further improved and demonstrated a high level of performance for the segmentation of clear and hazy MS samples.

1.3 Problem Statement and Motivation

Although the maxillary sinus plays a critical role for a healthy respiratory system, accurate and efficient segmentation of the maxillary sinus from CT images remains challenging. Traditional methods such as manual and semi-automatic segmentation are time-consuming, labor-intensive, and subject to high variability, which limits their practical utility in clinical settings. Automatic segmentation methods, particularly those based on deep learning, have shown promising results, but further research is required to assess their suitability for maxillary sinus segmentation, especially given the complexity of the sinus anatomy and inter-individual variations.

Additionally, while some studies have explored the use of different deep learning models for MS segmentation, few have directly compared the performance of multiple models or utilized advanced architectures.

It is aimed to implement and evaluate a plain U-Net model and three U-Net variants with different encoders - VGG16, ResNet50, and Inception ResNet v2 - for maxillary sinus segmentation.

By comparing the performance of these models and exploring their applicability for the task of maxillary sinus segmentation, it is hoped to contribute to the development of efficient, reliable, and automated methods for maxillary sinus segmentation. This could potentially lead to improvements in diagnostic accuracy, treatment planning, and overall patient care in clinical settings related to maxillofacial radiology and beyond.

1.4 Objectives

The primary objectives of this study are:

1. It is aimed to employ and assess four different deep learning models: : a U-Net model, and three variations of the U-Net model using VGG16, ResNet50, and InceptionResNet v2 as encoders for maxillary sinus segmentation. The models' performance is evaluated based on their ability to accurately segment the maxillary sinus from CT images.
2. After the initial segmentation using the plain U-Net and its variants, it is aimed to further segment the air section and mucosa inflammation sections within the MS.
3. The study seeks to compute the volume of the maxillary sinus and opacification ratios, which are critical for diagnosing sinus-related disorders and understanding sinus health.

4. By demonstrating the potential of deep learning models for maxillary sinus segmentation, the study hopes to contribute to the advancement of diagnostic processes in maxillofacial radiology.

1.5 Outline of the Thesis

This chapter provides an overview of the MS and a literature survey for the segmentation of MS. The subsequent chapter, Chapter 2, discusses the related work regarding general medical image segmentation, with a particular emphasis on U-Net and transfer learning strategies for medial image segmentation. Chapter 3 delineates the research methodology, including comprehensive information on the preparation of the dataset, the models, and the methodologies employed in the study. Chapter 4 details all the findings from each test, contrasting them with other outcomes, and includes an analysis of these results. Chapter 5 concludes the thesis and provides suggestions for future research.

CHAPTER 2
RELATED WORKS

2.1 Medical Image Segmentation

Image segmentation is a key process in computer vision and image processing that involves distinguishing the desired structures from the background and from one another. Each segment usually corresponds to different objects or regions in the image [47]. Medical image segmentation is a specialized sub-field within image segmentation that focuses on the partitioning of medical images into meaningful regions. This process is critical in many applications of medical imaging, aiding in diagnosis, study of anatomical structure, planning for surgical procedures, and monitoring of disease conditions [48].

Traditional methods for image segmentation have been employed for years. Some of the commonly used techniques are thresholding, edge detection, region-growing, clustering methods, and watershed transformations [49], [50]. Thresholding is one of the simplest forms of image segmentation. It involves separating an image into two or more segments based on the pixel intensity levels. It leverages intensity values to differentiate between tissues or organs [51]–[53]. In this method, boundaries of objects within an image are detected by identifying areas where the pixel intensity changes sharply. Common techniques include the Sobel, Prewitt, and Canny edge detectors. It has been used in many studies in medical imaging [54]–[56]. Region growing is a classical image segmentation technique used to extract meaningful regions from an image. It is particularly useful in medical image segmentation due to its ability to capture regions of interest based on some similarity criterion. Its combination with other traditional and neural network based segmentation methods or usage in a multi-step segmentation pipeline has found wide covarage in the literature [37], [38], [57], [58]. Clustering methods, like K-means, Fuzzy C-means, and Expectation-Maximization algorithms, are unsupervised techniques that group similar data points together based on certain features or properties. In the context of medical image segmentation, clustering aims to group together pixels or voxels (in 3D imaging) that have similar characteristics [59]–[62]. The watershed transformation is a popular technique used in image segmentation, inspired by geographical watershed principles [63], [64]. To improve the results of the watershed transformation in medical imaging, it is often combined with other techniques or used as part of a multi-step segmentation process [39], [65], [66]. However, these methods often require manual fine-tuning and are not always robust to the variations seen in clinical practice. The advent of deep learning has revolutionized the field of medical image segmentation [67]. Neural networks, particularly CNNs, have shown remarkable results in tasks such as tumor detection, organ segmentation, and lesion identification, outperforming traditional techniques in many aspects. CNNs have the ability to learn complex patterns and structures within images, making them well-suited for the task of medical image segmentation [68], [69].

Among deep learning architectures, U-Net has emerged as a leading choice for biomedical image segmentation due to its symmetric expansive path, which aids in precise localization [70], [71]. First made to figure out the structure of neurons in certain types of microscope images, U-Net has been adopted and adapted for various other medical image segmentation tasks, including maxillary sinus segmentation [44], [72], [73].

Transfer learning, the practice of reusing pre-trained models on new tasks, has also gained popularity in the domain of medical image segmentation. By leveraging models pre-trained on large datasets, transfer learning allows for improved performance, even with smaller dataset sizes [67], [74]. This approach is particularly useful in the medical field, where acquiring large annotated datasets can be challenging due to privacy issues and the effort required for expert annotation. Tajbakhsh et al. [75] discuss the benefits and challenges of fine-tuning pre-trained CNNs for medical image analysis tasks. They look into how transfer learning might be used in the medical imaging and highlight the differences between regular and medical images, which might change how knowledge is shared or transferred.

2.2 U-Net for Medical Image Segmentation

U-Net is a CNN architecture that was introduced in 2015 [72]. It was specifically designed for biomedical image segmentation, making it a highly influential and widely used model in medical imaging. The model is termed U-Net due to its "U" shape, where the left side of the U encodes the input image into feature representations, and the right side decodes these features back into the segmented image.

U-Net has been employed for precise tumor segmentation in various types of medical scans such as MRI, CT, and Ultrasound. This aids in the early detection of malignancies and the development of treatment plans [76]–[78]. In addition to tumor detection, U-Net is commonly used to segment different organs and anatomical structures such as the liver, heart, and lungs. This assists in the detailed analysis and visualization for surgeries and other medical procedures [79]–[81]. The ability of U-Net to capture fine details makes it suitable for tasks such as blood vessel segmentation, which is essential in studying vascular diseases [82], [83].

The U-Net’s effectiveness lies in its ability to learn from a relatively small amount of labeled data and generalize well to new, unseen data. Moreover, it enables end-to-end training and utilizes expansive paths with skip connections to recover spatial information lost during the downsampling process. These unique characteristics have made U-Net a popular choice for biomedical image segmentation tasks. Moreover, U-Net architecture has been utilized in several studies for MS segmentation [44], [46], [73], [84].

2.3 Transfer Learning in Medical Image Segmentation

While U-Net's architecture offers a powerful means for segmenting biomedical images, training the network from scratch requires a substantial amount of data and computational resources. Herein lies the importance of transfer learning, a method that leverages a model pre-trained on a large dataset, and fine-tunes it for a specific task [85].

The use of transfer learning in medical image segmentation provides several advantages. First, it enables the network to leverage a comprehensive set of learned features, reducing the amount of training data needed. Second, it accelerates the training process by starting with a pre-trained network rather than initializing weights randomly [86].

The integration of pre-trained models such as VGG16 [87], ResNet50 [88], and InceptionResNet v2 [89] as encoders in the U-Net architecture combines the strengths of these powerful models with the localization capabilities of U-Net. VGG16 is famous for being simple and good at identifying objects. ResNet50 is known for its advanced learning ability. InceptionResNet v2 blends the best features of two different systems, the Inception architecture and residual connections. The use of these models as encoders in U-Net leverages their ability to extract complex features, which is further enhanced by the expansive path of U-Net for accurate segmentation.

In the context of maxillary sinus segmentation, the integration of VGG16, ResNet50, and InceptionResNet v2 into the U-Net architecture can provide enhanced performance. By serving as encoders, these pre-trained models can bring a wealth of feature extraction capabilities to U-Net.

There is a growing body of research demonstrating the effectiveness of these integrated models for various segmentation tasks [90]–[93], but literature focusing specifically on maxillary sinus segmentation remains scarce.

CHAPTER 3
MATERIALS AND METHODS

3.1 Dataset

This study incorporated a total of 71 head CT scans from both male and female subjects. The scans were obtained from Gaziantep University Sahinbey Research and Practice Hospital. The dataset consisted of many different sizes of MS. The sizes of the MS in the early and final layers of a CT volume are small, and they increase through the middle layers. Furthermore, the dataset contained not only clear sinuses but also maxillary sinuses with varying degrees of opacification. Examples of maxillary sinuses of different sizes and opacification levels from our dataset are shown in Figure 3.1.

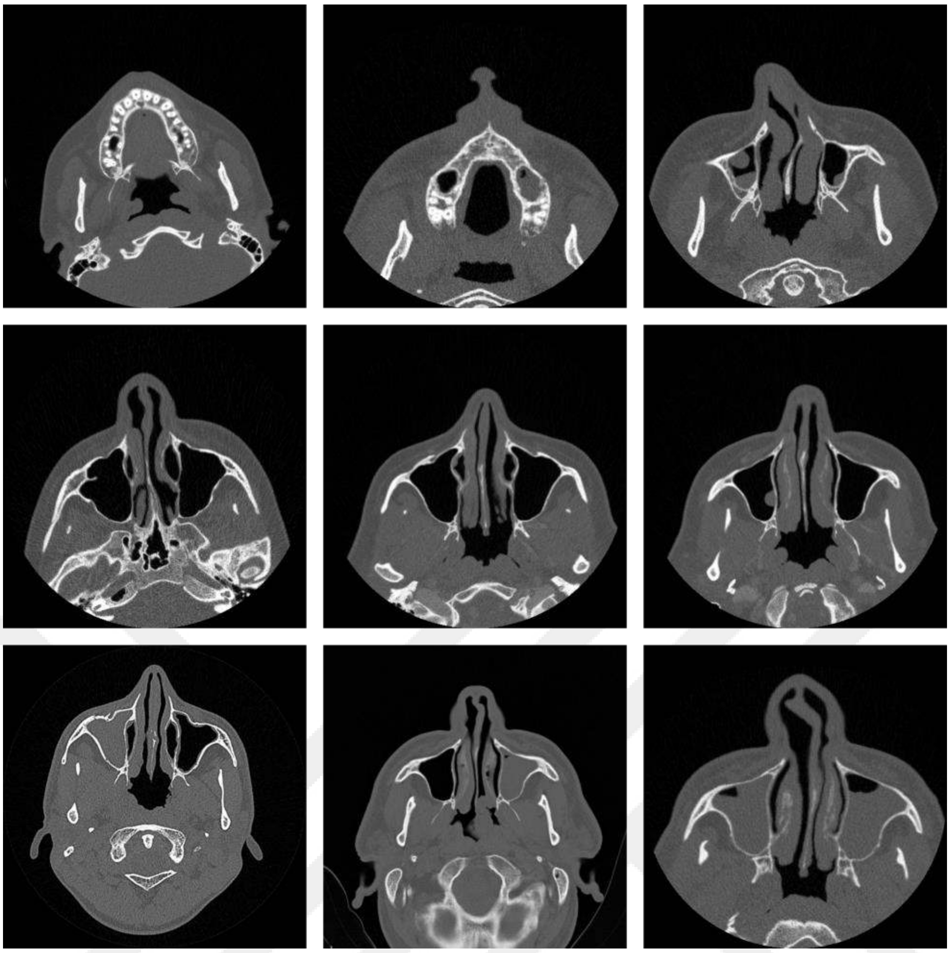

Figure 3.1 Examples of CT images from the dataset with different sizes and opacification levels

Axial, coronal, and sagittal views of the maxillary sinuses are shown in Figure 3.2. Slices in the axial plane were used for segmentation, considering the morphology of the maxillary sinus and its surrounding structures. The exclusion criteria for participant selection in this study included cases where the radiograph quality was inadequate due to the presence of artifacts, instances where an abnormality was detected in the maxillary sinus, or a history of prior surgery in the maxillary sinus. These factors were considered to ensure that the selected participants had reliable and consistent radiographic data and were free from any pre-existing conditions that could potentially impact the study results.

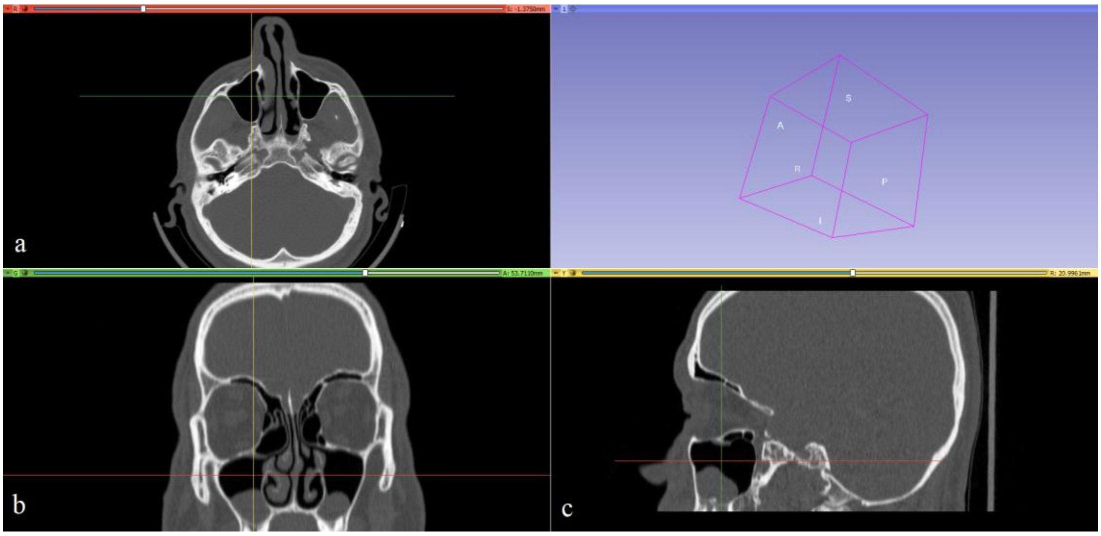

Figure 3.2 Axial, coronal, and sagittal views of the maxillary sinus.

3.1.1 Ground-truth Labelling

The CT scans, exported as Digital Imaging and Communications in Medicine (DICOM) files, were imported into 3D Slicer 4.11, an open-source medical image analysis software application [94]. The ground truth datasets were labeled for training and testing the model on a layer-by-layer basis using the 3D Slicer 4.11. The labelling process involved the following steps:

1. Contouring One Side: Utilizing the Segment Editor tool of the 3D Slicer, the contour of one side of the maxillary sinus was manually delineated.

2. Filling and Mask Creation: The region enclosed by the contour was automatically filled in to generate segmentation masks for that side of the sinus.

3. Contouring the Other Side: Subsequently, the contour of the maxillary sinus on the other side was also delineated manually.

4. Finalizing Labeling: The enclosed area of the other contour was then filled in, marking the completion of the labeling process for one slice.

The labelling steps are shown in Figure 3.3. These steps were applied to every maxillary sinus containing slices in the CT scans, and a complete maxillary sinus volume was created, as shown in Figure 3.4. The masked regions were also checked and edited in the coronal and sagittal planes to avoid any inconsistencies. Bony structures inside the sinus such as the septum were not included in the segmentation region, and additional edits were made as needed to remove those structures from the segmentation region.

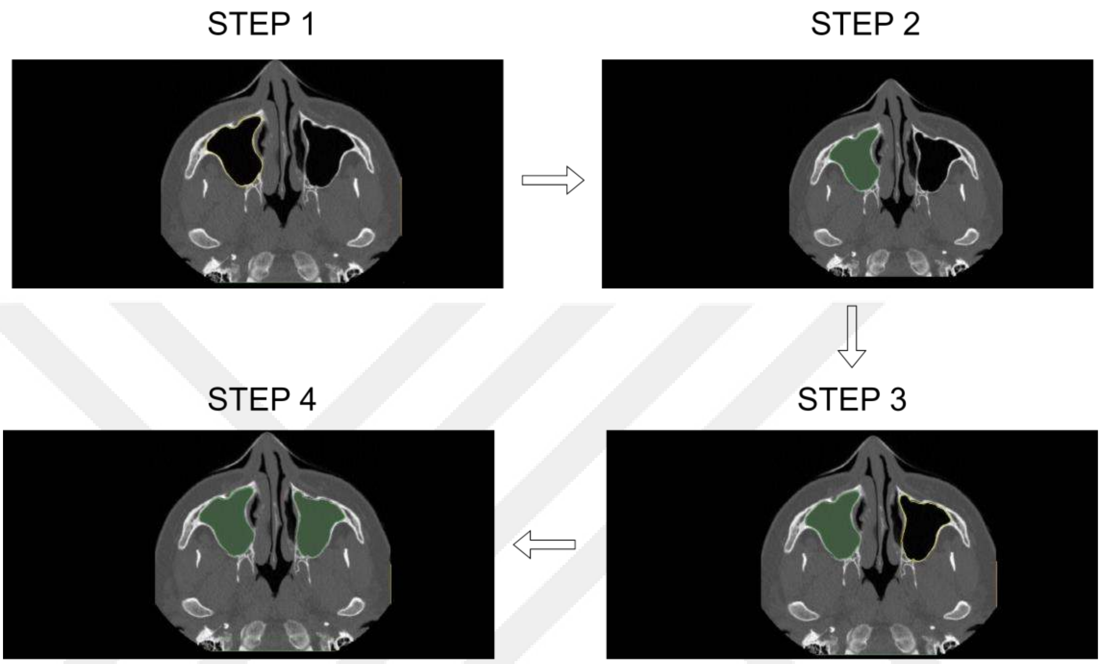

Figure 3.3 Labelling steps.

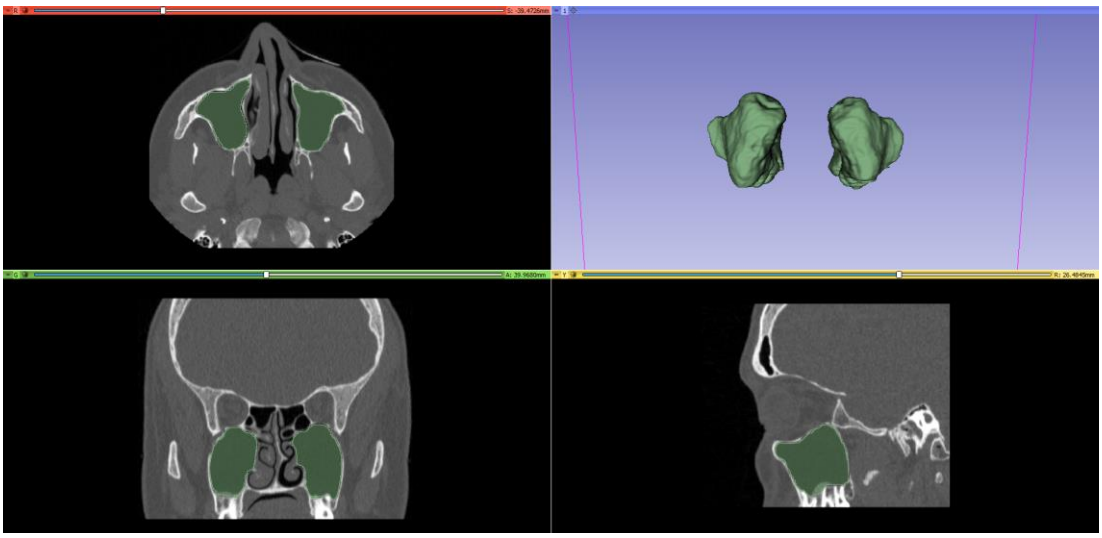

Figure 3.4 Completed labelling process for a sample CT volume.

