

Article

The Sphenoid Sinus as a Biometric Marker: AI-Based Automated Segmentation in CT Imaging for Forensic Identification

Victoriia Alekseeva ^{1,2,3,4,*}, Marcus Krüger ¹, Florian Zwicker ¹, Tom Graner ¹, Vlad Krasnikov ⁵, Parsa Lavasanifar ¹, Marcus Frohme ¹, Vitaliy Gargin ^{3,6,7} and Alina Nechyporenko ^{1,5}

- ¹ The Department of Molecular Biotechnology and Functional Genomics, Technical University of Applied Science (TH Wildau), 15745 Wildau, Germany; marcus.krueger@th-wildau.de (M.K.); florian_gabriel.zwicker@th-wildau.de (F.Z.); tom.graner@th-wildau.de (T.G.); pala4314@th-wildau.de (P.L.); marcus.frohme@th-wildau.de (M.F.); nechyporenko@th-wildau.de (A.N.)
- ² Department of Otorhinolaryngology, Kharkiv National Medical University, 61022 Kharkiv, Ukraine
- ³ Department of Professionally Oriented Disciplines, Kharkiv International Medical University, 61001 Kharkiv, Ukraine; vitgarg@ukr.net
- ⁴ Department of Clinical Disciplines, University of Medicine and Social Sciences, 61002 Kharkiv, Ukraine
- ⁵ Department of Computer Modeling and Intelligent Technologies, Kharkiv National University of Radio Electronics, 61023 Kharkiv, Ukraine; krasnikov.vlad.v@gmail.com
- ⁶ Department of Pathological Anatomy, Kharkiv National Medical University, 61022 Kharkiv, Ukraine
- ⁷ Department of Dentistry, Medical Faculty, V.N. Karazin Kharkiv National University, 61022 Kharkiv, Ukraine
- * Correspondence: vik13052130@gmail.com

Abstract

Reliable personal identification remains a major challenge in forensic medicine, particularly in cases involving decomposition, thermal injury, or absence of DNA and dental records. In this context, anatomically protected and morphologically unique structures such as the sphenoid sinus may serve as valuable biometric markers. The aim of this study was to evaluate the forensic applicability of sphenoid sinus morphology using computed tomography (CT), automated segmentation, and three-dimensional (3D) computational analysis. The proposed framework integrates CT-based image preprocessing, automated 3D segmentation using nnU-Net architecture, and geometric comparison of reconstructed sphenoid sinus models through point cloud analysis and deep learning approaches. Morphological variability, spatial configuration, and structural stability of the sphenoid sinus were analyzed as discriminative biometric features. The GSA-Net-based identification model demonstrated high recognition performance, achieving Top-1 accuracies exceeding 97.8% and Top-3 accuracies up to 100% in controlled datasets. The results support the concept that the sphenoid sinus possesses sufficient individuality and anatomical preservation to enable reliable ante-mortem and post-mortem identification. The study highlights the potential of integrating automated segmentation and AI-driven 3D analysis into forensic workflows and emphasizes the importance of standardized imaging protocols and larger annotated datasets for future clinical and forensic implementation.



Academic Editor: Manohar Das

Received: 15 May 2026

Revised: 7 August 2026

Accepted: 19 August 2026

Published: 25 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: sphenoid sinus; forensic identification; computed tomography; 3D reconstruction; nnU-Net; artificial intelligence; point cloud analysis; biometric marker

1. Introduction

Reliable personal identification remains a fundamental task in forensic medicine [1], particularly in situations where conventional identification methods, such as fingerprint analysis, dental comparison, or DNA profiling, cannot be applied. This typically occurs

in cases involving advanced decomposition, severe thermal damage, or the absence of antemortem reference data. Under these conditions, increasing attention has been directed toward anatomical structures that are resistant to external influences while exhibiting sufficient individual variability to support forensic identification [2].

Among the anatomical structures investigated for this purpose, the paranasal sinuses have attracted considerable interest, with the frontal sinus being the most extensively studied [3]. Previous investigations by Uthman et al. [4], Belaldavar et al. [5], and Souza et al. [6] demonstrated that frontal sinus morphology provides reliable individualization because of its high anatomical variability. However, the forensic applicability of the frontal sinus is limited by its absence in a proportion of individuals and by morphological alterations resulting from trauma or surgical interventions.

In contrast, the sphenoid sinus offers several advantages as a potential anatomical biometric while remaining comparatively understudied. Located within the central skull base, it is relatively protected from environmental influences and mechanical damage [7]. At the same time, it exhibits substantial interindividual variability in volume, shape, internal septation, pneumatization, and overall morphology [8,9]. These anatomical characteristics remain largely stable after skeletal maturation, making the sphenoid sinus a promising individualized marker for forensic identification [10].

Computed tomography (CT) plays a central role in the evaluation of paranasal sinus anatomy in forensic practice [11]. Multidetector CT (MDCT) remains the reference imaging modality because it provides high-resolution isotropic volumetric datasets suitable for accurate morphometric analysis and three-dimensional reconstruction of osseous structures [12]. Cone-beam CT (CBCT) is widely used in dental and maxillofacial imaging because of its lower radiation dose and high spatial resolution for bone; however, increased image noise and reduced soft-tissue contrast may limit segmentation accuracy [13]. In specialized forensic applications, postmortem CT (PMCT) and micro-computed tomography provide highly detailed visualization of anatomical structures, particularly in cadaveric investigations [14]. Magnetic resonance imaging (MRI) can complement CT by providing superior soft-tissue contrast for evaluating surrounding anatomical relationships and pathological changes, although it is less suitable for cortical bone assessment [15]. Nevertheless, differences in voxel size, reconstruction algorithms, and intensity calibration among imaging systems introduce variability across datasets, highlighting the need for standardized preprocessing procedures.

Various methods have been proposed for sphenoid sinus segmentation and morphometric analysis. Manual segmentation remains the anatomical reference standard because of its high accuracy but is labor-intensive and strongly dependent on expert experience [16]. Semi-automated software platforms, including Materialise Mimics [17] and 3D Slicer [18], accelerate segmentation using region-growing and threshold-based algorithms, although manual refinement is frequently required in anatomically complex regions.

Recent advances in artificial intelligence have substantially improved automated sphenoid sinus segmentation and identification. Bilous et al. [19] presented an automated CT-based identification approach using sphenoid sinus geometry as an anatomical biometric. Compared with two-dimensional slice-based analysis, three-dimensional approaches preserve the complete anatomical complexity of the sphenoid sinus, including its spatial configuration, depth, curvature, and internal structures, thereby providing more discriminative features for person identification. Fully automated reconstruction methods based on fuzzy c-means clustering, mathematical morphology, and stacked convolutional autoencoders have further advanced this field [20]. Deep learning architectures, including U-Net [16] and nnU-Net [21], have demonstrated high segmentation accuracy. In addition, our previous work explored the potential of AI-based approaches for sphenoid sinus

identification [22] and nnU-Net [23,24], enabling highly accurate automated segmentation, commonly evaluated using the Dice similarity coefficient [25]. Furthermore, attention-based neural networks operating on three-dimensional point clouds have demonstrated excellent identification performance under controlled conditions [26]. However, the performance of these methods remains highly dependent on the availability of high-quality annotated datasets and standardized imaging protocols.

Although previous studies have demonstrated the feasibility of sphenoid sinus-based identification, several important limitations remain. These include the lack of large publicly available annotated datasets, variability in imaging protocols between scanners, insufficient validation using postmortem or degraded datasets, and the tendency of existing studies to address isolated stages of the workflow rather than integrated end-to-end forensic identification systems. Addressing these challenges is essential for the reliable implementation of automated sphenoid sinus-based identification in routine forensic practice.

The present study aims to overcome these limitations by developing a unified framework for identification based on sphenoid sinus morphology derived from CT imaging. The proposed pipeline integrates automated 3D segmentation using nnU-Net with point cloud-based comparison of geometrical structures to enable robust ante-mortem to post-mortem person identification based on the stable and highly individualized morphology of the sphenoid sinus.

By integrating anatomical analysis with modern computational techniques, this work aims to establish a foundation for scalable and objective forensic identification systems based on deep cranial structures.

2. Materials and Methods

To enable the identification of a person using the sphenoid sinus structure, the following pipeline (Figure 1) was designed:

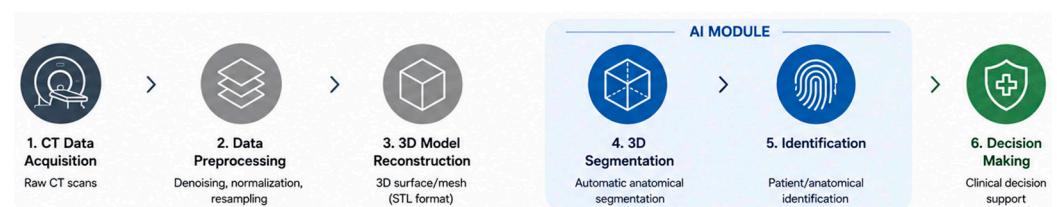


Figure 1. Forensic Identification Pipeline.

2.1. Study Design and Ethical Considerations

This study was conducted as a retrospective analysis of anonymized cranial computed tomography (CT) datasets. All procedures involving human data were performed in accordance with the ethical standards of the institutional and national research committees and with the principles of the Declaration of Helsinki (2013 revision).

The study was approved by the Bioethics Committee of Kharkiv National Medical University (Committee Protocol No. 5, 11 November 2018) and by the Academic Council of Kharkiv National Medical University (Protocol No. 3, 25 January 2019). The study involved a retrospective analysis of existing CT scans, with no additional interventions, and all data were fully anonymized.

Due to the retrospective nature of the study and the use of fully anonymized imaging data, the requirement for individual informed consent was waived where applicable. For datasets obtained from collaborating institutions (Municipal Non-profit Enterprise “City Clinical Hospital for Emergency Medical Services named after prof. O.I. Meschaninov” of Kharkiv City Council and Merefian Region Hospital), data usage agreements were established, ensuring compliance with local data protection regulations and ethical standards.

All CT scans were irreversibly anonymized prior to analysis, and no personally identifiable information was accessible to the researchers at any stage of the study.

