Towards Automatic Computation of the SkullCode: A Novel Biometric Approach for Cranial Identification

Antonio Aragon-Molina[1], Sofia Alemanno[3], Giulia Caccia[2,3], Carlo Campobasso[3],
Cristina Cattaneo[2], Danilo de Angelis[2], Ruggero Donida Labati[1], Daniele Gibelli[2],
Andrea Palamenghi[2], Vincenzo Piuri[1], Fabio Scotti[1]

[1]*Department of Computer Science, Università degli Studi di Milano, Milan, Italy*
[2]*Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milan, Italy*
[3]*Department of Experimental Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy*

{antonio.aragon, giulia.caccia, cristina.cattaneo, danilo.deangelis, ruggero.donida,
daniele.gibelli, andrea.palamenghi, vincenzo.piuri, fabio.scotti}@unimi.it
{sofia.alemanno, carlopietro.campobasso}@unicampania.it

***Abstract*—One of the main activities of forensic anthropologists consists of identifying living or deceased individuals from the analysis of physiological characteristics. In this context, recent studies evaluated the possibility of using novel physiological traits. Furthermore, recent works aim at realizing decision support tools for forensic scientists based on artificial intelligence techniques. In our previous work, computer scientists and forensic anthropologists collaborated to design a novel biometric recognition approach based on non-metric cranial traits, resulting in a 35-digit representation of these features (referred to as SkullCode). In this paper, we take a step towards the automated detection of SkullCode features. Specifically, we explore the possibility of classifying one of its traits by using deep neural networks. Our approach involves the extraction of two-dimensional slices from the skull's region of interest, followed by a deep neural network that classifies the presence or absence of the trait. To validate our method, in collaboration with Italian hospitals, we created a dataset of 952 annotated samples from 84 individuals. The best obtained accuracy is equal to 88.87%. This promising result highlights the potential of our method for real-world applications and brings us closer to developing a fully automated identification system based on SkullCode features.**

***Index Terms*—Skull, identification, non-metric cranial traits, automatic extraction**

I. Introduction

In forensic anthropology, personal identification has frequently relied on qualitative and metric comparisons of skeletal traits between antemortem and postmortem records, typically performed by human experts. The human cranium contains a wide array of anatomical variants known as non-metric traits [1], such as accessory foramina, bipartite bones, and sutural variations. Non-metric cranial traits have often been considered too variable to be used as reliable biometric features for individual recognition [2]. Nevertheless, recent advancements in forensic science suggest that, when analyzed collectively, these traits hold significant potential for biometric identification [1].

In our previous work [3], we introduced SkullCode, a binary encoding system for representing the presence of non-metric cranial traits. SkullCode encodes these traits as a 35-bit binary string, with each bit corresponding to a specific trait (1 for present, 0 for absent). This binary string serves as a discriminative feature vector suitable for biometric recognition. Specifically, a dissimilarity score between two SkullCodes can be computed using the Hamming distance. In our previous study, the feature extraction process was performed by forensic anthropologists through a visual analysis of Computed Tomography (CT) scans. This process is therefore time-consuming and subjective.

Recent studies have demonstrated that artificial intelligence technologies can be effectively used to analyze skeletal data [4] and to design decision-support systems for forensic applications [5]. However, to the best of our knowledge, no studies have yet explored artificial intelligence approaches to analyze non-metric cranial traits.

In this paper, we propose an automated method for extracting SkullCode features from CT scans. As a preliminary study, this paper focuses on a single trait within the SkullCode: *Pneumatization of the crista galli*. First, we present a method for generating labeled datasets using images obtained from CT scans. Then, we demonstrate how deep neural networks can effectively be trained to detect the presence of non-metric cranial traits. Our approach can be extended to classify additional features, advancing the automation of SkullCode extraction for forensic identification.

Our contribution is twofold:

1) We propose the first method in the literature for detecting the presence of non-metric cranial traits from CT scans.