3.1.2 Preprocessing

The DICOM files of the CT scans and their corresponding generated masks of maxillary sinuses were extracted and converted to PNG format for the model training and testing phases. Only the CT slices containing MS were included in the dataset, excluding any layers without MS. After these exclusions, the dataset comprised 3947 CT slices containing MS.

For the purpose of training and testing, the dataset was randomly divided into an 8:2 ratio, forming the training and test sets respectively. Further, the training set was subdivided into an 8:2 ratio to create the training and validation sets. The training set was utilized to train the models, the validation set assessed the models' performance during the training phase and facilitated the selection of the best performing model, and the test set evaluated the final performance of the models.

Before feeding into the model, all images were normalized to a range between 0 and 1. The original size (512 ×512 pixels) of all input images used for training, validation, and testing was maintained without applying any resizing. Due to memory constraints, the dataset did not undergo any data augmentation. Samples of final CT images and corresponding ground truth masks from the dataset are showed in Figure 3.5.

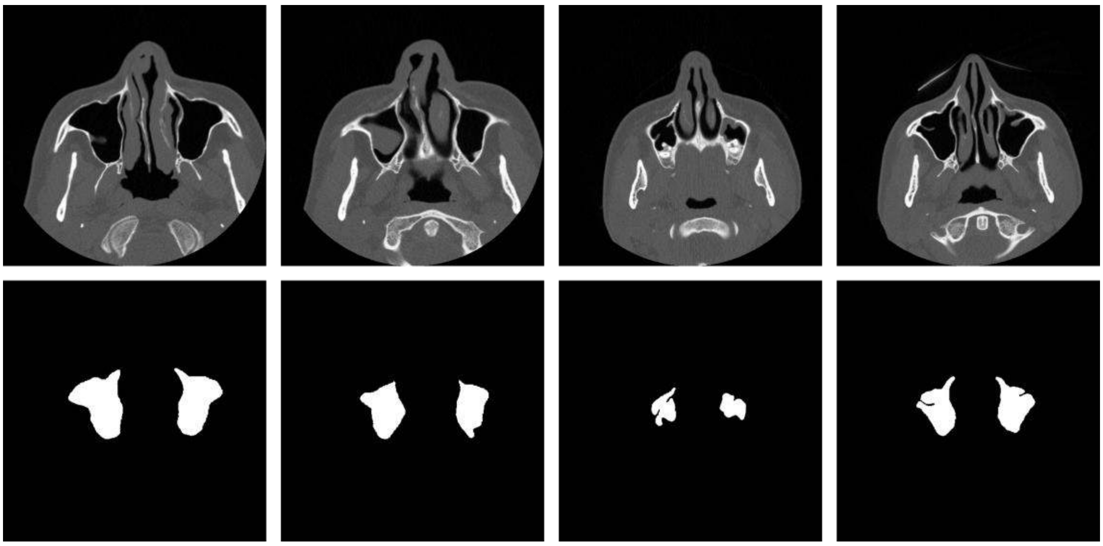

Figure 3.5 Samples of CT images and corresponding ground truth masks from the dataset.

3.2 Network Architectures

Several networks were created for maxillary sinus segmentation in CT images. The U-Net architecture was used as the baseline model in the conducted experiments. With its release in 2015, U-shaped structures and modified versions have found extensive applications in medical imaging and other fields [70]–[72]. U-Net's defining features are its encoder-decoder configuration and the integration of these structures via skip connections.

The contracting path, or the encoder, uses a series of convolutional blocks and pooling layers, a common feature in conventional convolutional network architectures. These layers progressively decrease the size of the feature maps while enlarging the receptive field. The receptive field size depends on the filter sizes and stride lengths used in the convolutional layers, with each layer's receptive field being influenced by its predecessor. As it goes deeper through the layers of the CNN, the receptive field size increases. Thus, the encoder path of the network can capture broader contextual information and extract higher-level features from the input image, which are essential for image-segmentation tasks.

The layers positioned at the deepest part of the network, known as the bottleneck layer, act as a channel between the encoder and decoder. They transfer crucial information for precise segmentation. This is called the bottleneck layer because the size of the feature maps is reduced to their minimum size, while the number of feature maps is higher compared to the layers in the encoder and decoder. Through the contracting path, the network progressively learns to extract increasingly intricate and meaningful features, reaching the most abstract image representations at the bottleneck layer. It compacts the information coded in the feature maps by shrinking the spatial dimensions while boosting the channel count, capturing the most abstract, high-level features.

The expansive path, or decoder, in U-Net creates segmentation maps by gradually upsampling the low-resolution feature maps. It starts with up-convolution or transpose convolution layers, enhancing the spatial dimension of the feature maps. These upsampled maps are concatenated with the feature maps from the contracting path. Following the concatenation, convolutional blocks comprising several convolutional and activation layers further process the merged feature maps. For semantic segmentation tasks, a 1×1 convolutional layer typically capped with a softmax or sigmoid activation function serves as the final layer, mapping the extracted features to the desired number of output channels. Figure 3.6 illustrates the original U-Net architecture to better visualize these concepts.

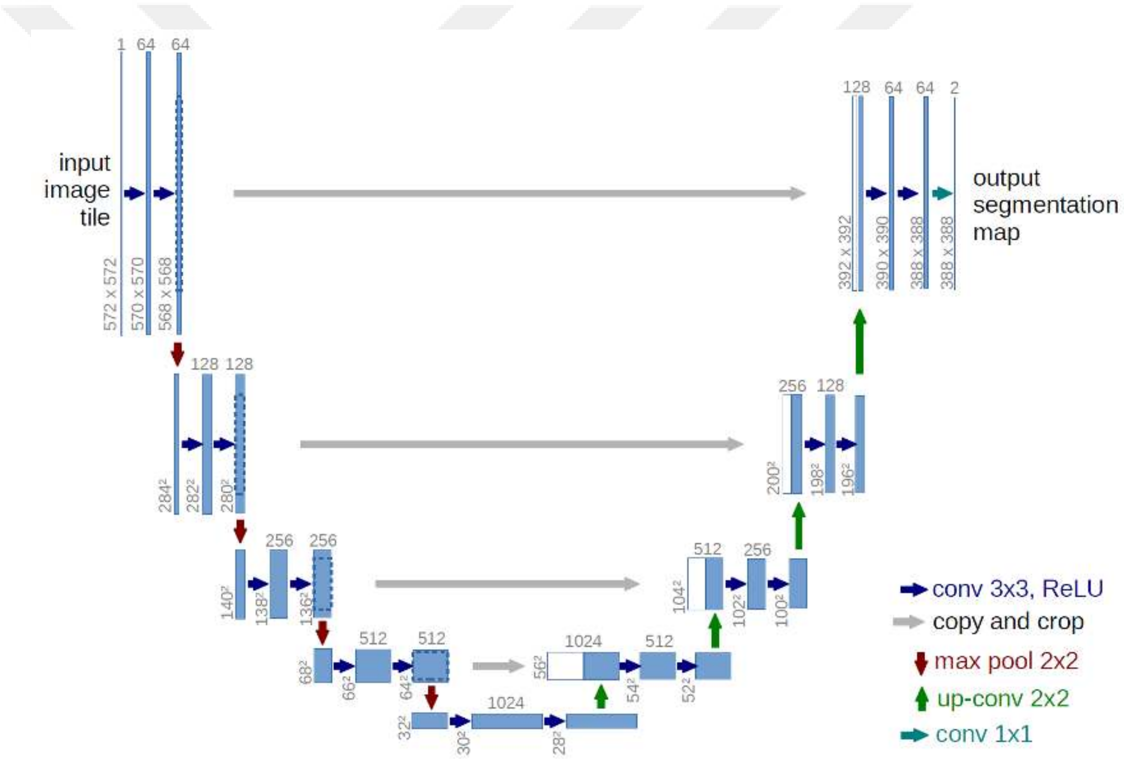

Figure 3.6 U-Net architecture from original paper [72].

In this work, the U-Net module was modified for maxillary sinus segmentation. Moreover, the segmentation of maxillary sinus was also performed using three different pre-trained transfer learning models: VGG16 [87], ResNet50[88], InceptionResNet v2 [89]. The models were used as encoders for the U-Net model and the decoders were constructed in accordance with the models.

3.2.1 U-Net

The proposed U-Net model, depicted in Figure 3.7, is comprised of two main parts: an encoder and a decoder. Channel depths, feature map sizes, and performed operations for every layer were provided in the figure. There are four levels with different feature map sizes in the encoder and decoder parts, and another level in the bottleneck layer with the smallest feature map size.

Every level features two convolutional blocks, each made up of a 3x3 convolution, batch normalization, and ReLU layers. In the encoder, a Max Pooling layer (2×2, stride 2) is used at each transition between levels to reduce the resolution of the feature maps by half. In the decoder part, transposed convolution layers are used at each level transition to halve the channels width while doubling the feature map size.

The feature maps resulting from these transposed convolutions are then concatenated with the corresponding feature maps from the encoder. Finally, a 1×1 convolutional layer is applied to condense the feature maps to a single channel. This is followed by a sigmoid activation function, which generates a binary segmentation map.

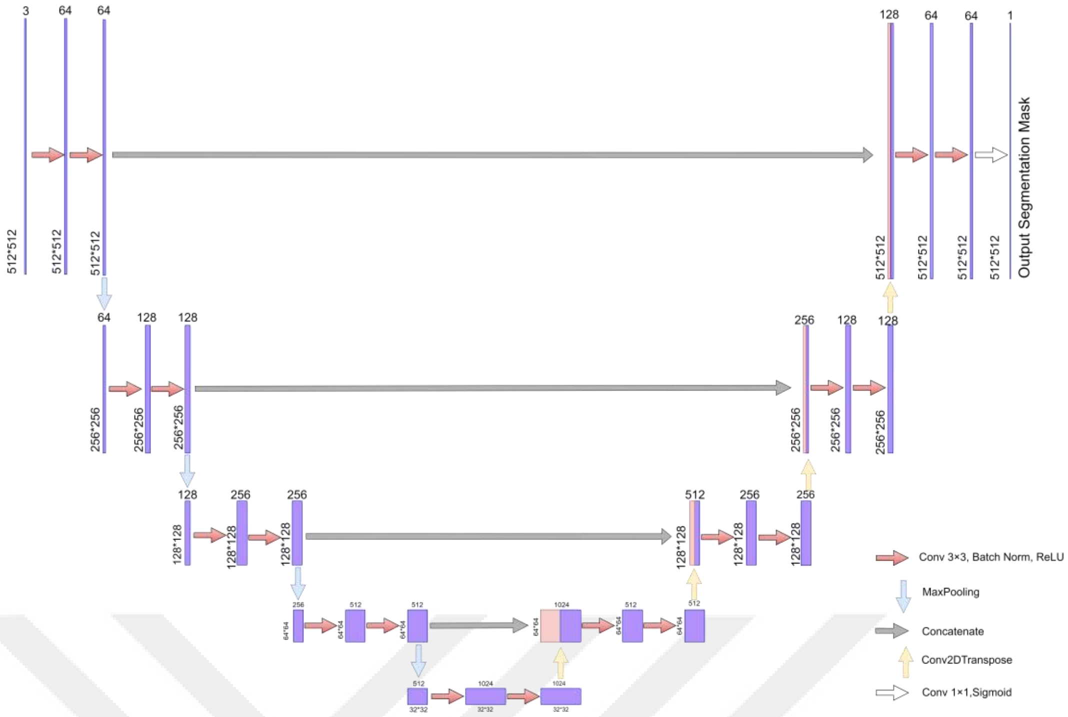

Figure 3.7 Proposed U-Net model.

3.2.2 U-Net with VGG16 Encoder

The VGG architecture was first presented in the 2014 research paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" by authors Karen Simonyan and Andrew Zisserman [87]. This deep learning structure comprises multiple convolutional layers followed by fully connected layers, typically designed with 16 or 19 layers in total. VGG stands for the Visual Geometry Group, a team from Oxford University, and 16 refers to 13 convolutional layers and three fully connected trainable layers. The core idea underpinning the VGG network is to enhance the network's depth using smaller 3×3 convolution filters. This strategy enables an increase in the number of layers while maintaining a reasonable amount of trainable parameters. In this work, the VGG16 network served as a pre-trained encoder within the VGG16-UNet model. The fully connected layers were excluded, which substantially reduces the number of parameters needing training from 138 million to a more manageable 14.7 million. Each convolutional layer of the VGG16 was accompanied by batch normalization and ReLU layers. Five Max Pooling layers across a 2×2 pixel window with a stride of 2, were applied, which effectively halving the resolution. In the decoder part of the VGG16-UNet model, there were 13 convolutional layers corresponding to those in the encoder part. Similar to the encoder, batch normalization and ReLU layers followed each convolutional layer. The decoder also incorporated five upsampling layers, which doubled the resolution of the feature map, mirroring the effect of the max-pooling layers in the encoder. Additionally, skip connections were used to concatenate the output of the upsampling layer with the corresponding same-sized convolution layer in the encoder. The sigmoid activation function was used at the output layer for the binary classification, distunguishing between background and maxillary sinus containing pixels. A visual representation of the VGG16-UNet architecture can be seen in Figure 3.8.

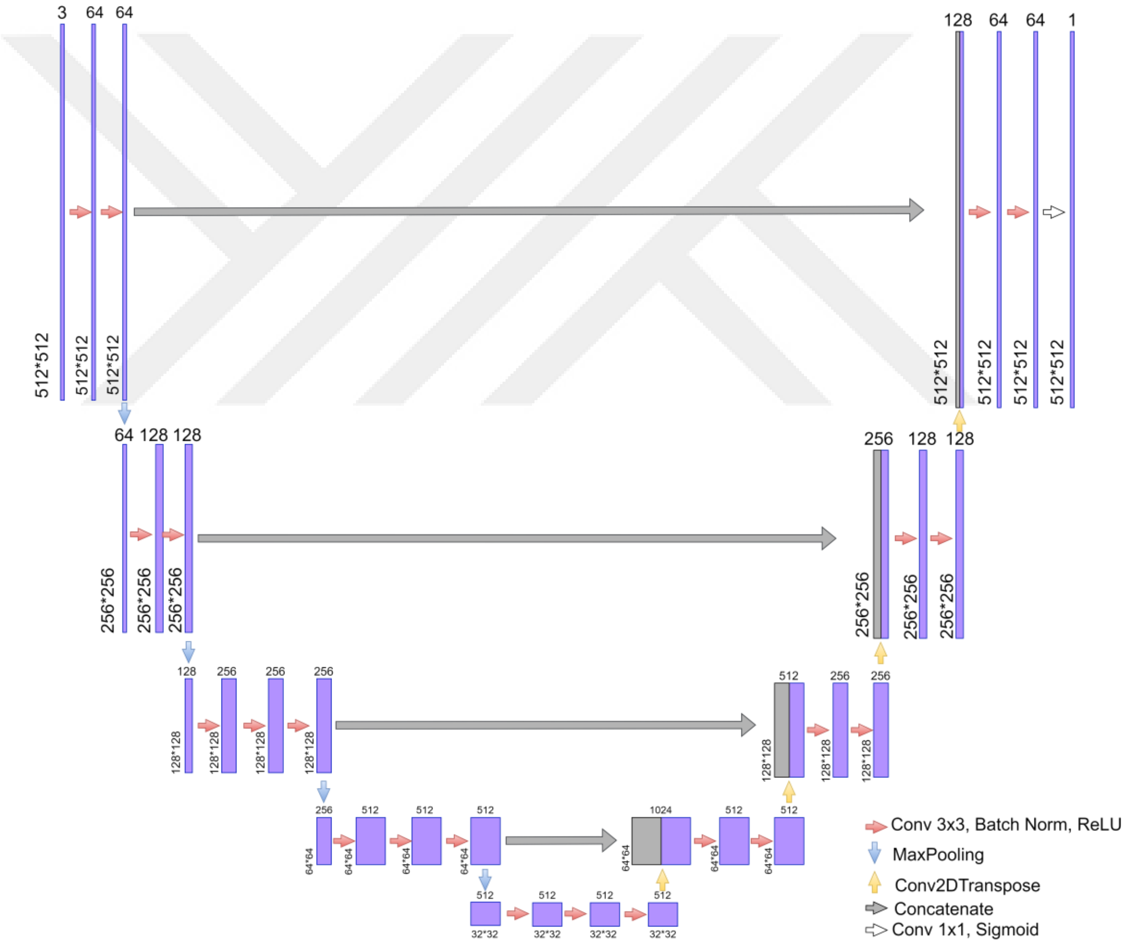

Figure 3.8 VGG16-UNet model.

3.2.3 U-Net with ResNet50 Encoder

ResNets (Residual Networks) is another popular CNN architecture, such as VGGNet. It was introduced and achieved outstanding performance in the ILSVRC and COCO competitions in 2015 [88]. ResNet and VGGNet differ from each other in terms of the way they handle the flow of information. VGGNet relies on the straightforward stacking of convolutional and pooling layers, and each layer learns to transform its input into a desired output representation, which eventually causes vanishing gradient problems as the model becomes deeper.

ResNet solved this problem by means of Residual Learning concept. In traditional deep learning (Figure 3.9-a), a network learns the desired underlying mapping $f(x)$ directly from the input x to the output. As networks get deeper, learning this direct mapping can become increasingly challenging, leading to the vanishing or exploding gradient problem. Residual learning (Figure 3.9-b), on the other hand, focuses on learning the residual or difference between the desired underlying mapping $f(x)$ and the input x. This residual function is represented as $g(x) = f(x) - x$.

