applied system innovation

MDPI

Article

Automated Identification from CT Using Sphenoid Sinus Geometry as an Anatomical Biometric

Nataliya Bilous [1,2,*], Vladyslav Malko [1], Dmytro Tkachenko [1] and Marcus Frohme [2]

[1] Computer Science Faculty, Kharkiv National University of Radio Electronics, 14 Nauky Ave., 61166 Kharkiv, Ukraine; vladyslav.malko@nure.ua (V.M.); dmytro.tkachenko@nure.ua (D.T.)
[2] Division Molecular Biotechnology and Functional Genomics, Technical University of Applied Sciences, 1 Hochschulring, 15745 Wildau, Germany; mfrohme@th-wildau.de
* Correspondence: nataliya.bilous@nure.ua

Abstract

Reliable identification of deceased individuals may be difficult when conventional biometric methods such as facial recognition, fingerprint analysis, or DNA profiling cannot be applied. In such cases, medical imaging records acquired during a person's lifetime may serve as an alternative source of identifying information. Certain anatomical structures visible in computed tomography (CT), including the sphenoid sinus, exhibit considerable inter-individual variability while remaining relatively stable within the same individual. This study investigates the feasibility of using sphenoid sinus morphology as an anatomical biometric for automated identification from head CT scans. Identification is formulated as a ranking problem in which a query CT examination is compared with a reference database using geometric descriptors derived from segmentation masks, reducing dependence on CT intensity values. The dataset consisted of CT scans from 816 individuals acquired in two patient positioning modes: Head First Supine (HFS) and Head First Prone (HFP). Several deep learning architectures, including YOLOv8 variants, YOLO11L-seg, UNet++, DeepLabV3+, HRNet, and SegFormer-B2, were evaluated for sphenoid sinus segmentation. Based on F1-score performance and cross-mode stability, YOLO11L-seg was selected and further trained to construct a database of binary masks representing individual sphenoid sinus anatomy. Identification was performed using pairwise mask comparison based on the Intersection over Union (IoU) metric. To reduce the influence of segmentation artifacts and slice-level variability, the final similarity score for each candidate was computed as the average of the four highest IoU values across slice comparisons. Individuals were ranked according to similarity, and identification was considered successful if the correct subject appeared among the top five candidates and exceeded a predefined similarity threshold. The proposed approach achieved Top-5 identification accuracies of 97.27% for HFP and 87.67% for HFS acquisitions. These results demonstrate the feasibility of using sphenoid sinus geometry as a stable anatomical biometric for automated identification. The key contribution of this study is the introduction of a ranking-based identification framework that utilizes anatomical biometrics derived from CT data for reliable patient matching.

Keywords: sphenoid sinus morphology; CT image analysis; deep learning segmentation; YOLO architecture; automated identification

Check for updates

Academic Editor: Teen-Hang Meen

Received: 16 March 2026
Revised: 25 April 2026
Accepted: 26 April 2026
Published: 29 April 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The digitization of medical imaging has led to the creation of large archives of diagnostic images integrated into picture archiving and communication systems (PACS) and elec-

tronic medical records. With the growth in data volume, the increase in inter-institutional exchanges, and repeated examinations, correct patient identification has become particularly important. Errors in matching studies to patients may occur due to human error, partial loss of metadata, technical failures, or incorrect transmission of information between systems. Such errors can potentially lead to diagnostic inaccuracies, misinterpretation of disease progression, and clinical decisions based on another patient's data. Traditional identification mechanisms in medical information systems rely primarily on text and numerical attributes such as name, date of birth, unique identifiers, and medical record numbers. However, this metadata may be incomplete, duplicated, or incorrectly entered. In recent decades, biometric methods based on physiological and behavioral characteristics, such as fingerprints, iris, face, or voice, have been actively developed. These methods are highly effective in authentication tasks, but their integration into medical imaging workflows is limited because they are not part of the imaging procedure itself. In contrast, anatomical structures visible in medical images can potentially be considered intrinsic biometric markers. Computed tomography of the head provides detailed information about cranial bone structures, which exhibit significant inter-individual variability. Among these structures, the sphenoid sinus is of particular interest due to its complex geometry and individualized morphology. Although deep learning-based segmentation methods have been extensively studied in medical imaging, most existing work focuses on diagnostic applications such as pathology detection, lesion assessment, or surgical planning. In contrast, patient identification based on segmented anatomical geometry remains less explored in lightweight and practically scalable formulations. Prior studies have demonstrated the feasibility of sphenoid-sinus-based identification, including geometric comparison and three-dimensional forensic matching. However, such approaches often rely on computationally intensive 3D reconstruction or may not fully reflect practical identification scenarios involving comparison against a database of candidate patients. An additional methodological aspect relevant to this problem is patient positioning during CT acquisition. Differences between Head First Supine (HFS) and Head First Prone (HFP) modes may affect slice distribution, the number of informative sections, and the geometric representation of anatomical structures. Since the proposed approach relies on segmentation-derived morphology, evaluating its behavior across positioning modes is important for assessing robustness under varying acquisition conditions. This paper addresses the problem of automated patient identification based on head CT data using sphenoid sinus segmentation and subsequent geometric matching of binary masks. The proposed approach does not rely on intensity features, manual feature engineering, or complex registration procedures. Instead, identification is formulated as a ranking task in which the similarity between a query CT examination and a reference database is determined based on the spatial overlap of segmented structures. Unlike prior studies that formulate identification-related tasks as classification problems or rely on intensity-based image features, the present study addresses CT-based identification as a database-level ranking problem. In this setting, a query CT examination is compared against a reference database, and candidates are ordered according to geometric similarity. The proposed framework relies exclusively on segmentation-derived morphology of the sphenoid sinus, thereby avoiding dependence on raw CT intensity values, manual feature engineering, or complex registration procedures. The main contribution of this study is not the development of a new segmentation architecture, but the formulation and validation of a lightweight anatomical biometric framework for CT-based identification. In this framework, segmentation serves as an enabling step, while identification is performed through direct geometric comparison of binary masks. The main contributions of this study are as follows:

- An anatomical biometric framework for CT-based identification based on sphenoid sinus geometry;
- Formulation of identification as a ranking-based matching problem, reflecting practical identification scenarios;
- Segmentation-driven morphological representation that does not rely on image intensity during identification;
- Evaluation of robustness under varying acquisition conditions, demonstrating the feasibility of the approach.
- Proof-of-concept validation on retrospective clinical CT data.

In contrast to many previous approaches that rely on computationally intensive three-dimensional reconstruction, the proposed method focuses on lightweight geometric comparison of two-dimensional segmentation masks, enabling efficient large-scale identification. Accordingly, the objective of this study is to investigate the feasibility of using sphenoid sinus morphology, represented through segmentation masks, for automated patient identification from head CT scans, and to evaluate a two-dimensional ranking-based matching framework under different patient positioning conditions.

2. Review of Literature

This section reviews existing approaches to anatomical biometrics in medical imaging, with a focus on automated comparison of morphological structures for identification tasks. Particular attention is given to the use of paranasal sinuses as individualized anatomical markers in CT-based identification. Previous studies confirm that the sphenoid sinus exhibits significant inter-individual variability while maintaining relative morphological stability within the same individual. In [1], an automated approach to personal identification based on sphenoid sinus segmentation and geometric comparison of masks was proposed, demonstrating the feasibility of using this structure as an anatomical biometric marker. Similar findings are reported in [2], where detailed morphometric analysis highlights variability in pneumatization patterns and structural configuration of the sinus. Although these studies establish the potential of sphenoid sinus geometry for identification, they primarily focus on demonstrating anatomical uniqueness rather than addressing scalable identification scenarios. Further development in this area has been associated with the transition from two-dimensional analysis to full three-dimensional representations. In [3], Dong et al. proposed a fully automated forensic identification framework based on 3D reconstruction and point cloud registration using the ICP algorithm, achieving high Rank-1 and Rank-2 accuracy. Similarly, ref. [4] demonstrated that 3D model superposition reduces the influence of positional variation and improves reproducibility. These studies suggest that volumetric representations provide a more stable basis for biometric comparison. However, such approaches require additional reconstruction steps, increased computational resources, and more complex processing pipelines, which may limit their applicability in large-scale identification systems. Works [5–7] demonstrate statistically significant differences in sinus morphology between population groups, supporting the presence of measurable anatomical variability. However, these studies are primarily focused on classification tasks such as sex determination and are not directly applicable to patient identification in database matching scenarios. Recent advances in machine learning and deep learning have significantly improved the accuracy of anatomical structure extraction from medical images. Studies such as [8,9] demonstrate that convolutional neural networks enable reliable segmentation of complex structures in MRI and CT data. At the same time, segmentation is typically treated as an end task rather than as part of a downstream identification pipeline. Additional research, including [10] explores attention-based feature fusion mechanisms that improve structural consistency in segmentation tasks.

These developments provide a technological foundation for extracting stable geometric representations of anatomical structures. A comprehensive overview of segmentation approaches for paranasal sinuses is presented in [11], where manual, semi-automatic, and fully automated methods are compared. Automated approaches are shown to improve reproducibility and reduce inter-observer variability, making them suitable for large-scale analysis. However, these methods are generally evaluated in isolation and are not explicitly integrated into scalable identification frameworks operating at the database level. Alternative approaches that do not rely on explicit segmentation have also been proposed. In [12], identification is performed using local invariant features (AKAZE) extracted from individual CT slices, demonstrating that feature-based matching can achieve high accuracy in large databases. While such methods avoid segmentation, they remain dependent on image intensity characteristics and local descriptors, which may be sensitive to acquisition variability. The methodological foundations of large-scale matching and ranking are discussed in [13–15], where identification is formulated as a ranking problem using similarity metrics and top-k selection strategies. These approaches enable efficient comparison across large candidate sets and provide robustness to minor variations in imaging conditions. Additionally, ref. [16] highlights the importance of standardized anatomical representations and quantitative similarity measures for building generalizable identification systems. Compared with prior sphenoid-sinus-based identification studies, the present work differs in its practical formulation and computational design. Three-dimensional approaches based on volumetric reconstruction and point-cloud registration provide detailed spatial representations and have demonstrated strong forensic identification potential. However, they usually require additional reconstruction steps, more complex registration procedures, and higher computational overhead. In contrast, the proposed framework operates directly on two-dimensional segmentation masks and formulates identification as a database-level ranking task. This design is intended to provide a lightweight and reproducible alternative for large-scale candidate retrieval, while preserving the morphological information needed for anatomical comparison. Despite the demonstrated potential of sphenoid sinus-based identification, several limitations remain in existing studies. First, many approaches formulate the problem as classification rather than database-level matching, which may not fully reflect real-world identification scenarios. Second, intensity-based and feature-based methods remain sensitive to variations in acquisition parameters and reconstruction settings. Third, approaches relying on three-dimensional reconstruction introduce additional computational complexity that may limit scalability in large databases. In this context, there remains a need for a lightweight and reproducible identification framework that operates directly on segmentation-derived geometry, avoids dependence on intensity information, and supports efficient ranking-based comparison across large datasets. Addressing this gap forms the basis of the present study.

