A DEEP LEARNING DECISION SUPPORT SYSTEM FOR FUNCTIONAL ENDOSCOPIC SINUS SURGERY USING COMPUTED TOMOGRAPHY SCANS

C. Gruber[1] & L.E. Burger[1*]

ARTICLE INFO

Contact details
∗   Corresponding author
     eldonburger@sun.ac.za

Author affiliations
1   Department of Industrial Engineering, Stellenbosch University, South Africa

ORCID® identifiers
L.E. Burger
https://orcid.org/0000-0002-2689-3586

ABSTRACT

Chronic sinusitis is a common disease that significantly affects quality of life. To treat chronic sinusitis, functional endoscopic sinus surgery (FESS) is frequently considered. FESS alleviates chronic sinusitis symptoms by restoring natural sinus drainage. Otolaryngologists rely on computed tomography (CT) reports to establish whether FESS is appropriate. To enhance sinus CT reports and improve decision-making, a segmentation model is developed. Both 2D and 3D segmentation models were compared, with the 3D model achieving marginally better results. The model accurately segments the sinus system, including the nasal cavity, achieving a mean Dice coefficient of 0.889 ± 0.028. The resulting 3D visualisation of the segmented sinus system enables quick identification of opacified regions, helping otolaryngologists to make informed decisions about the appropriateness of FESS. This automated approach reduces the time required to compile reports, improves the precision of clinical evaluations, and ultimately enhances patient care.

OPSOMMING

Chroniese sinusitis is 'n algemene siekte wat lewenskwaliteit aansienlik beïnvloed. Om chroniese sinusitis te behandel, word funksionele endoskopiese sinuschirurgie (FESS) gereeld oorweeg. FESS verlig chroniese sinusitis simptome deur die herstel van natuurlike sinus dreinering. Otolaryngoloë maak staat op rekenaartomografie (RT) verslae om vas te stel of FESS toepaslik is of nie. 'n Segmentasiemodel is ontwikkel om sinus-RT-verslae en besluitneming te verbeter. Beide 2D- en 3D-segmentasiemodelle is vergelyk, met die 3D-model wat marginale beter resultate behaal het. Die model segmenteer die sinusstelsel akkuraat, insluitend die neusholtes, en bereik 'n gemiddelde Dicekoëffisiënt van 0.889 ± 0.028. Die gevolglike 3D-visualisering van die gesegmenteerde sinusstelsel maak vinnige identifikasie van ondeursigtige streke moontlik, wat otolaryngoloë help om ingeligte besluite oor die toepaslikheid van chirurgie te neem. Hierdie outomatiese benadering verminder die tyd wat nodig is om verslae saam te stel, verbeter die akkuraatheid van kliniese evaluasies en verbeter uiteindelik pasiëntsorg.

1. INTRODUCTION

The sinuses consist of four pairs of interconnected air-filled cavities: the frontal, ethmoid, maxillary, and sphenoid sinuses [1], [2]. These cavities, located in various parts of the skull, connect to the nasal cavity and play a crucial role in maintaining respiratory health through mucociliary clearance [2], [3]. Mucociliary clearance relies on mucus production and ciliary motion. Epithelial cells line the interior of the sinuses and nasal cavity and secrete mucus, a gel-like substance with sticky characteristics that traps breathed-in particles such as bacteria, germs, and pollutants [2], [3]. Hair-like structures called cilia extend from the surfaces of the epithelial cells and perform coordinated rhythmic movements that sweep the mucus particles out of the sinuses and into the throat or nasal cavity for expulsion [2], [3], [4].

When mucus in the sinuses cannot drain freely and frequently, it stagnates and becomes a breeding ground for bacteria, increasing the risk of sinus infections, which can lead to sinusitis [2], [5]. Sinusitis is a common condition that affects people of all ages [2], [6]. Patients with sinusitis can experience fatigue, difficulty sleeping, and decreased productivity, all of which can affect their ability to work, study, or engage in social activities [7]. Sinusitis can also lead to anxiety, depression, and other psychological symptoms that have a further impact on the overall well-being of a patient [7].

Messerklinger's research on mucociliary clearance and its role in sinusitis laid the groundwork for functional endoscopic sinus surgery (FESS), a procedure designed to restore natural sinus drainage and alleviate chronic sinusitis symptoms [1], [2], [3], [8], [9], [10]. FESS is considered when a patient with chronic sinusitis has recurrent symptoms despite receiving appropriate medication [10], [11]. When an otolaryngologist contemplates surgical intervention, computed tomography (CT) scans are obtained [12]. The images obtained from these scans provide critical information such as the extent of the disease, the opacification of the sinus drainage pathways, the location of surgically pertinent anatomic structures, and anatomic variations [1], [4], [12], [13], [14], [15], [16].

To derive insights from sinus CT scans, radiologists must meticulously examine the multiple two-dimensional (2D) images or slices that constitute a sinus CT scan. Owing to the worldwide shortage of radiologists, the comprehensive assessment of CT scans may not be performed in depth or promptly [17], [18], [19]. South Africa in particular has only 1.2 radiologists per 100,000 individuals, which is in stark contrast to Europe, where there was an average of 12.8 radiologists per 100,000 individuals in 2020 [18], [20]. In extreme cases, radiologists are expected to examine one image every three to four seconds in an eight-hour workday [17], [18].