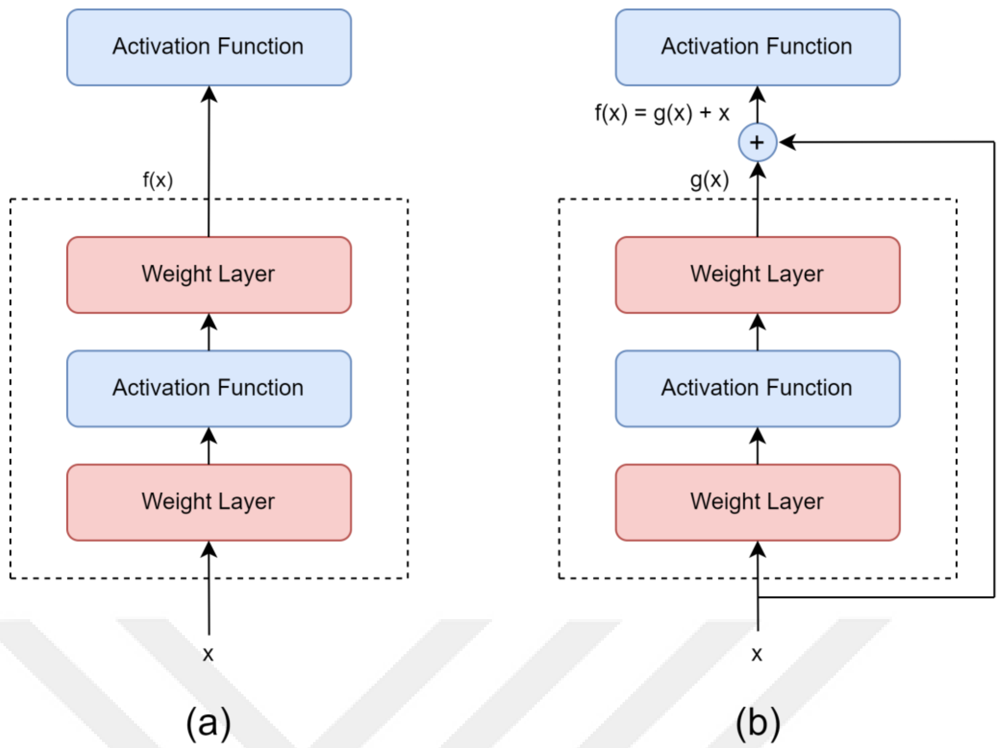

Figure 3.9 (a) Traditional and (b) residual learning.

ResNet-50 specifically consists of 50 layers, and it is part of the family of ResNet architectures which have various depths like 18, 34, 50, 101, and 152 layers. ResNet-50 architetcure is constructed with multiple residual blocks, as shown in Figure 3.10. In ResNet-50, each residual block is composed of three layers: a 1×1 convolution, a 3×3 convolution, and another 1x1 convolution. If the input and output of the residual block have the same shape, the shortcut connection directly adds the input to the output, as shown in Figure 3.10. In cases where the input and output of the residual block have different shapes, a linear transformation (i.e., a 1×1 convolution) is applied to the shortcut to match the dimensions, as shown in Figure 3.10. This is known as a projection shortcut, as opposed to an identity shortcut where the shapes are already compatible, and no transformation is needed.

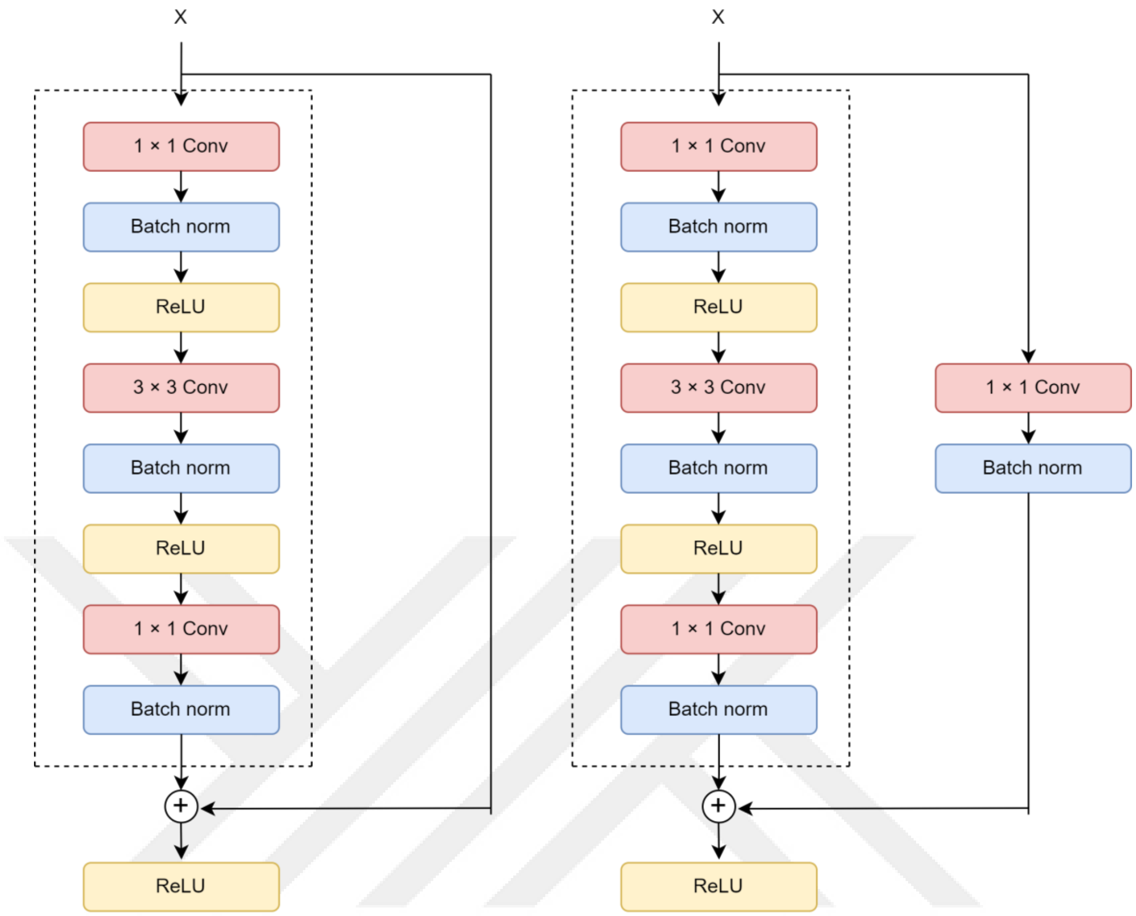

Figure 3.10 Residual blocks of ResNet50 model.

In this study, the ResNet-50 architecture was integrated as a pre-trained encoder for the U-Net architecture, except for the final fully convolutional layer of the model. The resulting hybrid architecture was termed as ResNet50-UNet. The ResNet50-UNet architecture's encoder part was made up of 4 stage residual blocks. It starts with a single convolutional layer with 64 filters, followed by a max-pooling operation. Stage 1 was repeated three times and consists of a specific sequence of convolutional layers with dimensions: (1×1, 64), (3×3, 64), and (1×1, 256). The numbers represent the filter size and the number of filters, respectively. Stage 2 was repeated four times, this stage employs a series of (1×1, 128), (3×3, 128), and (1×1, 512) convolutional layers. Stage 3 was an expanded stage, repeated six times, comprised of (1×1, 256), (3×3, 256), and (1×1, 1024) convolutional layers. The final stage, repeated four instances, includes convolutional layers with configurations of (1×1, 512), (3×3, 512), and (1×1, 2048). Throughout these stages, batch normalization and ReLU activation were applied following each convolutional layers. The only exception was the final 1×1 convolutional layer, specifically designed to match the depth of the final feature map with the segmentation mask's depth. The model concludes with the application of a sigmoid activation function to the final feature map. This step converts the output into a binary segmentation mask. The architecture of the ResNet50-UNet is shown in Figure 3.11.

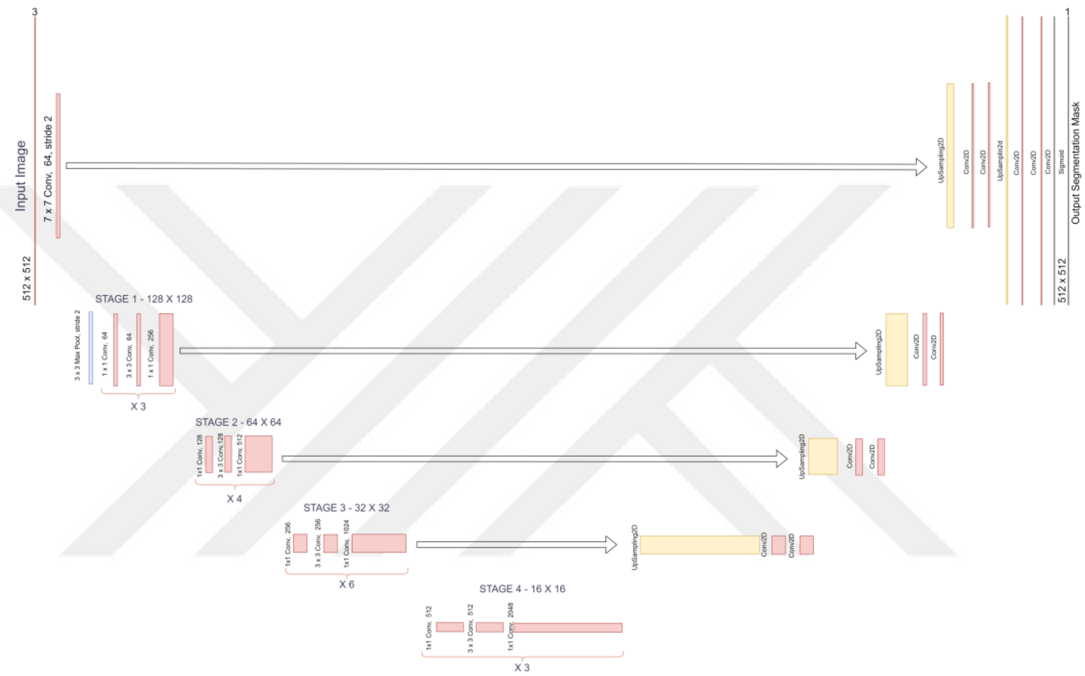

Figure 3.11 ResNet50-UNet model.

3.2.4 U-Net with Inception-ResNet v2 encoder

Inception-ResNet v2 is a CNN architecture that is a combination of Inception and ResNet [89]. Inception-ResNet v2 combines the principles of both architectures. The key idea is to introduce residual connections into the Inception modules. The Inception module is a key building block in the Inception family of architectures, including Inception-ResNet v2. It is designed to capture information at various scales and levels of abstraction. The design of the Inception module is based on the idea that pictures have many levels of details and shapes [95]. For segmentation of the maxillary sinus, an Inception module could be helpful. By using different filters and pooling methods at the same time, it can create a detailed picture of various features.

The network typically begins with a "stem" that consists of a few convolutional and pooling layers. The purpose of the stem is to reduce the spatial dimensions of the image and increase the number of channels, preparing the input for subsequent complex processing. Inception-Residual blocks, a sequence of blocks that combines Inception modules with residual connections, make up the majority of the model. Reduction blocks are used to change the feature map size between the stages of Inception-Residual blocks.

The Inception-ResNet v2 was used as the encoder part of the U-Net architecture in this study. It was a pretrained version on ImageNet dataset [96]. The general schematic of the encoder is roughly depicted in Figure 3.12. It was mainly composed of Inception-ResNet-A,B,C blocks and reduction blocks between them. The structure of the Inception-ResNet blocks is depicted in details in Figure 3.13. They include multiple parallel branches with different convolutional operations. Each branch consists of different combinations of convolutions with varying kernel sizes and pooling operations. The outputs of these branches were concatenated along the channel dimension and the Inception module’s output was fed into a residual connection. It is really a deep and complex model. The decoder section was composed of convolutional and upsampling layers similar with the ResNet50-UNet model.

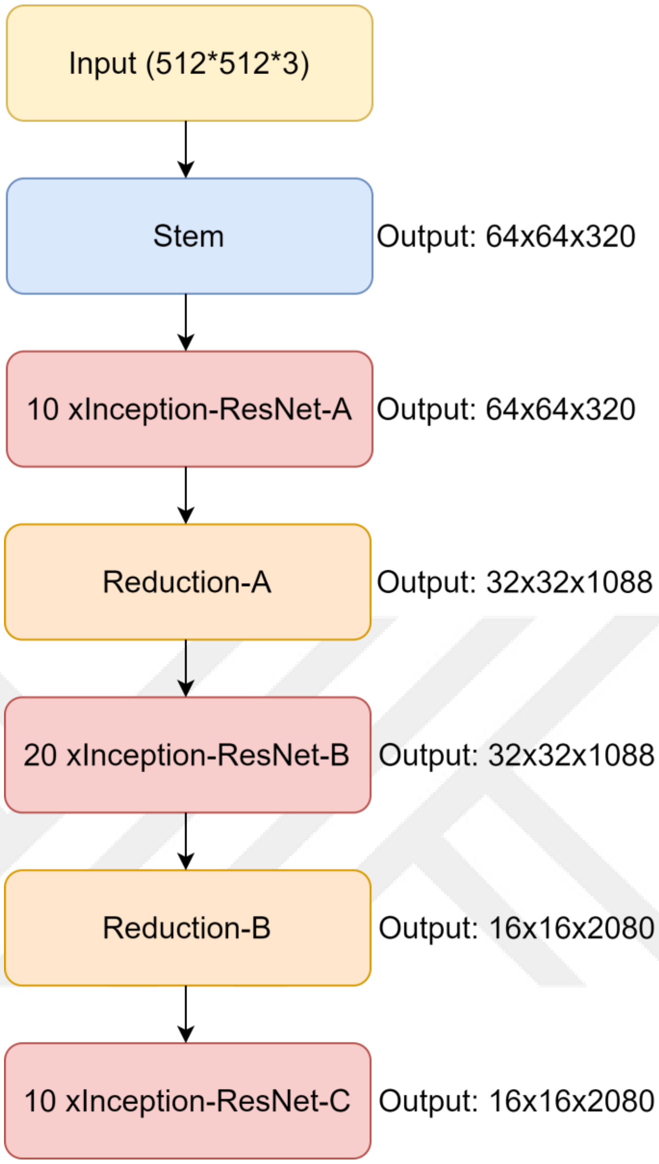

Figure 3.12 Inception-ResNet v2-UNet encoder schematic

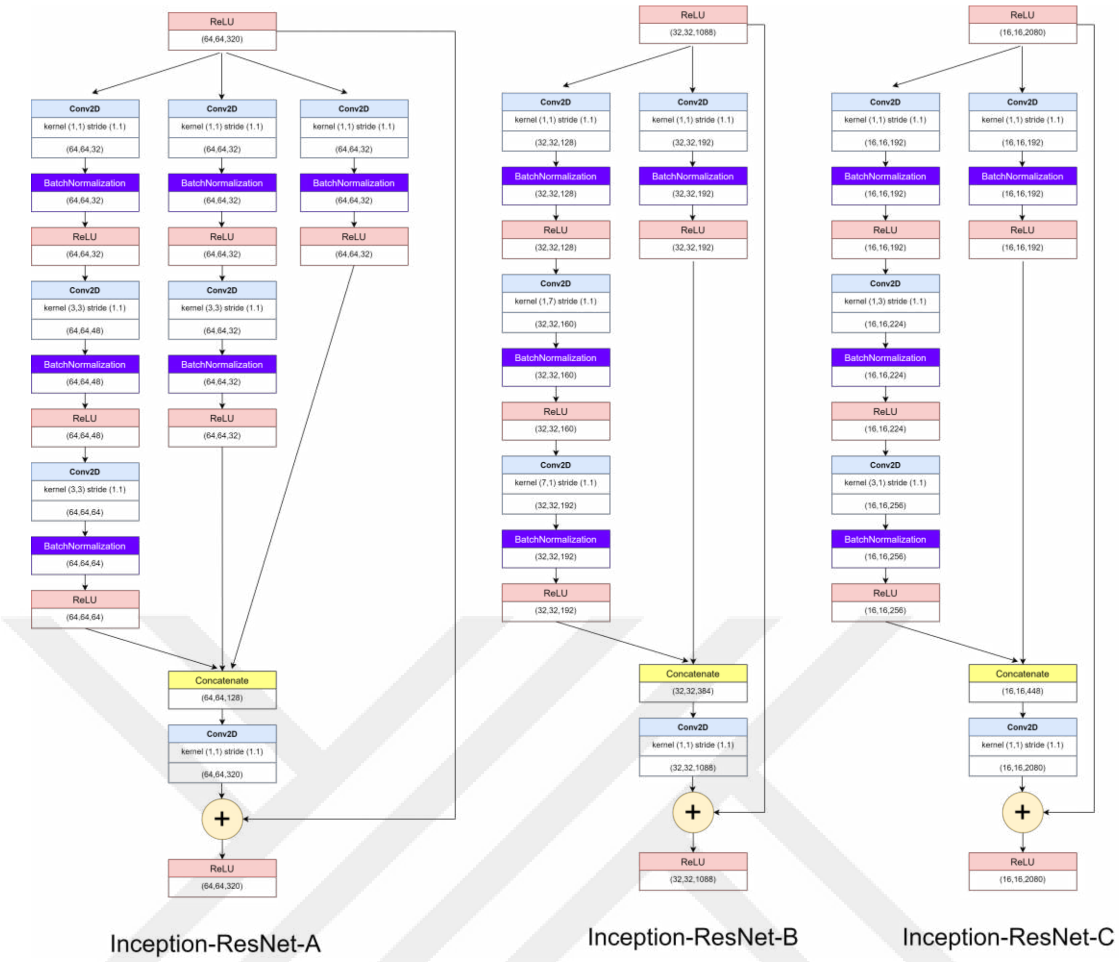

Figure 3.13 Inception-ResNet -A,B,C blocks.

3.3 Training and Validation Settings

In this section, it is presented the training and validation settings utilized in the development and evaluation of the proposed Maxillary Sinus segmentation models. Four distinct models were trained: a plain U-Net model, and three additional models utilizing pre-trained encoders (VGG16, ResNet50, and InceptionResNet v2) integrated into the U-Net architecture. All training experiments were performed using Python 3 and the Keras library with TensorFlow backend on Google Colab. Tesla A100 GPUs were utilized for accelerated computation, enabling efficient training of deep neural networks.

During model training, all models were optimized using the Adam optimizer with an initial learning rate of 1e-3. A batch size of 6 was employed. To avoid overfitting and ensure the models achieved optimal performance, early stopping was implemented during training. Specifically, if the validation loss did not decrease for 20 epochs, the training process was halted, and the model with the best validation performance was retained. Additionally, a learning rate scheduler was employed to dynamically adjust the learning rate. The learning rate was reduced by a factor of 10 if the validation loss did not decrease for 10 consecutive epochs until the lower bound of 1e-7 was reached.

The training process was set to run for a maximum of 100 epochs. However, with the early stopping mechanism in place, the actual number of epochs varied based on the convergence speed and generalization performance of each model. The plain U-Net model was halted at at epoch 85, while the other three models lasted until the maximum number of epochs.

The performance of the models is evaluated using several metrics: Dice Coefficient, Intersection over Union (IoU), precision, and recall. The dice coefficient is a statistical tool that measures the similarity between the predicted segmentation and the ground truth segmentation of the maxillary sinus. A Dice score of 1 represents a perfect segmentation by the model, whereas a score of 0 indicates no overlap at all between the predicted and actual segmentation. The Dice score is particularly useful when evaluating segmentation models for datasets with small objects or class imbalances because it takes into account both the true positive and false positive predictions. The Dice score is calculated as:

$$Dice\ Score = \frac{2 * |Intersection|}{|Prediction| + |Ground\ Truth|} \tag{3.1}$$

where:

- |Intersection| is the number of pixels that are correctly classified as both foreground (maxillary sinus) in both the predicted and ground truth masks.
- |Prediction| is the total number of pixels classified as foreground in the predicted mask.

- |Ground Truth| is the total number of pixels classified as foreground in the ground truth mask.

Figure 3.14 illustrates the Dice scores achieved by each of the trained models on both the training and validation sets. The Dice scores were calculated after each epoch during the training process, providing insights into the convergence behavior and generalization performance of the models.