3. Methods

3.1. Datasets and Experimental Design

3.1.1. Source of CT Data

The research included head computed tomography (CT) scans from 816 patients in which the sphenoid sinus region was fully covered by the scan. Each CT examination consisted of a sequence of axial slices with a spatial resolution of 512 × 512 pixels. Although head CT examinations typically capture the entire cranial anatomy, including multiple bone structures of the skull, the present study focuses specifically on the sphenoid sinus region. This choice is motivated by its high inter-individual variability combined with relative morphological stability within a single patient, which are essential properties for biometric identification. By restricting the analysis to this anatomically consistent and structurally

distinctive region, the influence of less informative or highly variable anatomical areas is reduced, allowing a more controlled evaluation of identification performance. Depending on the anatomical coverage and scanning protocol, each study contained approximately 10–30 axial slices in which the sphenoid sinus region was visible. Consequently, the dataset comprised 12,350 individual CT images used for segmentation and subsequent analysis. The images were used in their original format without any additional intensity preprocessing, in order to preserve the geometric and morphological characteristics of the sphenoid sinus, which are critical for subsequent segmentation and identification tasks (Figure 1).

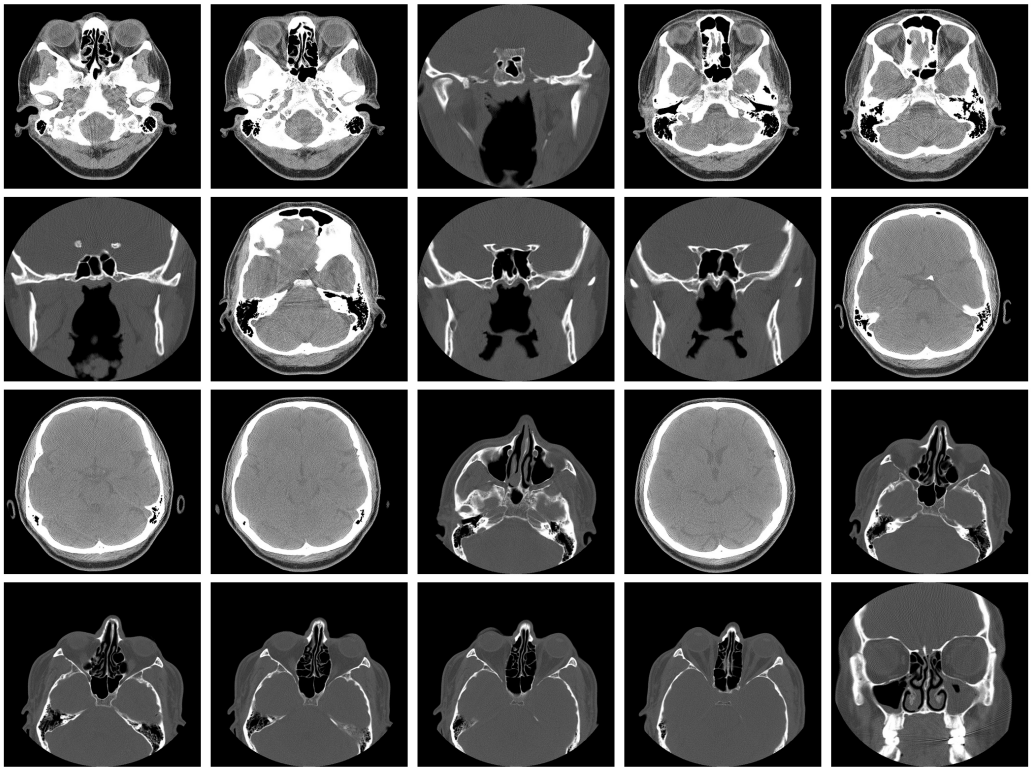

Figure 1. Examples from CT scan dataset.

To ensure a correct evaluation of model generalization and to eliminate data leakage, the dataset was divided into training, validation, and test subsets separately for each positioning mode at the patient level, with no overlap of patients between subsets. The data distribution across subsets was designed to preserve a representative balance of slice counts and anatomical variability. In the Head First Supine (HFS) group, the dataset included 486 patients with a total of 5832 CT slices (approximately 12 slices per patient). The training subset contained 3960 slices from 330 patients, the validation subset included 936 slices from 78 patients, and the test subset consisted of 936 slices from 78 patients. This partitioning ensured a consistent distribution of samples across subsets while maintaining strict patient-level separation. In the Head First Prone (HFP) group, 330 patients were available, with a total of 6518 CT slices (approximately 19–20 slices per patient). The training subset included 4545 slices from 231 patients, the validation subset contained 1008 slices from 51 patients, and the test subset consisted of 965 slices from 48 patients. A comparable distribution strategy was applied to maintain consistency between subsets and support unbiased model evaluation. Hounsfield scale normalization, histogram equalization, scaling, or image bit depth conversion were not applied. This approach was deliberately chosen in order to preserve the original geometry of the bone structures, since minor intensity transformations can indirectly affect the segmentation results and the shape of the binary mask. CT data

were collected from the Kharkiv Research Institute of General and Emergency Surgery and the Central District Hospital of Merefa (under an institutional agreement No. 173/10 18 dated 18 October 2018 on scientific and practical cooperation). All CT data were processed in anonymized form, and no patient identifiers were available during analysis.

3.1.2. Data Preparation

All CT scans were categorized according to the patient's positioning in the tomograph. Two standard clinical modes were considered:

- Head First Supine (HFS)—lying on the back, head first;
- Head First Prone (HFP)—lying on the stomach, head first.

The distribution according to positioning modes was of fundamental importance for the experimental design. Patient orientation affects the spatial location of anatomical structures relative to the tomograph coordinate system, which can lead to changes in the number of informative slices and their geometric representation. Even if one patient was examined in both modes, these sets were considered as independent samples. This approach allowed us to separately evaluate the stability of segmentation and the identification algorithm under changing orientation conditions. Although scans from the same patient acquired in different positioning modes are not biologically independent, they were treated as independent acquisition conditions. This design allows evaluation of the robustness of the segmentation and identification pipeline under variations in orientation. However, in this study, HFS and HFP acquisitions are treated as distinct acquisition conditions rather than independent identity samples. Since the identification stage relies on geometric similarity between segmentation masks rather than statistical learning of identity labels, this design does not introduce label leakage or bias into the ranking procedure. Instead, it allows evaluation of the robustness of the proposed framework under variations in patient orientation and slice distribution. Since identification is based on geometric similarity of segmentation masks rather than patient-level labels, this does not introduce bias in the ranking-based identification stage or lead to label leakage between subsets, since matching is performed at the level of geometric similarity rather than class labels. Since the sphenoid sinus is present only in a limited region of axial slices, a subset of relevant slices containing the target anatomical structure was predefined for each study. At the annotation stage, only those sections in which the sinus was clearly visualized were included for consideration, which made it possible to reduce the influence of irrelevant images on the training process and increase the stability of the identification algorithm, since further pairwise comparison of masks was performed only for informative sections. Informative slices were defined as axial slices in which the sphenoid sinus cavity was fully or predominantly visible and could be reliably delineated. Slice selection was performed manually during the annotation stage based on visual anatomical criteria. Only slices containing a sufficiently complete representation of the sinus structure were included in further processing. Ground truth segmentation masks were generated manually by two trained annotators with domain expertise in anatomical structure identification. Each CT study was annotated independently and subsequently reviewed. In cases of disagreement, a consensus was reached through joint discussion. The annotation process involved slice-by-slice delineation of the sphenoid sinus region. To ensure consistency, ambiguous cases were reviewed and resolved during the consensus stage. This procedure ensured reproducible and reliable reference masks for model training and evaluation. The annotation protocol remained consistent across all subsets and was applied prior to model training. For each positioning mode, separate subsets of data for training, validation, and testing were formed at the patient level, ensuring no overlap between subsets. This approach prevented data leakage and enabled a correct assessment of model generalization. At the initial stage of the

study, fixed training, validation, and test subsets were defined and used for comparative evaluation of different segmentation architectures under identical experimental conditions. After selecting the final architecture, the training process was refined using an extended yet carefully curated subset of the dataset, while the test subset remained unchanged to preserve the independence and objectivity of the final evaluation. This two-stage design enabled fair model comparison and reliable assessment of the identification framework.

3.1.3. Ethical and Data Handling Considerations

All studies were used in anonymized form. Personal identifiers were removed before the experimental sample was formed. Access to the data was restricted to the research group. Processing was performed exclusively for scientific purposes without the possibility of identifying individuals outside of algorithmic analysis.