2.2. Dataset Composition

The study population consisted of 477 adult subjects (238 males and 239 females) aged 18–95 years who underwent clinically indicated imaging of the paranasal sinuses at Kharkiv National Medical University and affiliated clinical institutions. Participants were distributed across the following age groups: 18–44 years ($n = 138$), 45–59 years ($n = 147$), 60–74 years ($n = 117$), 75–89 years ($n = 45$), and ≥ 90 years ($n = 30$). For the MDCT examinations of the paranasal sinuses, a standardized protocol was used with a slice thickness of 0.5–2.0 mm, tube voltage of 120 kVp, and automatic tube current modulation (approximately 100–250 mA), reconstructed using a high-resolution bone algorithm with tube current modulation (approximately 100–250 mA), reconstructed using a high-resolution bone algorithm with a standard 512×512 matrix size. For the CBCT examinations, the field of view was 13×16 cm with isotropic voxels (approximately 0.2–0.4 mm, depending on the acquisition protocol), optimized for high-resolution visualization of the paranasal sinuses and craniofacial skeleton.

The resulting masks were stored as binary volumes (foreground = 1, background = 0) and exported in NIfTI (.nii.gz) format while strictly preserving the original spacing, origin, and orientation metadata of the CT images. Consequently, the spatial resolution of each generated mask was completely identical (1:1) to the native resolution of its corresponding CT scan (e.g., isotropic voxels of 0.2–0.4 mm for CBCT and slice thickness of 0.5–2.0 mm for MDCT). This exact spatial alignment enabled direct compatibility and full-resolution processing within downstream deep learning frameworks such as nnU-Net.

All examinations included complete visualization of the sphenoid sinus and met the predefined image quality criteria. The datasets used for segmentation, identification, validation, and testing are summarized in Table 1.

Table 1. Dataset distribution across experimental tasks.

Task	Dataset	Number of Cases
Overall dataset	Cranial CT scans	477
Segmentation (nnU-Net)	Total curated dataset	384
	Training	224
	Validation	56
	Internal test	68
	Independent held-out test	36
Identification (GSA-Net)	Training	428 models
	Test	45 models from 42 individuals
Temporal stability assessment	Repeated CT examinations	2 patients (4 CT scans) *

* **Note:** In addition to the primary identification experiments, repeated CT examinations from two patients (two scans per patient, acquired approximately two years apart) were analyzed to provide a preliminary assessment of intra-individual temporal stability.

Both imaging protocols were routinely used for clinical assessment of the paranasal sinuses, providing sufficient spatial resolution for accurate three-dimensional segmentation and biometric analysis of the sphenoid sinus.

The example of CT is presented in Figure 2. The scans were characterized by heterogeneous acquisition parameters, including differences in slice thickness, image resolution, scanner type, and patient positioning.

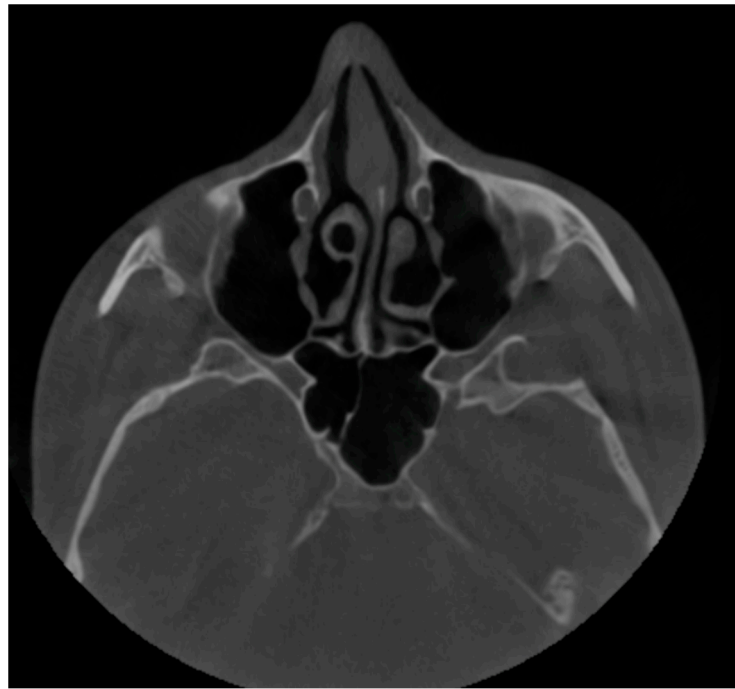


Figure 2. Example of a CT from the dataset.

Therefore, the dataset included CT scans of 473 individuals, representing a wide range of anatomical variability, imaging conditions, and image acquisition geometry, ensuring the reliability and generalizability of the developed methodology. In addition, the following inclusion and exclusion criteria were applied.

Inclusion Criteria: CT scans of the paranasal sinuses; CBCT scans including the sinus region; Scans acquired from both modern and older-generation CT systems; images with varying spatial resolution and reconstruction parameters. **Exclusion Criteria:** Significant structural alteration of the sinuses, including extensive destruction of sinus walls; Severe postoperative changes or large surgical defects; Tumors or mass lesions deforming sinus anatomy; Complete or near-complete obliteration of the sinus cavity; Mycetoma or other space-occupying lesions; Extreme inflammatory changes with loss of anatomical landmarks; Mild to moderate mucosal thickening affecting anatomical assessment.

2.3. Data Harmonization and Preprocessing

One of the key technical challenges was the variability of CT data. Differences in acquisition settings lead to inconsistencies in image intensity and spatial resolution, which could adversely affect both segmentation and downstream analysis. To ensure comparability across heterogeneous CT datasets, as well as compatibility with deep learning-based segmentation models, a standardized preprocessing pipeline was developed. Slices lacking valid spatial position information (`ImagePositionPatient`) or showing inconsistent slice thickness were excluded from further processing. These slices typically corresponded to cranial regions outside the anatomical area of interest, and their removal prevented geometric inconsistencies during volumetric reconstruction. Intensity normalization was performed using Otsu-based foreground detection to exclude surrounding air from the intensity statistics. Subsequently, voxel intensities were normalized using percentile clipping (2nd–98th percentile of the foreground voxels), reducing scanner-dependent grayscale variability while limiting the influence of extreme intensity values and preserving relevant anatomical structures. In addition, all datasets were spatially standardized by harmonizing image origin and orientation, ensuring consistency between CT volumes and corresponding segmentation masks. Voxel spacing in the z-direction was recalculated from the `ImagePosi-`

tionPatient coordinates rather than relying solely on the SliceThickness attribute, thereby preserving the true spatial relationship between adjacent slices. Finally, the processed data were exported in both DICOM and NIfTI formats [27,28], enabling clinical compatibility and efficient use in deep learning frameworks such as nnU-Net.

To reduce scanner-dependent variability arising from the use of different CT platforms, all image volumes were subjected to a unified preprocessing workflow, including voxel spacing harmonization, image orientation standardization, intensity normalization, and conversion to a common NIfTI representation before training and evaluation. This preprocessing minimized technical differences between MDCT and CBCT datasets while preserving anatomically relevant information required for automated segmentation.

2.3.1. Segmentation and 3D Modelling Workflow

Semi-automatic segmentation of the sphenoid sinus was evaluated using two software environments, 3D Slicer v.5.8.1 and Mimics (Materialise) v.27.0, Leuven, Belgium [17,18], to generate reference ground truth data for model training. Both platforms were capable of producing anatomically plausible segmentations; however, notable differences in workflow efficiency, usability, and practical segmentation quality were observed. While 3D Slicer required more manual initialization and iterative refinement, Mimics enabled a faster and more robust workflow through its Dynamic Region Growing functionality, allowing more consistent separation of the sphenoid sinus from adjacent anatomical structures. In contrast, it should also be noted that Mimics is commercial software, whereas 3D Slicer is open-source, making it more widely available for public use.

Based on this comparative evaluation, Mimics was selected as the primary annotation tool for dataset generation, as it proved more suitable for efficient large-scale creation of training data. Reference segmentations were generated following a standardized workflow in Mimics. The annotation process involved an experienced physician and a biomedical engineer, while challenging cases were reviewed together with the supervising physician and manually refined where necessary. Accordingly, the final reference masks used in this study were generated in Mimics by importing the DICOM datasets and applying Dynamic Region Growing with image-dependent sensitivity settings, shown in Figure 3. Initial segmentations were subsequently reviewed slice-by-slice in the axial, coronal, and sagittal planes. Whenever leakage into adjacent anatomical structures or incomplete delineation of the sphenoid sinus was observed, manual corrections were performed using the editing and mask-splitting tools. Additional functions such as cavity filling were applied where appropriate to improve mask completeness while preserving the anatomical boundaries of the sphenoid sinus. After manual refinement, the binary segmentation masks were converted into three-dimensional surface models using the standard surface reconstruction functionality provided by Mimics and subsequently exported as STL files for further processing. No additional mesh smoothing, remeshing, or geometric post-processing was applied prior to STL export. Following segmentation, every case underwent a final quality-control procedure consisting of visual inspection in axial, coronal, and sagittal views as well as the corresponding 3D reconstruction. Segmentations were assessed for anatomical plausibility, continuity of the sphenoid sinus cavity, exclusion of adjacent anatomical structures, and complete coverage of the sinus. Masks showing leakage, discontinuities, or missing regions were manually corrected before export for downstream processing. Additional options such as cavity filling were used to improve mask completeness. Where leakage into adjacent regions occurred, manual refinement and mask-splitting tools were applied (Figure 4). Final segmentations were visually verified in multiplanar views and 3D rendering before export for downstream processing.