To guarantee that sinus CT assessments could be performed comprehensively and promptly, the effectiveness of current sinus CT scan reporting must be improved. This study aims to enhance the effectiveness and efficiency of sinus CT scan assessments by developing a machine-learning sinus segmentation model that represents the sinus system as a three-dimensional (3D) model with supporting annotations. By analysing the 3D model rather than individual 2D slices, the effectiveness of sinus CT scan assessments is improved.

Towards the aim of developing a 3D model: (i) current sinus CT reporting practices are reviewed to identify the requirements of manual assessment procedures, (ii) models that have previously been proposed to automate sinus CT reporting are identified and compared in order to identify their shortcomings and (iii) machine learning models are proposed, implemented, and evaluated to determine their applicability in improving sinus CT scan assessments.

Unlike related work, this study specifically considers the segmentation of sinus drainage paths and directly compares 2D and 3D segmentation models. The proposed solution could segment sinus drainage paths, a key area that requires examination but that has been largely ignored by previous automation efforts. The 3D segmentation models only marginally outperform the 2D segmentation models, suggesting that computationally efficient 2D models may be a viable option for effective sinus CT scan assessments.

The paper is structured as follows: A detailed review of current sinus CT reporting practices is provided in Section 2. Machine learning models that have previously been proposed to automate sinus CT reporting are evaluated in Section 3. Machine learning models are then proposed and evaluated, with the experimental setup discussed in Section 4 and the results in Section 5. Section 6 concludes the paper with a summary of the main findings and limitations of this study.

2. SINUS COMPUTED TOMOGRAPHY ASSESSMENT

The assessment of sinus CT scans is crucial to determine whether FESS should be performed, as a surgical intervention is not without risks and costs [8], [21]. Patients who undergo FESS may experience minor complications such as bleeding, infection, crusting, tooth, lip numbness, and/or disease recurrence, or more serious complications such as optic nerve damage, meningitis, and carotid vascular injury [4], [22]. Thomas *et al.* [23] evaluated the cost and operation time of 1,477 endoscopic sinus surgeries and found that the total operation cost ranged from 2,100 to 4,600 United States Dollars (USD). Given the risks and costs associated with FESS, the comprehensive assessment of CT scans is essential to ensure that decisions are well-informed and unbiased.

Despite the importance of comprehensive assessment, no universally accepted standard for radiology reporting on sinus CT scans exists [12]. Deutschmann *et al.* [12] surveyed Canadian otolaryngologists and found that sinus CT radiologic reporting did little to assist with clinical evaluation. More recently, Cadd *et al.* [24] determined that only seven out of 129 otorhinolaryngology surgeons surveyed in Australia considered CT reporting practices useful. Radiologists typically receive limited formal education in reporting, learning through periodic correction and imitation of other reports [25]. Consequently, CT scan reports vary, based on the training and expertise of the radiologist.

To improve the state of current sinus CT scan reporting, radiologists should report on:
1. the extent of sinus opacification [4], [12], [13], [14];
2. the opacification of sinus drainage pathways [1], [4];
3. any anatomical variants that predispose patients to recurrent diseases and that would have a significant impact on potential surgical interventions [4], [12], [16], [24]; and
4. observations of polyps, cysts, deviated nasal septums, concha bullosa, and bone thinning [12].

For instance, radiologists should report any obstruction of the ostiomeatal complex, a key sinus drainage pathway, since an obstruction can prevent effective mucociliary clearance, potentially leading to sinus infections [3].

Efforts to enhance the quality of sinus CT scan reporting have led to the adoption of standardised formats, the most prominent being the Lund and Mackay scoring format [26], which requires radiologists to grade six areas of the sinuses on both the left and right sides, using a scorecard.

The maxillary sinus, the sphenoid sinus, the frontal sinus, the posterior ethmoid, and the anterior ethmoid are scored using an ordinal scale with three categories: a score of zero indicates no abnormality, a score of one indicates partial opacification, and a score of two indicates complete opacification [26], [27]. The ostiomeatal complex is scored using an ordinal scale with two categories: zero indicates no obstructions and two indicates obstruction [26], [27]. An example of this scoring system is illustrated in Table 1.

Table 1: Lund and Mackay scorecard

Radiological grading	Right sinus	Left side
Maxillary sinus	0-2	0-2
Sphenoid sinus	0-2	0-2
Frontal sinus	0-2	0-2
Posterior ethmoids	0-2	0-2
Anterior ethmoids	0-2	0-2
Ostiomeatal complex	0 or 2	0 or 2

While prominent, the Lund and Mackay scorecard has been criticised for having insufficient levels to track the progression or regression of disease, leading to the development of several modified versions of the Lund and Mackay scorecards [26]. For instance, Kennedy *et al.* [28] proposed reporting opacification using a score from zero to five to allow finer grading.

While the Lund and Mackay scoring format and similar grading systems have aided in the standardisation and interpretation of radiology reports, they do not reduce the workload of radiologists. To assign a score to each of the twelve areas, a radiologist must still meticulously evaluate the multiple slices that constitute a sinus CT scan. Owing to the limited range of the scale used, a vast range of conditions could be assigned similar scores. To avoid misinterpretation, an otolaryngologist would have to review the accompanying CT scan to understand the severity of the assigned score.