3.2. Segmentation Model Architecture and Training

3.2.1. Formalization of the Segmentation Problem

Segmentation of the sphenoid sinus is formulated as a binary semantic segmentation problem. Let the input image be denoted as $I \in R^{\{H \times W\}}$, where $H = W = 512$. The goal is to construct a mapping $f(I; \theta) \rightarrow M$, where $M \in \{0,1\}^{\{H \times W\}}$ is a binary mask of the sphenoid sinus, and θ is a set of neural network parameters. Each pixel is assigned to one of two classes: 1—sphenoid sinus, 0—background. The peculiarity of the task lies in the complex individually variable form of the structure, the presence of thin contours, heterogeneous pneumatization, and partial volumetric effects at the border between bone tissue and air cavity. This creates increased requirements for the accuracy of boundary localization and the stability of the mask shape between adjacent slices. The segmentation model must ensure high spatial consistency of the predicted contours, minimal fragmentation of the area, and the ability to generalize for different patient positioning modes. Since the mask is subsequently used as the main object of comparison, even minor systematic distortions in geometry can affect the identification result.

3.2.2. Architectural Approaches

To select the optimal architecture, a comparative analysis of the YOLOv8s-seg [17], YOLOv8x-seg, YOLO11L-seg [18], and UNet++ [19] models was performed. The selection was based on their active use in medical segmentation tasks and differences in network depth, number of parameters, and decoder structure. The YOLO family of models implements an end-to-end architecture with a feature extractor, multi-scale context aggregation, and a segmentation head. This approach allows for the effective combination of local and global features, which is critical for the correct reproduction of the complex geometry of the sphenoid sinus. The compact version of YOLOv8s-seg is characterized by fewer parameters, while YOLOv8x-seg has increased network depth and width. The YOLO11L-seg model is focused on a more complete spatial-contextual representation and potentially provides higher contour stability. UNet++ uses a nested skip-connection structure with dense feature aggregation at different depth levels. This architecture traditionally performs well in medical segmentation tasks, especially for structures with complex boundaries. Comparing these approaches allowed us to evaluate the impact of architectural complexity on the quality of geometry reproduction. In addition to the selected architectures, widely used state-of-the-art segmentation models such as DeepLabV3+ [20], HRNet [21], and the transformer-based SegFormer [22] were also included in the evaluation. These models represent different architectural paradigms, including convolutional networks with atrous spatial pyramid pooling, high-resolution feature representations, and transformer-based global context modeling. Their inclusion enables a comprehensive comparison across mod-

ern segmentation strategies and ensures that the conclusions drawn from the experiments are not limited to a specific class of architectures.

3.2.3. Training Protocol

Training was performed separately for HFS and HFP positioning modes, which made it possible to exclude the mixing of the influence of patient orientation with the characteristics of the model itself. Each architecture was trained for 150 epochs with a batch size of 32 and a fixed input resolution of 512 × 512 pixels, which corresponded to the output format of CT images. Parameter optimization was performed using the Adam method with an adaptive cosine decay learning rate schedule. This strategy ensured stable convergence and avoided sharp fluctuations in the loss function. To reduce the risk of overfitting, validation error control with early stopping was applied, with a patience of 20 epochs based on validation loss. The loss function was formed as a combination of binary cross-entropy and an IoU-oriented component. In general, it is written as

$$L = \lambda_1 * \ BCE\left(M_{pred}, M_{\ gt}\right) + \lambda_2 * \ \left(1 - IoU\left(M_{\ pred}, M_{\ gt}\right)\right) \quad (1)$$

where M_{pred} is the predicted mask, $M_{\ gt}$ is the ground truth mask, and λ_1 and λ_2 are weighting coefficients. The weighting coefficients were set to $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$. This combination allows us to simultaneously minimize pixel error and optimize the global overlap of the segmentation area. Data augmentation included small geometric transformations such as random rotations within ±5°, translations within ±5%, and slight scaling variations. Strong deformations or intensity transformations were not applied in order to preserve anatomical geometry.

3.2.4. Evaluation Metrics and Selection of the Final Model

Segmentation performance was evaluated primarily using the F1-score, defined as the harmonic mean of precision and recall. Precision measures the proportion of correctly predicted positive pixels among all predicted positives, while recall quantifies the proportion of correctly predicted positive pixels among all ground truth positives. In addition, the Intersection over Union (IoU) metric was used to assess the spatial overlap between predicted masks and reference annotations. Following the comparative analysis of the evaluated architectures, the model demonstrating the most stable segmentation performance across both positioning modes was selected. In this context, stability was considered alongside average metric values, as consistent contour reconstruction across slices is critical for reliable downstream identification. The selected model was further trained on an extended dataset while preserving an independent test subset to ensure unbiased evaluation. The final trained model was used to generate segmentation masks for the entire reference database, forming the geometric basis for the anatomical identification algorithm described in the next section.

3.2.5. Implementation Details

All experiments were implemented in Python 3.10 using the PyTorch 2.7.1 (CUDA 11.8) deep learning framework. The YOLO-based segmentation models (YOLOv8s-seg, YOLOv8x-seg, and YOLO11L-seg) were implemented using the Ultralytics library, while UNet++ was implemented as a standard encoder–decoder architecture with nested skip connections. Training and inference were performed on a high-performance computing server equipped with an NVIDIA A100-SXM4-80GB GPU, ensuring stable training dynamics and efficient processing of high-resolution medical images. All input CT slices were resized to a fixed resolution of 512 × 512 pixels, corresponding to the original acquisition format. The models were trained with a batch size of 32 using the Adam optimizer with an initial learning

rate of 1×10^{-3}. A cosine decay learning rate schedule was applied to promote stable convergence. Each model was trained for 150 epochs, with early stopping based on validation loss to prevent overfitting. The loss function was defined as a weighted combination of binary cross-entropy and an IoU-based term, enabling both pixel-level accuracy and global shape consistency. Data augmentation was limited to mild geometric transformations, such as small rotations and translations, in order to preserve anatomical correctness. More aggressive augmentations, including elastic deformations or intensity transformations, were intentionally avoided to prevent distortion of the sphenoid sinus morphology. No preprocessing steps, such as Hounsfield unit normalization, histogram equalization, or contrast enhancement, were applied to the CT images. This design choice ensured preservation of the original anatomical geometry and avoided potential distortions affecting segmentation quality and subsequent identification. During inference, segmentation masks were generated directly from model outputs without post-processing operations such as smoothing, morphological filtering, or connected-component correction. This ensured that the geometric representation used for identification remained fully consistent with the learned model predictions. The identification stage was implemented separately from segmentation. Pairwise IoU computations between masks were performed using vectorized array operations, enabling efficient evaluation across the entire database. The overall pipeline was fully automated and did not require manual intervention at any stage.

3.3. Reference Mask Database and Patient Identification Algorithm

3.3.1. Construction of the Reference Segmentation Database

After completion of the segmentation stage and selection of the final model, the network was applied to all CT examinations included in the study cohort. For each examination, all axial slices containing the sphenoid sinus region were processed. The output of the model for every slice was a binary segmentation mask representing the spatial geometry of the sphenoid sinus without any post-processing or morphological refinement. All masks were stored at the original spatial resolution of 512×512 pixels in order to preserve anatomical fidelity. No resizing, smoothing, contour correction, or connected-component filtering was applied, since any artificial modification of the predicted region could alter the geometric characteristics that serve as the basis for subsequent identification. Each mask was therefore treated as a discrete spatial representation of the anatomical structure. The reference database was organized hierarchically at four levels: patient, study, series, and slice. For each slice, the system stored a reference to the original CT image and the corresponding binary mask. This structure ensured full traceability of the data and reproducibility of all pairwise comparisons performed during the identification procedure. From a formal perspective, for each patient k a set of masks $B_k = \left\{B_k^1, B_k^2, \ldots, B_k^{n_k}\right\}$ was defined, where n_k denotes the number of informative slices containing the sphenoid sinus. The complete reference database was therefore represented as the union of all patient-specific mask sets.

3.3.2. Mathematical Formulation of the Identification Procedure

Patient identification was formulated as a ranking problem. Let a new CT examination be represented by a set of segmentation masks $A = \left\{A^1, A^2, \ldots, A^m\right\}$, where m denotes the number of slices containing the sphenoid sinus in the new study. The objective of the algorithm was to determine which patient in the reference database exhibited the highest geometric similarity with A. For each patient k, pairwise similarity values were computed between every mask A^i and every mask B_k^j belonging to that patient. Similarity was quantified using the Intersection over Union metric, defined as $IoU\left(A^i, B_k^j\right) = \frac{\left|A^i \cap B_k^j\right|}{\left|A^i \cup B_k^j\right|}$, where

$|\cdot|$ denotes the number of pixels in the corresponding set. The Intersection over Union (IoU) metric was used as a geometric similarity measure between binary segmentation masks. Although IoU is commonly used for evaluating segmentation accuracy, in this study it is applied to quantify the morphological similarity between anatomical structures extracted from different CT examinations. This metric provides a normalized measure of spatial overlap between two binary masks. Because the number of informative slices may differ between examinations, full pairwise comparison was performed between sets A and B_k. The resulting similarity values formed a distribution for each patient. To obtain a stable and representative similarity score, the four highest IoU values were selected, and their arithmetic mean was computed. This choice was empirically validated during preliminary experiments, where alternative aggregation strategies (Top-3 and Top-5) were evaluated. The Top-4 configuration provided the best balance between robustness and sensitivity to slice-level variability. The integrated similarity coefficient S_k for patient k was therefore defined as the average of the four maximal overlap values. This strategy emphasized the most geometrically consistent correspondences while reducing the influence of local discrepancies caused by slice variability, minor anatomical differences, or segmentation noise. All patients in the reference database were ranked in descending order of S_k. The resulting ordered list represented the output of the identification system. The overall identification pipeline, including segmentation, mask comparison, similarity aggregation, and ranking, is illustrated in Figure 2.