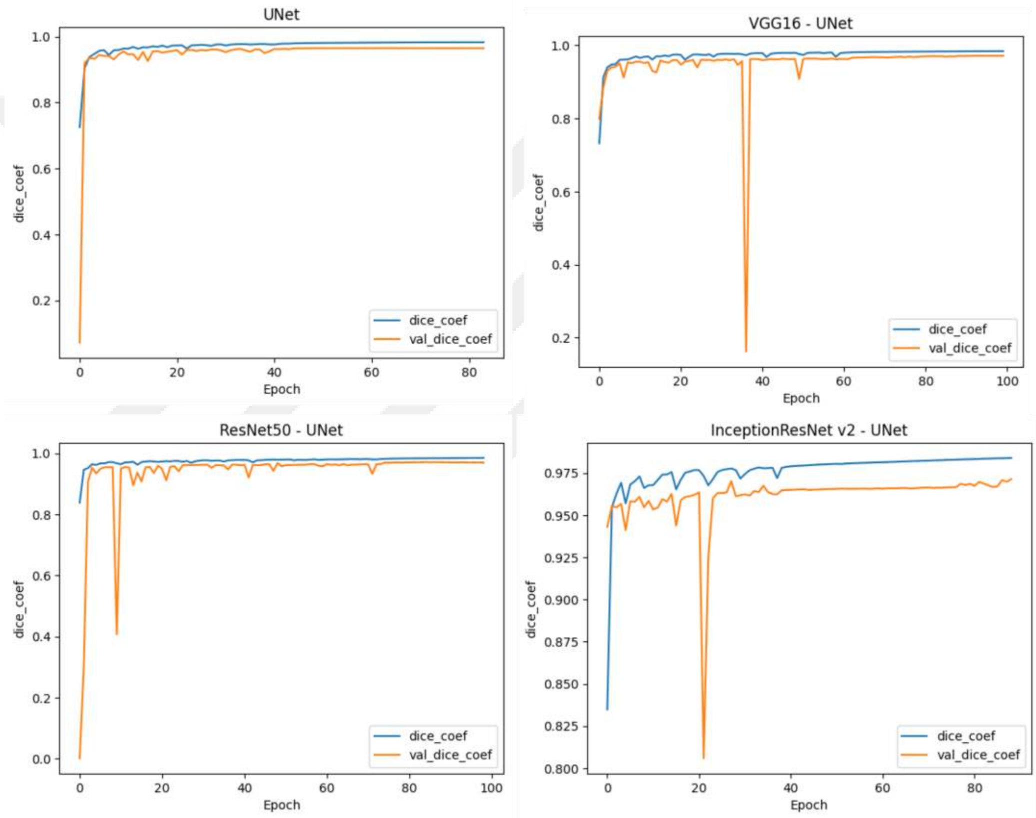

Figure 3.14 Comparison of training and validation Dice Coefficients.

IoU is another metric used to to quantify the correspondence between the predicted and actual segmentations. It measures the overlap between the prediction and the ground truth divided by the union of both areas. An IoU score close to 1 signifies almost perfect overlap, and thus, higher accuracy of segmentation. The IoU score is calculated as:

$$IoU = \frac{|Intersection|}{|Prediction| + |Ground\ Truth| - |Intersection|} \quad (3.2)$$

where:

- |Intersection| is the count of pixels that were correctly identified as maxillary sinus (foreground) in both the predicted and ground truth masks.
- |Prediction| is the total count of pixels that were identified as maxillary sinus (foreground) in the predicted mask.
- |Ground Truth| is the total count of pixels that are actually maxillary sinus (foreground) in the ground truth mask.

The IoU scores achieved by each of the trained models on both the training and validation sets are illustrated at Figure 3.15.

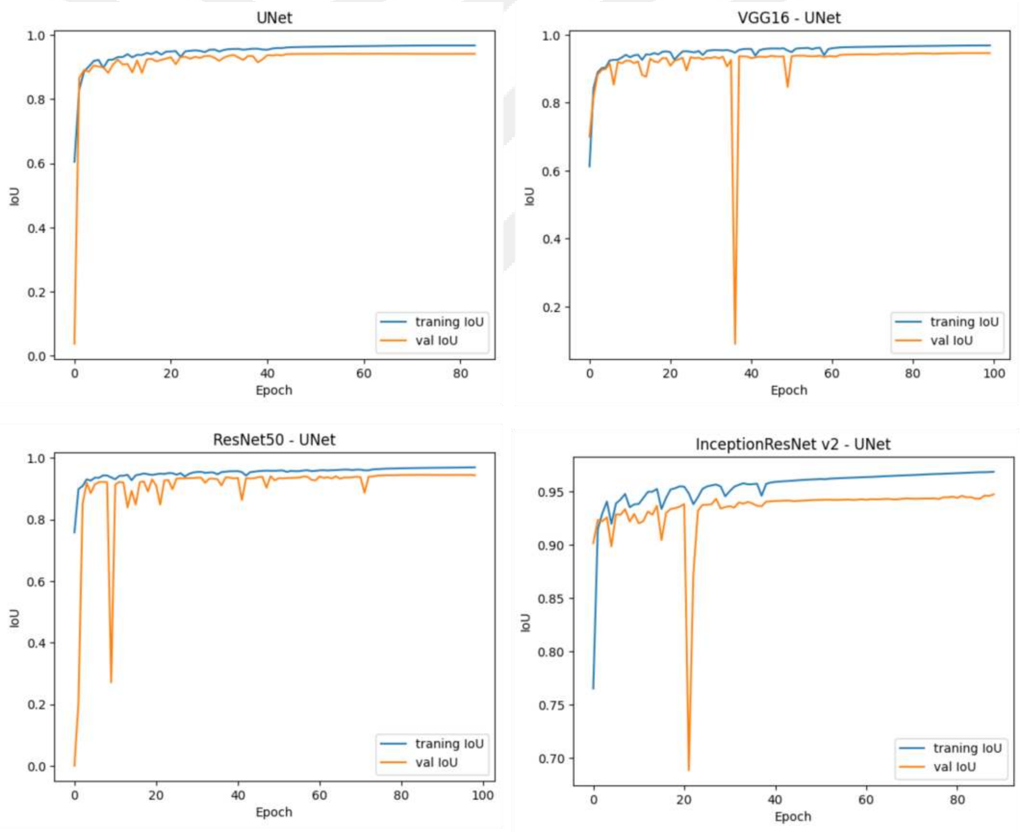

Figure 3.15 Comparison of training and validation IoU.

Recall (or sensitivity) calculates the proportion of actual positives that are correctly identified. In the context of this study, it signifies the fraction of maxillary sinus pixels that are correctly identified out of all the actual maxillary sinus pixels in the CT images. A high recall means the proposed model has effectively identified most of the maxillary sinus pixels, minimizing the chance of false negatives. Recall is calculated as:

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

where:

- TP, or True Positives, represent the pixels that were correctly classified as part of the maxillary sinus.
- FN, or False Negatives, represent the pixels which are a part of the maxillary sinus but were incorrectly classified by the model.

A high recall score suggests that the model has successfully identified the majority of the relevant pixels, whereas a low recall score indicates that the model has failed to detect a significant portion of the maxillary sinus.

The recall metric is especially critical in tasks where it is vital to correctly identify every instance of the target class. This is often the case in medical imaging, where missing an instance of an organ or anomaly could result in loss of important information. However, recall should not be used as the only measure of a model's performance. It is often analyzed alongside other metrics like precision, IoU, and Dice score to give a more complete picture of the model's accuracy. For instance, a model might have a high recall but low precision, meaning it correctly identifies many maxillary sinus pixels but also incorrectly identifies many non-sinus pixels as sinus. This results in a balancing act between sensitivity (recall) and specificity (precision). Figure 3.16 displays the recall scores achieved by each of the trained models on both the training and validation sets.

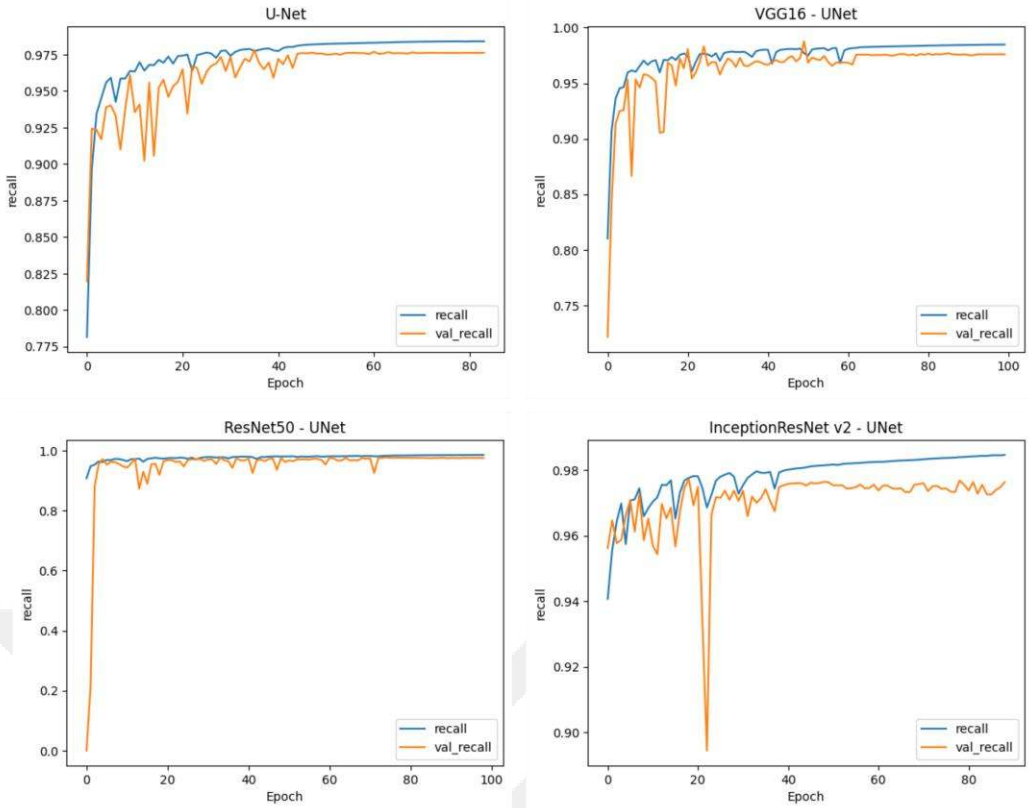

Figure 3.16 Comparison of training and validation Recall metrics.

Precision measures the proportion of correctly predicted positive observations out of the total predicted positives. In this context, it quantifies the fraction of pixels that are correctly classified as maxillary sinus out of all the pixels the model identifies as maxillary sinus. A high precision indicates a lower false-positive rate, implying that when the proposed model predicts a pixel to be a part of the maxillary sinus, it is highly likely to be correct. The formula for computing precision is:

$$Precision = \frac{TP}{TP + FP} \tag{3.4}$$

where:

- TP, or True Positives, are the pixels that the model correctly identified as being part of the maxillary sinus (foreground) in both the predicted and the actual (ground truth) masks.

- FP, or False Positives, are the pixels that the model incorrectly identified as being part of the maxillary sinus (foreground) in the predicted mask, when they are not part of the maxillary sinus in the ground truth mask.

The precision scores achieved by each of the trained models on both the training and validation sets are displayed at Figure 3.17.

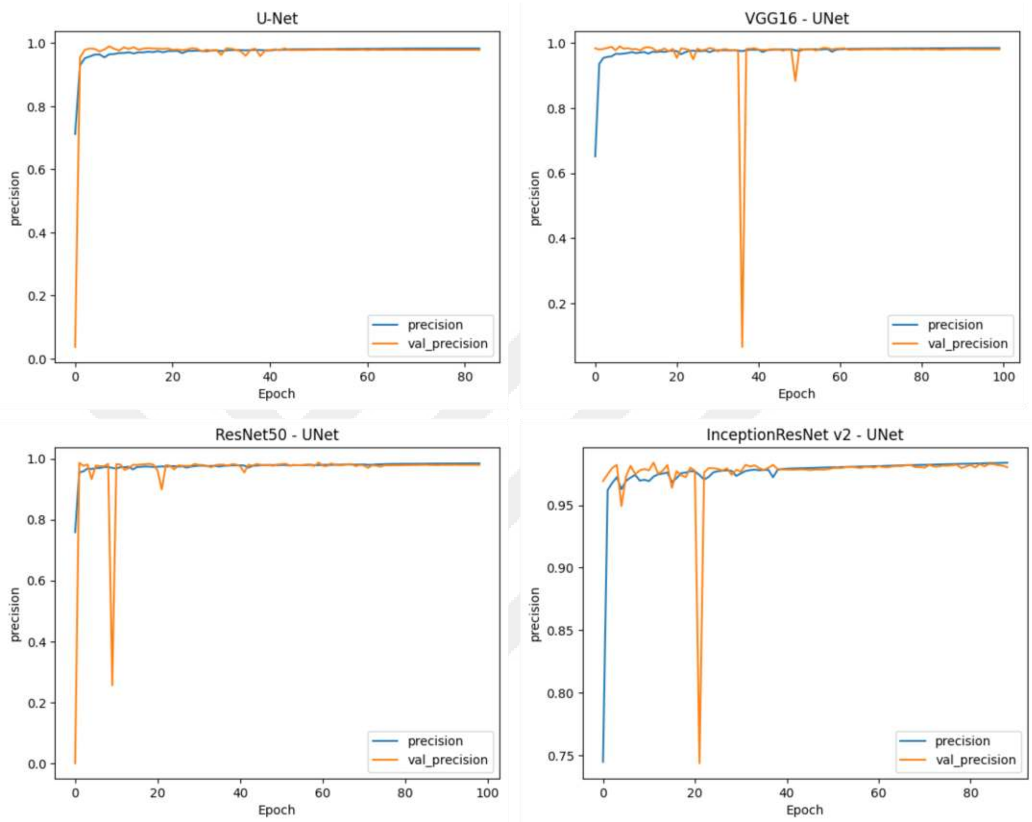

Figure 3.17 Comparison of training and validation Precision metrics.

3.4 Volume Measurement and Opacification Ratio Calculation

The volume and opacification ratio of the MS can be crucial in medical diagnosis and treatment. Understanding the volume of the maxillary sinus can assist in diagnosing various sinus conditions, including chronic sinusitis. The opacification ratios can indicate the presence of inflammation, infection, or tumors [3], [4]. In procedures like sinus surgery or dental implant placement, knowing the exact volume and condition of the maxillary sinus is essential. It helps surgeons and dentists to plan the procedure more precisely and minimize potential risks [12]. Monitoring changes in the maxillary sinus volume and opacification over time can be a critical part of evaluating the effectiveness of treatments like antibiotics or surgical interventions. The volume of the maxillary sinus can vary significantly between individuals. Understanding these variations can be essential for personalized medicine and tailoring treatments to the specific patient [97]. In the context of scientific research, studying the maxillary sinus's volume and opacification ratios can contribute to a deeper understanding of sinus pathology and the development of new diagnostic techniques or treatments.

To calculate the volume of the Maxillary Sinus, a two- dimensional ROI was extracted from each CT slice. The area of the segmented region in each slice was multiplied by the slice thickness to obtain the volume contribution of that particular slice [36], [41]. The total volume of the Maxillary Sinus was then determined by summing up the volume contributions from all slices, as shown in Eq. 3.5:

$$\mathrm{V} = \sum_{i=1}^{n} \mathrm{A}(s_i)\,\Delta s \tag{3.5}$$

where V is the total volume of the MS, n is the number of MS containing slices, $\mathrm{A}(s_i)$ is the area of segmented region in the i-th slice, Δs is the slice thickness. The area of the segmented region in each slice was measured by multiplying the number of pixels in the segmented region by the pixel size. The CT images used in the study had varying pixel sizes and slice thickness values. These values were automatically extracted from the metadata of the DICOM files of the CT images. The ground truth volumes are automatically calculated by 3D Slicer. It provides accurate volume measurements for reference.

To segment the air section (black) and mucosa inflammation region (gray) in all slices, thresholding was applied using Otsu's method [98]. Otsu’s thresholding is a widely used image segmentation technique for automatically determining the optimal threshold value to separate the pixels of a grayscale image into two classes: foreground and background. The threshold value is chosen such that the pixel intensities are divided into two groups in a manner that maximizes the difference between the two groups, making the resulting segmentation more accurate. The generated masks were applied to the original CT images, and only the pixels within the mask regions were extracted. Thresholding was just applied to maxillary sinus patches. A few samples of the final segmentation results are shown in Figure 3.18. Opacification ratio was calculated as the percentage of the volume of the inflamated region to the total sinus volume. It provides a measure of the extent to which the sinus is affected by inflammation.

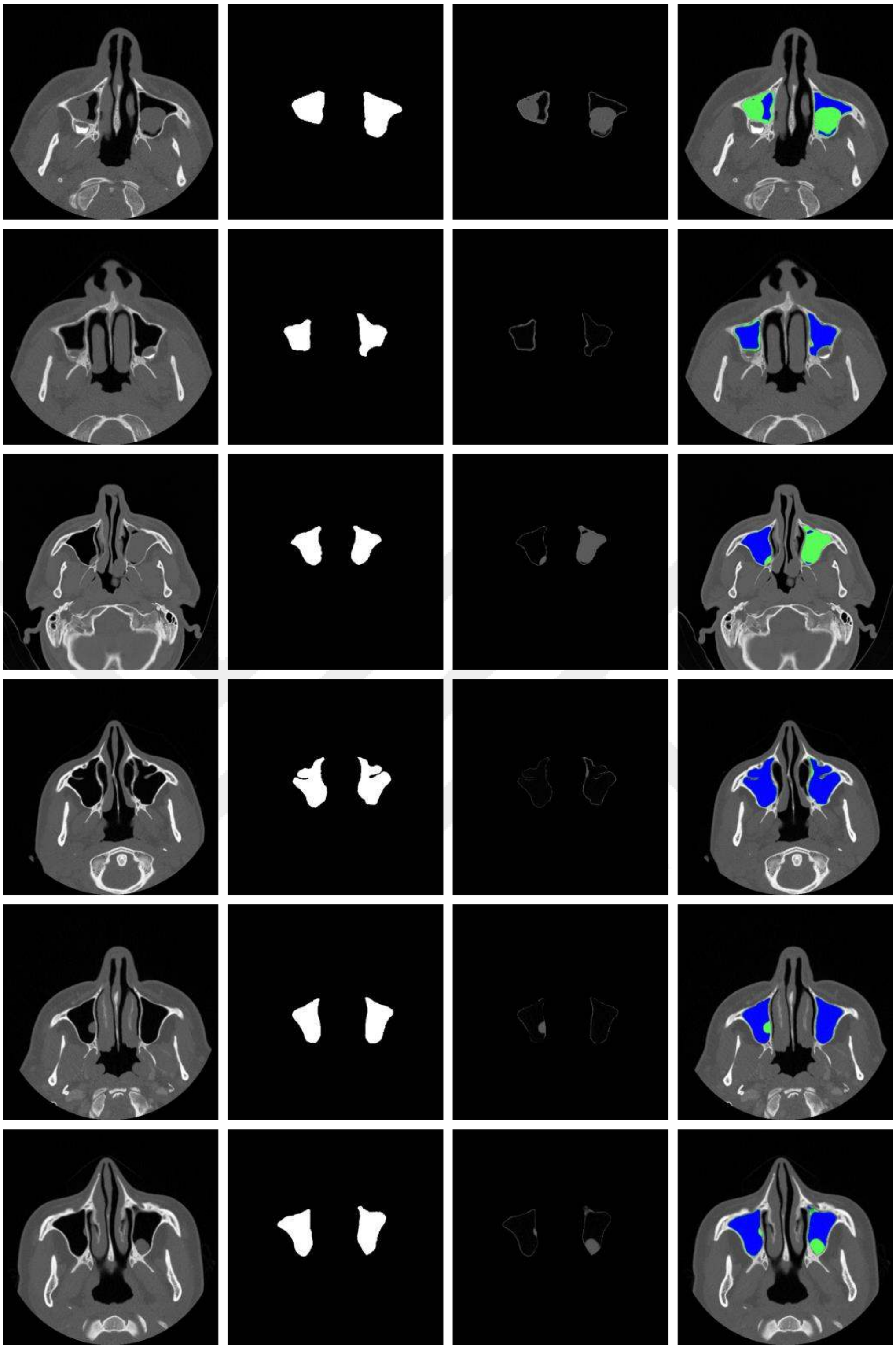

Figure 3.18 First Column: CT slices, Second Column: automatic segmentation results produced by the CNN models, Third Column: extracted maxillary Sinus patches, Fourth Column: segmentation results with air (blue) and opacification (green) seperated based on their pixel intensity values.