3. AUTOMATED SINUS COMPUTED TOMOGRAPHY ASSESSMENT

Machine learning models have shown promising results in a range of medical applications, including the analysis of medical images to improve the accuracy and efficiency of diagnostic procedures [29]. Machine learning models, particularly those suitable for image-based tasks such as image recognition, image detection, and image segmentation, offer promising solutions that address some of the limitations of current sinus CT radiologic reporting. Image classification models can be used to classify images into known categories; image detection models can be used to locate features in an image; and image segmentation models can be used to divide an image automatically into distinct segments or regions.

Convolutional neural networks (CNNs) are typically used for image-based tasks, since they are translation-invariant and can automatically and adaptively learn features directly from images. By automatically learning features from images, the need for hand-crafted features is eliminated. CNNs employ multiple convolutional, pooling, and non-linear layers in succession, allowing the models to learn increasingly complex features. Features are usually learned using supervised learning, a process in which a model is trained using a data set that consists of input-output pairs. Several studies have investigated using supervised learning to develop CNNs in order to analyse sinus CT scans automatically. The relevant studies that have been identified are discussed and compared in the subsections that follow, grouped by image task.

3.1. Object classification

Chowdhury *et al*. [13] developed an object classification model using a data set of 2D coronal CT slices from 239 patients to classify whether the ostiomeatal complex was *opened* or *closed*. The CNN could accurately distinguish between the two categories, achieving an area-under-the-curve (AUC) performance of 0.87. While accurate, the CNN only covered a single area of the Lund and Mackay report, and assumed that the ostiomeatal complex could be evaluated from a single 2D slice. Using a single 2D slice can lead to the loss of crucial information, as the sinus system is inherently 3D and requires the selection of an appropriate slice.

Ozbay and Tunc [5] focused on classifying the sinus system as either *normal* or *abnormal*, using a three-step process. Five 2D slices that include the full view from the front of the head are first selected from a 3D sinus CT scan. The sinus system is then identified in each image using Otsu's method, and cropped. Otsu's method separates pixels into two classes by finding a threshold that minimises the intra-class variance [30]. A CNN is then trained on the cropped image to classify images as *normal* or *abnormal*. When the individual predictions are combined for the five images that have been considered, the model achieves an accuracy of 0.98 on a held-out test set of 67 patients. Ozbay and Tunc [5] did not specify the criteria that were used to distinguish between *normal* and *abnormal* sinus CT scans. The criteria to decide whether FESS should be performed are poorly defined [31], [32], [33], and the use of FESS varies significantly geographically [6], [10], [34], [35]. Consequently, image classification models could propagate the bias of noisy categorisation.

When the criteria used to assign labels to images are not defined, it becomes difficult to establish the significance of the results. It is straightforward to develop a system that classifies an image as *abnormal* when any opacification is detected. On the other hand, determining whether FESS should be performed is a difficult task that must account for several factors. To avoid any label ambiguity and bias, object detection and segmentation models could be used. These models provide specific information about the structures considered as opposed to distinct categories.

3.2. Object detection

Laura *et al*. [36] developed an object detection model to identify and locate key structures of the sinus. The object detection model was trained to detect seven different objects from 2D slices, namely the frontal sinus, the sphenoid bone, the left ethmoid bone, the right ethmoid bone, the left maxillary sinus, the right maxillary sinus, and the nasal cavity. To create 3D objects, the detected objects from the 2D slices were combined to form irregular polyhedrons.

The model was developed and evaluated using 57 CT scans. Global performance values were not provided; instead, recall and precision graphs were provided for each of the seven detected objects. Overall, the model had good performance, covering all the areas of the Lund and Mackay reporting format. However, the clinical evaluation of the detected objects was not covered in the study.

3.3. Object segmentation

Iwamoto *et al*. [37] proposed a two-step process to segment the maxillary sinus automatically. A region of interest is first identified from each 2D CT slice using a probabilistic atlas – a statical representation of the anatomical variability of the maxillary sinus. A CNN is then used to segment the maxillary sinus from the region of interest. The CNN they used was developed using 80 sinus CT scans and evaluated on 20 sinus CT scans using the Dice coefficient.

The Dice coefficient measures the agreement between a manually segmented image and an image segmented by a model by considering the degree of overlap between the images – that is:

$$\text{Dice coefficient} = \frac{2|V_{model} \cap V_{manual}|}{|V_{model}| + |V_{manual}|} \qquad (1)$$

where V_{model} is the set of predicted pixels or voxels and V_{manual} is the set of manually annotated pixels or voxels. A Dice coefficient of one indicates perfect agreement between the manually segmented image and the predicted segmented image, while a dice coefficient of zero indicates no overlap. The model proposed by Iwamoto *et al*. [30] achieved a dice coefficient of 0.83, which indicated that the segmentation results mostly agreed with the manual segmentation.

Xu *et al*. [38] proposed a two-step process to segment the maxillary sinus automatically. A classification model is first used to determine whether each image contains the maxillary sinus. Next, the images that contain the maxillary sinus are segmented individually. The model was developed using 35 sinus CT scans and evaluated on 26 sinus CT scans. The classification model they used to detect the maxillary sinus achieved a test accuracy of 97%. Three different CNN segmentation models were evaluated for segmenting the maxillary sinus: a U-Net [39], a V-Net [40], and a V-Net based on edge supervision. These segmentation models achieved similar Dice coefficients of 0.93, 0.93, and 0.94 respectively. Although effective in segmentation, the studies by Iwamoto *et al*. [37] and by Xu *et al*. [38] covered only one area of the Lund and Mackay report and excluded clinical evaluation, similar to the study of Laura *et al*. [36].