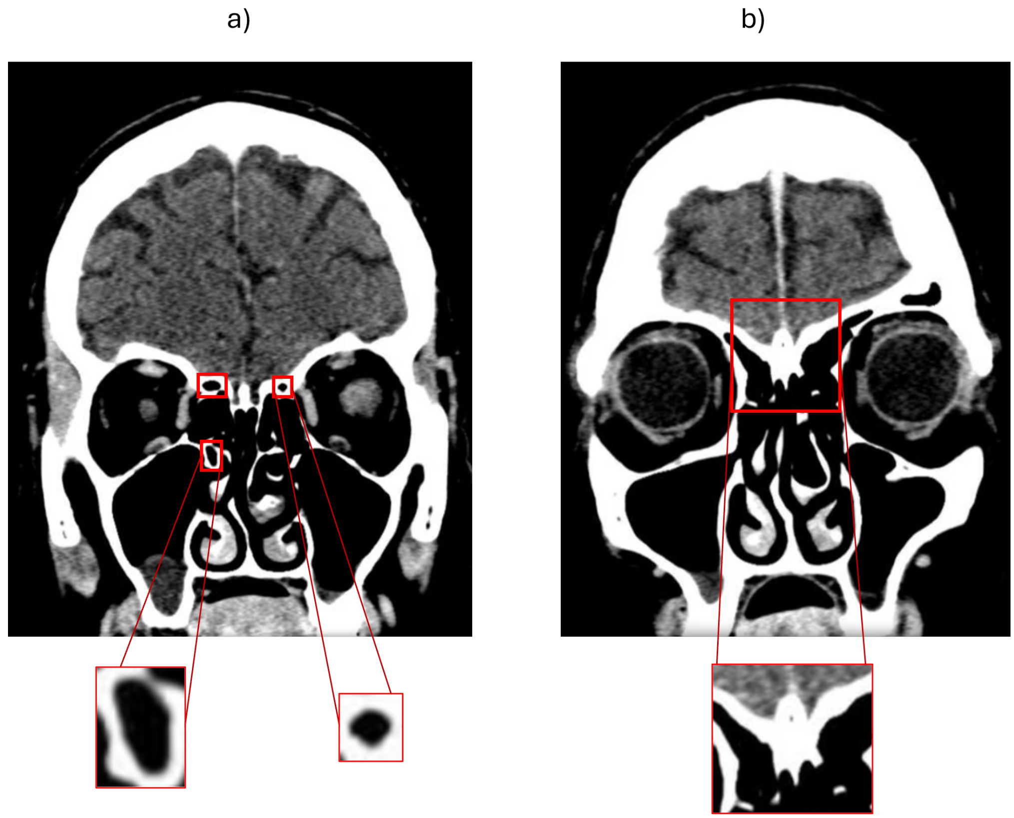

Fig. 1. Selection of the Region of Interest (ROI): (a) Cropping the area around the feature without considering the surrounding anatomical context may lead to confusion. (b) Cropping the feature along with key surrounding anatomical structures provides the necessary contextual information for accurate identification. In our study, we apply the approach (b). Image courtesy of IMAIOS [6].

2) We created a labeled dataset of 952 images extracted from 84 CT scans by a team of computer scientists and anthropologists.

This paper is structured as follows. Section II summarizes the related work. Section III explains the materials and methods. Section IV details the experiments performed. Section V concludes the work.

II. Related work

In the literature, there are relatively few studies on the automatic identification of global characteristics of the skull. Among the most commonly studied features are biological sex and ethnicity, typically derived from three-dimensional (3D) representations of cranial data. These methods may rely on projections from different skull views [7], depth estimations from 3D data [8], or the direct application of 3D neural networks [9]. Other studies have used craniometric datasets as input for estimating biological sex [10].

Research on the analysis of local skull regions is also limited. Some studies have focused on brain tumor detection using two-dimensional (2D) image slices, primarily obtained from Magnetic Resonance Imaging (MRI) scans [11]. These studies frequently employ 2D Convolutional Neural Networks (CNNs) for classification tasks [12], [13]. In addition, spatio-temporal models have been explored to track the evolution of features over time, enabling dynamic and context-rich detection [14]. Other segmentation-focused studies utilize single or multiple views of regions of interest to improve the segmentation accuracy, as demonstrated in [13], [15], [16].

