# CMInet: Automated Assessment of Physiological Cerebellar Mobility in Upright and Supine MRI

Jiamin Wang[1,†], Shiying Ke[2,†], Xiaoying Yang[2], Fuhua Jia[2], Kun Yan[3], Tianxiang Cui[2,*], Yulin Wang[2,*], and Chengbo Wang[2,*]

[1]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[2]Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo, China
[3]Department of Radiology, Ningbo No. 2 Hospital, Ningbo, China

***Abstract*—Chiari type I malformation (CM-I) is characterized by the caudal herniation of cerebellar tonsils through the foramen magnum. Diagnosis traditionally relies on measuring tonsillar position relative to the McRae line, a manual process prone to inter-observer variability. While AI has advanced skeletal landmark prediction in CT, applying these methods to MRI remains challenging due to low osseous contrast, creating a dependency on radiation-based imaging. Furthermore, existing models often encounter a physical perception bottleneck, struggling to map high-dimensional sensor data to precise spatial physical coordinates. To address this and ensure scientific rigor, we developed CMInet—the first fully MRI-based framework integrating cerebellar segmentation with direct osseous landmark regression—by adhering to a standardized and reproducible research framework. This approach, tailored for the unique challenges of multi-positional MRI analysis, eliminates the need for supplementary CT scans and streamlines clinical workflows. Leveraging a rare, hardware-limited paired dataset from specialized upright MRI, our model effectively captures subtle gravity-dependent structural changes. The proposed framework delivers an objective tool for anatomical assessment, reducing manual workload while establishing a critical normative baseline to distinguish physiological mobility from pathological herniation.**

***Index Terms*—Chiari I malformation, MRI, Deep learning, Cerebellar segmentation, McRae line**

## I. Introduction

Chiari type I malformation (CM-I) is defined as the caudal herniation of the cerebellar tonsils through the foramen magnum [1]. With a prevalence of 0.1–1%, CM-I imposes a significant clinical burden due to associated neurological disability [2]. As illustrated in Fig. 1, cerebellar height is measured relative to the McRae line, defined as the maximum perpendicular distance when the tonsils intersect the line, or the shortest distance to the margin otherwise. Current diagnostic protocols rely on manual measurement [3], a labor-intensive process prone to inter-observer variability of 0.5–3.1 mm. Even with double reading, discrepancies of up to 7.7 mm and an inter-class correlation of only 0.83 persist [4]. These limitations underscore the critical need for automated methods to enhance efficiency, reproducibility, and clinical reliability [5].

Existing automated approaches reveal a significant modality gap: while CT offers robust osseous landmark detection, MRI is the standard for cerebellar segmentation [6]. However, delineating osseous boundaries in MRI is non-trivial due to a signal-to-noise ratio markedly lower than that of CT [7]. Consequently, standard segmentation networks trained on general medical images [8]–[10] often fail to capture these subtle osseous landmarks in MRI, and few studies have jointly achieved landmark localization and cerebellar segmentation within this modality. To bridge this gap, we propose CMInet, a framework that integrates deep learning–based basion–opisthion (McRae line) detection with cerebellar segmentation, followed by computer vision algorithms for tonsillar height estimation. This system enables the quantification of cerebellar height and volume across postures, establishing normative baselines for posture-dependent changes.

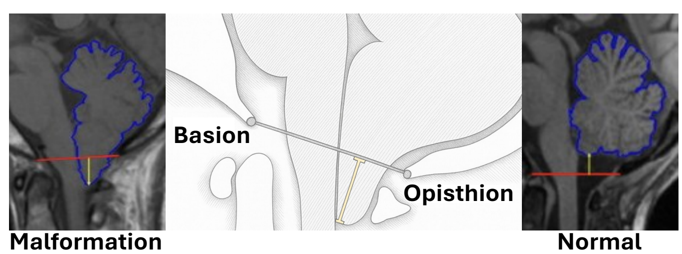


Fig. 1. Manual measurement of tonsillar height [11]

The proposed approach leverages a unique dataset acquired from a rotatable cryogen-free 1.5T MRI system, enabling both supine and upright imaging under stable field conditions [12]. High field strength, a 12-channel head coil, and 3D MP-RAGE sequences provided superior resolution and contrast, revealing posture-dependent cerebellar changes invisible to conventional supine MRI [13]. This work makes three key contributions:

1) We introduce a unified automated framework for cerebellar height measurement on MRI. This framework

†Equal contribution.
* Corresponding authors.

This Project is Supported by the Ningbo Science and Technology Bureau (Ningbo Commonweal Programme, Project ID 2025S114; Ningbo Young Scientific and Technological Innovation Leading Talent Scheme, Project ID 2025QL055).

jointly incorporates upright and supine acquisitions to quantify posture-specific cerebellar morphology without ionizing radiation.

2) We design CMInet as the first network tailored for cerebellar height estimation, demonstrating the feasibility of AI-based measurement of cerebellar height and volume directly from MRI.
3) We construct the first paired upright–supine MRI dataset for cerebellar analysis. Motivated by the observation that Chiari symptoms often worsen in the upright posture [14], and leveraging rare rotatable MRI hardware, we provide a normative baseline for investigating gravity-dependent structural changes, which is a resource currently unavailable in standard multi-center datasets.