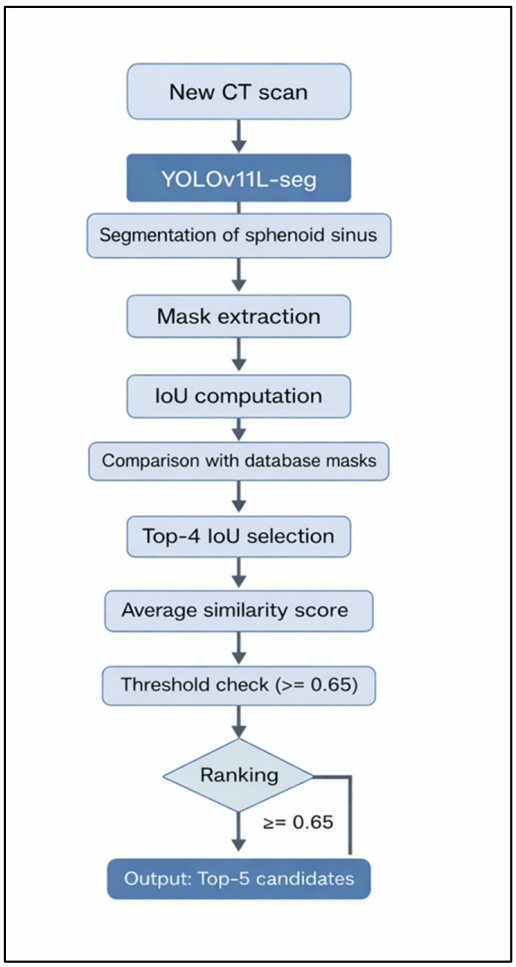

Figure 2. Schematic representation of the mask comparison and ranking process.

In this framework, segmentation masks serve as compact representations of individual anatomical morphology.

3.3.3. Computational Considerations and Algorithmic Stability

The computational complexity of the identification procedure depends on the number of masks in the new examination and the total number of masks stored in the reference database. For each patient k, the comparison requires evaluation of $m \times n_k$ IoU computations. In practice, IoU calculation for binary masks of fixed resolution is computationally efficient, as it reduces to element-wise logical operations and pixel counting. The implementation was optimized using vectorized array operations, enabling efficient processing even when the size of the reference database increases. Since the method relies exclusively on geometric properties of binary masks, it reduces dependence on CT intensity distributions, reconstruction parameters, or contrast variations between repeated examinations. Another practical advantage of the proposed approach is the compact representation of anatomical structures. Binary segmentation masks require significantly less storage space than volumetric three-dimensional models or polygonal meshes typically generated in 3D reconstruction workflows. This makes it possible to construct large identification databases with reduced storage requirements and lower computational overhead for large-scale comparisons. This design choice enhances robustness to acquisition variability and reduces susceptibility to intensity-based artifacts. The complete identification pipeline operates automatically. Once a new CT examination is provided, segmentation is performed, similarity coefficients are computed for all patients in the database, and a ranked candidate list is generated without manual intervention. The modular structure of the system allows expansion of the reference database without modification of the core algorithm, supporting scalability and long-term applicability in medical image management systems.

4. Results

4.1. Segmentation Performance Evaluation

The first stage of the experimental evaluation focused on a quantitative and qualitative comparison of convolutional neural network architectures for sphenoid sinus segmentation. The assessment was conducted separately for the two positioning modes, Head First Supine (HFS) and Head First Prone (HFP), in order to exclude confounding effects related to patient orientation. All models were trained and evaluated using identical data splits and hyperparameters, ensuring that observed differences in performance were attributable exclusively to architectural characteristics. Segmentation quality was primarily quantified using the F1-score, defined as the harmonic mean of precision and recall:

$$F1_{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2}$$

where

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}, \tag{3}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{4}$$

and *True Positives*, *False Positives*, and *False Negatives* denote the number of true positive, false positive, and false negative pixels, respectively. In addition to F1-score, the Intersection over Union (IoU) metric was considered to evaluate spatial overlap between predicted masks and reference annotations:

$$IoU = \frac{\left| M_{pred} \cap M_{gt} \right|}{\left| M_{pred} \cup M_{gt} \right|} \tag{5}$$

It should be noted that the IoU metric is used in two different contexts in this study. In the segmentation stage, IoU is computed between predicted masks and ground truth annotations to evaluate model accuracy. In the identification stage, IoU is used as a similarity measure between segmentation masks from different CT examinations. These two uses represent distinct concepts and serve different purposes within the proposed framework. Both metrics were computed at the slice level and averaged across the independent test subsets for each positioning mode. In addition to segmentation accuracy, model complexity was analyzed in terms of the number of trainable parameters (Params) and floating-point operations (FLOPs). These metrics provide complementary information about the computational cost of each architecture and are important for assessing their practical suitability for large-scale deployment. While F1-score reflects segmentation quality, Params and FLOPs characterize model compactness and inference complexity, respectively. In addition, Intersection over Union (IoU) was included as a complementary metric to provide a more direct measure of spatial overlap between predicted and ground truth masks. Their joint analysis makes it possible to evaluate the trade-off between segmentation accuracy and computational efficiency. Table 1 presents the F1-score and IoU values obtained for all evaluated architectures under HFS and HFP positioning modes. To address the need for broader comparison with modern segmentation approaches, the evaluation was extended to include representative state-of-the-art architectures from different design paradigms, including DeepLabV3+, HRNet, and the transformer-based SegFormer. All models were evaluated under a unified experimental protocol, including identical input resolution, training conditions, and dataset splits, ensuring that the comparison reflects architectural differences rather than variations in experimental setup. The reported IoU values correspond to standard segmentation IoU computed against ground truth masks.

Table 1. Comparison of segmentation performance (F1-score, segmentation IoU) and model complexity (Params, FLOPs) across evaluated architectures under HFS and HFP conditions.

Model	Params (M)	FLOPs (G)	HFS F1-Score	HFP F1-Score	HFS IoU	HFP IoU
YOLOv8s-seg	11.2	28.5	0.885	0.79	0.83	0.74
YOLO11L-seg	25.3	89.7	0.93	0.83	0.88	0.78
YOLOv8x-seg	68.2	257.3	0.93	0.80	0.88	0.75
UNet++	9.1	44.8	0.901	0.81	0.85	0.76
DeepLabV3+	43.5	180.0	0.91	0.80	0.86	0.75
HRNet-W48	63.2	210.0	0.915	0.81	0.87	0.76
SegFormer-B2	27.6	62.3	0.918	0.81	0.87	0.76

To further assess segmentation stability beyond mean performance metrics, the variance of F1-score across slices was analyzed for the top-performing models identified in Table 1, including YOLO11L-seg, YOLOv8x-seg, and SegFormer-B2. These models were selected based on their high segmentation accuracy and architectural diversity. The results are summarized in Table 2.

Table 2. Variance of F1-score across slices for selected high-performing models.

Model	HFS Variance	HFP Variance
YOLO11L-seg	0.0021	0.0018
YOLOv8x-seg	0.0031	0.0027
SegFormer-B2	0.0029	0.0025

The variance analysis indicates that YOLO11L-seg demonstrates the lowest variability across slices among the evaluated architectures. This reflects improved spatial consistency of predicted contours, which is critical for the stability of the subsequent identification

stage. The complexity analysis shows that the evaluated models differ substantially in computational cost. YOLOv8s-seg represents the most compact configuration, with the smallest number of parameters and the lowest FLOPs, but also demonstrates the lowest segmentation accuracy and IoU values. At the opposite end, YOLOv8x-seg exhibits the highest computational complexity due to its increased depth and representational capacity, resulting in higher computational cost without a significant improvement in performance. YOLO11L-seg provides a more favorable balance between segmentation quality and model complexity. Compared with YOLOv8x-seg, it achieves comparable F1-score values across both positioning modes while requiring lower computational resources and demonstrating improved spatial overlap as reflected in higher IoU values. UNet++, although competitive in terms of parameter count, showed slightly lower segmentation accuracy and reduced cross-mode stability. The additional state-of-the-art architectures, including DeepLabV3+, HRNet-W48, and SegFormer-B2, demonstrate competitive performance, with F1-scores in the range of 0.91–0.918 for HFS and around 0.80–0.81 for HFP, and corresponding IoU values following similar trends. Despite differences in architectural design, including convolutional, high-resolution, and transformer-based approaches, the overall segmentation accuracy remains within a relatively narrow range across all evaluated models. This observation indicates that, for the considered anatomically constrained task, increasing architectural complexity does not lead to proportional improvements in segmentation quality. Instead, the stability and consistency of the predicted anatomical contours, as well as their spatial agreement with ground truth masks, play a more critical role for subsequent identification. From this perspective, YOLO11L-seg offers the most suitable compromise between accuracy, robustness, and computational efficiency, which justifies its selection as the final segmentation backbone in the proposed identification framework. Under the HFS positioning mode, the highest F1-score (0.93) was achieved by both YOLO11L-seg and YOLOv8x-seg, with corresponding IoU values of 0.88, while SegFormer-B2 (0.918), HRNet-W48 (0.915), and DeepLabV3+ (0.91) also demonstrated competitive performance. However, qualitative analysis revealed that YOLO11L-seg provided more stable contour delineation across anatomically complex slices, whereas YOLOv8x-seg occasionally produced boundary irregularities in peripheral sinus regions. In the HFP mode, YOLO11L-seg achieved the highest F1-score of 0.83 with IoU of 0.78, followed by UNet++ (0.81/0.76), HRNet-W48 (0.81/0.76), SegFormer-B2 (0.81/0.76), YOLOv8x-seg (0.80/0.75), DeepLabV3+ (0.80/0.75), and YOLOv8s-seg (0.79/0.74). Although absolute F1 and IoU values were lower in HFP compared to HFS, the relative ranking of models remained consistent. The reduced segmentation performance in HFP is likely influenced by differences in slice distribution and anatomical projection associated with prone positioning. To further analyze segmentation consistency, the variance of F1-score across slices was computed for the best-performing model in each positioning mode. Let $F_{1,i}$ denote the slice-level F1-score for slice i, and let N be the number of slices in the test subset. The empirical variance was estimated as:

$$Var(F1) = \frac{1}{N-1} \times \sum_{i=1}^{N} (F_{1,i} - \mu)^2 \quad (6)$$

where μ is the mean F1-score across slices. Lower variance indicates improved stability of the segmentation contour across anatomical variations. YOLO11L-seg demonstrated reduced variance compared to alternative architectures, confirming improved spatial consistency. Qualitative evaluation supported the quantitative findings. In slices characterized by irregular pneumatization patterns or partial sinus coverage, higher-capacity models preserved anatomical continuity and avoided mask fragmentation. In contrast, smaller models occasionally produced disconnected mask components or underestimated sinus boundaries. Figures 3 and 4 illustrate representative examples of segmentation outputs

produced by YOLO11L-seg and UNet++ in both positioning modes. The visual comparison highlights improved boundary adherence and reduced false positive regions in the YOLO11L-seg predictions.