CHAPTER 4
RESULTS and DISCUSSION

4.1 Performance Comparison of the Models

In this study, a plain U-Net model was employed alongside three U-Net variations incorporating VGG16, ResNet50, and InceptionResNet v2 as encoders for MS segmentation. The models were trained and validated on a specially labelled dataset of CT images, emphasizing the maxillary sinus region. All the models were evaluated on test dataset according to Dice, IoU, Precision, and Recal metrics. Table 4.1 shows the performance of the models. The Inception ResNet v2 – U-Net model emerged as the top performer among the models tested. Its success underscores the potential benefits of integrating sophisticated deep learning architectures within the U-Net framework. The use of pre-trained models (VGG16, ResNet50, InceptionResNet v2) as encoders in the U-Net architecture demonstrated improved performance compared to the plain U-Net model. This validates the hypothesis that transfer learning can boost the segmentation task, bringing with it inherent knowledge within these pre-trained models. Despite being outperformed by its more complex counterparts, the plain U-Net model exhibited satisfactory results. Its relatively simpler architecture and satisfactory performance highlight its suitability for applications where computational resources may be constrained or where a more straightforward approach is desired.

Table 4.1 Performance comparison of the models.

MODEL	Dice	IoU	Precision	Recall
U-Net	0.95021	0.91870	0.96592	0.94825
VGG16-UNet	0.95494	0.92448	0.96884	0.95257
ResNet50-UNet	0.95203	0.92144	0.96635	0.94962
Inception ResNet v2 - UNet	0.96671	0.94004	0.96771	0.97050

4.2 Visual Assessment of Segmentation Results

Figure 4.1-11 show examples of segmentation results of 2D slices containing MS with bone structures, opacified MS, and FP, FN results from all models, respectively. Segmentation results can be easily evaluated visually by comparing them with the original CT images and GT masks found in the images. The segmentation results samples including FP and FN pixels for Inception-ResNet v2-UNet model were not provided because of the lack of the visually detectable samples.

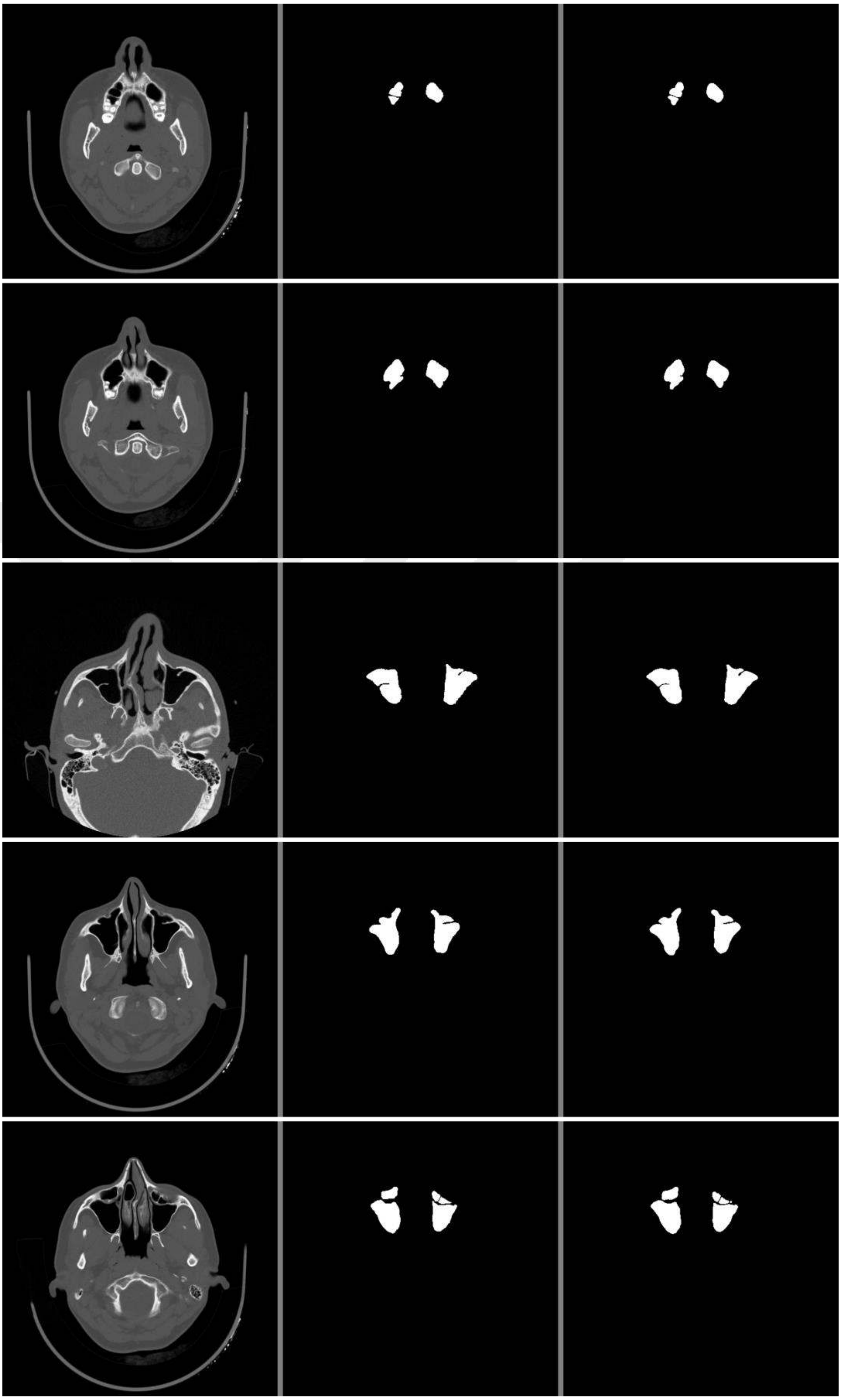

Figure 4.1 Examples of segmentation results for CT slices with bony structures of the plain U-Net model.

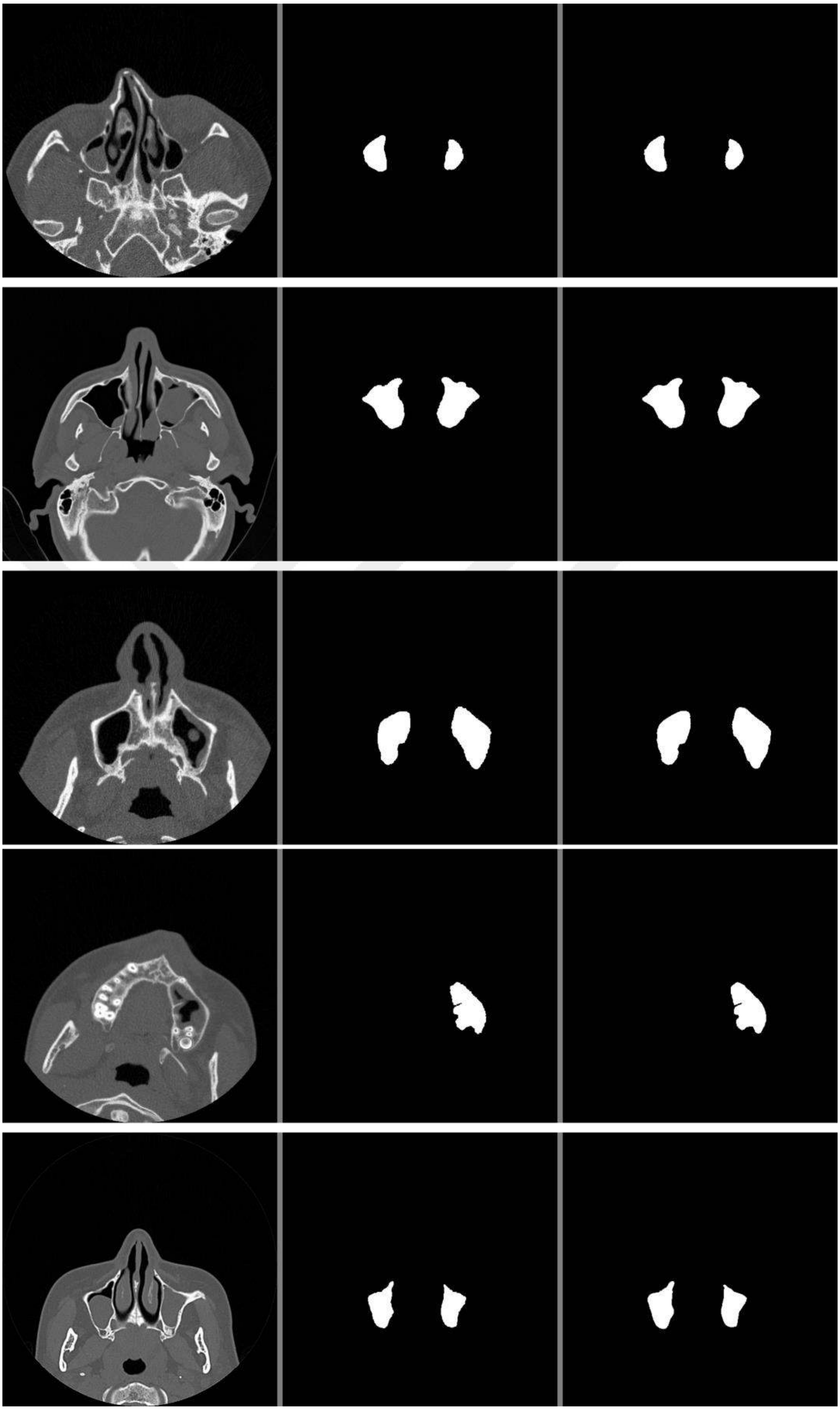

Figure 4.2 Examples of segmentation results for CT slices with opacified MS of the plain U-Net model.

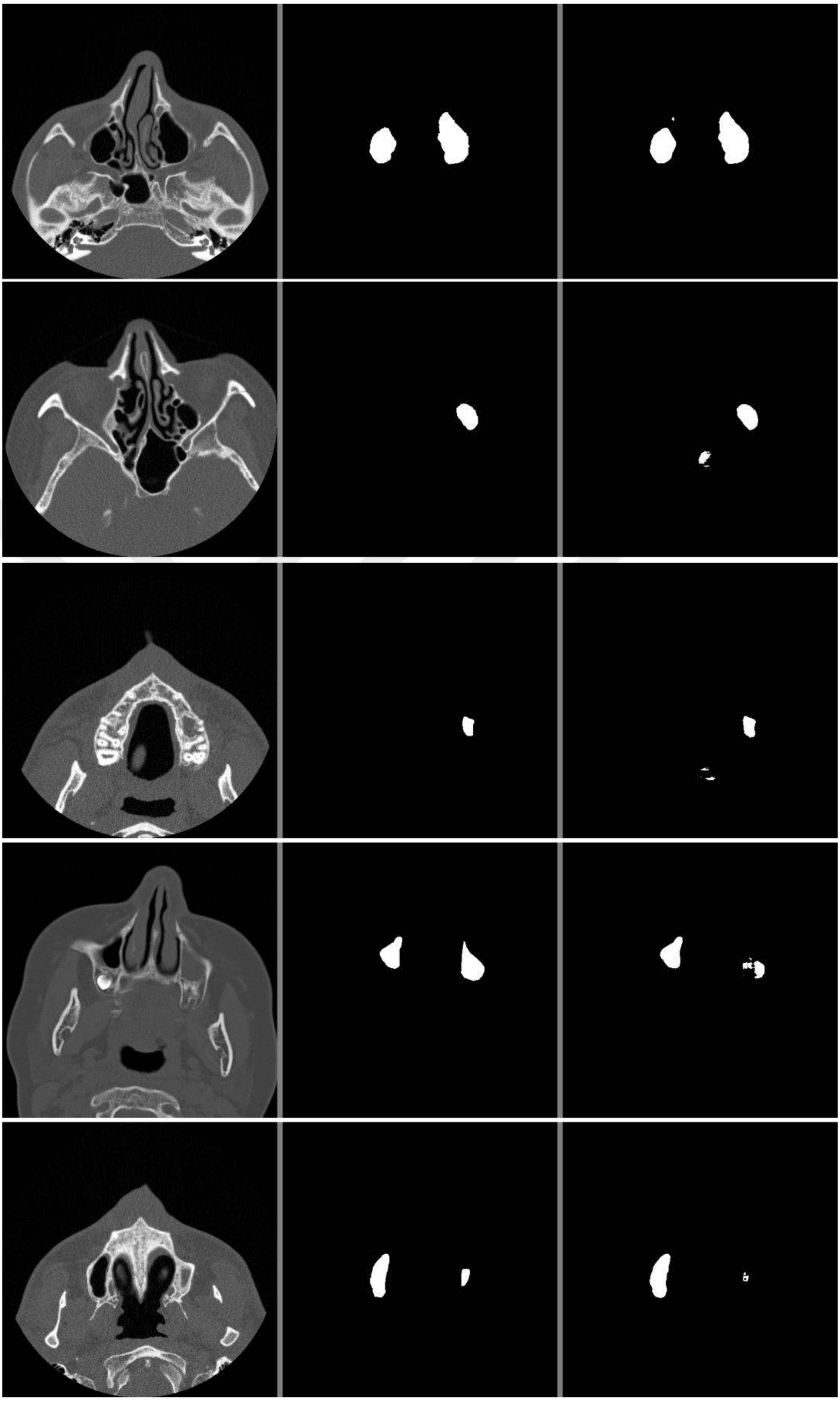

Figure 4.3 Examples of segmentation results for CT slices with FP and FN pixels of the plain U-Net model.

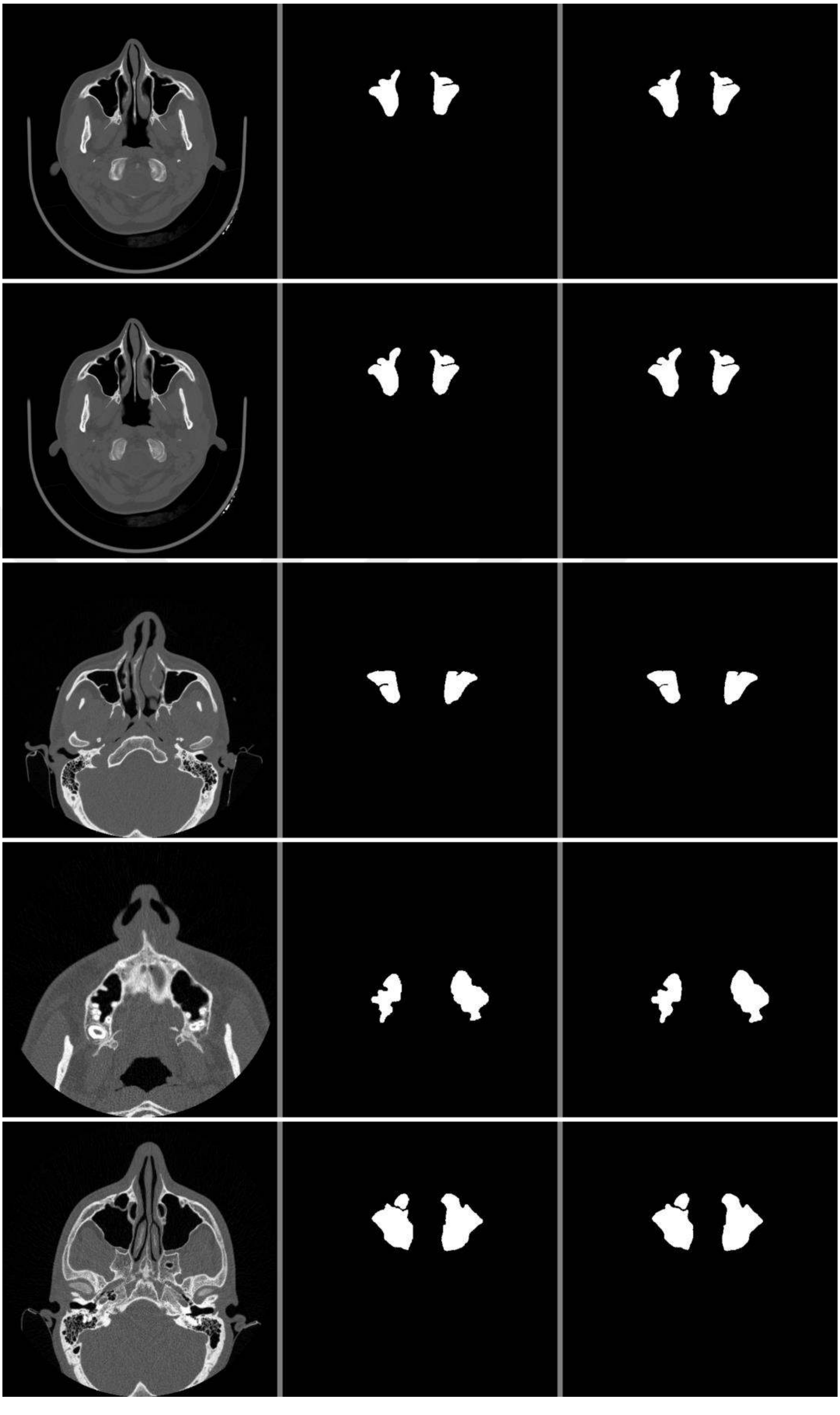

Figure 4.4 Examples of segmentation results for CT slices with bony structures of the VGG16-UNet model.