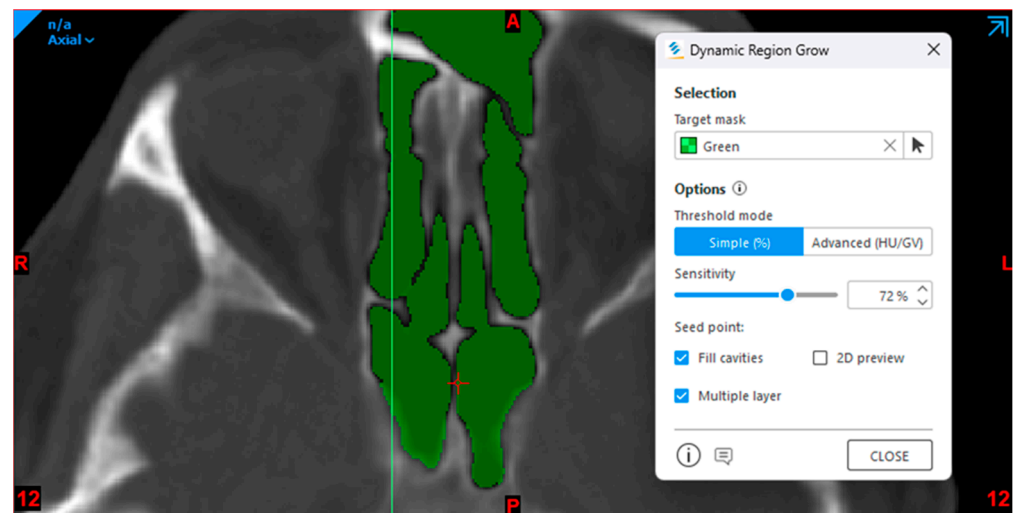


Figure 3. Exemplary presentation of the Dynamic Region Grow function in Mimics (Materialise). Several seed points are used to initially label the sphenoid sinus (and surrounding structures) and then refine the selection. Grayscale sensitivity at 72%; fill cavities and multiple layer labeling are active. L (left), R (right), A (anterior), P (posterior), segmented anatomical structure (green), 12 (CT-scan slice number).

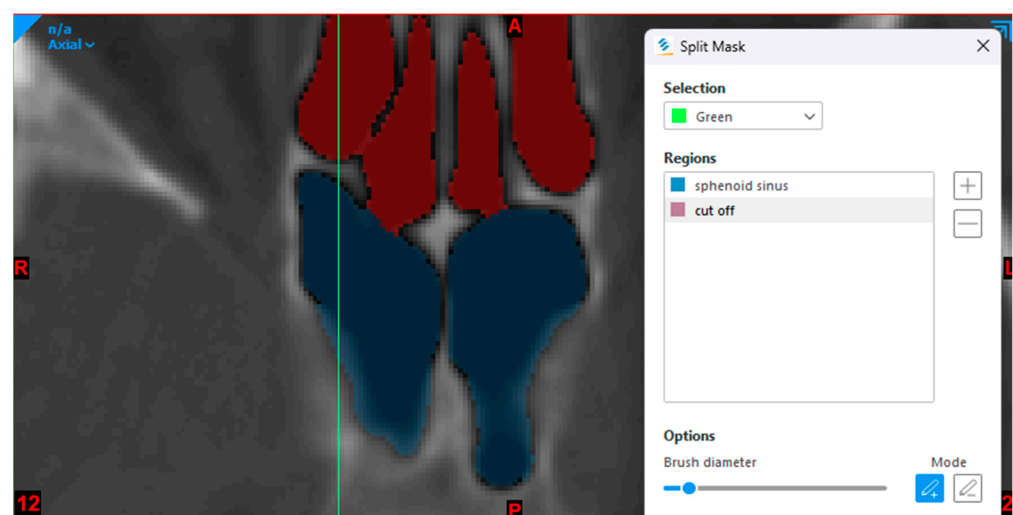


Figure 4. Exemplary presentation of the Split Mask function in Mimics (Materialise). In some slices, the sphenoid sinus (blue) and the cut-off (red) are marked, especially in connected parts, and then the process is automatically completed by Mimics.

2.3.2. Voxelization and Mask Generation

To enable the use of the created segmentations for deep learning, the exported STL surface meshes were converted into voxel-based binary masks aligned with the corresponding CT volumes. Since the preceding preprocessing steps ensured a consistent spatial coordinate system between the CT data and STL models, no additional registration or geometric transformation was required. This ensured voxel-wise correspondence between CT images and reference masks for supervised deep learning.

Voxelization was performed using a ray-casting approach implemented in Python with the trimesh library. Rays were cast along the axial direction through the STL surface, and ordered ray–surface intersections were used to identify voxel intervals located inside the segmented object. To improve computational efficiency, ray casting was restricted to the bounding box of each STL model and processed in parallel blocks. Small boundary corrections were applied at the superior and inferior mesh limits when the STL surface coin-

cided with the first or last CT slice. These corrections compensated for discretization effects introduced during surface mesh generation and prevented small gaps at the boundaries of the voxelized masks.

The resulting masks were stored as binary volumes (foreground = 1, background = 0) and exported in NIfTI (.nii.gz) format while preserving the original spacing, origin, and orientation metadata of the CT images. This enabled direct compatibility with downstream deep learning frameworks such as nnU-Net.

2.4. 3D Segmentation

This is crucial in forensic contexts to ensure absolute certainty regarding the anatomical structure used as a biometric marker. In this respect, CNNs, especially the U-Net architecture, enable anatomical isolation and consistency of the geometry of the sphenoid sinus region, and all comparisons refer to the same anatomical entity.

2.4.1. U-Net Architecture and Its Limitations

Initially, a standard 3D U-Net architecture [29] was considered for the volumetric segmentation task. While the 3D U-Net is a well-established baseline in medical image segmentation, it requires manual tuning of hyperparameters, network topology, and pre-processing pipelines tailored to the specific dataset. In our preliminary experiments [30,31], the standard 3D U-Net struggled with the high anatomical variance in the sinus cavities, leading to suboptimal geometric precision and training instability. Furthermore, adapting the architecture to fully utilize the available high-end hardware (e.g., maximizing batch sizes across multiple GPUs) required complex manual engineering, making it less robust for generating consistent results.

2.4.2. nnU-Net Framework

To overcome these limitations, we employed the nnU-Net (no-new-Net) framework [32], a self-configuring method for deep learning-based biomedical image segmentation. The nnU-Net automatically adapts its architecture, preprocessing, and training scheme to the provided dataset, establishing a highly robust and reproducible pipeline. For our experiments, we utilized the 3d_fullres configuration, which processes the CT volumes at their original resolution to preserve fine anatomical details of the sinus structures.

Preparing the Dataset

While the overall project dataset continued to expand, a strictly controlled subset of 384 CT scans was maintained for these specific segmentation experiments. Utilizing this fixed dataset size was essential to establish a consistent baseline, ensuring fair and reliable comparisons across the different training configurations during the hyperparameter optimization phase:

- Phase 1 (No Clean): The initial raw datasets (DS003/DS004) containing original, partially faulty training images and masks.
- Phase 2 (Clean Baseline): A strictly curated dataset resulting from systematic data cleaning, where all defective masks were manually removed to eliminate label noise.
- Phase 3 & Final (DS006): The final curated dataset used for advanced hyperparameter optimization.

To ensure a robust and unbiased evaluation, the 384 scans were strictly divided prior to any model training. Two distinct testing sets were isolated:

- Internal Testing Set (68 cases): Separated to evaluate the baseline generalization and parameter tuning during the development phase.

- Independent Held-out Set/“External” Test (36 cases): A strictly isolated subset removed from the general dataset pool at the very beginning. The network was completely blinded to this data, serving as an ultimate stress test to verify the final model’s generalization capabilities on unseen cases.

The remaining 280 scans were utilized for model development and were randomly split into a training set (80%) and a validation set (20%) to continuously monitor convergence and prevent overfitting.

To sum up, the dataset underwent several curation stages, reflected in internal dataset identifiers. DS003 contains newly created expert reference masks; DS004 is a merged dataset combining masks from multiple sources (an 85/15 train/test split was used at this stage). During quality review, cases with defective segmentation masks, where the reference annotation was unreliable or unusable, were identified and removed from the pool, yielding the cleaned dataset DS006, which was used for all hyperparameter-optimization experiments reported here. nnU-Net’s built-in 5-fold cross-validation on DS006 partitions the dataset into five approximately equal folds, each serving once as the validation set. For hyperparameter optimization, a single fold (fold 0, the default nnU-Net partition) was used, yielding 240 training and 61 validation cases. The final reported configuration (foreground oversampling raised to 50%, trained on cleaned DS006) corresponds to Run 7 in internal experiment tracking.

Optimization of Hyperparameters

Throughout the experimentation phase, we optimized key hyperparameters to maximize geometric accuracy and stability. As a result, the final optimized configuration (Run 7) incorporates the following adjustments:

- Foreground oversampling (OS50): To address the class imbalance between the background and the fine sinus structures, the foreground oversampling was increased from the nnU-Net default (33%) to 50%. This forced the model to focus on the target sinus structures in half of the training batches.
- The batch size equals 2, which corresponds to the default configuration.

All segmentation models were trained with nnU-Net v2 (PyTorch 2.8.0+cu128, cuDNN 9.10.2) in the 3d_fullres configuration on a single NVIDIA A100-SXM4 (80 GB) GPU. The self-configured network was a 7-stage PlainConvUNet (feature maps 32–320) with instance normalization, leaky-ReLU, and deep supervision, operating on $12 \times 384 \times 384$ patches at a spacing of $3.0 \times 0.386 \times 0.386$ mm. Optimization used SGD with Nesterov momentum (0.99) and weight decay 3×10^{-5} for 1000 epochs of 250 iterations each (batch size 2), with a polynomial learning-rate schedule (initial lr 0.01). The loss was the sum of soft Dice and cross-entropy, applied with deep supervision. Intensities were normalized with nnU-Net’s CT normalization, on top of our preprocessing (Otsu masking, 2nd/98th-percentile normalization, 16-bit rescaling); the unmodified nnU-Net default augmentation pipeline was used. For the best configuration (Run 7), foreground oversampling was raised from the nnU-Net default of 33% to 50% via a custom trainer (nnUNetTrainer_Oversample50).

Multi-GPU Deployment

To facilitate the massive increase in batch size, a distributed training setup across four NVIDIA A100 (80 GB) GPUs was deployed. The increased batch size exposed the model to higher anatomical variance per update step, significantly stabilizing the gradients and reducing outliers.