Some studies have also addressed skull fracture detection using 2D slices from CT scans [17], [18]. As an example, the study presented in [19] employs a slice-based approach to isolate the skull from surrounding tissues for further analysis.

To the best of our knowledge, no prior work has specifically addressed the identification of non-metric cranial traits from 2D slices.

III. Materials and Methods

We propose a novel automatic method for classifying CT scan slices of the skull into two categories based on the presence of a non-metric cranial trait: present (class 1) and absent (class 0). The method is based on a binary classifier implemented using deep neural networks, which takes as input local regions extracted from CT scan slices.

To validate our approach, we created a labeled dataset. Specifically, an expert anthropologist manually annotated the CT scans using our custom software, which then extracted local regions centered on the annotated coordinates.

A. Collection of the CT scans dataset

For the dataset collection, we collaborated with Italian hospitals to gather a total of 84 CT scans from 84 individuals, including 44 males and 40 females of unknown ethnicity. The scans have been performed between 2015 and 2020 [20].

B. Dataset Annotation

The annotation task was performed by anthropologists using a custom Python-based tool. Each trait was identified within the scan from a specific view (axial, sagittal, or coronal) by examining the skull from the appropriate angle. The tool allows users to load patient data, interactively switch between views, scroll through slices, zoom in and out, and annotate the center and bounding box of the trait.

The following traits were annotated:

- *Pneumatization of the crista galli*,
- *Pneumatization of the clinoid processes*,
- *Pneumatization of the greater wings*.

For this preliminary study, only one trait was selected for classification: *pneumatization of the crista galli*, as it was the only feature consistently present in a sufficient number of patients. Future studies based on a larger dataset are expected to allow for the inclusion and evaluation of additional non-metric cranial traits.

The trait *pneumatization of the crista galli* appears as a 3D cavity in a specific region of the skull located between the eyes. When CT scans are sliced into 2D images, the trait presents as a circular-like structure that varies in size and shape across different slices. However, the skull contains numerous similar cavities, making it challenging to correctly identify the trait based solely on appearance.

To distinguish *pneumatization of the crista galli* from other cavities, we rely on the consistent wing-shaped bone structure that surrounds the feature. This anatomical configuration remains constant across slices and provides a reliable reference for localization.

Because the traits appeared in multiple slices, annotating each slice individually would have been time-consuming. Instead, only the first and last slices in which the trait was visible were annotated. Assuming a roughly spherical shape for the region surrounding a non-metric cranial trait, it is reasonable to infer that the trait would also be present in the intermediate slices. Therefore, we estimated the center of the region of interest in each intermediate slice by linearly interpolating between the manually annotated centers of the first and last slices.

Out of a total of 84 patients, 7 exhibited the trait *pneumatization of the crista galli*.

C. Creation of the samples

The input samples for the classifier consist of local image regions automatically cropped from CT slices. These image crops include contextual information from the surrounding anatomy to ensure accurate identification of the non-metric cranial trait. As illustrated in Fig. 1, regions that are too small do not allow for proper differentiation of the target trait from other skull cavities of similar shape.

TABLE I
SLICES DISTRIBUTION FOR THE CLASS 1 INDIVIDUALS

Patient ID	**Number of slices**
2	19
29	11
35	32
40	20
53	11
69	11
70	12
Total	116

Different procedures were used to obtain samples for class 1 and class 0.

For class 1, we cropped square local regions from the annotated slices. The center of each crop corresponds to the annotated center coordinates of the trait. Since the CT scans were acquired using different equipment and settings, the crop size was defined as a percentage of the skull width. Specifically, for the *pneumatization of the crista galli*, we extracted square crops with side lengths equal to 35% of the skull's width. This percentage was empirically determined to ensure that both the non-metric cranial trait and its anatomical context were adequately captured. The complete data acquisition pipeline for class 1 is illustrated in Fig. 2.