Unlike the studies by Iwamoto *et al*. [37] and Xu *et al*. [38], Jung *et al*. [41] included the clinical evaluation of the maxillary sinus. The authors incorporated the clinical evaluation by differentiating between *air* and *lesion*, and performed 3D segmentation directly using a 3D U-Net [39]. The 3D U-Net was trained on 83 cone-beam CT (CBCT) scans and evaluated on 20 internal and 20 external CBCT scans. The model achieved a Dice coefficient of 0.93 for the *air* class and a Dice coefficient of 0.76 for the *lesion* class on the internal data set. On the external data set, however, the model performed poorly, achieving a Dice coefficient of 0.97 for the *air* class and a Dice coefficient of 0.54 for the *lesion* class. Furthermore, it was noted that the performance of the model diminished for patients with severe maxillary sinusitis. As part of the study, the authors established that CNN-assisted segmentation could reduce the manual segmentation time by more than 50% [41].

Humphries *et al*. [42] developed a 3D CNN based on the Tiramisu [43] architecture. The model was trained on 140 CT scans to segment the combined sinus cavities. The output of the model was used to estimate the opacification percentage of the combined sinus cavities. CT pixels with values of between -500 and +200 Hounsfield (HU) units were assumed to represent opacification. The CNN achieved a Dice coefficient of 0.93, and the estimated percentage opacification had a strong linear relationship with Lund and Mackay scores, indicating that opacification calculations could be used as a potential substitute for Lund and

Mackay scores. Although Humphries *et al*. [42] compared their results with Lund and Mackay scores, the authors did not consider sinus drainage pathways such as the ostiomeatal complex.

3.4. Comparison of related work

Table 2 compares the approaches identified above against the areas of the Lund and Mackay reporting format.

Table 2: Comparison of related work

	Maxillary sinus			Sphenoid sinus			Frontal sinus			Posterior ethmoids			Anterior ethmoids			Ostiomeatal complex		
	Classification	Detection	Segmentation	Classification	Detection	Segmentation	Classification	Detection	Segmentation	Classification	Detection	Segmentation	Classification	Detection	Segmentation	Classification	Detection	Segmentation
[5]	x		x	x		x	x		x	x		x	x		x	x		x
[13]																x		
[36]		x			x			x			x			x			x	
[37]			x															
[38]			x															
[41]	x		x															
[42]	x		x	x		x	x		x	x		x	x		x			

x indicates areas considered

Four of the seven identified approaches covered only one area of the Lund and Mackay report. Laura *et al*. [36] developed an object detection model that covered all the areas of the Lund and Mackay report, but excluded clinical evaluation. Humphries *et al*. [42] developed a segmentation model that included clinical evaluation of all the areas of the Lund and Mackay report, but excluded the ostiomeatal complex.

Five of the seven studies considered here used 2D images instead of 3D images directly. Since none of the papers directly compared 2D CNNs with 3D CNNs, it remains unclear which approach would yield better results. The complexity of the sinus anatomy is more accurately captured in 3D owing to its inherent 3D structure, but more computational resources would be required to develop 3D CNN models.

The data sets employed to develop sinus CT models were typically small, ranging from 35 to 239 CT scans. This limitation is to be expected, as the annotation of 3D CT scans is a time-consuming task – even more so for segmentation, which requires the annotation of each voxel.

While existing studies have made important strides in automating the analysis of sinus CT scans, there remains a need for a holistic model that can comprehensively evaluate the sinus system. This study aims to determine whether the sinus system, inclusive of sinus drainage pathways, can be accurately segmented, and whether more accurate results are obtained when employing 3D models as opposed to 2D models. The result is used to develop a model that can construct a 3D visualisation of the sinus system that indicates opacified areas.

4.   METHODOLOGY

The data collection and preparation, the selection of evaluation measures, the evaluation methodology employed, and the machine learning models considered are discussed in this section. '

4.1. Data collection

Data was collected from 35 patients who underwent sinus CT scan examinations in South Africa. Each CT scan was initially stored as a series of Joint Photographic Experts Group (JPEG) images in the coronal view, the preferred perspective for sinus CT scans [32]. Unique identifiers were generated for each patient, and all metadata were removed to protect the patients' privacy.

The number of slices in a single CT scan ranged from 72 to 156. Slices that did not contain the sinus system and nasal cavity were manually removed. After removing the empty slices, each CT scan was standardised to 70 slices by selecting slices at about equal intervals and/or removing slices at regular intervals. The slices were then stacked and saved as 3D images in the Neuroimaging Informatics Technology Initiative (NIfTI) file format. The annotated segmentation mask was created in 3D Slicer, a free open-source software for 3D medical image segmentation and image processing [44]. Voxel intensities were normalised to a range between zero and one, and the images and labels were resized to a height and width of 352 pixels.