2.5. Person Identification

Segmentation serves to isolate the sphenoid sinus as the anatomically relevant structure, but it is the subsequent identification step, based on registration and similarity analysis, that determines whether two sets of scans correspond to the same individual. Without this step, the evaluation would remain limited to structural extraction, whereas meaningful forensic interpretation depends on quantifying anatomical correspondence rather than merely aligning surrounding cranial geometry.

Point cloud-based methods have gained attention for comparing three-dimensional anatomical structures, using registration algorithms like Iterative Closest Point or learning-based approaches (e.g., PointNet) [33] to quantify geometric similarity.

GSA-Net-Based Identification

GSA-Net (Geometric Self-Attention Network), introduced by Li et al. [26], was developed for human identification from CT-derived 3D point clouds of the sphenoid sinus. Its geometric feature-extraction module combines local and non-local information represented in Euclidean and eigenvalue spaces, while the neighborhood self-attention module incorporates positional information to capture long-range contextual relationships. ArcFace loss is used to promote intra-class compactness and inter-class separation.

GSA-Net was selected as the identification architecture for the present pipeline because the original task-specific benchmark directly compared it with PointNet, PointNet++, DGCNN, GS-Net, PointPN, PointNeXt, CurveNet, Point Transformer, and manual visual assessment. GSA-Net achieved 99.55% Top-1 and 100% Top-3 accuracy among 220 test individuals and outperformed the evaluated point-cloud baselines by approximately 3% in Top-1 accuracy [26]. For additional task-specific context, previous studies reported Rank-1 accuracy of 96.24% for ICP-based matching [34] and 93.94% for MVSS-Net [23], while geometric 3D–3D superimposition was evaluated using RMS-based separation of matching and non-matching pairs [35]. Because these studies used different cohorts, acquisition settings, gallery sizes, and evaluation protocols, their results provide contextual evidence rather than direct baselines for the present cohort.

In our work, GSA-Net was implemented in PyTorch, Python 3.11, PyTorch 2.11 [36], and Nvidia A100. Adam was used as the optimizer for 2000 epochs with $\beta_1 = 0.9$ and $\beta_2 = 0.99$, a learning rate of 0.001 with a decay rate of 0.7 every 40 epochs. The minimum learning rate was set to 1×10^{-6} . A batch size of 8 was chosen. For all individuals except two, only a single CT examination was available. Two individuals had repeated CT examinations (two scans per individual acquired approximately 5 months apart), which were used exclusively for the temporal stability assessment. These repeated scans were not included in the training or testing process and were used solely as a proof-of-concept validation of the identification approach. No CT scans from the same individual were distributed across different dataset partitions. Each query was therefore generated by applying a random rigid rotation to the same CT-derived surface as its unrotated gallery entry; five rotations were evaluated per model. These rotated point clouds are technical transformations of the same reconstruction, not independent CT examinations. Consequently, this experiment assesses orientation robustness and same-acquisition closed-set retrieval rather than identification across independently acquired repeat CT scans. As proposed in [26] we calculated the cosine similarity distance with the baseline, which is the query to be identified. The top 3 scores were listed as Top-1, Top-2, and Top-3. A Top-1 match was achieved when the patient's highest score was the correct point cloud. A Top-3 match was achieved when the patient's correct match was either the best, second-best, or third-best score. For the two patients with two CT-Scans and thus two point clouds, a match was achieved if one of the two correct point clouds was a top-1 or top-3 score, respectively.

As an outlook, future work will explore an automated shape-similarity pipeline for sphenoid sinus comparison. Each segmented 3D model is centered at its centroid, and a fixed number of approximately evenly spaced points are sampled across the surface using a uniform surface sampling method. The radial distance from the centroid to each sample point yields a one-dimensional shape signature that is invariant to rigid transformations and can be compared across models of differing resolution. This approach is conceptually related to the shape distributions framework of Osada et al. [37], in which surface point statistics—most notably the distribution of pairwise distances (D2)—are used as compact, robust signatures for 3D shape retrieval. We plan to evaluate both centroid-based (D1) and pairwise (D2) variants on our sphenoid dataset and report classification performance in a follow-up study.

3. Results

This section is structured around three complementary results: the segmentation accuracy of the nnU-Net model on the primary dataset (Section 3.1) and the identification of the sphenoid sinus under standard conditions (Section 3.2). Together, these analyses characterize both the baseline performance of the pipeline and its behavior under conditions representative of real-world CT data variability.

3.1. Description of the Evaluation Results for nnU-Net

The segmentation performance of the nnU-Net was evaluated across different dataset phases and hyperparameter configurations. Performance was quantified using the Dice Similarity Coefficient (DSC) for volumetric overlap, together with distance-based metrics, including the 95th percentile Hausdorff Distance (HD95) and the Average Symmetric Surface Distance (ASSD), to assess boundary accuracy and contour agreement.

The Dice Similarity Coefficient (DSC) is an overlap-based metric used to measure the volumetric similarity between the predicted segmentation and the reference manual segmentation. The DSC ranges from 0 (no overlap) to 1 (perfect agreement) and was calculated as follows:

$$DSC(X, Y) = \frac{2 |X \cap Y|}{|X| + |Y|}, \quad (1)$$

where X represents the set of voxels corresponding to the ground truth (manual segmentation), Y represents the set of voxels corresponding to the predicted segmentation, $|X|$ and $|Y|$ denote the total number of voxels in each segmentation mask, and $|X \cap Y|$ indicates the number of overlapping voxels shared between both masks.

The 95th percentile Hausdorff Distance (HD95) was used to evaluate boundary precision between the segmented surfaces. Unlike the conventional Hausdorff Distance, the HD95 uses the 95th percentile of distances, thereby reducing sensitivity to isolated outliers and noise. The metric was defined as follows:

$$HD95 = \max(d_H(S(X), S(Y)), d_H(S(Y), S(X))), \quad (2)$$

where $S(X)$ and $S(Y)$ denote the sets of surface points extracted from the ground truth and predicted segmentations, respectively, and $d_H(A, B)$ represents the 95th percentile of the shortest Euclidean distances from all points on surface A to surface B .

The Average Symmetric Surface Distance (ASSD) was used to quantify the mean spatial discrepancy between segmentation boundaries. ASSD computes the average distance from each surface point of the predicted segmentation to the nearest point on the ground

truth surface and vice versa, thereby providing a symmetric measure of contour accuracy. The ASSD was calculated as follows:

$$ASSD(X, Y) = \frac{1}{|S(X)| + |S(Y)|} \left(\sum_{x \in S(X)} d(x, S(Y)) + \sum_{y \in S(Y)} d(y, S(X)) \right), \quad (3)$$

where $S(X)$ and $S(Y)$ represent the extracted surface point sets of the ground truth and predicted segmentations, respectively; $|S(X)|$ and $|S(Y)|$ denote the total number of surface points in each segmentation; x and y are individual surface coordinates; and $d(x, S(Y))$ represents the shortest Euclidean distance from a point x to the surface $S(Y)$.

The full quantitative results are summarized in Table 2 and Figure 5.

Table 2. Quantitative evaluation of the segmentation models on internal and independent held-out test datasets across different optimization phases. Values are presented as Mean (Median) $\pm$ Standard Deviation.

Test Scenario (Data)	Dice Coefficient [%]	HD95 [mm]	ASSD [mm]
Internal Test Data (68 Cases)			
Phase 1: Baseline (DS004, No Clean)	93.71 (96.37) $\pm$ 8.03	2.45 (0.45) $\pm$ 6.18	0.42 (0.13) $\pm$ 0.91
Phase 2: Baseline (DS004, Clean)	94.46 (96.62) $\pm$ 8.30	1.29 (0.42) $\pm$ 2.00	0.25 (0.09) $\pm$ 0.49
Phase 3: Intermediate (DS006, Clean + OS50)	94.99 (96.77) $\pm$ 7.13	1.71 (0.42) $\pm$ 5.45	0.35 (0.09) $\pm$ 1.29
Final: Optimized (DS006, OS50, Batch 8)	94.90 (96.65) $\pm$ 6.77	1.36 (0.49) $\pm$ 1.82	0.24 (0.13) $\pm$ 0.40
Independent Held-out Set (36 Cases)			
Phase 1: Baseline (DS003, No Clean)	94.63 (96.29) $\pm$ 4.63	1.51 (0.82) $\pm$ 2.05	0.25 (0.13) $\pm$ 0.29
Phase 2: Baseline (DS003, Clean)	95.16 (96.73) $\pm$ 7.15	1.14 (0.42) $\pm$ 1.78	0.21 (0.09) $\pm$ 0.42
Phase 3: Intermediate (DS006, Clean + OS50)	96.28 (96.73) $\pm$ 1.97	0.71 (0.43) $\pm$ 0.66	0.14 (0.10) $\pm$ 0.15
Final: Optimized (DS006, OS50, Batch 8)	95.84 (96.34) $\pm$ 2.06	0.97 (0.69) $\pm$ 0.82	0.17 (0.13) $\pm$ 0.15



Figure 5. Training and validation loss curves. Pseudo dice and pseudo dice moving average curves.

The optimization process yielded the following key observations:

- Impact of data cleaning: Training with the raw, uncleaned datasets (Phase 1) yielded an internal mean DSC of 93.71%. However, geometric precision was severely compromised by label noise, reflected in a high internal mean HD95 of 2.45 mm. Removing the

defective masks (Phase 2) immediately stabilized the network, reducing the internal HD95 error drastically to 1.29 mm.

- Impact of oversampling (peak performance): Applying 50% foreground oversampling (Phase 3) led to the highest absolute peak performance on the independent held-out set, achieving a DSC of 96.28% and a precise HD95 of 0.71 mm. However, this aggressive oversampling caused a localized false positive on a single internal image (SINUS_1620). This temporarily increased the internal mean HD95 to 1.71 mm, indicating a susceptibility to severe boundary outliers, even though the median remained completely stable at 0.42 mm.
- Final optimization (trade-off for stability): In medical image segmentation, worst-case outliers are often more critical than marginal gains in average overlap. To counteract the internal instability observed in Phase 3, the final model combined the cleaned dataset with 50% oversampling and a batch size of 8. This successfully mitigated the outlier problem. Internal geometric stability improved significantly (mean HD95 dropped from $1.51 (0.82) \pm 2.05$ to $0.97 (0.69) \pm 0.82$).