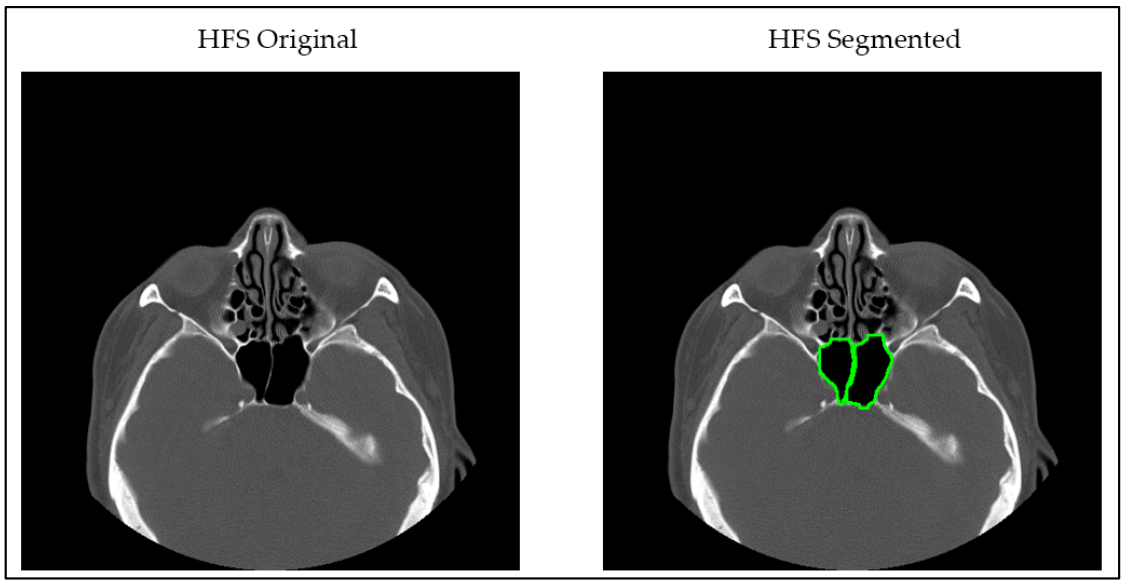

Figure 3. Example of sphenoid sinus segmentation using YOLO11L-seg.

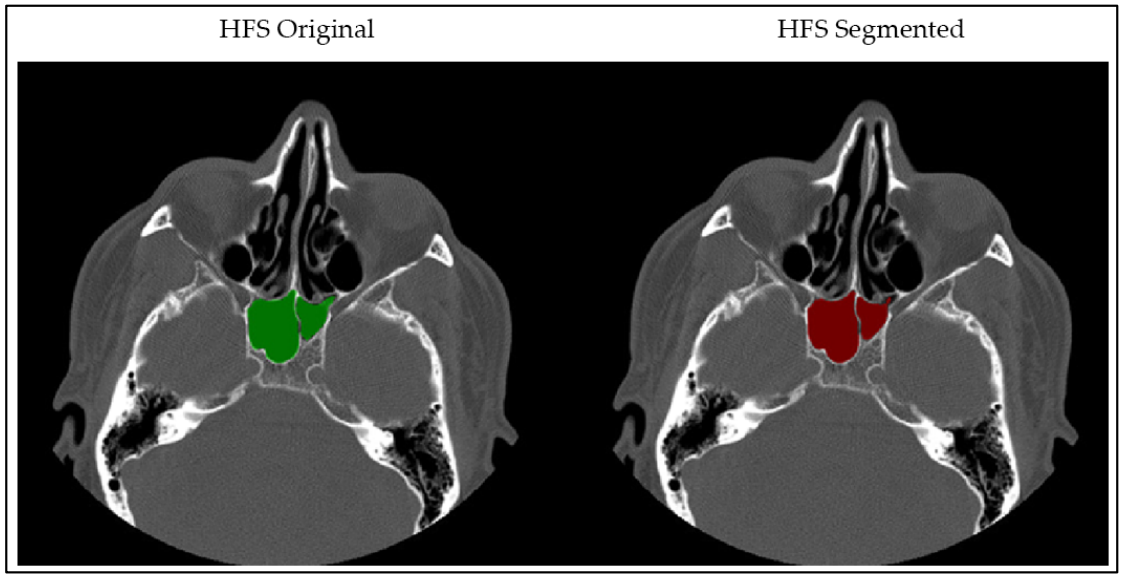

Figure 4. Example of sphenoid sinus segmentation using UNet++.

Overall, the comparative evaluation demonstrates that architectural depth and representational capacity significantly influence segmentation accuracy for anatomically complex structures. Although multiple architectures achieved comparable mean F1-scores under certain conditions, YOLO11L-seg exhibited superior cross-mode stability and contour consistency, justifying its selection for extended training and subsequent identification experiments. From the perspective of ablation analysis, these results confirm the necessity of selecting a segmentation backbone that combines high contour accuracy with cross-mode stability. Since the identification stage relies entirely on the geometric fidelity of predicted masks, the superior stability of YOLO11L-seg directly contributes to the reliability of the subsequent ranking-based matching procedure.

4.2. Extended Training of the Final Segmentation Model

Following the architectural comparison stage, YOLO11L-seg was selected as the final segmentation backbone and subjected to additional training on a refined subset of the dataset. The objective of this phase was to improve robustness, reduce generalization error, and increase stability of segmentation performance across anatomical variability and acquisition conditions. At this stage, the model was further refined using a controlled subset of the dataset, focusing on representative and challenging samples to improve robustness and generalization. For the HFS positioning mode, a subset of 1505 images was used for training and 431 images for validation. For the HFP positioning mode, 2107 images were used for training and 578 for validation. Importantly, the independent test subsets formed during the initial comparison stage were preserved unchanged in order to guarantee unbiased final evaluation. The optimization process remained identical to the initial stage: training was conducted for 150 epochs with a batch size of 32 and a fixed input resolution of 512 × 512 pixels. The learning rate schedule followed a cosine decay strategy. Let η_0 denote the initial learning rate and T the total number of epochs. The learning rate at epoch t was computed as:

$$\eta(t) = \eta_0 \times \frac{1 + \cos\left(\frac{\pi t}{T}\right)}{2} \tag{7}$$

This scheduling approach ensures smooth convergence and reduces oscillatory behavior near local minima. To evaluate the impact of extended training, we analyzed both mean performance metrics and their dispersion across slices. Let $F1_{before}$ and $F1_{after}$ denote the average F1-scores before and after extended training, respectively. The relative improvement Δ was computed as:

$$\Delta = \frac{F1_{after} - F1_{before}}{F1_{before}} \tag{8}$$

Although the absolute increase in F1-score was moderate, the primary effect of extended training was observed in reduced variance and improved contour stability, particularly in anatomically challenging slices. The reduction in performance variance across slices indicates improved generalization. Table 3 summarizes segmentation metrics of YOLO11L-seg before and after extended training.

Table 3. Segmentation performance of YOLO11L-seg before and after dataset expansion.

Positioning Mode	Stage	Mean F1
HFS	Initial training	0.93
HFS	Extended training	0.945
HFP	Initial training	0.83
HFP	Extended training	0.85

Although the mean F1-score values remained close to those observed in the initial stage, qualitative inspection demonstrated smoother boundaries and fewer disconnected mask fragments after extended training. This suggests that the additional data primarily improved structural coherence rather than raw pixel-level accuracy. Training dynamics were further analyzed by examining convergence curves. Let $L_{train}(t)$ and $L_{val}(t)$ denote the training and validation loss at epoch t. Stable convergence was characterized by monotonic decrease of $L_{val}(t)$ without divergence from $L_{train}(t)$. No evidence of overfitting was observed, as the validation loss plateaued without a subsequent increase in later epochs. Extended training was performed on a server equipped with an NVIDIA A100-SXM4-80GB GPU, which ensured stable gradient computation and allowed full-batch processing without memory bottlenecks. The computational environment contributed to

reproducibility and deterministic behavior during optimization. The final trained version of YOLO11L-seg was subsequently applied to all CT examinations in the database to generate the definitive segmentation masks used in identification experiments. By increasing the diversity of training samples and preserving identical test subsets, the extended training stage strengthened the reliability of the segmentation component without introducing evaluation bias. Figure 5 presents representative learning curves for the extended training stage, illustrating stable optimization behavior and convergence consistency.

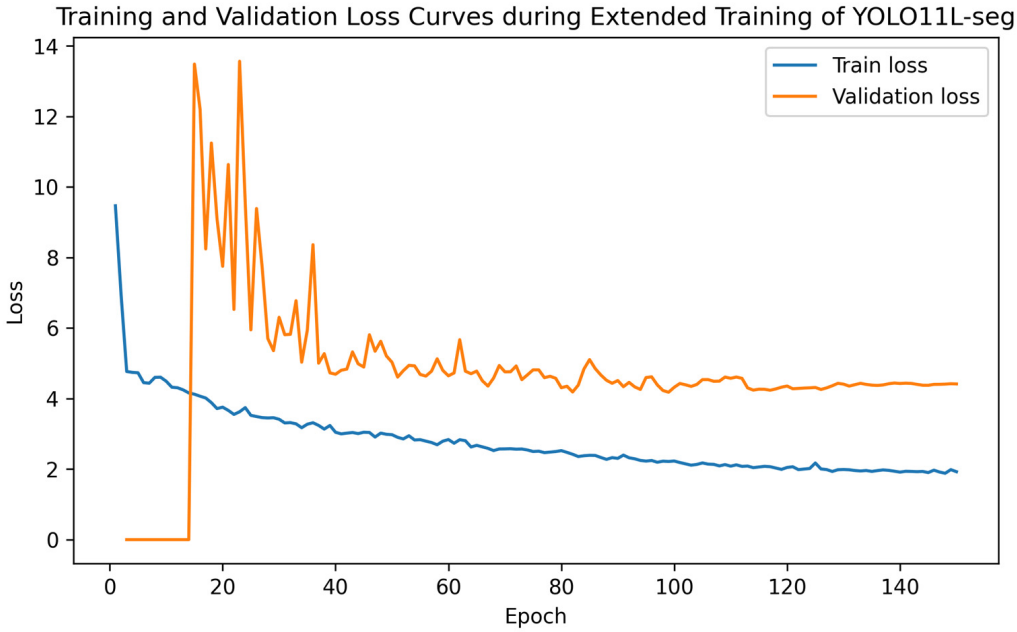

Figure 5. Training and validation loss curves during extended training of YOLO11L-seg.