For class 0, we empirically identified a set of skull slices where the trait could potentially appear if present. To ensure consistency, we first reoriented the scans, removed black margins, and extracted slices corresponding to 40% to 65% of the skull's depth. The crop dimensions matched those of class 1, with side lengths equal to 35% of the skull's width.

Class 1 includes 116 samples obtained from 7 patients, while class 0 includes 836 samples obtained from 77 patients. Table I summarizes the number of annotated slices per patient.

Data loading, orientation, and extraction were performed using the 3D Slicer Python framework [21].

Since the CT scans were acquired using different machines, the extracted crops for both classes were resized to 224×224 pixels.

D. Deep neural networks

The proposed cranial trait detection method employs a binary classifier that takes a single input image from the dataset and classifies it as belonging to either class 1 or class 0. We evaluated different deep neural network architectures, each pre-trained on large-scale image datasets. The selected models are designed to accept input images with a resolution of 224×224 pixels, which aligns with the dimensions of our processed data. These networks were originally trained on the ImageNet dataset [22], which comprises $14,197,122$ labeled images across a wide range of categories.

The considered deep neural network architectures are AlexNet [9], Vgg16 [23], MobileNetV2 [24], ResNet18, ResNet34, and ResNet50 [25].

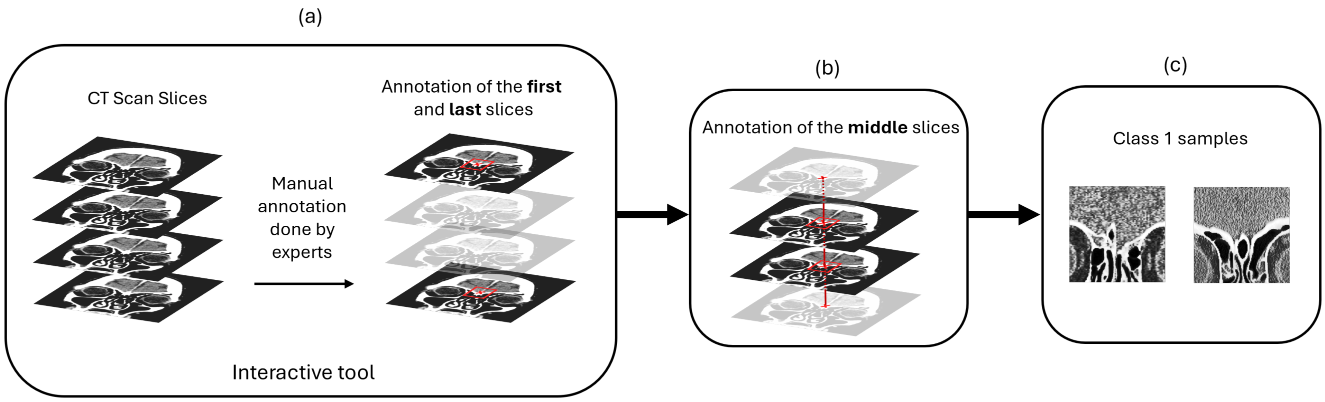

Fig. 2. Overview of the data acquisition pipeline for the class 1 (trait present) samples. (a) Annotation of slices using our custom Python tool, under the supervision of a forensic anthropologist. (b) Annotation of the feature's center for each slice in which the feature is visible. (c) Automatic extraction of the local trait region.

- *AlexNet*: A pioneering deep CNN with 5 convolutional layers and 3 fully connected layers, known for winning the ImageNet competition in 2012.
- *VGG16*: A deep CNN with 16 layers, using 3×3 convolution filters and 2×2 max-pooling layers, known for its simple architecture and high accuracy.
- *MobileNetV2*: A lightweight CNN that uses depthwise separable convolutions and inverted residual blocks, designed for efficient mobile and edge device deployment while maintaining high accuracy.
- *ResNet18*: A residual network with 18 layers that utilizes skip connections to mitigate the vanishing gradient problem, enabling deeper and more efficient training.
- *ResNet34*: A deeper version of ResNet18, with 34 layers, utilizing residual connections to combat the vanishing gradient problem, allowing for deeper models while maintaining high training efficiency.
- *ResNet50*: A deeper residual network with 50 layers, featuring bottleneck architecture for improved performance, enabling very deep networks without degrading performance due to vanishing gradients.