3.2. Identification of Sphenoid Sinus

3.2.1. Identification of the Sphenoid Sinus Under Standard Conditions

GSA-Net was trained on 428 3D sphenoid sinus models and evaluated on an independent validation set comprising 45 3D models representing 43 individuals; two individuals contributed two CT examinations each, originating from a total cohort of 473 patients.

For biometric identification under standard conditions, the 3D sphenoid sinus models were evaluated in a closed-set, multi-candidate matching configuration. Each probe model was compared against the candidate models in the evaluation gallery, and the resulting GSA-Net similarity scores were ranked according to their similarity to the probe. The rank of the correct identity was subsequently used to assess identification performance. The system achieved a Top-1 identification accuracy of 97.8% and a Top-3 accuracy of 100% (see Supplementary Material, Results_GSA-Net_identification). Additionally, the Cumulative Matching Characteristic (CMC) was calculated to further evaluate the identification performance and support the previously reported Top-1 and Top-3 accuracies. The CMC results confirmed the high identification performance, with a Rank-1 identification rate of 97.8% and a Rank-3 rate of 100%. The CMC remained at 100% for all subsequent ranks, further supporting the previously obtained Top-1 and Top-3 results.

In our work, GSA-Net was implemented in PyTorch, Python 3.11, PyTorch 2.11 [36], and trained on an Nvidia A100 GPU. Hyperparameters and training procedure strictly follow the recommendations of Li et al. [26], the developers of the original GSA-Net. Adam was used as the optimizer for 2000 epochs with $\beta_1 = 0.9$ and $\beta_2 = 0.99$, a learning rate of 0.001 with a decay rate of 0.7 every 40 epochs. The minimum learning rate was set to 1×10^{-6} . A batch size of 8 was chosen. The dropout rate was set to 0.5, and 428 individual 3D models of the sphenoid sinus were used to train GSA-Net. 45 additional 3D models were left out for evaluation purposes.

Figures 6 and 7 illustrate representative genuine and competing candidate comparisons. For patient 1112, comparison with the corresponding sphenoid sinus model yielded a GSA-Net similarity score of approximately 0.8021. Among the remaining gallery candidates, the second-highest similarity score was obtained for patient 1193. This example illustrates the multi-candidate ranking principle, in which the similarity of a probe to the corresponding identity is evaluated in relation to competing candidate identities in the gallery.

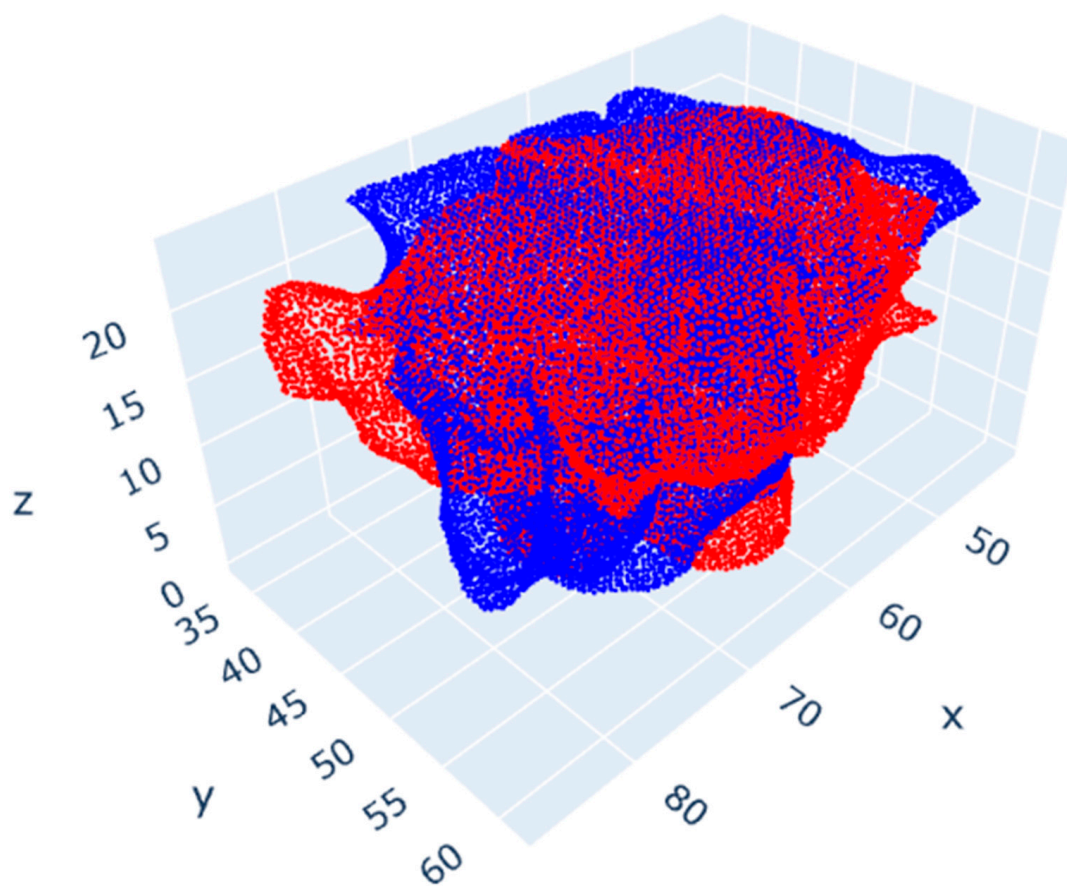


Figure 6. Point Cloud projections for the 3D models of the sphenoid sinus of patient 1112 in blue and rotated in red. Score provided by GSA-Net~0.8021.

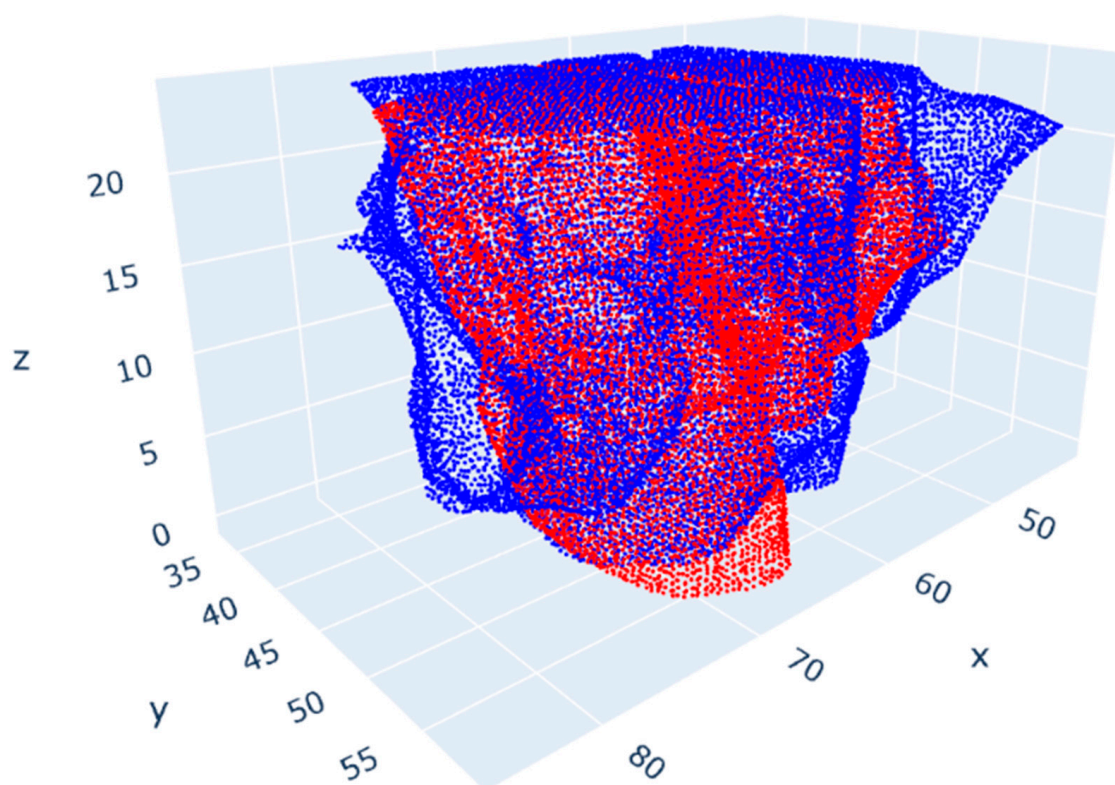


Figure 7. Point Cloud projections for the 3D models of the sphenoid sinus of patient 1112 in blue and patient 1193 rotated in red. Score provided by GSA-Net~0.4572.

Although the obtained identification performance is promising, the biometric evaluation comprises only 42 independent subjects and should therefore be interpreted as a preliminary proof-of-concept rather than as a definitive estimate of population-level forensic identification performance. Furthermore, a fully independent forensic gallery–probe evaluation would ideally require separately acquired CT examinations of the same individuals, with one examination serving as the gallery sample and an independently acquired examination serving as the probe. As repeated CT examinations were not systematically available for the subjects included in the present dataset, this constitutes a limitation of the current study. Future work will therefore focus on validation using a substantially larger independent cohort and, where available, longitudinal or repeated CT acquisitions of the same individuals.

3.2.2. Identification of the Sphenoid Sinus Under Controlled Geometric Perturbation

The resulting transformed 3D models were used exclusively for robustness and sensitivity analysis. Since these perturbed models were derived from the same underlying 3D sphenoid sinus meshes, they were not considered independent biometric samples and were not interpreted as an increase in the effective size of the biometric evaluation set. Accordingly, the experiments presented in this section assess the robustness of the identification approach to controlled geometric perturbations rather than representing additional independent forensic identification experiments.

This distinction is particularly relevant to the forensic interpretation of the results. A fully independent forensic gallery–probe evaluation would ideally require separately acquired CT examinations of the same individuals, with one examination serving as the gallery sample and another independently acquired examination serving as the probe. The controlled perturbations investigated here instead address a complementary question: whether the identity-related information encoded by the sphenoid sinus representation remains sufficiently stable when the same underlying anatomical structure is affected by defined variations in segmentation, mesh processing, image resolution, or centroid localization.