4.3. Patient Identification Performance

After construction of the sphenoid sinus segmentation database and completion of extended model training, the patient identification experiment was conducted to evaluate the discriminative capacity and robustness of the proposed mask-based matching strategy. The experiment was designed to simulate repeated CT acquisitions of the same patient under realistic variability conditions, including moderate image perturbations within the sphenoid sinus region. The identification evaluation was performed separately for the HFS and HFP positioning modes. For HFS, 43 patients were selected from the database; for HFP, 44 patients were selected. For each original CT examination, ten modified variants were generated by introducing controlled artificial perturbations within the segmented sphenoid sinus region. These perturbations were intended to simulate realistic variations observed in follow-up examinations, including minor geometric distortions, local intensity changes, and noise artifacts. As a result, 430 modified examinations were generated for HFS and 440 for HFP. Each modified examination was treated as an independent query sample. The final YOLO11L-seg model was applied to generate segmentation masks for all relevant slices. The identification algorithm then compared each predicted mask A^i against all reference masks B_k^j stored in the database. Similarity between masks was quantified using the Intersection over Union metric:

$$IoU\left(A^i,\ B_k^j\right) = \frac{\left\{\left|A^i \cap B_k^j\right|\right\}}{\left\{\left|A^i \cup B_k^j\right|\right\}} \tag{9}$$

For each patient k in the reference database, the four highest IoU values obtained across all slice-to-slice comparisons were selected. The integrated similarity coefficient S_k was computed as the arithmetic mean of these four maximal values:

$$S_k = \frac{1}{4}\sum_{r=1}^{4} IoU_r^{(k)} \tag{10}$$

where $IoU_r^{(k)}$ denotes the r-th largest IoU value for patient k. This aggregation strategy emphasizes the most stable geometric correspondences and reduces the influence of isolated mismatches. Identification was considered successful if S_k exceeded the predefined threshold $\tau = 0.65$, and the true patient was ranked among the top five candidates returned by the system. The threshold value was determined empirically during preliminary experiments in order to balance identification sensitivity and the probability of false positive matches. The ranking was performed in descending order of S_k. Table 4 summarizes the quantitative results of the identification experiment.

Table 4. Patient identification results for modified CT examinations.

Orientation	Average IoU (%)	Correctly Identified Patients (%)
HFS	94.08	87.67
HFP	97.10	97.27

In addition to Top-5 identification accuracy, Rank-1 performance was evaluated to provide a stricter assessment of identification quality. Rank-1 accuracy reached 84% for HFS and 92% for HFP, indicating that in the majority of cases, the correct patient was ranked as the most similar candidate. These results demonstrate that the proposed method not only successfully retrieves the correct individual within the candidate set but also consistently assigns the highest similarity score to the true match. The reported IoU values in Table 4 correspond to similarity scores between masks of the same individual, which are expected to be higher than segmentation IoU measured against ground truth annotations. The results demonstrate high identification stability in both positioning modes. For HFP, the correct identification rate reached 97.27%, indicating strong geometric distinctiveness of the sphenoid sinus representation. In the HFS group, the identification rate was 87.67%, which, although lower than in HFP, still confirms the discriminative potential of the proposed method. The difference in performance between positioning modes can be explained by structural properties of the datasets rather than by limitations of the identification algorithm itself. In the HFP projection, examinations typically contained a larger number of informative slices that included the sphenoid sinus region. This facilitated a more reliable selection of the four highest IoU correspondences and increased stability of S_k. In contrast, HFS examinations often contained fewer informative slices, which constrained the aggregation step and slightly reduced final identification accuracy. To further analyze ranking behavior, the separation margin between the correct patient's similarity score and the highest competing score was evaluated. Let S_{true} denote the similarity coefficient of the true patient and $S_{\max_other}$ denote the maximum similarity coefficient among all other patients. The separation margin ΔS was defined as:

$$\Delta S = S_{true} - S_{\max_other} \tag{11}$$

To quantitatively assess ranking stability, summary statistics of the separation margin ΔS were computed for both positioning modes. The results are presented in Table 5.

Table 5. Statistical characteristics of the separation margin ΔS.

Metric	HFS	HFP
Mean ΔS	0.18	0.25
Std ΔS	0.07	0.05

The positive mean values of ΔS indicate consistent dominance of the correct patient in the ranking, while the relatively low standard deviation reflects stable separation between correct and competing candidates across different query samples. In the majority of evaluated cases, ΔS remained significantly above zero, demonstrating that the geometric descriptor derived from segmentation masks provides sufficient discriminative capacity for reliable patient matching. Figure 6 presents representative examples of identification outcomes, including ranking lists and corresponding mask comparisons for correctly matched cases.

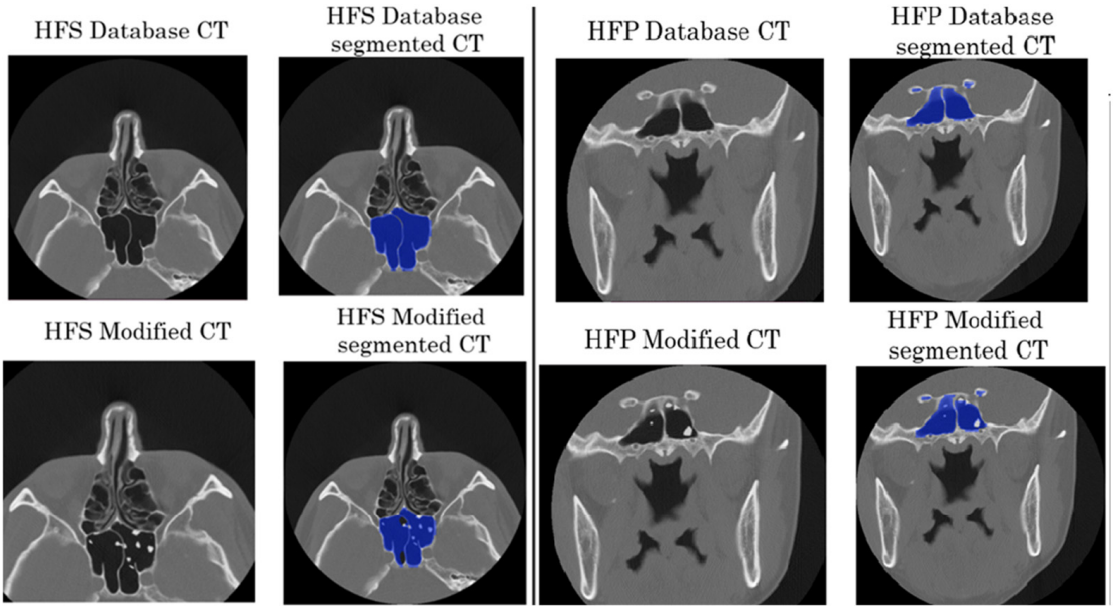

Figure 6. Representative identification results and ranking outputs for modified CT examinations.

Overall, the experimental findings confirm that sphenoid sinus geometry, when represented as segmentation masks and compared using IoU-based similarity aggregation, constitutes a stable and discriminative anatomical biometric suitable for automated patient identification in CT imaging environments. The key finding of this experiment is that high identification accuracy can be achieved using a lightweight two-dimensional mask-based representation without relying on three-dimensional reconstruction or intensity-based features. In contrast to existing approaches that employ volumetric matching or learned feature representations, the proposed method demonstrates that comparable identification performance can be obtained through direct geometric comparison of segmented anatomical structures, while maintaining lower computational complexity and improved robustness to acquisition variability.

4.4. Ablation Study and Sensitivity Analysis

To further verify the effectiveness of the core components of the proposed framework, an ablation study was conducted. The analysis focused on three key design choices: the segmentation backbone used to construct the reference mask database, the Top-K IoU aggregation strategy applied during identification, and the similarity threshold used to determine successful matches. These experiments were designed to quantify the individual

contribution of each component to the overall identification performance and to assess the stability of the proposed formulation with respect to variations in threshold selection, slice aggregation, and ranking behavior under perturbations.

4.4.1. Sensitivity to the Similarity Threshold

Identification success was defined by the condition $S_k \geq \tau$ and correct inclusion of the true patient within the top five ranked candidates. In the primary experiment, τ was fixed at 0.65. To assess threshold sensitivity, we analyzed identification performance for $\tau \in [0.60, 0.75]$ with step 0.05.

Let $Acc(\tau)$ denote identification accuracy as a function of the threshold. Formally,

$$Acc(\tau) = \frac{1}{N}\sum I(S_{true} \geq \tau) \tag{12}$$

where N is the total number of modified examinations and $I(\cdot)$ is the indicator function. The analysis demonstrated that accuracy remained stable within the interval $\tau \in [0.60, 0.70]$ for both positioning modes, with only minor degradation observed when τ exceeded 0.72 in the HFS subset. This indicates that the chosen threshold is not a fine-tuned parameter but lies within a plateau of stable performance. Table 6 summarizes identification accuracy for different threshold values.

Table 6. Identification accuracy (%) as a function of similarity threshold τ.