We chose these networks due to their proven effectiveness in image classification tasks and their generalization ability across different datasets.

IV. Experimental Results

This preliminary study aims at assessing whether the proposed method is capable of learning effective feature representations. In this section, we present the validation strategy, training procedure, and obtained results.

A. Validation strategy

A key challenge arises from the spatial proximity of CT slices from the same patient's skull. As a result, images from a single individual tend to be highly similar, which could lead to overly optimistic performance estimates if such samples are included in both the training and test sets.

To mitigate this issue, we partitioned the dataset at the patient level, ensuring that all slices from a given individual were assigned exclusively to either the training or test set.

Additionally, to address class imbalance and ensure a representative distribution of samples, we applied stratified sampling based on class proportions. This approach preserved the overall ratio of class 0 and class 1 in each subset.

To evaluate the model's generalization capability across the dataset, we employed 7-fold cross-validation. Given that the dataset includes 7 individuals of class 1, each fold contained slices from 1 individual of class 1 and slices from 11 individuals of class 0.

B. Training procedure

During the training process, the deep neural networks have been optimized using the Adam Optimizer, with a learning rate set to 1×10^{-4} and a batch size of 512. We experimented with two versions of the loss function. The first version uses the standard Binary Cross Entropy, which treats all classes equally and does not account for class imbalance. In this case, the loss function calculates the error based on the predictions without modifying the weights of different classes.

To address the issue of class imbalance, we also explored a second version of the loss function, where we introduced class weights. These weights are designed to adjust the importance given to each class during the training process. Specifically, we calculated the weights using the following formula:

$$w_c = \frac{N}{K \cdot N_c} \ , \tag{1}$$

where w_c is the class weight for class c, N is the total number of samples, K is the number of classes, and N_c is the number of samples in class c.

We evaluated different data augmentation techniques to enhance the model performance, including histogram equalization, Gaussian noise, and horizontal flipping. However, none of these augmentations led to significant improvements in the model's performance, so they were excluded from the final testing phase.

TABLE II
PERFORMANCE COMPARISON WITHOUT APPLYING CLASS WEIGHTS TO THE LOSS FUNCTION

Model	Accuracy (%)	TPR	TNR
AlexNet	85.40	0.086	0.961
VGG16	87.71	0.052	0.992
MobileNetV2	86.24	0.017	0.980
ResNet18	84.77	0.147	0.945
ResNet34	88.87	0.121	0.995
ResNet50	87.82	0.086	0.988

TABLE III
PERFORMANCE COMPARISON APPLYING CLASS WEIGHTS TO THE LOSS FUNCTION

Model	Accuracy (%)	TPR	TNR
AlexNet	70.69	0.780	0.700
VGG16	76.68	0.800	0.760
MobileNetV2	74.26	0.910	0.720
ResNet18	82.14	0.410	0.880
ResNet34	71.85	0.530	0.750
ResNet50	86.76	0.390	0.930

C. Results

We evaluated the performance of our method using the following figures of merit: Accuracy, True Positive Rate (TPR), and True Negative Rate (TNR).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \,, \tag{2}$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \,, \tag{3}$$

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \,. \tag{4}$$

Where:

- TP (True Positive) is the number of instances where the model correctly predicts a positive class.
- TN (True Negative) is the number of instances where the model correctly predicts a negative class.
- FP (False Positive) is the number of instances where the model incorrectly predicts a positive class when the actual class is negative.
- FN (False Negative) is the number of instances where the model incorrectly predicts a negative class when the actual class is positive.

Table II summarizes the results obtained without applying class weights to the loss function, while Table III presents the results with class weights applied.