All perturbations were generated in mesh space from 43 post-segmentation STL surfaces representing 41 individuals and were applied independently to each original mesh rather than cumulatively, yielding $43 \times 4 \times 4 = 688$ perturbed queries. Four perturbation families were evaluated: (i) surface-normal noise, defined as $v_i = v_i + \varepsilon_i n_i$, where v_i and n_i denote the original vertex position and corresponding unit surface normal, respectively, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, with $\sigma \in \{0.01, 0.04, 0.08, 0.12\}$ mm; (ii) quadric decimation retaining a fraction $q \in \{0.95, 0.90, 0.80, 0.75\}$ of the original mesh faces; (iii) isotropic mesh voxelization at pitch $p \in \{0.3, 0.4, 0.5, 0.8\}$ mm, followed by solid-grid filling, marching-cubes reconstruction, and transformation from voxel-index space to the original physical coordinate system; and (iv) global translation, defined as $v_i = v_i + t$, where $t = mu$, $u \sim \text{Unif}(S^2)$, and $m \sim \mathcal{U}(0, r)$, with $r \in \{0.5, 1.0, 1.5, 2.0\}$ mm. One realization was generated for each mesh-parameter combination; randomized operations used fixed level-specific seeds. These operations are synthetic mesh-level stress tests: voxelization does not reproduce an independent CT acquisition, while global translation assesses positional invariance rather than error in an internally estimated centroid. Identification performance for each perturbation family and parameter level is reported in Tables 2 and 3.

Table 3. Identification accuracy under geometric perturbation.

Condition	n	Top-1 Accuracy	Top-3 Accuracy	Mean Top-1 Score	Mean Score Change vs. Original
Unperturbed source queries	43	100.0%	100.0%	0.887	0.000
All perturbed meshes	688	79.4%	86.8%	0.811	−0.076
Segmentation noise	172	99.4%	100.0%	0.873	−0.014
Mesh resolution reduction	172	95.3%	100.0%	0.833	−0.054
Voxel-size resampling	172	23.3%	47.7%	0.657	−0.230
Centroid localization error	172	99.4%	99.4%	0.883	−0.004

To assess whether the identification model preserves the identity signal after controlled geometric distortion, 43 unperturbed source meshes representing 41 individuals were evaluated together with 688 perturbed variants. All queries were matched against the complete gallery of 45 models representing 43 individuals. Each original mesh was processed through four perturbation families, with four parameter levels per family. Identification performance was reported as top-1 and top-3 accuracy, following the ranking-based evaluation strategy used in comparable 3D biometric identification studies.

Because perturbation suffixes were appended to the query identifiers in the result file, correctness was recomputed by mapping each distorted file back to the identity label of its corresponding unperturbed mesh. Thus, for example, a query named 1112_noise_0.08mm.stl was treated as a correct identification when one of the returned candidates had identity 1112.

Under baseline conditions, the 43 unperturbed source queries were identified with 100.0% top-1 accuracy and 100.0% top-3 accuracy (43/43 cases), with a mean top-1 similarity score of 0.887.

Across all distorted variants combined, top-1 accuracy decreased to 79.4% (546/688), while top-3 accuracy remained higher at 86.8% (597/688). The mean top-1 score decreased from 0.887 for unperturbed meshes to 0.811 under distortion, corresponding to an average score reduction of 0.076 (see Table 3).

The model remained highly stable under three of the four tested distortion types. Segmentation noise had almost no effect on the rank ordering of candidates: 171/172 distorted meshes were correctly identified at top-1, and all cases were recovered within the top-3 candidates. Mesh resolution reduction produced a larger but still limited decrease in top-1 performance (164/172, 95.3%); however, every mesh was still correctly included in the top-3 candidate list (see Table 4). Centroid localization error also showed minimal influence within the tested range, with only one failure at the maximum 2.0 mm offset.

Voxel-size resampling was the dominant source of performance degradation. In contrast to the other perturbation types, which together retained 98.1% top-1 accuracy and 99.8% top-3 accuracy, voxel reconstruction reduced top-1 accuracy to 23.3% and top-3 accuracy to 47.7%. This indicates that the descriptor is much more sensitive to surface changes introduced by voxelization and marching-cubes reconstruction than to small surface-normal noise, moderate mesh decimation, or global centroid displacement.

The top-3 metric provides additional practical value for a forensic identification workflow. Across all perturbed meshes, 51 cases were not ranked first but were still present within the first three returned candidates. This behavior is relevant because forensic identification is rarely interpreted in isolation from case context. If external information is available, such as the set of missing persons or whether a higher-ranked candidate is already known to be alive or otherwise accounted for, a top-3 candidate list can still support a targeted and defensible identification workflow.

Table 4. Top-1 and Top-3 identification accuracy according to perturbation level.

Perturbation	Level	Top-1 Accuracy	Top-3 Accuracy	Mean Top-1 Score
Segmentation noise	0.01 mm	100.0%	100.0%	0.871
Segmentation noise	0.04 mm	100.0%	100.0%	0.881
Segmentation noise	0.08 mm	97.7%	100.0%	0.879
Segmentation noise	0.12 mm	100.0%	100.0%	0.860
Mesh resolution	0.95	95.3%	100.0%	0.847
Mesh resolution	0.90	97.7%	100.0%	0.847
Mesh resolution	0.80	93.0%	100.0%	0.818
Mesh resolution	0.75	95.3%	100.0%	0.819
Voxel resampling	0.3 mm	25.6%	60.5%	0.669
Voxel resampling	0.4 mm	18.6%	51.2%	0.640
Voxel resampling	0.5 mm	27.9%	46.5%	0.677
Voxel resampling	0.8 mm	20.9%	32.6%	0.642
Centroid error	0.5 mm	100.0%	100.0%	0.885
Centroid error	1.0 mm	100.0%	100.0%	0.885
Centroid error	1.5 mm	100.0%	100.0%	0.897
Centroid error	2.0 mm	97.7%	97.7%	0.866

At the end of the identification pipeline, a qualified physician performs the final evaluation and makes a decision. To support these functionalities and enable the identification workflow, a web application is currently under development. At present, a database and API infrastructure have been implemented for data management and integration of the AI models developed in this work.

4. Discussion

The combination of CT data preprocessing, accurate and fast creation of 3D models of the sphenoid sinus, deep learning segmentation, and point cloud-based identification using GSA-Net enables a robust identification pipeline for person recognition. In practical forensic workflows, the proposed AI framework is intended to function as a decision-support tool rather than a replacement for forensic experts. Following CT image acquisition, the automated segmentation model can rapidly delineate the sphenoid sinus and generate a standardized three-dimensional model. The identification module can then compare the extracted anatomical features with reference datasets and provide similarity scores or ranked candidate matches. The final identification decision, however, remains the responsibility of the forensic expert, who integrates the AI results with other available evidence, including dental records, anthropological findings, DNA analysis, fingerprints, and contextual case information. By automating the time-consuming segmentation process and providing objective quantitative analysis, the proposed system has the potential to improve workflow efficiency, reduce observer variability, and support more consistent forensic decision-making, particularly in mass disaster victim identification and humanitarian forensic investigations.

Moreover, the results achieved in the work at all phases of the pipeline ensure a reliable identification process.

The training curves and quantitative results in Table 2 demonstrate that the nnU-Net framework achieved stable and high-quality segmentation performance across all experimental phases, with consistent convergence behavior and no evidence of overfitting. The progressive improvement from Phase 1 to the final optimized configuration highlights the critical impact of data quality and training strategy: in particular, data cleaning substantially reduced geometric errors, as reflected by the marked decrease in HD95, while foreground oversampling improved sensitivity to fine sinus structures. However, the inter-

mediate configuration also revealed that aggressive optimization can introduce instability in the form of outliers, emphasizing the importance of balancing peak performance with robustness. The final model represents this trade-off effectively, achieving strong Dice scores alongside improved geometric consistency (HD95 and ASSD), which is particularly relevant for downstream tasks such as shape-based identification. Overall, the results confirm that nnU-Net provides a reliable and adaptable framework for segmenting anatomically complex structures like the sphenoid sinus, while also underlining the importance of dataset curation and balanced training configurations for robust real-world performance.

The results achieved in this study using GSA-net to identify the correct sphenoid sinus confirm the results shown in Li et al. and thus show that, with proper 3D models of the sphenoid sinus, an identification is possible. Achieving a Top-1 accuracy of 97.8% is an impressive result, but it still highlights that there can be mismatches and thus wrong identifications. The Top-3 accuracy achieved of 100% shows that even when the top match is wrong, the correct person is still in the top 3 scores. When providing the matching scores to the proper authorities, additional information about the whereabouts of the other matches can provide the insights needed to identify the correct match. Accordingly, the observed 97.8% Top-1 and 100% Top-3 accuracy should not be interpreted as identification performance between independently acquired ante-mortem/post-mortem or repeat CT examinations.

Creating one single 3D model of a sphenoid sinus from a CT scan took about 24 min for our team. The automated generation takes seconds, thus significantly improving the time required by a factor of 10^3 , to build a large database for identification purposes and, finally, the time required for a human to identify the correct person.

A central aspect of the robustness evaluation was the definition of a realistic distortion envelope, representing the range of variability expected in clinical CT acquisition and processing. All perturbations were limited to values consistent with imaging standards and expert assessment of segmentation variability, ensuring that the results reflect practical, real-world conditions. Within this envelope, the pipeline demonstrated stable identification performance, indicating suitability for deployment under typical operating scenarios.