To increase the number of training instances artificially, data augmentation was randomly applied during training. The data augmentation techniques included flipping images along the y-axis, modifying the pixel intensities of images within a range of ±0.1, scaling within a range of ±10%, and rotating images within a range of ±0.05 radians. Each augmentation technique had a 50% probability of being applied during training. Model performance was assessed without data augmentation, with each data augmentation technique applied individually and for a combination of the three augmentation techniques.

4.2. Evaluation measures

The performance of each model being considered was evaluated using five-fold cross-validation. The data set was randomly divided into five non-overlapping folds of equal size. Each fold was used once as a validation set. The mean Dice coefficient across the five validation sets was reported along with the standard deviation. Although the Dice coefficient is the most common segmentation measure used in medical segmentation studies [46], the Dice coefficient has several limitations. Notably, it assumes that the ground truth is accurate. Identifying the sinus system from CT scans is straightforward for healthy patients, but is difficult for patients with sinus infections, where differentiating between mucus and inflamed tissue can be problematic. Consequently, the ground truth may be inaccurate. In addition to the Dice coefficient, model performance was manually validated by comparing the ground truth and predicted segmentations, as recommended by Müller *et al*. [45].

4.3. Model description

A 2D and a 3D U-Net were considered, since the related work did not directly compare 2D segmentation models against 3D segmentation models. In contrast, different 3D segmentation models have been compared against one another and have achieved similar Dice coefficients. The U-Net architecture selected in this study consisted of an encoder-decoder structure with skip connections between corresponding blocks of the encoder and decoder. The encoder path reduces the spatial dimension of the input image, while the decoder path restores the spatial dimension of the representation to the original size. When the spatial resolution of a tensor is increased, both the input from the preceding layer and feature maps from the encoder path, connected via skip connections, are considered. These connections help to capture the finer details, making a U-Net suitable for medical image segmentation tasks that require high precision.

The encoder of the U-Net model consists of five sequentially connected blocks. Each block contains two convolutional layers, each followed by a batch normalisation layer, a dropout layer, and a parametric rectified linear unit (PReLU) activation. After these two sets, there is another convolutional layer, followed by a batch normalisation layer, a dropout layer, and a PReLU activation. Finally, each block includes a residual layer that adds the input of the block to its output. The convolutional layers extract features, the PReLU functions introduce non-linearity, batch normalisation layers improve convergence and stability during training, while dropout with a probability of 20% helps to prevent overfitting. The first convolutional layer of the first three blocks uses a stride of two to down-sample the input. The number of channels is doubled in each block, starting from 16 and increasing to 256.

The decoder of the U-Net model consists of four sequentially connected blocks. Each block in the decoder consists of transposed convolutional layers followed by a batch normalisation layer, a dropout layer, and a PReLU activation. In addition, the first three blocks include a second convolutional layer, followed by a batch normalisation layer, a dropout layer, and a PReLU activation. The last layer of each block is a residual layer that adds the input of the block to its output. The last block employs only a single convolutional layer that produces a tensor with the same shape as the initial input. A sigmoid activation function is applied to transform the outputs within the range of zero to one. Values equal to or greater than 0.5 are considered to form part of the sinus system. The model architecture for the 3D U-Net is summarised in Table 3.

Table 3: 3D U-Net architecture

Layer	Output shape	Number of parameters
Input	352, 352, 80	
Encoder		
Block 1	16, 176, 176, 40	9 618
Block 2	32, 88, 88, 20	30 610
Block 3	64, 44, 44, 10	121 986
Block 4	128, 22, 22, 5	885 634
Block 5	256, 22, 22, 5	2 688 770
Decoder		
Skip connection	384, 22, 22, 5	
Block 6	64, 44, 44, 10	774 530
Skip connection	128, 44, 44, 10	
Block 7	32, 88, 88, 20	113 522
Skip connection	64, 88, 88, 20	
Block 8	16, 176, 176, 40	34 658
Skip connection	32, 176, 176, 40	
Block 9	352, 352, 80	896

The 2D and 3D U-Net employ the same architecture, except for the dimensionality of the convolutions and operations; the 2D U-Net employs 2D convolutions and operations, while the 3D U-Net employs 3D convolutions and operations. The 2D U-Net is about 65% smaller than the 3D U-Net, significantly reducing the computational resources required for training. For the 2D U-Net, the images from a single CT scan are processed individually, and the outputs are combined to form a coherent 3D volume. The final 3D volume is then used to highlight areas of opacification by considering the HU units of the corresponding CT scan voxels.

Both the 2D and the 3D models were trained to minimise the Dice loss, which is computed by subtracting the Dice coefficient from one. To minimise the loss, the adaptive moment estimation (ADAM) [46] optimisation algorithm was used with β_1 set to 0.9 and β_2 set to 0.999. The learning rate was selected to ensure convergence. The training of each model was stopped after 100 epochs or when the validation loss did not decrease in three consecutive epochs. When the training was completed, the model was returned to the state with the best validation loss. The performance of the model was then determined using the held-out validation data.

To reduce computational requirements, training with a limited batch size and sliding window inference was considered. Using a batch size smaller than four produced poor performance, which could likely be attributed to increased noise in the gradients used to update the model's parameters. Training with a sliding window in a patch-like manner led to a fragmented understanding of the sinus and nasal cavity structures: some regions were well-defined, while others were overlooked. Consequently, training was performed using a batch size equal to or greater than four and without sliding window inference.