The comparison between Table II and Table III shows that applying class weights improves the classifier's ability to detect the minority class (class 1), as indicated by an increase in the TPR across all models. Without class weighting, models tend to achieve higher overall accuracy and TNR, but at the cost of significantly reduced performance in identifying class 1 samples. In the absence of class weights, ResNet34 achieved the highest overall accuracy (88.87%) and TNR (0.995). Using class weights, models such as MobileNetV2 demonstrated a significant improvement in TPR (0.910), while achieving inferior accuracy (74.26%) and TNR (0.720).

These results are encouraging and highlight the importance of selecting the loss function based on the specific objectives of the classification task. In scenarios where detecting the minority class is critical, applying class weights can significantly enhance model sensitivity at the expense of other performance metrics.

V. CONCLUSIONS

This paper presents the first study in the literature focused on the automatic detection of non-metric cranial traits from Computed Tomography (CT) scans. Our method represents an initial step toward the automatic extraction of a feature vector, termed SkullCode, designed for identity recognition in forensic applications.

We constructed a dataset of annotated CT scans from 84 individuals. We evaluated our method using 7-fold cross-validation and compared its performance under two different loss functions. The best-performing model in terms of overall accuracy was ResNet34, achieving an accuracy of 88.87%. Using class weights, MobileNetv2 demonstrated a notable improvement in True Positive Rate (TPR = 0.910), indicating enhanced sensitivity to the minority class. These promising results support the feasibility of automatic detection of SkullCode traits from medical imaging.

Future work will focus on expanding the dataset, improving model architectures, and incorporating additional features to increase the robustness and generalizability of the SkullCode framework across diverse populations and imaging conditions. Furthermore, ongoing research will aim to include data from various imaging modalities and to annotate and classify a broader range of non-metric cranial traits.

ACKNOWLEDGMENTS

This work was supported in part by the European Union – Next-GenerationEU – National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment N. 1.1, Call PRIN 2022 PNRR D.D. 1409 14-09-2022 – FRANTIC – FRequency of ANatomical Traits for personal IdentifiCation – CUP: 53D23004890006. This work was supported in part also by the EC under project Chips JU EdgeAI (101097300) and by project SERICS (PE00000014) under the MUR NRRP funded by the EU - NGEU. Project EdgeAI is supported by the Chips Joint Undertaking and its members including top-up funding by Austria, Belgium, France, Greece, Italy, Latvia, Netherlands, and Norway. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union, the Chips Joint Undertaking, or the Italian MUR. Neither the European Union, nor the granting authority, nor Italian MUR can be held responsible for them.

References

[1] A. Palamenghi, A. Borlando, D. D. Angelis, C. Sforza, C. Cattaneo, and D. Gibelli, "Exploring the potential of cranial non-metric traits as a tool for personal identification: the never-ending dilemma," *International Journal of Legal Medicine*, vol. 135, pp. 2509–2518, Nov 2021.

[2] E. Caloro, M. Cè, D. Gibelli, A. Palamenghi, C. Martinenghi, G. Oliva, and M. Cellina, "Artificial intelligence (AI)-based systems for automatic skeletal maturity assessment through bone and teeth analysis: A revolution in the radiological workflow?," *Applied Sciences*, vol. 13, no. 6, 2023.

[3] A. Palamenghi, A. Aragon-Molina, G. Caccia, D. Mazzarelli, S. Alemanno, R. Donida Labati, F. Scotti, V. Piuri, C. P. Campobasso, C. Cattaneo, D. De Angelis, and D. Gibelli, "From traditional to innovative: implications of cranial non-metric traits in personal identification," *International Journal of Legal Medicine*, 2025.

[4] A. Thurzo, H. S. Kosnáčová, V. Kurilová, S. Kosmeľ, R. Beňuš, N. Moravanský, P. Kováč, K. M. Kuracinová, M. Palkovič, and I. Varga, "Use of advanced artificial intelligence in forensic medicine, forensic anthropology and clinical anatomy," *Healthcare*, vol. 9, no. 11, 2021.