τ	HFS Accuracy (%)	HFP Accuracy (%)
0.60	88.10	96.8
0.65	87.67	97.27
0.70	86.20	96.10
0.75	82.30	94.20

The monotonic but moderate decline in $Acc(\tau)$ for larger τ confirms that the similarity distribution between true and non-matching patients is well separated. From an ablation perspective, these results demonstrate that the threshold value $\tau = 0.65$ provides the most appropriate operating point for the proposed system. Lower threshold values slightly increase sensitivity but also increase the risk of accepting false positive matches, whereas higher threshold values reduce false positives at the cost of missed correct identifications. Therefore, $\tau = 0.65$ was retained as the most balanced choice for reliable identification.

4.4.2. Influence of the Number of Informative Slices

One of the key structural differences between HFS and HFP datasets was the number of slices containing clearly visible sphenoid sinus regions. Let m denote the number of informative slices in a query examination. Since the similarity coefficient S_k is computed from the four largest IoU values, sufficient slice coverage is critical for stable aggregation. To analyze this dependency, we evaluated the correlation between identification success and m. The empirical probability of correct identification conditioned on m was estimated as:

$$P(correct|m) = \frac{N_{correct}(m)}{N_{total}(m)} \tag{13}$$

where $N_{correct}(m)$ denotes the number of correctly identified cases with m informative slices. The analysis revealed that identification probability increased with m and plateaued once $m \geq 6$–8 slices. In HFP examinations, the average number of informative slices was higher than in HFS, explaining the superior identification rate observed in Table 3. In cases where $m < 4$, the aggregation strategy becomes less robust because the four highest

IoU values are drawn from a limited candidate pool. To investigate the effect of the number of selected slices on identification performance, additional experiments were conducted using Top-K aggregation strategies, where $K \in \{3, 4, 5\}$. The results are presented in Table 7. The selection of $K = 4$ was empirically motivated by the need to balance robustness and discriminative power. Using fewer slices (Top-3) increases sensitivity to local segmentation errors and incomplete anatomical coverage, while including more slices (Top-5) introduces less informative comparisons, reducing the contribution of the most representative anatomical regions. For this experiment, a subset of 10 patients per positioning mode was used for controlled evaluation of aggregation strategies.

Table 7. Influence of the number of selected slices on patient identification performance.

Number of Selected Slices	Number of Patients	Variations per Patient	Orientation	Correctly Identified Patients (%)
3	10	10	HFS	85
3	10	10	HFP	91
4	10	10	HFS	88
4	10	10	HFP	97
5	10	10	HFS	71
5	10	10	HFP	82

When three slices were used, a decrease in identification accuracy was observed compared to the Top-4 configuration. This behavior can be explained by the limited amount of information contributing to the final similarity score. With fewer slices, the influence of local segmentation inaccuracies, partial sinus visualization, and slice-specific anatomical variability becomes more pronounced, leading to reduced stability of the aggregated similarity measure. The best performance was achieved when four slices were used. In this case, the similarity score is computed from a sufficient number of highly representative matches while avoiding the inclusion of less informative slices. This configuration provides an optimal balance between robustness and selectivity, ensuring that the most consistent geometric correspondences dominate the final decision. In contrast, increasing the number of selected slices to five resulted in a noticeable decrease in identification accuracy. This effect is associated with the inclusion of less reliable slice comparisons in the aggregation process. Such slices may contain partial anatomical structures, lower segmentation quality, or increased variability, which reduces the discriminative power of the similarity score and introduces additional noise into the ranking process. A consistent trend is observed across both positioning modes. However, the impact of slice selection is more pronounced in the HFS group, where the number of informative slices per study is typically lower. In contrast, the HFP group maintains higher accuracy across all configurations due to a larger pool of informative slices, increasing the probability of selecting stable and representative matches. This experiment also confirms that the aggregation strategy is not a secondary implementation detail, but a core component of the identification framework. The observed performance degradation for both Top-3 and Top-5 settings indicates that the proposed method is sensitive to the way slice-level similarities are summarized, and that the Top-4 configuration provides the most effective balance between selectivity and robustness.

4.4.3. Ranking Margin Stability Under Perturbations

To assess discriminative robustness, the separation margin ΔS between the true patient and the strongest competing candidate was analyzed statistically. Let ΔS_i denote the margin for the i-th query examination. The mean and standard deviation of ΔS were computed as:

$$\mu_\Delta = \frac{1}{n} \times \sum \Delta S_i \tag{14}$$

$$\sigma_{\Delta} = \sqrt{\frac{1}{N-1}\sum(\Delta S_i - \mu_{\Delta})^2} \quad (15)$$

In both positioning modes, μ_{Δ} remained positive and substantially greater than zero, indicating consistent dominance of the correct match. The standard deviation σ_{Δ} was lower in HFP than in HFS, reflecting greater ranking stability when more informative slices were available. Additionally, perturbation robustness was evaluated by comparing similarity coefficients before and after artificial modifications. Let $S_{original}$ and $S_{modified}$ denote similarity coefficients computed using original and perturbed examinations, respectively. The relative deviation was defined as:

$$R = \frac{\left|S_{modified} - S_{original}\right|}{S_{original}} \quad (16)$$

The average R remained low across both positioning modes, confirming that moderate image perturbations do not significantly alter geometric matching outcomes. Overall, the robustness analysis demonstrates that the proposed identification framework exhibits stable behavior with respect to threshold variation, slice coverage differences, and moderate structural perturbations. The system maintains consistent ranking dominance of the true patient and shows predictable degradation only under constrained slice availability conditions. These findings further support the suitability of sphenoid sinus geometry as a reliable anatomical biometric for CT-based patient identification.

Taken together, the experiments presented in this section provide an ablation analysis of the proposed framework. The results demonstrate that the identification performance is jointly determined by three key components. First, the segmentation backbone directly affects the geometric fidelity of the masks, which in turn defines the discriminative power of the identification stage. Models with lower segmentation stability lead to reduced identification accuracy. Second, the aggregation strategy plays a critical role. As shown in Table 7, the Top-4 configuration provides the best balance between robustness and selectivity, while both smaller and larger values of K lead to degraded performance. Third, the similarity threshold defines the operating regime of the system. As demonstrated in Table 6, the selected value τ = 0.65 ensures a stable trade-off between sensitivity and robustness. These results confirm that the proposed identification framework is internally consistent and that each design component contributes significantly to the overall system performance.

5. Discussion

The results support the hypothesis that the geometry of the sphenoid sinus can serve as a stable, individualized anatomical feature suitable for patient identification based on CT images. However, these findings should be interpreted as a proof-of-concept demonstrating feasibility under controlled experimental conditions rather than as definitive validation of a generalized identification system. Although high identification accuracies were achieved, these results should not be interpreted as establishing a universal benchmark for future studies. The reported performance is specific to the experimental conditions, including dataset composition, acquisition protocols, and the selected anatomical structure. Instead, the results should be considered as evidence of feasibility and as an indication of the potential of sphenoid sinus morphology as an anatomical biometric. Further validation on independent and larger-scale datasets is required before defining standardized performance references. Unlike intensity characteristics, which can vary depending on reconstruction parameters, noise level, or contrast, the shape and spatial arrangement of bone structures have significantly higher morphological stability within an individual. It is this property that underlies the concept of using anatomy as a biometric marker. The segmentation

stage demonstrated that modern deep learning architectures can reproduce the complex geometry of the sphenoid sinus with high accuracy. Not only is the average segmentation accuracy important, but also its stability between individual slices. After extended training of the model, more consistent contour reproduction was observed in cases of complex anatomical structure and variable pneumatization. The reduction in quality fluctuations between slices indicates an improvement in the model's generalization ability and its resistance to anatomical variations. The identification algorithm, based on aggregating the largest mask overlap values, implements a strategy of partial alignment of geometric features. This approach reduces the impact of local inconsistencies between individual slices and focuses on the most representative and stable areas of the anatomical structure. In practical terms, this means that the system does not depend on individual segmentation inaccuracies or minor differences between slices, but forms an integral assessment based on the most reliable matches. From the perspective of biometric systems, the results indicate a consistent separation between the same patient and other individuals. Direct comparison with prior work is limited due to differences in problem formulation. While most existing methods focus on classification or intensity-based matching, the proposed approach is based on ranking-driven geometric comparison of anatomical structures, which makes direct metric-level comparison not fully applicable. An important advantage of the proposed approach is its methodological simplicity compared with commonly used workflows based on full three-dimensional reconstruction of anatomical structures. In many forensic and radiological studies, reconstruction of paranasal sinuses is performed using specialized software such as Materialise Mimics 20.0, which requires manual interaction and significant computational resources. In contrast, the proposed method operates directly on two-dimensional segmentation masks and relies on simple geometric similarity metrics, which substantially reduces computational complexity while preserving the anatomical information necessary for identification. At the same time, it should be acknowledged that three-dimensional approaches provide more complete spatial representations of anatomical structures and have demonstrated strong performance in forensic identification tasks. The choice between 2D and 3D representations therefore reflects a trade-off between computational efficiency and geometric completeness. The proposed approach differs from existing deep learning-based methods by focusing on explicit anatomical structures rather than learned image features. This also makes the proposed framework easier to reproduce and more suitable for deployment in routine identification workflows than methods requiring full 3D reconstruction. In most cases, the similarity coefficient for the correct patient significantly exceeded the corresponding values for other candidates. This confirms that the geometry of the sphenoid sinus has sufficient discriminatory power for use in automated identification systems. An important observation is the difference between the HFS and HFP modes. The lower level of identification in the HFS group is not related to algorithmic limitations, but to the characteristics of the source data. In HFP mode, studies typically contained a larger number of informative slices with clear visualization of the sphenoid sinus, which increased the stability of the integral similarity assessment. In contrast, in HFS, the number of such slices was often smaller, which limited the ability to select the most representative matches and slightly reduced the overall accuracy. From a clinical point of view, the proposed approach can be considered as an additional level of verification of patient matching in medical image storage and exchange systems. The method does not require the use of personal data or external biometric parameters, as it is based solely on anatomical features. This opens up opportunities for integration into PACSs as an automatic mechanism for detecting potential labeling errors or mixing of studies from different patients. However, the study has several important limitations that should be considered when interpreting the results. First, the evaluation was performed

using synthetic perturbations of CT scans rather than real repeated examinations, which may not fully capture the variability observed in clinical practice. Second, although HFS and HFP scans were treated as distinct acquisition conditions, they originate from the same anatomical structures and therefore do not represent fully independent identity samples. Third, the annotation process, while carefully controlled, remains subject to inherent variability associated with manual delineation of complex anatomical boundaries. Fourth, the identification framework relies on a heuristic Top-4 aggregation strategy and an empirically selected similarity threshold, which, although validated experimentally, are not derived from a formal optimization procedure. Finally, the study was conducted on a controlled dataset with consistent acquisition conditions, and therefore further validation on multi-center and more heterogeneous datasets is required to assess generalizability. A promising direction for further research is the combination of several bone structures into a single combined anatomical descriptor, which could potentially increase the reliability of the system. In summary, the results of the study demonstrate that the sphenoid sinus has the properties of a stable morphological marker suitable for automated identification. The combination of deep segmentation and geometric comparison of masks ensures resistance to moderate changes in scanning conditions and confirms the validity of the concept of "anatomy as biometrics" in the context of computed tomography. The proposed approach introduces a new direction in medical image analysis by treating anatomical structures as intrinsic biometric identifiers.