5.   RESULTS

The 3D U-Net was trained using a batch size of four and a learning rate of 0.0005, while the 2D U-Net was trained using a batch size of 256 and a learning rate of 0.005. The batch size could be increased for the 2D U-Net, as the 2D U-Net consisted of about 1.6 million trainable parameters compared with the 4.7 million or so parameters of the 3D U-Net. The mean Dice coefficient, along with the standard deviation of the models and data augmentation strategies that were considered, is provided in Table 4.

Table 4: Average validation performance per model

| Model	Flip image	Modify pixel intensities	Scale and rotate images	Mean Dice coefficient
2D U-Net				0.869 ± 0.064
3D U-Net				0.871 ± 0.053
3D U-Net	✓			0.859 ± 0.072
3D U-Net		✓		**0.889 ± 0.028**
3D U-Net			✓	0.878 ± 0.061
3D U-Net	✓	✓	✓	0.882 ± 0.029

Best performance highlighted in bold

The 2D U-Net and the 3D U-Net without data augmentation achieved a mean Dice coefficient of 0.869 ± 0.064 and 0.871 ± 0.053 respectively. Using data augmentation to increase the variability in the training data had a limited impact on segmentation performance while roughly doubling the training time. The 3D U-Net with the best performance used pixel intensity modification, achieving a mean Dice coefficient of 0.889 ± 0.028. Intuitively, modifying the pixel intensities could help to introduce variations similar to opacification, which could improve the ability of the model to generalise to different levels of sinus opacification.

An example of a single slice of the segmented sinus system generated by the model is illustrated in Figure 1. The model successfully segments the nasal cavity. However, in some instances, voxels that do not represent the sinus system are incorrectly categorised as part of the sinus system.
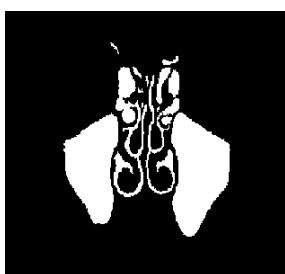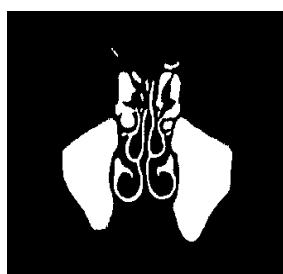
(a) Manual segmentation   **(b) Automated segmentation**

Figure 1: Comparison of manual segmentation vs automated segmentation

Once segmented, the sinus system can be visualised and the opacified areas highlighted, based on the corresponding HU units of the original CT scan, as illustrated in Figure 2. In addition, the exact amount of opacification can be calculated and presented alongside the model output.

6.   CONCLUSION

Current sinus CT reports provide limited assistance to otolaryngologists in deciding whether FESS should be performed. Improving sinus CT reporting requires innovative solutions that consider real-world constraints such as the global shortage of radiologists. To improve sinus CT reporting standards, a machine learning model was developed that can automatically segment the sinus system. Unlike previous work, the segmentation model included the nasal cavity and sinus drainage paths, which play a key role in effective mucociliary clearance.

Figure 2: 3D segmented sinus CT with opacified regions indicated in red

A 2D segmentation model was compared directly with a 3D segmentation model to determine whether 3D segmentation models would provide superior results. The 3D U-Net marginally outperformed the 2D U-Net, which contradicted the large performance gain reported by Avesta *et al*. [47] in brain segmentation. The use of data augmentation to improve model performance was explored, as annotating 3D sinus CT scans is both time-consuming and costly. Data augmentation had a marginal impact on the model's performance.

The best model, the 3D U-Net trained with modified pixel intensities, achieved a mean Dice coefficient of 0.889 ± 0.028, indicating that the complete sinus system can be accurately segmented. The performance achieved is lower than the 0.93 Dice coefficient reported by Humphries *et al*. [39], who used around four times more training data and did not consider the nasal cavity. When interpreting the results, it is important to consider that the Dice coefficient assumes that the manually annotated label is accurate.

Future work should focus on addressing the limitations of the current study, which include the limited data set size and the integration of the model into clinical practice. To enhance reporting, it should be investigated whether CNNs could be used to identify anatomical variants automatically that predispose patients to recurrent sinusitis and/or that would have a significant impact on potential surgical interventions as well as the identification of polyps, cysts, deviated nasal septums, concha bullosa, and bone thinning.

REFERENCES

[1]   K. Rao, "Computed tomography of paranasal sinus pathologies with functional endoscopic sinus surgery/nasal endoscopy correlation," *Clinical Rhinology: An International Journal*, vol. 1, no. 8, pp. 15–19, 2015. doi: 10.5005/jp-journals-10013-1222

[2]   S. J. Zinreich, "Functional anatomy and computed tomography imaging of the paranasal sinuses," *American Journal of Medical Science*, vol. 316, no. 1, pp. 2–12, 1998. doi: 10.1016/S0002-9629(15)40365-9

[3]   M. T. Gaffey and M. M. Homsi, *Sinus endoscopic surgery*. Treasure Island, FL: StatPearls Publishing, 2022.