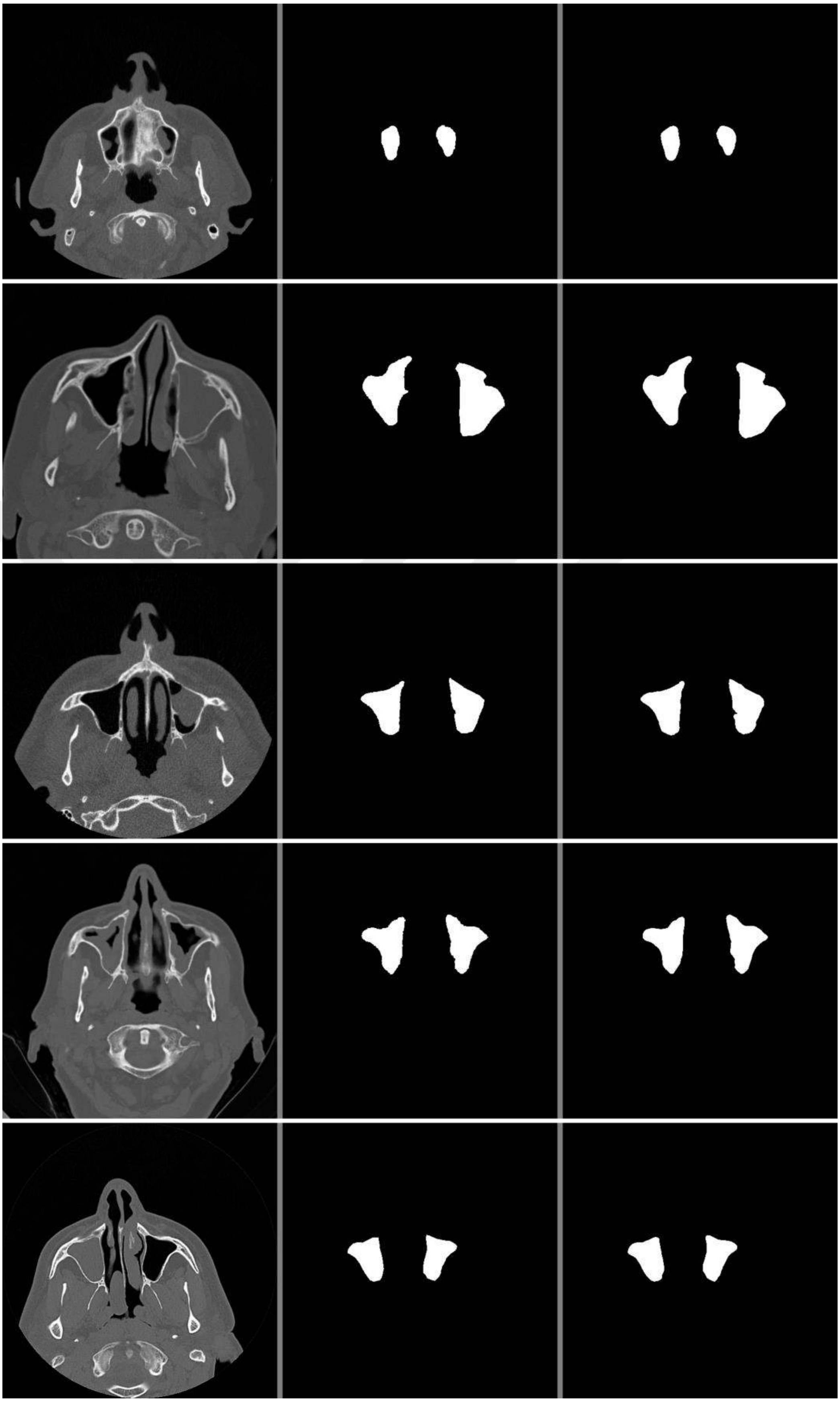

Figure 4.5 Examples of segmentation results for CT slices with opacified MS of the VGG16-UNet model.

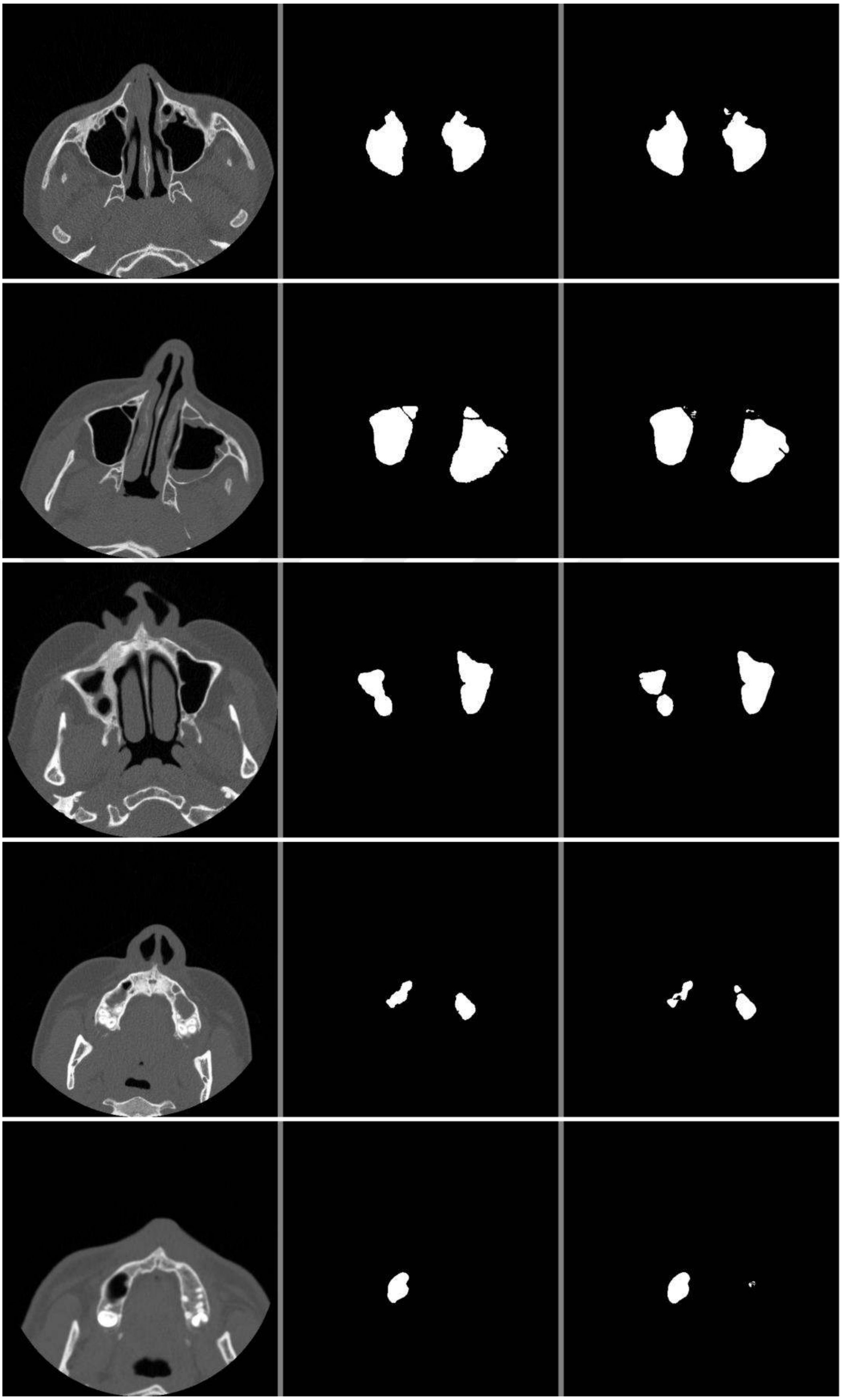

Figure 4.6 Examples of segmentation results for CT slices with FP and FN pixels of the VGG16-UNet model.

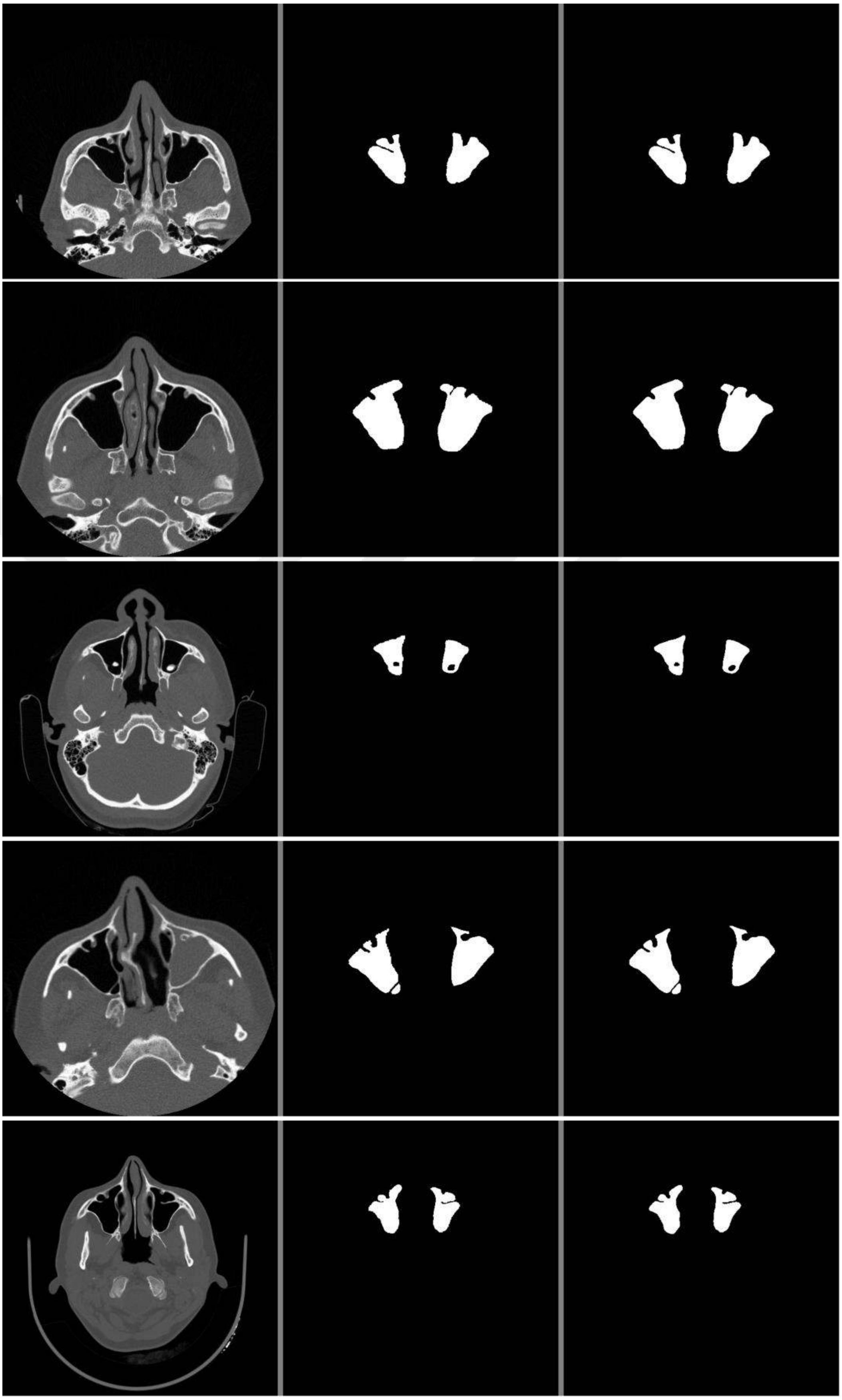

Figure 4.7 Examples of segmentation results for CT slices with bony structures of the ResNet50-UNet model.

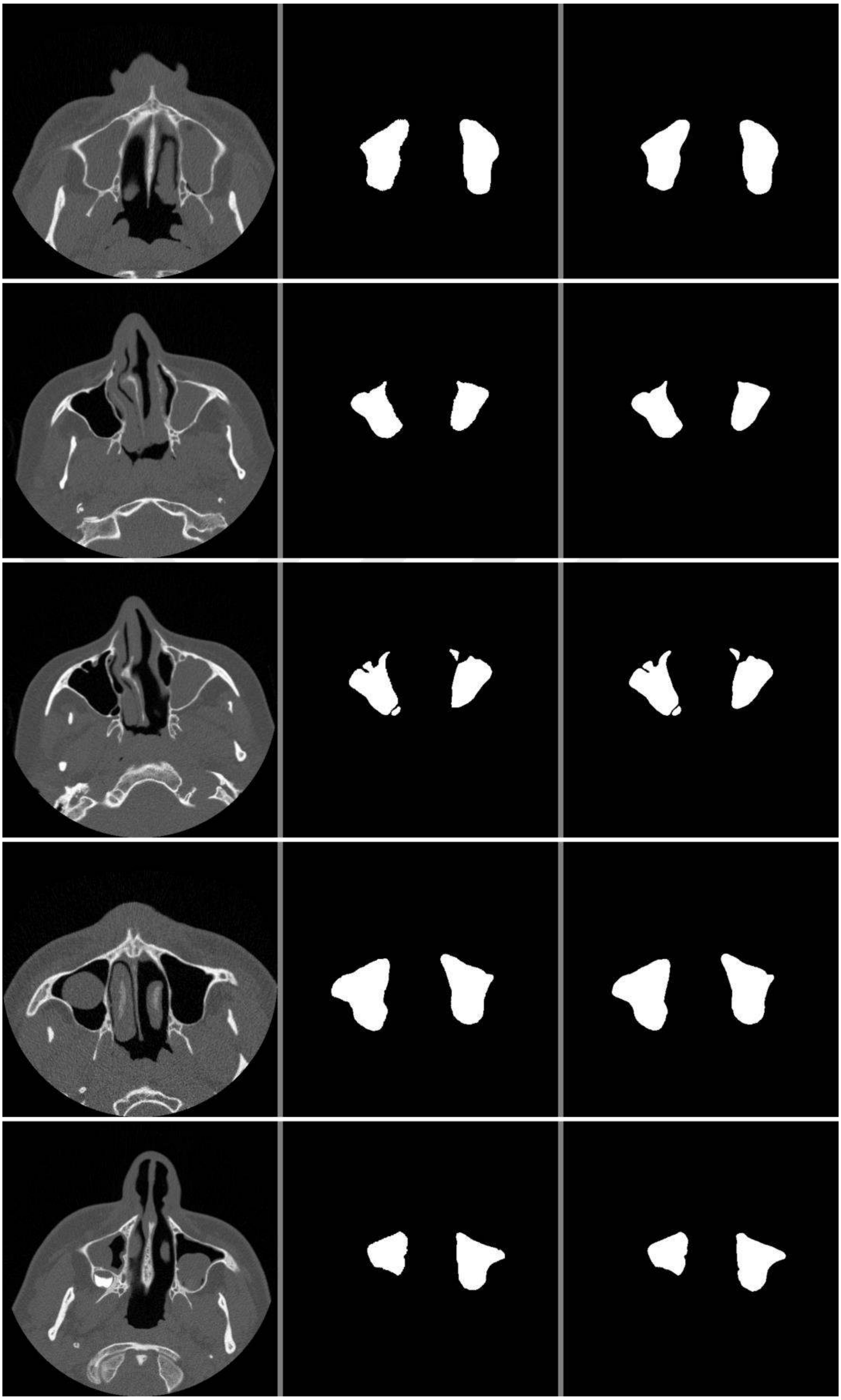

Figure 4.8 Examples of segmentation results for CT slices with opacified MS of the ResNet50-UNet model.

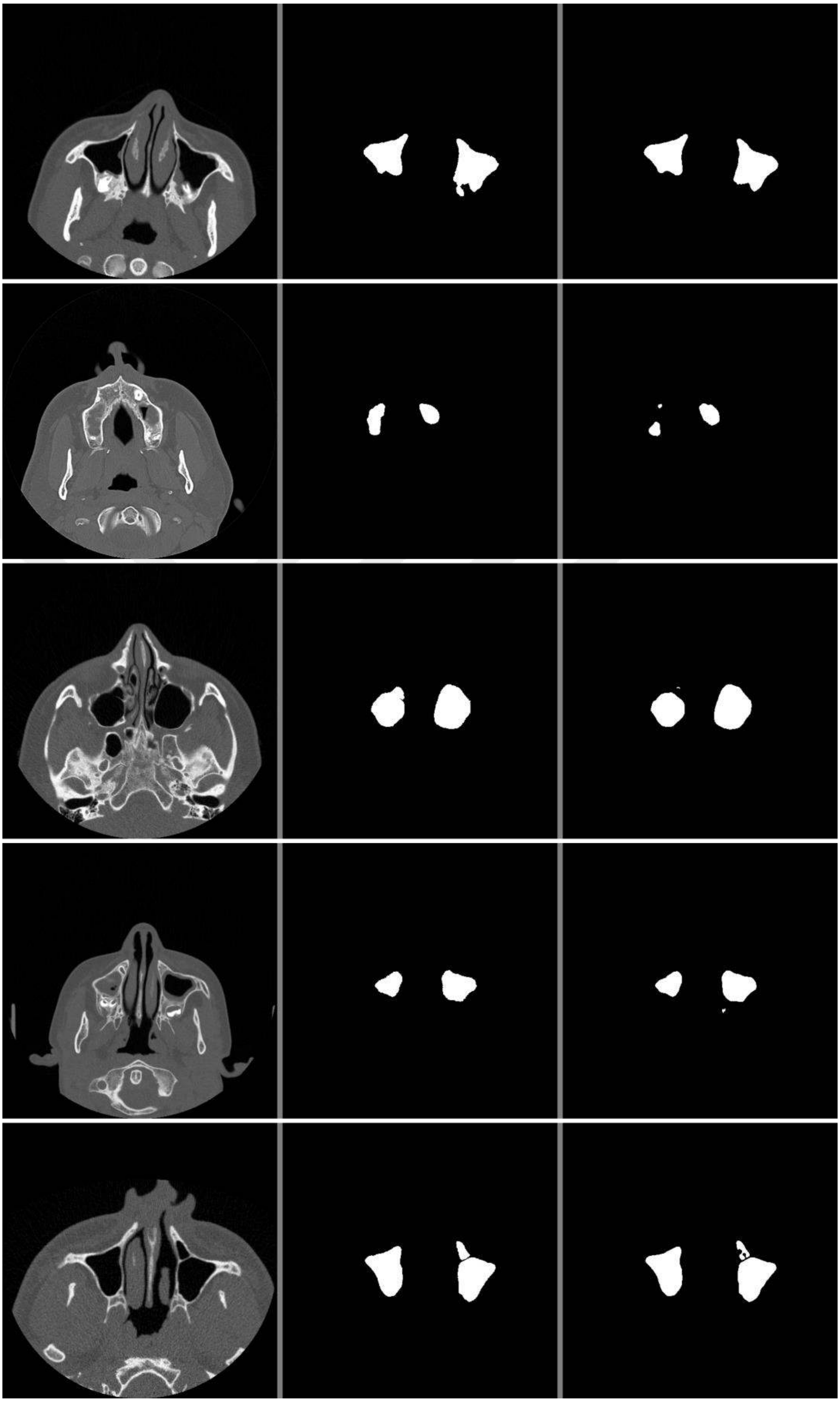

Figure 4.9 Examples of segmentation results for CT slices with FP and FN pixels of the ResNet50-UNet model.