[5] N. Galante, R. Cotroneo, D. Furci, G. Lodetti, and M. B. Casali, "Applications of artificial intelligence in forensic sciences: Current potential benefits, limitations and perspectives," *International Journal of Legal Medicine*, vol. 137, pp. 445–458, Mar. 2023.

[6] A. Micheau and D. Hoa, "e-anatomy atlas." https://www.imaios.com. DOI: https://doi.org/10.37019/e-anatomy10.37019/e-anatomy.

[7] H. Zhang, Q. Li, X. Wang, Q. Wu, C. Ma, M. Zhou, and G. Geng, "A novel deep neural network for identification of sex and ethnicity based on unknown skulls," *Pattern Recognition*, p. 111450, 2025.

[8] L. Pengyue, X. Siyuan, J. Yi, Y. Wen, L. Xiaoning, G. Guohua, and W. Shixiong, "Aninet: a deep neural network for skull ancestry estimation," *BMC Bioinformatics*, vol. 22, no. 1, p. 550, 2021.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[10] A. Del Bove and A. Veneziano, "A generalised neural network model to estimate sex from cranial metric traits: A robust training and testing approach," *Applied Sciences*, vol. 12, no. 18, 2022.

[11] Y. Xie, F. Zaccagna, L. Rundo, C. Testa, R. Agati, R. Lodi, D. N. Manners, and C. Tonon, "Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): Review, challenges, and future perspectives," *Diagnostics*, vol. 12, no. 8, 2022.

[12] S. Kumar Mishra, D. Kumar, G. Kumar, and S. Kumar, "Multi-classification of brain mri using efficientnet," in *2022 Proc. of the International Conference for Advancement in Technology (ICONAT)*, pp. 1–6, 2022.

[13] V. Rajinikanth, S. Kadry, and Y. Nam, "Convolutional-neural-network assisted segmentation and svm classification of brain tumor in clinical mri slices," *Information Technology and Control*, vol. 50, no. 2, pp. 342–356, 2021.

[14] S. Chatterjee, F. A. Nizamani, A. Nürnberger, and O. Speck, "Classification of brain tumours in mr images using deep spatiospatial models," *Scientific Reports*, vol. 12, no. 1, p. 1505, 2022.

[15] B. Divya, R. P. Nair, K. Prakashini, R. Girish Menon, P. Litvak, P. Mandava, D. Vijayasenan, and S. Sumam David, "A more generalizable dnn based automatic segmentation of brain tumors from multimodal low-resolution 2d mri," in *2021 Proc. of the IEEE 18th India Council International Conference (INDICON)*, pp. 1–5, 2021.

[16] Z. Liu, C. Ma, W. She, and X. Wang, "Transmvu: Multi-view 2d u-nets with transformer for brain tumour segmentation," *IET Image Processing*, vol. 17, no. 6, pp. 1874–1882, 2023.

[17] Y. M. Liew, J. H. Ooi, R. R. Azman, D. Ganesan, M. I. Zakaria, A. S. Mohd Khairuddin, and L. K. Tan, "Innovations in detecting skull fractures: A review of computer-aided techniques in ct imaging," *Physica Medica*, vol. 124, p. 103400, 2024.

[18] A. Mangrulkar, S. Rane, and V. Sunnapwar, "Automated skull damage detection from assembled skull model using computer vision and machine learning," *International Journal of Information Technology*, vol. 13, pp. 1785–1790, 2021.

[19] A. Fatima, T. Madni, F. Anwar, *et al.*, "Automated 2d slice-based skull stripping multi-view ensemble model on nfbs and ibsr datasets," *Journal of Digital Imaging*, vol. 35, pp. 374–384, 2022.

[20] A. Palamenghi, M. Cellina, M. Cè, A. Cappella, C. Sforza, and D. Gibelli, "Correlation analysis on anatomical variants of accessory foramina in the sphenoid bone for oncological surgery," *Cancers*, vol. 15, no. 22, 2023.

[21] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. M. Fennessy, M. Sonka, J. Buatti, S. R. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging*, vol. 30, pp. 1323–1341, Nov 2012.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.