[4]   J. K. Hoang, J. D. Eastwood, C. L. Tebbit, and C. M. Glastonbury, "Multiplanar sinus CT: A systematic approach to imaging before functional endoscopic sinus surgery," *American Journal of Roentgenology*, vol. 194, no. 6, pp. 527–536, 2010. doi: 10.2214/AJR.09.3584

[5]   S. Ozbay and O. Tunc, "Deep learning in analysing paranasal sinuses," *Elektronika ir Elektrotechnika*, vol. 28, no. 3, pp. 65–70, 2022. doi: 10.5755/J02.EIE.31133

[6]   D. Hastan *et al*., "Chronic rhinosinusitis in Europe — An underestimated disease. A GA^{2}LEN study," *Allergy: European Journal of Allergy and Clinical Immunology*, vol. 66, no. 9, pp. 1216–1223, 2011. doi: 10.1111/J.1398-9995.2011.02646.X

[7]   R. J. Schlosser *et al*., "Depression-specific outcomes after treatment of chronic rhinosinusitis," *Journal of the American Medical Association Otolaryngology — Head and Neck Surgery*, vol. 142, no. 4, pp. 370-376, 2016. doi: 10.1001/jamaoto.2015.3810

[8]   A. Al-Mujaini, U. Wali, and M. Al Khabori, "Functional endoscopic sinus surgery: Indications and complications in the ophthalmic field," *Oman Medical Journal*, vol. 24, no. 2, pp. 70–80, 2009. doi: 10.5001/omj.2009.18

[9]   A. López and S. A. Martinson, "Respiratory system, mediastinum, and pleurae," *Pathologic Basis of Veterinary Disease*, vol. 1, no. 6, pp. 471–560, 2017. doi: 10.1016/B978-0-323-35775-3.00009-6

[10]   L. Rudmik *et al.*, "Defining appropriateness criteria for endoscopic sinus surgery during management of uncomplicated adult chronic rhinosinusitis: A RAND/UCLA appropriateness study," *International Forum of Allergy & Rhinology*, vol. 6, no. 6, pp. 557–567, 2016. doi: 10.1002/ALR.21769

[11]   K. C. Welch and J. A. Stankiewicz, "A contemporary review of endoscopic sinus surgery: Techniques, tools, and outcomes," *Laryngoscope*, vol. 119, no. 11, pp. 2258–2268, 2009. doi: 10.1002/lary.20618

[12]   M. W. Deutschmann *et al.*, "Radiologic reporting for paranasal sinus computed tomography: A multi-institutional review of content and consistency," *Laryngoscope*, vol. 123, no. 5, pp. 1100–1105, 2013. doi: 10.1002/lary.23906

[13]   N. I. Chowdhury, T. L. Smith, R. K. Chandra, and J. H. Turner, "Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks," *International Forum of Allergy & Rhinology*, vol. 9, no. 1, pp. 46–52, 2019. doi: 10.1002/ALR.22196

[14]   C. J. Massey, L. Ramos, D. M. Beswick, V. R. Ramakrishnan, and S. M. Humphries, "Clinical validation and extension of an automated, deep learning–based algorithm for quantitative sinus CT analysis," *American Journal of Neuroradiology*, vol. 43, no. 9, pp. 1318–1324, 2022. doi: 10.3174/ajnr.A7616

[15]   T. Gregurić *et al.*, "Imaging in chronic rhinosinusitis: A systematic review of MRI and CT diagnostic accuracy and reliability in severity staging," *Journal of Neuroradiology*, vol. 48, no. 4, pp. 277–281, 2021. doi: 10.1016/j.neurad.2021.01.010

[16]   M. Emirzeoglu, B. Sahin, S. Bilgic, M. Celebi, and A. Uzun, "Volumetric evaluation of the paranasal sinuses in normal subjects using computer tomography images: A stereological study," *Auris Nasus Larynx*, vol. 34, no. 2, pp. 191–195, 2007. doi: 10.1016/j.anl.2006.09.003

[17]   A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Review Cancer*, vol. 18, no. 8, pp. 500–510, 2018. doi: 10.1038/s41568-018-0016-5

[18]   The Royal College of Radiologists, "UK workforce census 2020 report: Clinical radiology," London: The Royal College of Radiologists, No. BFCR(21)3, Apr. 2021. [Online]. Available: https://www.rcr.ac.uk [Accessed: Jun. 14, 2023].

[19]   P. Rajpurkar *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[20]   R. Schoeman and M. Haines, "Radiologists' experiences and perceptions regarding the use of teleradiology in South Africa," *South African Journal of Radiology*, vol. 27, no. 1, pp. 2647–2649, 2023. doi: 10.4102/sajr.v27i1.264

[21]   S. G. Mistry, D. R. Strachan, and E. L. Loney, "Improving paranasal sinus computed tomography reporting prior to functional endoscopic sinus surgery — An ENT-UK panel perspective," *The Journal of Laryngology and Otology*, vol. 130, no. 10, pp. 962–966, May 2016. doi: 10.1017/S0022215116008902

[22]   K. C. McMains, "Safety in endoscopic sinus surgery," *Current Opinion in Otolaryngology & Head and Neck Surgery*, vol. 16, no. 3, pp. 247–251, 2008. doi: 10.1097/MOO.0b013e3282fdccad