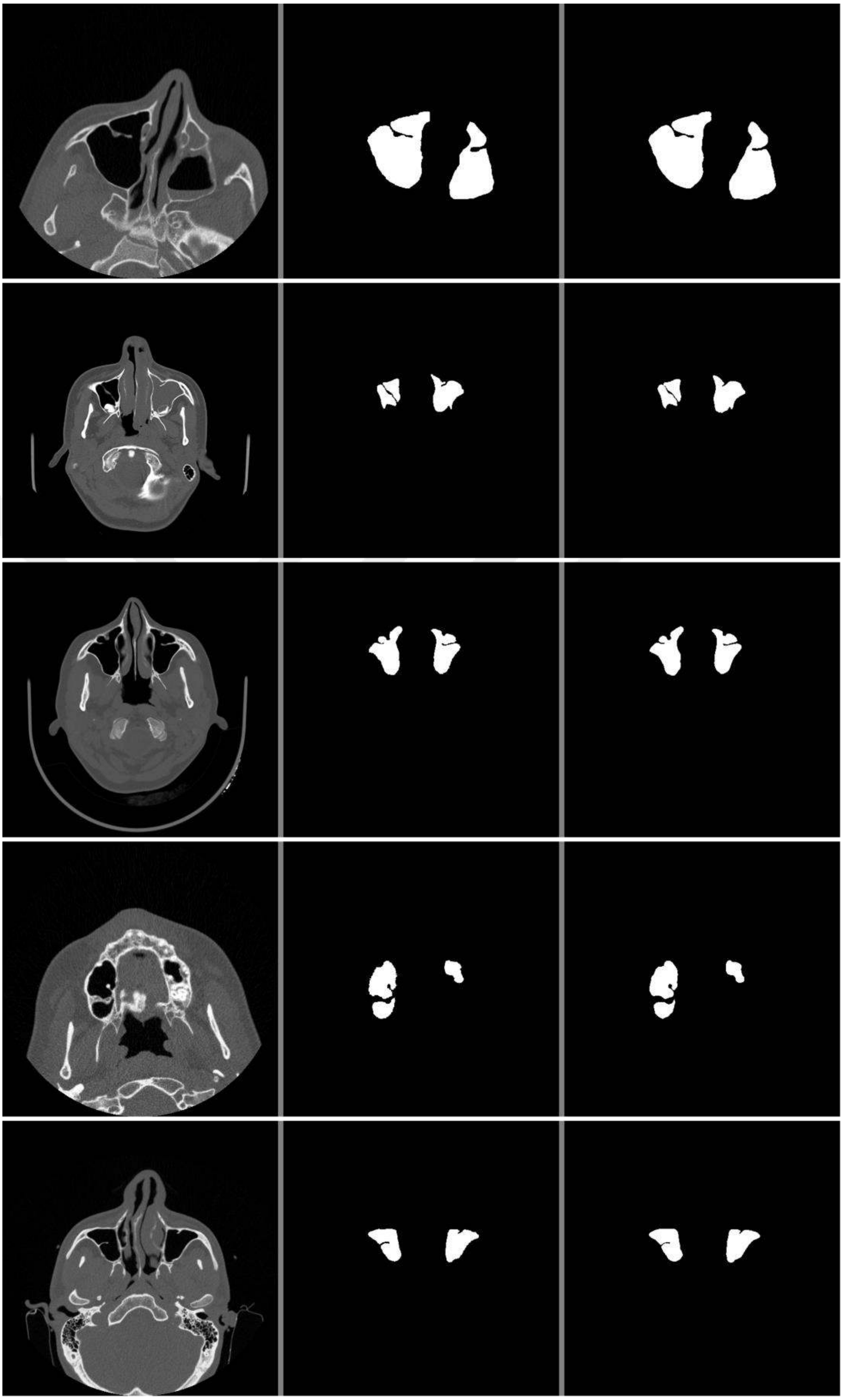

Figure 4.10 Examples of segmentation results for CT slices with bony structures of the Inception-ResNet v2-UNet model.

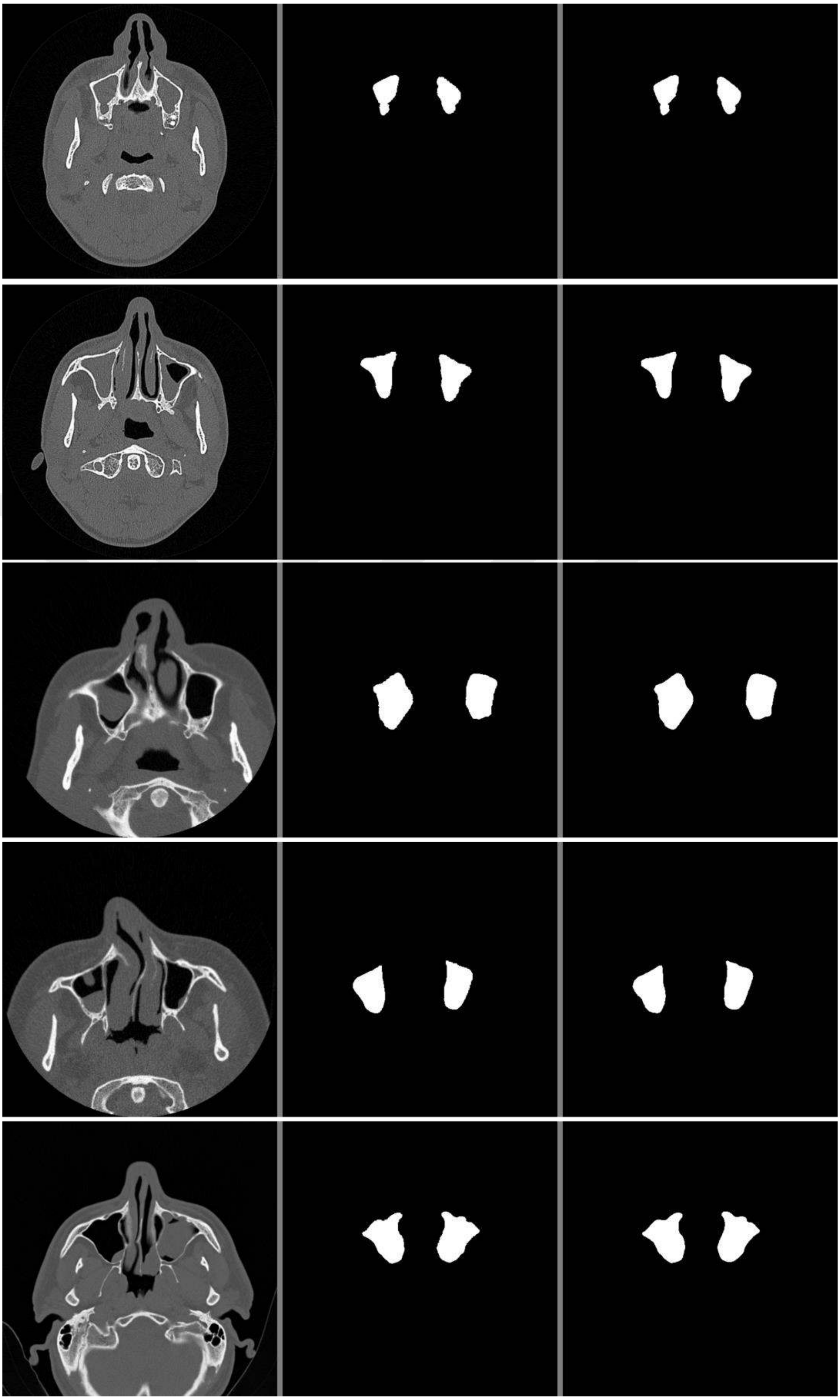

Figure 4.11 Examples of segmentation results for CT slices with opacified MS of the Inception-ResNet v2-UNet model.

4.3 Comparison of the Calculated Volumes and Opacification Ratios

Automatic segmentation results of eight complete CT scans produced by all the models were used for volume measurement and opacification ratio calculation. For volume comparison, samples were selected from studies of different sexes, age groups, and different opacity levels. Ground truth volumes were calculated automatically in 3D Slicer platform to campare the models' performances. The final ground truth volumes, calculated total and opacified volumes, and opacification ratios for some samples are compared in Table 4.2 - 5 .

Table 4.2 The total and opacified volumes of MS and opacification ratios of segmentation results produced by the plain U-Net model for some samples.

Sample No	Gender	Age	GT Vol. (cm^3)	Total Vol. (cm^3)	Op. Vol. (cm^3)	Op. Ratio (%)
1	F	22	34.78	33.06	2.62	7.92
2	F	20	33.63	33.43	3.13	9.36
3	F	37	41.52	42.08	4.39	10.43
4	F	26	21.67	22.19	1.18	5.32
5	M	19	48.49	48.35	15.42	31.89
6	M	26	50.98	50.77	40.96	80.68
7	F	30	22.29	22.51	12.17	54.06
8	M	24	41.89	41.97	29.11	69.36

Table 4.3 The total and opacified volumes of MS and opacification ratios of segmentation results produced by the VGG16-UNet model for some samples.

Sample No	Gender	Age	GT Vol. (cm^3)	Total Vol. (cm^3)	Opac. Vol. (cm^3)	Op. Ratio (%)
1	F	22	34.78	35.04	2.61	7.44
2	F	20	33.63	33.44	3.11	9.30
3	F	37	41.52	42.18	4.40	10.43
4	F	26	21.67	21.30	2.18	10.23
5	M	19	48.49	48.38	15.48	31.89
6	M	26	50.98	50.59	40.78	80.61
7	F	30	22.29	21.18	12.26	57.88
8	M	24	41.89	41.84	28.98	69.27

Table 4.4 The total and opacified volumes of MS and opacification ratios of segmentation results produced by the ResNet50-UNet model for some samples.

Sample No	Gender	Age	GT Vol. (cm^3)	Total Vol. (cm^3)	Opac. Vol. (cm^3)	Op. Ratio (%)
1	F	22	34.78	34.94	2.57	7.35
2	F	20	33.63	33.50	3.15	9.40
3	F	37	41.52	41.88	4.25	10.15
4	F	26	21.67	21.58	2.24	10.38
5	M	19	48.49	48.35	15.49	32.04
6	M	26	50.98	50.74	40.93	80.66
7	F	30	22.29	20.38	11.04	54.17
8	M	24	41.89	42.02	29.17	69.42

Table 4.5 The total and opacified volumes of MS and opacification ratios of segmentation results produced by the Inception-ResNet v2-UNet model for some samples.

Sample No	Gender	Age	GT Vol. (cm^3)	Total Vol. (cm^3)	Opac. Vol. (cm^3)	Op. Ratio (%)
1	F	22	34.78	35.10	2.62	7.46
2	F	20	33.63	33.39	3.08	9.22
3	F	37	41.52	42.64	4.65	10.91
4	F	26	21.67	21.48	2.18	10.14
5	M	19	48.49	47.58	15.51	32.60
6	M	26	50.98	50.89	41.09	80.74
7	F	30	22.29	22.71	12.38	54.51
8	M	24	41.89	42.11	29.25	69.46

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

Accurate segmentation of the maxillary sinus from CT images is a critical and complex task that has immense significance in the field of maxillofacial radiology. This study makes a considerable contribution by evaluating and comparing the performance of a plain U-Net model with three U-Net variants, employing VGG16, ResNet50, and Inception-ResNet v2 as encoders.

The results highlighted the Inception-ResNet v2-UNet model as the most effective, with a DSC of 0.96671 and an IoU of 0.94004. The use of pre-trained encoders demonstrated their effectiveness in improving segmentation, affirming the importance of transfer learning in medical image analysis. Despite this, even the plain U-Net model demonstrated satisfactory results, with a DSC of 0.95021 and an IoU of 0.91870.

This study not only explored the basic segmentation of the maxillary sinus but also delineated the air section and mucosal inflammation within the sinus. The ability to calculate essential diagnostic metrics such as volume and opacification ratios was another remarkable achievement.

The study also acknowledged the computational constraints, showing that even simpler models like plain U-Net could be effective when resources are limited. It strikes a balance between complexity and efficacy and, provides options for various practical scenarios.

5.2 Future Works

The success of the models presented in this study illustrates the potential of deep learning in medical image segmentation, particularly for complex anatomical structures like the maxillary sinus. Although the results are promising, there is room for improvement and exploration. Expanding the dataset to include various age groups, ethnic backgrounds, and medical conditions could create a more robust model, capable of handling diverse clinical scenarios. Future studies could aim to integrate these models into existing radiological workflows. Close collaboration with clinicians and radiologists will be essential to ensure that the tools developed meet the real-world needs of medical practitioners. The methodologies developed could be extended to analyze other sinus regions. This could lead to new ways of diagnosing and planning treatment in many areas of medicine.

REFERENCES

[1] A. Whyte and R. Boeddinghaus, “The maxillary sinus: physiology, development and imaging anatomy,” *Dentomaxillofacial Radiol.*, vol. 48, no. 8, p. 20190205, Dec. 2019, doi: 10.1259/dmfr.20190205.

[2] G. W. Bell, B. B. Joshi, and R. I. Macleod, “Maxillary sinus disease: diagnosis and treatment,” *Br. Dent. J.*, vol. 210, no. 3, pp. 113–118, Feb. 2011, doi: 10.1038/sj.bdj.2011.47.

[3] D. L. Hamilos, “Chronic sinusitis,” *J. Allergy Clin. Immunol.*, vol. 106, no. 2, pp. 213–227, Aug. 2000, doi: 10.1067/mai.2000.109269.

[4] A. Ahovuo-Saloranta, U.-M. Rautakorpi, O. V. Borisenko, H. Liira, J. W. W. Jr, and M. Mäkelä, “Antibiotics for acute maxillary sinusitis,” *Cochrane Database Syst. Rev.*, no. 2, 2008, doi: 10.1002/14651858.CD000243.pub2.

[5] T. Norlander, M. Fukami, K. M. Westrin, P. Stierna, and B. Carlsöö, “Formation of mucosal polyps in the nasal and maxillary sinus cavities by infection,” *Otolaryngol. Neck Surg.*, vol. 109, no. 3, pp. 522–529, Sep. 1993, doi: 10.1177/019459989310900322.

[6] J. Składzień, J. A. Litwin, M. Nowogrodzka-Zagórska, and W. Wierzchowski, “Morphological and clinical characteristics of antrochoanal polyps: comparison with chronic inflammation-associated polyps of the maxillary sinus,” *Auris. Nasus. Larynx*, vol. 28, no. 2, pp. 137–141, Apr. 2001, doi: 10.1016/S0385-8146(00)00108-5.

[7] N. Bhattacharyya, “Factors affecting survival in maxillary sinus cancer,” *J. Oral Maxillofac. Surg.*, vol. 61, no. 9, pp. 1016–1021, Sep. 2003, doi: 10.1016/S0278-2391(03)00313-6.

[8] S. G. Press, “Odontogenic tumors of the maxillary sinus,” *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 16, no. 1, pp. 47–54, Feb. 2008, doi: 10.1097/MOO.0b013e3282f419da.

[9] M. Abdel-Aziz, H. El-Hoshy, K. Azooz, N. Naguib, and A. Hussein, “Maxillary sinus mucocele: predisposing factors, clinical presentations, and treatment,” *Oral Maxillofac. Surg.*, vol. 21, no. 1, pp. 55–58, Mar. 2017, doi: 10.1007/s10006-016-0599-5.

[10] F. Caylakli, H. Yavuz, A. C. Cagici, and L. N. Ozluoglu, “Endoscopic sinus surgery for maxillary sinus mucoceles,” *Head Face Med.*, vol. 2, no. 1, p. 29, Dec. 2006, doi: 10.1186/1746-160X-2-29.

[11] S. M. Kim, “Definition and management of odontogenic maxillary sinusitis,” *Maxillofac. Plast. Reconstr. Surg.*, vol. 41, no. 1, p. 13, Dec. 2019, doi: 10.1186/s40902-019-0196-2.

[12] T. Regnstrand, M. Ezeldeen, S. Shujaat, K. Ayidh Alqahtani, D. Benchimol, and R. Jacobs, “Three-dimensional quantification of the relationship between the upper first molar and maxillary sinus,” *Clin. Exp. Dent. Res.*, vol. 8, no. 3, pp. 750–756, Jun. 2022, doi: 10.1002/cre2.561.

[13] J. P. van den Bergh, C. M. ten Bruggenkate, F. J. Disch, and D. B. Tuinzing, “Anatomical aspects of sinus floor elevations,” *Clin. Oral Implants Res.*, vol. 11, no. 3, pp. 256–265, Jun. 2000, doi: 10.1034/j.1600-0501.2000.011003256.x.

[14] H.-L. Chan and H.-L. Wang, “Sinus Pathology and Anatomy in Relation to Complications in Lateral Window Sinus Augmentation,” *Implant Dent.*, vol. 20, no. 6, pp. 406–412, Dec. 2011, doi: 10.1097/ID.0b013e3182341f79.

[15] M.-J. Kim *et al.*, “Maxillary Sinus Septa: Prevalence, Height, Location, and Morphology. A Reformatted Computed Tomography Scan Analysis,” *J. Periodontol.*, vol. 77, no. 5, pp. 903–908, May 2006, doi: 10.1902/jop.2006.050247.
[16] Y. Ariji, T. Kuroki, S. Moriguchi, E. Ariji, and S. Kanda, “Age changes in the volume of the human maxillary sinus: a study using computed tomography,” *Dento Maxillo Facial Radiol.*, vol. 23, no. 3, pp. 163–168, Aug. 1994, doi: 10.1259/dmfr.23.3.7835518.
[17] D. Lorkiewicz-Muszyńska *et al.*, “Development of the maxillary sinus from birth to age 18. Postnatal growth pattern,” *Int. J. Pediatr. Otorhinolaryngol.*, vol. 79, no. 9, pp. 1393–1400, Sep. 2015, doi: 10.1016/j.ijporl.2015.05.032.
[18] A. Ikeda, M. Ikeda, and A. Komatsuzaki, “A CT study of the course of growth of the maxillary sinus: normal subjects and subjects with chronic sinusitis,” *ORL J. Oto-Rhino-Laryngol. Its Relat. Spec.*, vol. 60, no. 3, pp. 147–152, 1998, doi: 10.1159/000027584.
[19] J. M. Sánchez Fernández, J. A. Anta, “Morphometric Study of the Paranasal Sinuses in Normal and Pathological Conditions,” *Acta Otolaryngol. (Stockh.)*, vol. 120, no. 2, pp. 273–278, Jan. 2000, doi: 10.1080/000164800750001080.
[20] E. Sivaslı *et al.*, “Anatomic variations of the paranasal sinus area in pediatric patients with chronic sinusitis,” *Surg. Radiol. Anat.*, vol. 24, no. 6, pp. 399–404, Dec. 2002, doi: 10.1007/s00276-002-0074-x.
[21] P. Kuligowski, A. Jaroń, O. Preuss, E. Gabrysz-Trybek, J. Bladowska, and G. Trybek, “Association between Odontogenic and Maxillary Sinus Conditions: A Retrospective Cone-Beam Computed Tomographic Study,” *J. Clin. Med.*, vol. 10, no. 13, p. 2849, Jun. 2021, doi: 10.3390/jcm10132849.
[22] S. Bathla, R. Fry, and K. Majumdar, “Maxillary sinus augmentation,” *J. Indian Soc. Periodontol.*, vol. 22, no. 6, p. 468, 2018, doi: 10.4103/jisp.jisp_236_18.
[23] X. Qi, J. Zhong, and S. Cui, “Adaptive Localizing Region-Based Level Set for Segmentation of Maxillary Sinus Based on Convolutional Neural Networks,” *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, Nov. 2021, doi: 10.1155/2021/4824613.
[24] D. W. Kennedy, “Functional Endoscopic Sinus Surgery: Technique,” *Arch. Otolaryngol.*, vol. 111, no. 10, pp. 643–649, Oct. 1985, doi: 10.1001/archotol.1985.00800120037003.
[25] J. Kim and H. Jang, “A review of complications of maxillary sinus augmentation and available treatment methods,” *J. Korean Assoc. Oral Maxillofac. Surg.*, vol. 45, no. 4, p. 220, 2019, doi: 10.5125/jkaoms.2019.45.4.220.
[26] T. Starch-Jensen and J. D. Jensen, “Maxillary Sinus Floor Augmentation: a Review of Selected Treatment Modalities,” *J. Oral Maxillofac. Res.*, vol. 8, no. 3, Sep. 2017, doi: 10.5037/jomr.2017.8303.
[27] M. Cellina *et al.*, “Segmentation procedures for the assessment of paranasal sinuses volumes,” *Neuroradiol. J.*, vol. 34, no. 1, pp. 13–20, Feb. 2021, doi: 10.1177/1971400920946635.
[28] M. Pérez Sayáns *et al.*, “Volumetric study of the maxillary sinus in patients with sinus pathology,” *PLOS ONE*, vol. 15, no. 6, p. e0234915, Jun. 2020, doi: 10.1371/journal.pone.0234915.
[29] J. Luz, D. Greutmann, D. Wiedemeier, C. Rostetter, M. Rücker, and B. Stadlinger, “3D-evaluation of the maxillary sinus in cone-beam computed tomography,” *Int. J. Implant Dent.*, vol. 4, no. 1, p. 17, Dec. 2018, doi: 10.1186/s40729-018-0128-4.
[30] U. L. Demir, M. E. Akca, R. Ozpar, C. Albayrak, and B. Hakyemez, “Anatomical correlation between existence of concha bullosa and maxillary sinus volume,” *Surg. Radiol. Anat.*, vol. 37, no. 9, pp. 1093–1098, Nov. 2015, doi: 10.1007/s00276-015-1459-y.
[31] T. N. Andersen *et al.*, “Accuracy and precision of manual segmentation of the maxillary sinus in MR images—a method study,” *Br. J. Radiol.*, p. 20170663, Mar. 2018, doi: 10.1259/bjr.20170663.
[32] S. Pirner *et al.*, “CT-based manual segmentation and evaluation of paranasal sinuses,” *Eur. Arch. Otorhinolaryngol.*, vol. 266, no. 4, pp. 507–518, Apr. 2009, doi: 10.1007/s00405-008-0777-7.