6. Conclusions

This study investigates the feasibility of automated patient identification based on sphenoid sinus geometry extracted from CT images using deep learning-based segmentation. The experimental results demonstrate that a lightweight two-dimensional representation of anatomical structures, combined with a ranking-based matching strategy, can achieve high identification accuracy under controlled conditions. The proposed framework formulates identification as a database-level ranking problem and relies on direct geometric comparison of segmentation masks. This approach eliminates the need for intensity-based features, handcrafted descriptors, or computationally intensive three-dimensional reconstruction, while preserving the anatomical information necessary for reliable matching. The obtained results indicate that the sphenoid sinus possesses sufficient inter-individual variability and intra-individual stability to serve as a promising anatomical biometric. At the same time, the findings should be interpreted as a proof-of-concept rather than a fully validated identification system. The evaluation was conducted on a controlled dataset and included synthetic perturbations of CT examinations rather than real repeated scans. In addition, the proposed identification framework relies on heuristic design choices, including the Top-4 aggregation strategy and an empirically selected similarity threshold, which may require further refinement and formal optimization. From a practical perspective, the proposed approach demonstrates potential applicability as an auxiliary verification mechanism in medical imaging systems, particularly in scenarios where traditional metadata-based identification is unreliable or unavailable. The method may also have relevance in forensic contexts, where anatomical structures visible in CT scans can serve as alternative identifiers when conventional biometric data are missing. Future research should focus on validation using larger and more heterogeneous datasets, including multi-center data and real longitudinal examinations. Further improvements may be achieved by integrating multiple anatomical structures into a combined descriptor, enhancing robustness to acquisition variability, and exploring alternative aggregation strategies for similarity estimation. Overall, the results support the concept of "anatomy as biometrics" and demonstrate that segmentation-derived geometric representations can form the basis for automated iden-

tification systems in CT imaging, while highlighting the need for further validation and methodological refinement.

Author Contributions: Conceptualization, N.B. and V.M.; methodology, N.B.; software, V.M. and D.T.; validation, D.T., and V.M.; formal analysis, N.B. and M.F.; investigation, D.T., N.B. and V.M.; resources, M.F.; data curation, D.T. and N.B.; writing—original draft preparation, N.B. and V.M.; writing—review and editing, M.F.; visualization, V.M. and D.T.; supervision, N.B.; project administration, N.B.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was performed in the project "BIKO-UA" (Bilderkennung zur Identifikation von Kriegsopfern in der Ukraine/Image Recognition for the Identification of War Victims in Ukraine) [ID: 03DPS1241] supported in the program DATIpilot Innovationssprints by the (former) German Ministry of Education and Research.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the use of fully anonymized retrospective data. The original data collection was conducted within an academic research project at Kharkiv National Medical University in accordance with institutional procedures.

Informed Consent Statement: Patient consent was obtained in the original study. The present study uses only fully anonymized data and does not involve direct interaction with human subjects.

Data Availability Statement: The data used in this study are not publicly available due to privacy and ethical restrictions, but may be available from the corresponding author upon reasonable request and with permission from the data provider.

Acknowledgments: The work was a collaboration between Molecular Biotechnology and Functional Genomics, TH Wildau, and the research laboratory "Information Technologies in Learning and Computer Vision Systems", NURE. The authors are grateful to Viktoriia Alieksieieva for providing the CT raw data and fundamental training in labeling and annotation of the anatomical structures. During the preparation of this work, the authors used Grammarly (version 1.2.194) to improve the language of the manuscript. After using this tool, the authors carefully reviewed and edited the text and take full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goujon, A.; Neumayer, S.; Unser, M. Learning Weakly Convex Regularizers for Convergent Image-Reconstruction Algorithms. *SIAM J. Imaging Sci.* **2024**, *17*, 91–115. [CrossRef]
2. Auffret, M.; Garetier, M.; Diallo, I.; Aho, S.; Ben Salem, D. Contribution of the Computed Tomography of the Anatomical Aspects of the Sphenoid Sinuses to Forensic Identification. *J. Neuroradiol.* **2016**, *43*, 404–414. [CrossRef]
3. Dong, X.; Fan, F.; Wu, W.; Wen, H.; Chen, H.; Zhang, K.; Zhang, J.; Deng, Z. Forensic Identification from Three-Dimensional Sphenoid Sinus Images Using the Iterative Closest Point Algorithm. *J. Digit. Imaging* **2022**, *35*, 1034–1040. [CrossRef] [PubMed]
4. Alsalama, A.; Harous, S.; Elnagar, A. Paranasal Sinus Analysis Based on Deep Learning and Machine Learning Techniques: A Comprehensive Survey. *Intell. Syst. Appl.* **2025**, *27*, 200559. [CrossRef]
5. Ahmed, J.; Namrata; Sujir, N.; Shenoy, N.; Natarajan, S.; Muralidharan, A.; Shetty, A.C. A Comparative Analysis of Sphenoid and Frontal Sinuses Using Cone Beam Computed Tomography for Sex Determination. *J. Oral Biol. Craniofac. Res.* **2024**, *14*, 478–483. [CrossRef] [PubMed]
6. Koç, A. Are maxillary and sphenoid sinus volumes predictors of gender and age? A cone beam computed tomography study. *Cumhur. Dent. J.* **2020**, *23*, 348–355. [CrossRef]
7. Sivasamy, I.; Ramakarishna, P.; Johaley, S.; Sikdar, S.D.; Jain, S.K.; Diwan, R.K. Evaluation of Maxillary and Sphenoidal Sinuses' Volume and Bizygomatic Width Using Cone Beam Computed Tomography—A Promising Tool for Sex Determination. *J. Indian Acad. Oral Med. Radiol.* **2024**, *36*, 297–300. [CrossRef] [PubMed]
8. Bilous, N.; Komarov, O. Application of Deep Learning Methods for Brain Tumor Segmentation in MRI Images. In Proceedings of the 7th International Scientific and Technical Conference "Information Systems and Technologies" (IST-2018); Information Systems and Technologies Conference, Kharkiv, Ukraine, 10–15 September 2018; pp. 438–442.

9. Novikov, R.K. Investigation of CT Image Segmentation Methods for Human Identification. In Proceedings of the VII International Student Scientific Conference "Current Issues and Prospects for Conducting Scientific Research", Orléans, France, 19–20 November 2024; pp. 79–81.
10. Yuan, Y.; Cheng, Y.; Pan, B.; Jin, G.; Yu, D.; Ye, M.; Zhang, Q. A Multi-Modal Attention Fusion Framework for Road Connectivity Enhancement in Remote Sensing Imagery. *Mathematics* **2025**, *13*, 3266. [CrossRef]
11. Cellina, M.; Gibelli, D.; Cappella, A.; Toluian, T.; Pittino, C.V.; Carlo, M.; Oliva, G. Segmentation Procedures for the Assessment of Paranasal Sinuses Volumes. *Neuroradiol. J.* **2021**, *34*, 13–20. [CrossRef] [PubMed]
12. Heinrich, A. Automatic Personal Identification Using a Single CT Image. *Eur. Radiol.* **2024**, *35*, 2422–2433. [CrossRef] [PubMed]
13. Hramm, O.; Bilous, N.; Ahekian, I. Configurable Cell Segmentation Solution Using Hough Circles Transform and Watershed Algorithm. In *Proceedings of the 2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL)*; IEEE: Sozopol, Bulgaria, 2019; pp. 602–605.
14. Bilous, N.; Malko, V.; Frohme, M.; Nechyporenko, A. Comparison of CNN-Based Architectures for Detection of Different Object Classes. *AI* **2024**, *5*, 2300–2320. [CrossRef]
15. Whangbo, J.; Lee, J.; Kim, Y.J.; Kim, S.T.; Kim, K.G. Deep Learning-Based Multi-Class Segmentation of the Paranasal Sinuses of Sinusitis Patients Based on Computed Tomographic Images. *Sensors* **2024**, *24*, 1933. [CrossRef] [PubMed]
16. Song, D.; Yang, S.; Han, J.Y.; Kim, K.G.; Kim, S.T.; Yi, W.-J. Comparison of Segmentation Performance of Cnns, Vision Transformers, and Hybrid Networks for Paranasal Sinuses with Sinusitis on CT Images. *Sci. Rep.* **2025**, *15*, 32087. [CrossRef] [PubMed]
17. Jocher, G.; Chaurasia, A.; Qiu, J. *Ultralytics YOLO*, version v8; Ultralytics: Frederick, MD, USA, 2023.
18. Jocher, G.; Qiu, J. *Ultralytics YOLO*, version 11; Ultralytics: Frederick, MD, USA, 2024.
19. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *arXiv* **2018**, arXiv:1807.10165.
20. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* **2018**, arXiv:1802.02611. [CrossRef]
21. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 3349–3364. [CrossRef] [PubMed]
22. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *arXiv* **2021**, arXiv:2105.15203. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.