[23]   A. J. Thomas, E. D. McCoul, J. D. Meier, C. I. Newberry, T. L. Smith, and J. A. Alt, "Cost and operative time estimation itemized by component procedures of endoscopic sinus surgery," *International Forum of Allergy & Rhinology*, vol. 10, no. 6, pp. 755–761, 2020. doi: 10.1002/alr.22554

[24]   B. Cadd, H. Khatri, S. Mohr, and J. C. Oosthuizen, "Radiological reporting of CT paranasal sinuses: CLOSE to the mark? The Australian otolaryngologists' perspective," *Australian Journal of Otolaryngology*, vol. 5, pp. 1-10, 2022. doi: 10.21037/ajo-20-85

[25]   M. P. Hartung, I. C. Bickle, F. Gaillard, and J. P. Kanne, "How to create a great radiology report," *RadioGraphics*, vol. 40, no. 6, pp. 1658–1670, 2020. doi: 10.1148/rg.2020200020

[26]   T. Okushi *et al.*, "A modified Lund-Mackay system for radiological evaluation of chronic rhinosinusitis," *Auris Nasus Larynx*, vol. 40, no. 6, pp. 548–553, 2013. doi: 10.1016/j.anl.2013.04.010

[27]   V. Lund and I. Mackay, "Staging in rhinosinusitis," *Rhinology*, vol. 31, no. 1, pp. 183–184, 1994.

[28]   D. W. Kennedy *et al.*, "Treatment of chronic rhinosinusitis with high-dose oral Terbinafine: A double blind, placebo-controlled study," *Laryngoscope*, vol. 115, no. 10, pp. 1793–1799, 2005. doi: 10.1097/01.mlg.0000175683.81260.26

[29]   J. Kumari, E. Kumar, and D. Kumar, "A structured analysis to study the role of machine learning and deep learning in the healthcare sector with big data analytics," *Archives of Computational Methods in Engineering*, vol. 30, no. 6, pp. 3673–3701, 2023. doi: 10.1007/s11831-023-09915-y

[30] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979. doi: 10.1109/TSMC.1979.4310076

[31] M. C. Alanin and C. Hopkins, "Effect of functional endoscopic sinus surgery on outcomes in chronic rhinosinusitis," *Current Allergy and Asthma Reports*, vol. 20, no. 7, pp. 27–32, 2020. doi: 10.1007/s11882-020-00932-6

[32] M. Desrosiers *et al.*, "Canadian clinical practice guidelines for acute and chronic rhinosinusitis," *Allergy, Asthma and Clinical Immunology*, vol. 7, no. 2, 2011. doi: 10.1186/1710-1492-7-2

[33] C. Hanson and K. Lepule, "Management of acute and chronic sinusitis," *South African Pharmaceutical Journal*, vol. 84, no. 4, pp. 45-51, 2017. Available: https://hdl.handle.net/10520/ejc-mp_sapj_v88_n5_a7

[34] L. Rudmik, Z. M. Soler, and T. L. Smith, "Geographic variation of endoscopic sinus surgery in the United States," *Laryngoscope*, vol. 125, no. 8, pp. 1772–1778, 2015. doi: 10.1002/lary.25314

[35] G. Venkatraman, D. S. Likosky, D. Morrison, W. Zhou, S. R. G. Finlayson, and D. C. Goodman, "Small area variation in endoscopic sinus surgery rates among the Medicare population," *Journal of the American Medical Association Otolaryngology — Head and Neck Surgery*, vol. 137, no. 3, pp. 253–257, 2011. doi: 10.1001/archoto.2011.17

[36] C. O. Laura, P. Hofmann, K. Drechsler, and S. Wesarg, "Automatic detection of the nasal cavities and paranasal sinuses using deep neural networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging*, Venice, Italy, 2019, pp. 1154-1157. doi: 10.1109/ISBI.2019.8759481

[37] Y. Iwamoto *et al.*, "Automatic segmentation of the paranasal sinus from computer tomography images using a probabilistic atlas and a fully convolutional network," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany: IEEE, 2019, pp. 2789–2792. doi: 10.1109/EMBC.2019.8856703

[38] 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," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 9, pp. 1457–1465, 2020. doi: 10.1007/s11548-020-02228-6

[39] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision*, Stanford, California: IEEE, Oct. 2016, pp. 565-571. doi: 10.1109/3DV.2016.79

[40] 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, 688, 2021. doi: 10.3390/diagnostics11040688

[41] S. M. Humphries *et al.*, "Volumetric assessment of paranasal sinus opacification on computed tomography can be automated using a convolutional neural network," *International Forum of Allergy & Rhinology*, vol. 10, no. 11, pp. 1218–1225, 2020. doi: 10.1002/alr.22588

[42] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional Densenets for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, Hawaii, 2017, pp. 11–19.

[43] A. Fedorov *et al.*, "3D slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323-1341, 2012. doi: 10.1016/j.mri.2012.05.001

[44] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BioMed Central Research Notes*, vol. 15, no. 1, 210, 2022. doi: 10.1186/s13104-022-06096-y

[45] D. P. Kingma and J. Ba, "ADAM: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[46] A. Avesta, S. Hossain, M. Lin, M. Aboian, H. M. Krumholz, and S. Aneja, "Comparing 3D, 2.5D, and 2D approaches to brain image auto-segmentation," *Bioengineering*, vol. 10, no. 2, 181, 2023. doi: 10.3390/bioengineering10020181