[33] K. Tingelhoff *et al.*, “Analysis of manual segmentation in paranasal CT images,” *Eur. Arch. Otorhinolaryngol.*, vol. 265, no. 9, pp. 1061–1070, Sep. 2008, doi: 10.1007/s00405-008-0594-z.
[34] C.-F. J. Kuo and S.-C. Liu, “Fully Automatic Segmentation, Identification and Preoperative Planning for Nasal Surgery of Sinuses Using Semi-Supervised Learning and Volumetric Reconstruction,” *Mathematics*, vol. 10, no. 7, p. 1189, Apr. 2022, doi: 10.3390/math10071189.
[35] H. Shi, W. C. Scarfe, and A. G. Farman, “Maxillary Sinus 3D Segmentation and Reconstruction from Cone Beam CT Data Sets,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 1, no. 2, pp. 83–89, Aug. 2006, doi: 10.1007/s11548-006-0041-9.
[36] S. H. Cho *et al.*, “Factors for Maxillary Sinus Volume and Craniofacial Anatomical Features in Adults With Chronic Rhinosinusitis,” *Arch. Otolaryngol. Neck Surg.*, vol. 136, no. 6, p. 610, Jun. 2010, doi: 10.1001/archoto.2010.75.
[37] A. Farias Gomes, T. de Oliveira Gamba, M. C. Yamasaki, F. C. Groppo, F. Haiter Neto, and R. de F. Possobon, “Development and validation of a formula based on maxillary sinus measurements as a tool for sex estimation: a cone beam computed tomography study,” *Int. J. Legal Med.*, vol. 133, no. 4, pp. 1241–1249, Jul. 2019, doi: 10.1007/s00414-018-1869-6.
[38] K. Hung, L. Hui, A. W. K. Yeung, Y. Wu, R. T.-C. Hsung, and M. M. Bornstein, “Volumetric analysis of mucous retention cysts in the maxillary sinus: A retrospective study using cone-beam computed tomography,” *Imaging Sci. Dent.*, vol. 51, no. 2, p. 117, 2021, doi: 10.5624/isd.20200267.
[39] G. Giacomini *et al.*, “Computed tomography-based volumetric tool for standardized measurement of the maxillary sinus,” *PLOS ONE*, vol. 13, no. 1, p. e0190770, Jan. 2018, doi: 10.1371/journal.pone.0190770.
[40] J. Xu, S. Wang, Z. Zhou, J. Liu, X. Jiang, and X. Chen, “Automatic CT image segmentation of maxillary sinus based on VGG network and improved V-Net,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 9, pp. 1457–1465, Sep. 2020, doi: 10.1007/s11548-020-02228-6.
[41] C.-F. J. Kuo, Y.-S. Leu, D.-J. Hu, C.-C. Huang, J.-J. Siao, and K. B. P. Leon, “Application of intelligent automatic segmentation and 3D reconstruction of inferior turbinate and maxillary sinus from computed tomography and analyze the relationship between volume and nasal lesion,” *Biomed. Signal Process. Control*, vol. 57, p. 101660, Mar. 2020, doi: 10.1016/j.bspc.2019.101660.
[42] Z. Deng, B. Wang, and Z. Zhu, “BE-FNet: 3D Bounding Box Estimation Feature Pyramid Network for Accurate and Efficient Maxillary Sinus Segmentation,” *Math. Probl. Eng.*, vol. 2020, pp. 1–16, Jan. 2020, doi: 10.1155/2020/5689301.
[43] S.-K. Jung, H.-K. Lim, S. Lee, Y. Cho, and I.-S. Song, “Deep Active Learning for Automatic Segmentation of Maxillary Sinus Lesions Using a Convolutional Neural Network,” *Diagnostics*, vol. 11, no. 4, p. 688, Apr. 2021, doi: 10.3390/diagnostics11040688.
[44] Y. Zhang, K. Qian, Z. Zhu, H. Yu, and B. Zhang, “DBA-UNet: a double U-shaped boundary attention network for maxillary sinus anatomical structure segmentation in CBCT images,” *Signal Image Video Process.*, Dec. 2022, doi: 10.1007/s11760-022-02440-8.
[45] K. F. Hung, Q. Y. H. Ai, A. D. King, M. M. Bornstein, L. M. Wong, and Y. Y. Leung, “Automatic detection and segmentation of morphological changes of the maxillary sinus mucosa on cone-beam computed tomography images using a three-dimensional convolutional neural network,” *Clin. Oral Investig.*, vol. 26, no. 5, pp. 3987–3998, May 2022, doi: 10.1007/s00784-021-04365-x.
[46] H. Choi, K. J. Jeon, Y. H. Kim, E.-G. Ha, C. Lee, and S.-S. Han, “Deep learning-based fully automatic segmentation of the maxillary sinus on cone-beam computed tomographic images,” *Sci. Rep.*, vol. 12, no. 1, p. 14009, Aug. 2022, doi: 10.1038/s41598-022-18436-w.

[47] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: a review," *Artif. Intell. Rev.*, vol. 54, no. 1, pp. 137–178, Jan. 2021, doi: 10.1007/s10462-020-09854-1.
[48] M. A. Abdou, "Literature review: efficient deep neural networks techniques for medical image analysis," *Neural Comput. Appl.*, vol. 34, no. 8, pp. 5791–5812, Apr. 2022, doi: 10.1007/s00521-022-06960-9.
[49] E. Smistad, T. L. Falch, M. Bozorgi, A. C. Elster, and F. Lindseth, "Medical image segmentation on GPUs – A comprehensive review," *Med. Image Anal.*, vol. 20, no. 1, pp. 1–18, Feb. 2015, doi: 10.1016/j.media.2014.10.012.
[50] L. Lenchik *et al.*, "Automated Segmentation of Tissues Using CT and MRI: A Systematic Review," *Acad. Radiol.*, vol. 26, no. 12, pp. 1695–1706, Dec. 2019, doi: 10.1016/j.acra.2019.07.006.
[51] Y. Feng, H. Zhao, X. Li, X. Zhang, and H. Li, "A multi-scale 3D Otsu thresholding algorithm for medical image segmentation," *Digit. Signal Process.*, vol. 60, pp. 186–199, Jan. 2017, doi: 10.1016/j.dsp.2016.08.003.
[52] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *J. Electron. Imaging*, vol. 13, no. 1, pp. 146–165, Jan. 2004, doi: 10.1117/1.1631315.
[53] J. K. Leader *et al.*, "Automated lung segmentation in X-ray computed tomography: development and evaluation of a heuristic threshold-based scheme1," *Acad. Radiol.*, vol. 10, no. 11, pp. 1224–1236, Nov. 2003, doi: 10.1016/S1076-6332(03)00380-5.
[54] B. Dhruv, N. Mittal, and M. Modi, "Comparative Analysis of Edge Detection Techniques for Medical Images of Different Body Parts," in *Data Science and Analytics*, B. Panda, S. Sharma, and N. R. Roy, Eds., in Communications in Computer and Information Science. Singapore: Springer, 2018, pp. 164–176. doi: 10.1007/978-981-10-8527-7_15.
[55] M. N. Saad, Z. Muda, N. S. Ashaari, and H. A. Hamid, "Image segmentation for lung region in chest X-ray images using edge detection and morphology," in *2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014)*, Penang, Malaysia: IEEE, Nov. 2014, pp. 46–51. doi: 10.1109/ICCSCE.2014.7072687.
[56] M. Nikolic, E. Tuba, and M. Tuba, "Edge detection in medical ultrasound images using adjusted Canny edge detection algorithm," in *2016 24th Telecommunications Forum (TELFOR)*, Belgrade, Serbia: IEEE, Nov. 2016, pp. 1–4. doi: 10.1109/TELFOR.2016.7818878.
[57] Jianping Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1454–1466, Oct. 2001, doi: 10.1109/83.951532.
[58] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," in *Medical Imaging 2001: Image Processing*, SPIE, Jul. 2001, pp. 1337–1346. doi: 10.1117/12.431013.
[59] A. R. Khan, S. Khan, M. Harouni, R. Abbasi, S. Iqbal, and Z. Mehmood, "Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification," *Microsc. Res. Tech.*, vol. 84, no. 7, pp. 1389–1399, 2021, doi: 10.1002/jemt.23694.
[60] J. Shi, Y. Ye, D. Zhu, L. Su, Y. Huang, and J. Huang, "Comparative analysis of pulmonary nodules segmentation using multiscale residual U-Net and fuzzy C-means clustering," *Comput. Methods Programs Biomed.*, vol. 209, p. 106332, Sep. 2021, doi: 10.1016/j.cmpb.2021.106332.
[61] W. Cai, B. Zhai, Y. Liu, R. Liu, and X. Ning, "Quadratic polynomial guided fuzzy C-means and dual attention mechanism for medical image segmentation," *Displays*, vol. 70, p. 102106, Dec. 2021, doi: 10.1016/j.displa.2021.102106.
[62] J. Li *et al.*, "An EM-based semi-supervised deep learning approach for semantic segmentation of histopathological images from radical prostatectomies," *Comput. Med.*

Imaging Graph., vol. 69, pp. 125–133, Nov. 2018, doi: 10.1016/j.compmedimag.2018.08.003.
[63] S. Beucher, “The Watershed Transformation Applied to Image Segmentation,” *Scanning Microsc.*, vol. 1992, no. 6, Jan. 1992, [Online]. Available: https://digitalcommons.usu.edu/microscopy/vol1992/iss6/28
[64] J. B. T. M. Roerdink and A. Meijster, “The Watershed Transform: Definitions, Algorithms and Parallelization Strategies,” *Fundam. Informaticae*, vol. 41, no. 1–2, pp. 187–228, Jan. 2000, doi: 10.3233/FI-2000-411207.
[65] X. Ji, Y. Li, J. Cheng, Y. Yu, and M. Wang, “Cell image segmentation based on an improved watershed algorithm,” in *2015 8th International Congress on Image and Signal Processing (CISP)*, Shenyang, China: IEEE, Oct. 2015, pp. 433–437. doi: 10.1109/CISP.2015.7407919.
[66] G.-R. Kwon, D. Basukala, S.-W. Lee, K. H. Lee, and M. Kang, “Brain image segmentation using a combination of expectation-maximization algorithm and watershed transform,” *Int. J. Imaging Syst. Technol.*, vol. 26, no. 3, pp. 225–232, 2016, doi: 10.1002/ima.22181.
[67] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
[68] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, “Medical Image Analysis using Convolutional Neural Networks: A Review,” *J. Med. Syst.*, vol. 42, no. 11, p. 226, Nov. 2018, doi: 10.1007/s10916-018-1088-1.
[69] A. A. Adegun, S. Viriri, and R. O. Ogundokun, “Deep Learning Approach for Medical Image Analysis,” *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–9, May 2021, doi: 10.1155/2021/6215281.
[70] L. Liu, J. Cheng, Q. Quan, F.-X. Wu, Y.-P. Wang, and J. Wang, “A survey on U-shaped networks in medical image segmentations,” *Neurocomputing*, vol. 409, pp. 244–258, Oct. 2020, doi: 10.1016/j.neucom.2020.05.070.
[71] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.
[72] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science, vol. 9351. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
[73] N. Morgan, A. Van Gerven, A. Smolders, K. de Faria Vasconcelos, H. Willems, and R. Jacobs, “Convolutional neural network for automatic maxillary sinus segmentation on cone-beam computed tomographic images,” *Sci. Rep.*, vol. 12, no. 1, p. 7523, May 2022, doi: 10.1038/s41598-022-11483-3.
[74] D. Karimi, S. K. Warfield, and A. Gholipour, “Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations,” *Artif. Intell. Med.*, vol. 116, p. 102078, Jun. 2021, doi: 10.1016/j.artmed.2021.102078.
[75] N. Tajbakhsh *et al.*, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016, doi: 10.1109/TMI.2016.2535302.
[76] L. Li and H. Ma, “RDCTrans U-Net: A Hybrid Variable Architecture for Liver CT Image Segmentation,” *Sensors*, vol. 22, no. 7, Art. no. 7, Jan. 2022, doi: 10.3390/s22072452.
[77] M. Futrega, A. Milesi, M. Marcinkiewicz, and P. Ribalta, “Optimized U-Net for Brain Tumor Segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi and S. Bakas, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 15–29. doi: 10.1007/978-3-031-09002-8_2.
[78] T. Yang, J. Song, L. Li, and Q. Tang, “Improving brain tumor segmentation on MRI based on the deep U-net and residual units,” *J. X-Ray Sci. Technol.*, vol. 28, no. 1, pp. 95–110, Jan. 2020, doi: 10.3233/XST-190552.

[79] K. Wang *et al.*, “Automated CT and MRI Liver Segmentation and Biometry Using a Generalized Convolutional Neural Network,” *Radiol. Artif. Intell.*, vol. 1, no. 2, p. 180022, Mar. 2019, doi: 10.1148/ryai.2019180022.

[80] “3D Deeply-Supervised U-Net Based Whole Heart Segmentation | SpringerLink.” https://link.springer.com/chapter/10.1007/978-3-319-75541-0_24 (accessed Aug. 11, 2023).

[81] B. N. Narayanan and R. C. Hardie, “A Computationally Efficient U-Net Architecture for Lung Segmentation in Chest Radiographs,” in *2019 IEEE National Aerospace and Electronics Conference (NAECON)*, Dayton, OH, USA: IEEE, Jul. 2019, pp. 279–284. doi: 10.1109/NAECON46414.2019.9058086.

[82] Y. Cheng *et al.*, “Retinal blood vessel segmentation based on Densely Connected U-Net,” *Math. Biosci. Eng.*, vol. 17, no. 4, pp. 3088–3108, 2020, doi: 10.3934/mbe.2020175.

[83] Z. Li, M. Jia, X. Yang, and M. Xu, “Blood Vessel Segmentation of Retinal Image Based on Dense-U-Net Network,” *Micromachines*, vol. 12, no. 12, p. 1478, Nov. 2021, doi: 10.3390/mi12121478.

[84] F. Preda *et al.*, “Deep convolutional neural network-based automated segmentation of the maxillofacial complex from cone-beam computed tomography:A validation study,” *J. Dent.*, vol. 124, p. 104238, Sep. 2022, doi: 10.1016/j.jdent.2022.104238.

[85] V. Iglovikov and A. Shvets, “TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation.” arXiv, Jan. 17, 2018. Accessed: May 15, 2023. [Online]. Available: http://arxiv.org/abs/1801.05746

[86] M. Raghu, J. Kleinberg, C. Zhang, and S. Bengio, “Transfusion: Understanding Transfer Learning for Medical Imaging”.

[87] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv, Apr. 10, 2015. Accessed: May 15, 2023. [Online]. Available: http://arxiv.org/abs/1409.1556

[88] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[89] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.” arXiv, Aug. 23, 2016. Accessed: Jul. 20, 2023. [Online]. Available: http://arxiv.org/abs/1602.07261

[90] N. Sharma *et al.*, “U-Net Model with Transfer Learning Model as a Backbone for Segmentation of Gastrointestinal Tract,” *Bioengineering*, vol. 10, no. 1, p. 119, Jan. 2023, doi: 10.3390/bioengineering10010119.

[91] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, “Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks,” *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 59–69, Apr. 2019, doi: 10.1016/j.isprsjprs.2019.02.006.

[92] W. Gómez-Flores and W. Coelho De Albuquerque Pereira, “A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound,” *Comput. Biol. Med.*, vol. 126, p. 104036, Nov. 2020, doi: 10.1016/j.compbiomed.2020.104036.

[93] J. Wacker, M. Ladeira, and J. E. V. Nascimento, “Transfer Learning for Brain Tumor Segmentation.” arXiv, Nov. 26, 2020. Accessed: Aug. 12, 2023. [Online]. Available: http://arxiv.org/abs/1912.12452

[94] “3D Slicer image computing platform,” *3D Slicer*. https://slicer.org/ (accessed Mar. 21, 2023).

[95] C. Szegedy *et al.*, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[96] P. Iakubovskii, “qubvel/segmentation_models.” Mar. 23, 2023. Accessed: Mar. 24, 2023. [Online]. Available: https://github.com/qubvel/segmentation_models

[97] K. Avdeeva and W. Fokkens, “Precision Medicine in Chronic Rhinosinusitis with Nasal Polyps,” *Curr. Allergy Asthma Rep.*, vol. 18, no. 4, p. 25, 2018, doi: 10.1007/s11882-018-0776-8.
[98] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms,” *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

CURRICULUM VITAE

PERSONAL INFORMATION

SURNAME, Name : DEDEOĞLU, Ahmet Said

EDUCATION

Degree	Institution	Year of Graduation
B.Sc.	University of Gaziantep	2019

WORK EXPERIENCE

Year	Place	Enrollment
2020-present	University of Gaziantep	Research Assistant

PUBLICATIONS

[1] A. S. Dedeoğlu, S. Özbay and O. Tunç, "Automatic Segmentation of Maxillary Sinus with U-Net Model with Pre-trained Encoder," 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Istanbul, Turkiye, 2023, pp. 1-5, doi: 10.1109/HORA58378.2023.10156706.