This general robustness statement does not extend to the mesh-revoxelization condition, which reduced Top-1 and Top-3 accuracy to 23.3% and 47.7%, respectively. This finding should not be interpreted as evidence that GSA-Net is intrinsically unsuitable for CT data acquired at different voxel sizes. The present operation was performed after segmentation in mesh space: voxelization, solid-grid filling, and marching-cubes reconstruction replaced the native tessellation with a lattice-derived surface and changed local geometry, topology, vertex density, and the neighborhoods used for point-cloud feature extraction. Because GSA-Net relies on local Euclidean and eigenvalue-space features, the current implementation appears sensitive to this representation shift, even though global translation and scale are removed by centering and unit-sphere normalization. The result therefore demonstrates a limitation of the current mesh-based input representation rather than calibrated cross-scanner or cross-protocol performance. Future work will evaluate voxel-size robustness at the image level by resampling the original CT volumes to predefined spacings, re-running the same nnU-Net segmentation, reconstructing new meshes, and then testing GSA-Net with a fixed gallery and deterministic point sampling. This additional re-segmentation step will distinguish CT-resolution effects from sensitivity to mesh reconstruction itself.

Among the investigated perturbations, centroid localization error was expected to be critical due to its role in defining the radial distance-based shape fingerprint. However, only minimal performance degradation was observed, with nearly all cases correctly identified even at maximum displacement. This result should be interpreted carefully,

as the current implementation models centroid error as a global translation, potentially testing invariance to spatial shifts rather than true sensitivity of the descriptor. Future experiments should therefore directly perturb the reference point used for feature extraction to isolate this effect. Although the point clouds are already centered and unit-sphere-normalized before inference, the present perturbations operate only at the mesh level and do not capture upstream imaging variability, including reconstruction kernels, intensity differences, motion artifacts, or repeat CT acquisition. Robustness may be further improved by standardizing mesh resolution and surface-sampling density, training across multiple voxel and mesh resolutions, and incorporating geometric descriptors invariant to rotation, scale, and sampling density. Future work will evaluate these strategies under stronger stress tests and independent repeat CT examinations. A pre-specified confidence-gated hybrid could retain GSA-Net as the primary model above a validated confidence threshold and otherwise fuse or defer to an alternative descriptor or architecture; thresholds and fusion rules would be fixed on validation data before independent testing.

A detailed analysis of the 2.2% of cases that failed to achieve a Top-1 match revealed a distinct anatomical pattern of confusion. These failed cases predominantly featured sphenoid sinuses with relatively small volumes, smooth internal walls, and a lack of complex septation (typically presenting only a single, simple septum or incomplete septa), as illustrated in Figure 8. Because GSA-Net relies on geometric features extracted from the point cloud, sinuses with such simplified morphology yield less distinct spatial signatures, making them harder to isolate from anatomically similar “neighbors” (see Figure 8).

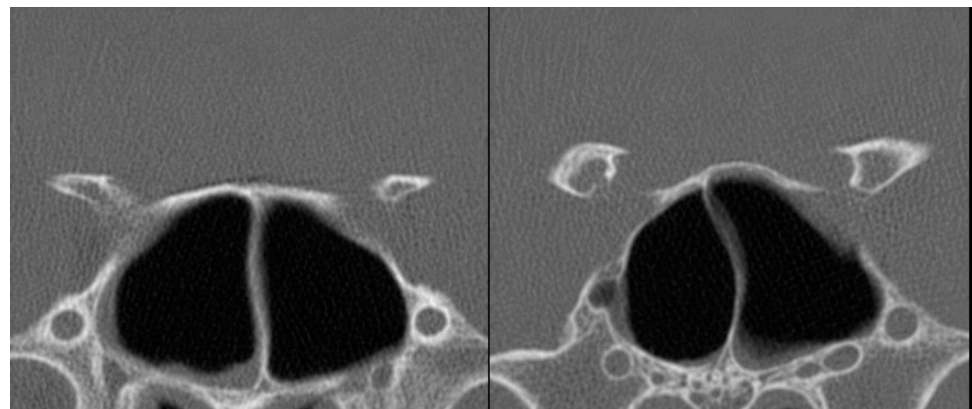


Figure 8. Example of anatomically similar “neighbors”.

Conversely, geometric perturbations (such as spatial noise and minor centroid shifts) had a negligible impact on identification accuracy. The only substantial processing limitation was voxel-size resampling. Figure 9 demonstrates the progressive loss of geometric detail following voxelization at different voxel sizes. Notably, the mechanism of this processing failure inherently aligns with the anatomical findings: re-voxelization acts as a morphological smoothing filter. It erases high-frequency topological details and artificially simplifies the sinus structure, effectively creating the same geometric similarities observed in naturally smooth, unseptated sinuses.

In a real forensic workflow, the proposed AI-based pipeline should be considered a decision-support system rather than an autonomous identification tool. Following CT acquisition, the sphenoid sinus can be automatically segmented, converted into a 3D point-cloud representation, and compared with available reference data to generate a ranked list of potential identity matches together with their similarity scores. This ranking can support forensic experts by rapidly narrowing the set of candidate identities, particularly when conventional identification characteristics are unavailable or compromised. Importantly, the final identification decision should remain with a qualified forensic expert, who evaluates

the AI-generated candidates in combination with visual assessment of the anatomical structures, image and segmentation quality, available ante- and post-mortem information, and other established forensic evidence. Cases characterized by low similarity scores, small differences between the highest-ranked candidates, or uncertain segmentation should be explicitly flagged for additional expert review rather than automatically classified.

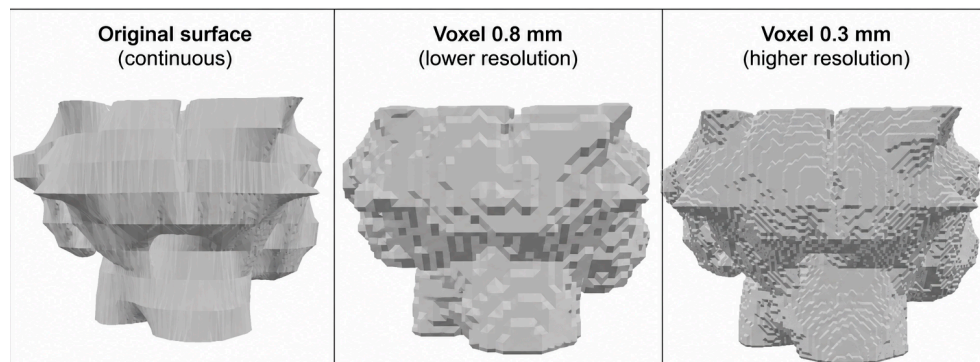


Figure 9. Comparison of the original surface model and its voxelized representations using 0.8 mm and 0.3 mm voxel sizes.

Future work will focus on validating the proposed approach using multicenter datasets acquired from different institutions and imaging systems.

It is also important to highlight the lack of external validation in this study. All CT data were collected from a single institution using 2 CT scanners. Consequently, although an independent held-out test set was used for internal evaluation, the generalizability of the proposed framework across different institutions, scanner types, acquisition protocols, and patient populations has not yet been established. Future studies will focus on multicenter validation using datasets acquired from diverse clinical settings.

Thus, the proposed system is intended to complement established forensic identification procedures by providing standardized, reproducible, and quantitative decision support while maintaining expert oversight and human responsibility for the final conclusion.

5. Conclusions

This study presented an integrated pipeline for forensic human identification based on the morphology of the sphenoid sinus extracted from cranial CT imaging. The framework combines a preprocessing stage, semi-automatic reference segmentation in Mimics, automated 3D segmentation with nnU-Net, and point-cloud-based identification using GSA-Net. By bringing these components together within a single workflow, the present work addresses a gap in the existing studies, in which segmentation and identification have most often been investigated in isolation.

The segmentation experiments, carried out on a curated subset of 384 CT volumes drawn from a broader cohort of 473 anonymized scans, indicated that the proposed configuration of nnU-Net combining systematic data cleaning, 50% foreground oversampling, and an increased batch size enabled by multi-GPU training produced anatomically plausible and geometrically stable delineations of the sphenoid sinus. On the independent held-out set, the model reached a Dice similarity coefficient of 96.28% and a 95th percentile Hausdorff distance of 0.71 mm, suggesting that the approach can cope reasonably well with the variability of clinical CT acquisitions, although some sensitivity to localized boundary outliers was still observed.

Using the resulting 3D representations, the GSA-Net-based identification stage achieved a Top-1 accuracy of 97.8% and a Top-3 accuracy of 100% across 45 models representing 43 individuals, evaluated under multiple rotations. These results are consistent

with previously reported findings and support the view that the sphenoid sinus carries sufficient individualizing information to be considered a useful anatomical marker in forensic identification. In addition, replacing manual 3D model creation, which required approximately 24 min per case, with automated segmentation reduced processing time to a few seconds, indicating that the pipeline may offer a practical basis for building larger reference databases under controlled conditions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/electronics15173805/s1>. Results_GSA-Net_identification.xlsx.

Author Contributions: Conceptualization, V.A., M.F. and A.N.; methodology, M.K., T.G. and F.Z.; software, F.Z. and P.L.; validation, M.K. and T.G.; formal analysis, F.Z.; investigation, M.K., P.L. and V.K.; data curation, V.K. and V.G.; writing—original draft preparation, V.A.; writing—review and editing, M.K., M.F. and V.G.; visualization, F.Z.; supervision, M.F. and A.N.; project administration, A.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Philipp Schwartz Initiative (PSI) of the Alexander von Humboldt Foundation, funded by the German Federal Foreign Office, through a scholarship awarded to V.A. This research was also supported by the DATIpilot Innovation Sprint Project BIKO-UA (FKZ 03DPS1241), “Bildererkennung zur Identifikation von Kriegsopfern in der Ukraine” (Image Recognition for the Identification of War Victims in Ukraine), funded by the German Federal Ministry of Education and Research (BMBF).

Data Availability Statement: The clinical datasets used in this study are not publicly available due to data protection regulations but may be provided in anonymized form upon reasonable request and with appropriate ethical approval.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
ASSD	Average Symmetric Surface Distance
CNN	Convolutional Neural Network
CT	Computed Tomography
DSC	Dice Similarity Coefficient
GSA-Net	Geometric Structure Analysis Network
HD95	95th Percentile Hausdorff Distance
ML	Machine Learning
MRI	Magnetic Resonance Imaging
nnU-Net	no-new-Net (self-configuring U-Net framework)
OS50	Oversampling factor 50
PMCT	Post-Mortem Computed Tomography
STL	Standard Tessellation Language
U-Net	Convolutional Neural Network for Biomedical Image Segmentation
3D	Three-Dimensional

References

1. Torimitsu, S.; Uemura, Y.; Tsuneya, S.; Tweeatsani, N.; Yoshida, M.; Yajima, D.; Inokuchi, G.; Motomura, A.; Chiba, F.; Hoshioka, Y.; et al. Forensic Personal Identification Using Two-Dimensional Computed Tomographic Imaging of the Frontal Sinus: A Comparative Analysis of Feature Matching Algorithms. *Int. J. Leg. Med.* **2025**, *139*, 2857–2867. [\[CrossRef\]](#)
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. Avent, P.R.; Campbell, J.L.; Garvin, H.M.; Butaric, L.N. Personal Identification Using Frontal Sinus Coding Methods: The Effect of Mixed Image Modality Comparisons. *J. Forensic Sci.* **2024**, *69*, 1155–1170. [CrossRef] [PubMed]
4. Uthman, A.T.; Al-Rawi, N.H.; Al-Naaimi, A.S.; Tawfeeq, A.S.; Suhail, E.H. Evaluation of Frontal Sinus and Skull Measurements Using Spiral CT Scanning: An Aid in Unknown Person Identification. *Forensic Sci. Int.* **2010**, *197*, 124.e1–124.e7. [CrossRef] [PubMed]
5. Belaldavar, C.; Kotrashetti, V.S.; Hallikerimath, S.R.; Kale, A.D. Assessment of Frontal Sinus Dimensions to Determine Sexual Dimorphism among Indian Adults. *J. Forensic Dent. Sci.* **2014**, *6*, 25–30. [CrossRef] [PubMed]
6. Souza, L.A., Jr.; Marana, A.N.; Weber, S.A.T. Automatic Frontal Sinus Recognition in Computed Tomography Images for Person Identification. *Forensic Sci. Int.* **2018**, *286*, 252–264. [CrossRef] [PubMed]
7. Uthman, A.T.; Alomar, A.; Almukhtar, A.; Jaber, R.; Essale, R.; Abdulsalam, R.; Al-Rawi, N.H. Accuracy of Sphenoidal Sinus Morphometry in Forensic Identification Using Cone Beam Computed Tomography. *J. Int. Dent. Med. Res.* **2021**, *14*, 1485–1491.
8. Kiciński, P.; Podgórski, M.; Grzelak, P.; Małachowska, B.; Polgaj, M. Authors' Classification of Sphenoid Sinus Pneumatizations into the Sphenoid Bone Processes. *J. Clin. Med.* **2025**, *14*, 5811. [CrossRef] [PubMed]
9. Tyr, A.; Zilg, B.; Gelius, T.; Möllby, R.; Heldring, N. Postmortem CT Analysis of Paranasal Sinuses Using an Experimental Model of Drowning. *Int. J. Leg. Med.* **2024**, *138*, 1401–1409. [CrossRef] [PubMed]
10. Gurlek Celik, N.; Akman, B. Analysis of Sphenoid Sinus and Ethmoid Sinus Volume and Asymmetry by Sex: A 3D-CT Study. *Surg. Radiol. Anat.* **2024**, *46*, 551–558. [CrossRef] [PubMed]
11. Bolliger, S.A.; Thali, M.J. Imaging and Virtual Autopsy: Looking Back and Forward. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, *370*, 20140253. [CrossRef] [PubMed]
12. Dedouit, F.; Savall, F.; Mokrane, F.Z.; Rousseau, H.; Crubézy, E.; Rougé, D.; Telmon, N. Virtual Anthropology and Forensic Identification Using Multidetector CT. *Br. J. Radiol.* **2014**, *87*, 20130468. [CrossRef] [PubMed]
13. Pauwels, R.; Beinsberger, J.; Collaert, B.; Theodorakou, C.; Rogers, J.; Walker, A.; Cockmartin, L.; Bosmans, H.; Jacobs, R.; Bogaerts, R.; et al. Effective Dose Range for Dental Cone Beam Computed Tomography Scanners. *Eur. J. Radiol.* **2012**, *81*, 267–271. [CrossRef] [PubMed]
14. Roberts, I.S.D.; Benamore, R.E.; Peebles, C.; Roobottom, C.; Traill, Z.C. Post-Mortem Imaging as an Alternative to Autopsy in the Diagnosis of Adult Deaths: A Validation Study. *Lancet* **2012**, *379*, 136–142. [CrossRef] [PubMed]
15. Hill, V.B.; Ulmer, J.L.; Mueller, W.M.; Yetkin, F.Z.; Haughton, V.M.; Prost, R.W. A Practical Review of Functional MRI Anatomy of the Language and Motor Systems. *AJNR Am. J. Neuroradiol.* **2019**, *40*, 1084–1090. [CrossRef] [PubMed]
16. Heo, J.H.; Kim, M.J.; Jang, S.J.; Lee, J.; Im, S.B.; Lee, S.; Na, J.Y.; Kim, Y.; Yoon, Y.; Kwon, J.H. Multi-Dimensional Deep Learning-Based Segmentation and Volumetric Analysis of Sphenoid Sinus Fluid on Postmortem CT Images. *Sci. Rep.* **2026**; in press.
17. Materialise. *Mimics Innovation Suite*; Materialise NV: Leuven, Belgium. Available online: <https://www.materialise.com/en/health-care/mimics-innovation-suite> (accessed on 18 August 2026).
18. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.-C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [CrossRef] [PubMed]
19. Bilous, N.; Malko, V.; Tkachenko, D.; Frohme, M. Automated Identification from CT Using Sphenoid Sinus Geometry as an Anatomical Biometric. *Appl. Syst. Innov.* **2026**, *9*, 89. [CrossRef]
20. Souadih, K.; Belaid, A.; Ben Salem, D.; Conze, P.-H. Automatic Forensic Identification Using 3D Sphenoid Sinus Segmentation and Deep Characterization. *Med. Biol. Eng. Comput.* **2020**, *58*, 291–306. [CrossRef] [PubMed]
21. 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]
22. Alekseeva, V.; Krüger, M.; Graner, T.; Weiß, S.; Frohme, M.; Nechyporenko, A. Exploring the Sphenoid Sinus as a Biometric Marker for Human Identification. In *Proceedings of the 5th International Workshop of IT-Professionals on Artificial Intelligence (ProFIT AI 2025), Liverpool, UK, 15–17 October 2025*; Chumachenko, D., Potapov, I., Yakovlev, S., Yakovlev, S., Eds.; CEUR-WS: Aachen, Germany, CEUR Workshop Proceedings 2025; 2025; Volume 4164, pp. 487–493.
23. Wen, H.; Wu, W.; Fan, F.; Liao, P.; Chen, H.; Zhang, Y.; Deng, Z.; Lv, W. Human Identification Performed with Skull's Sphenoid Sinus Based on Deep Learning. *Int. J. Leg. Med.* **2022**, *136*, 1067–1074. [CrossRef] [PubMed]
24. Gülşen, İ.T.; Kuran, A.; Evli, C.; Baydar, O.; Dinç Başar, K.; Bilgir, E.; Çelik, Ö.; Bayrakdar, İ.Ş.; Orhan, K.; Acu, B. Deep Learning Model for Automated Segmentation of Sphenoid Sinus and Middle Skull Base Structures in CBCT Volumes Using nnU-Net v2. *Oral Radiol.* **2026**, *42*, 139–147. [CrossRef] [PubMed]
25. Seghier, M.L. Image Segmentation Evaluation with the Dice Index: Methodological Issues. *Int. J. Imaging Syst. Technol.* **2024**, *34*, e23203. [CrossRef]
26. Li, X.; Zou, R.; Chen, H. Human Identification Based on Sphenoid Sinus in Point Cloud with Geometric Self-Attention Network. *Multimed. Tools Appl.* **2025**, *84*, 14719–14737. [CrossRef]

27. Winkler, A.M. The NIFTI File Format. Brainder 2012. Available online: <https://brainder.org/2012/09/23/the-nifti-file-format/> (accessed on 18 August 2026).
28. Bidgood, W.D.; Horii, S.C.; Prior, F.W.; Van Syckle, D.E. Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging. *J. Am. Med. Inform. Assoc.* **1997**, *4*, 199–212. [[CrossRef](#)] [[PubMed](#)]
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; LNCS; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
30. Nechyporenko, A.; Frohme, M.; Omelchenko, V.; Alekseeva, V.; Lupyr, A.; Gargin, V. Identification of Personality Based on the Sphenoid Sinus Structure Using Machine Learning. In Proceedings of the 6th International Conference on Informatics and Data-Driven Medicine (IDDM 2023), Bratislava, Slovakia, 17–19 November 2023; CEUR Workshop Proceedings 2023; Volume 3609, pp. 167–173.
31. Nechyporenko, A.; Frohme, M.; Alekseeva, V.; Gargin, V.; Sytnikov, D.; Hubarenko, M. Deep Learning Based Image Segmentation for Detection of Odontogenic Maxillary Sinusitis. In Proceedings of the 41st IEEE International Conference on Electronics and Nanotechnology (ELNANO 2022), Kyiv, Ukraine, 10–14 October 2022; pp. 339–342. [[CrossRef](#)]
32. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat. Methods* **2021**, *18*, 203–211. [[CrossRef](#)] [[PubMed](#)]
33. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 652–660. [[CrossRef](#)]
34. 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](#)]
35. Cappella, A.; Gibelli, D.; Cellina, M.; Mazzearelli, D.; Oliva, A.G.; De Angelis, D.; Sforza, C.; Cattaneo, C. Three-Dimensional Analysis of Sphenoid Sinus Uniqueness for Assessing Personal Identification: A Novel Method Based on 3D–3D Superimposition. *Int. J. Leg. Med.* **2019**, *133*, 1895–1901. [[CrossRef](#)] [[PubMed](#)]
36. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*; Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035. Available online: <https://pytorch.org/> (accessed on 18 August 2026).
37. Osada, R.; Funkhouser, T.; Chazelle, B.; Dobkin, D. Shape Distributions. *ACM Trans. Graph.* **2002**, *21*, 807–832. [[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.