## II. The Proposed CMInet Framework

### *A. Framework for Tonsillar Height and Cerebellar Volume Estimation*

We propose CMInet, a dual-branch framework designed for automated cerebellar landmark detection and volumetric analysis. As illustrated in Fig. 2, the system processes sagittal MRI scans acquired in both upright and supine positions to quantify anatomical changes under gravitational loading.

*1) Geometric Definition of Tonsillar Height:* The Landmark Prediction module first localizes the coordinates of the basion ($P_b = (x_b, y_b)$) and the opisthion ($P_o = (x_o, y_o)$). These points define the McRae line $L$, which serves as the stable osseous reference for all subsequent measurements. The line can be represented by the general equation $Ax+By+C=0$, derived from the two landmark coordinates.

Concurrently, the Segmentation module generates a binary cerebellum mask $M \in \{0,1\}^{H\times W}$. The tonsillar contour $\mathcal{C}$ is extracted from the caudal-most boundary of $M$. To determine the tonsillar height H, we first calculate the perpendicular distance $d(p_i, L)$ for every pixel $p_i(x_i, y_i)$ along the contour $\mathcal{C}$ relative to the McRae line:

$$d(p_i, L) = \frac{|Ax_i + By_i + C|}{\sqrt{A^2 + B^2}} \tag{1}$$

We establish a coordinate system where the McRae line represents the zero-reference plane ($H = 0$). Let $\mathcal{C}$ and $L$ denote the sets of all pixel coordinates comprising the tonsillar contour and the line, respectively. The final tonsillar height is conditionally determined based on the spatial relationship between these two sets:

$$H = s \times \begin{cases} +\min_{p_i \in \mathcal{C}} \left(d(p_i, L)\right) & \mathcal{C} \cap L = \emptyset \\ -\max_{p_i \in \mathcal{C}} \left(d(p_i, L)\right) & \mathcal{C} \cap L \neq \emptyset \end{cases} \tag{2}$$

where s represents the pixel resolution (mm/pixel). In the first case, where the tonsil is positioned normally above the foramen magnum, we select the *shortest* perpendicular distance to the line across all slices to quantify the cerebellar height. In the second case, where the tonsil is positioned below the standard, we select the *longest* perpendicular distance to the line across all slices to quantify the maximum extent of cerebellar descent. Consistent with clinical conventions, positive values indicate a position superior to the McRae line, while negative values indicate descent into the spinal canal.

*2) Volumetric Quantification:* The middle panel of Fig. 2 depicts the segmentation process. Sagittal MRI slices are processed to yield a dense probability map, which is binarized to suppress background noise. To quantify the total cerebellar volume $V_{cereb}$, we aggregate the segmented voxels across the entire slice stack. Let $M^{(k)}$ denote the binary mask for the $k$-th slice in a scan of $N$ slices. The volume is calculated as:

$$V_{cereb} = \sum_{k=1}^{N} \sum_{i,j} M_{i,j}^{(k)} \times v_x \times v_y \times v_z \tag{3}$$

where $v_x, v_y$ represent the in-plane pixel spacing and $v_z$ represents the slice thickness. The top-right panel of Fig. 2 displays the resulting 3D reconstructions, enabling posture-specific morphological analysis. This dual-metric approach allows CMInet to capture both linear displacement (tonsillar descent) and volumetric deformation (tissue compression) induced by gravity [15], [16].

### *B. Landmark Prediction (Basion and Opisthion)*

Unlike CT imaging, where osseous structures present high contrast, bone boundaries in MRI typically exhibit weak signals and low signal-to-noise ratios (SNR), rendering direct segmentation or heatmap-based detection challenging. To address this, we employ a regression-based network rather than a standard U-Net segmentation approach. As shown in Fig. 2, we retain the encoder for hierarchical feature extraction but replace the decoder with a lightweight channel-compression and fully connected (FC) scheme for direct coordinate regression. At the encoder's final stage, the feature tensor $T \in \mathbb{R}^{C\times H'\times W'}$ is flattened into a vector $z \in \mathbb{R}^d$ ($d = C{\cdot}H'{\cdot}W'$), which is then passed through two FC layers to produce the four-dimensional coordinate output.

The flattening operation is critical to our design. Unlike global average pooling, which aggregates and thus discards spatial details, flattening preserves the spatial arrangement of the feature map entries:

$$z = \mathrm{vec}(T), \quad z_n = T_{c,i,j} \tag{4}$$

where the index $n$ maps uniquely to spatial position $(i, j)$ and channel $c$. This bijective mapping allows the subsequent FC layers to learn location-specific weights across the feature field, effectively translating hierarchical spatial features directly into landmark coordinates. This modification offers three key benefits: (i) replacing the upsampling decoder with 1×1 convolutions and FC layers significantly reduces parameter count; (ii) flattening efficiently links spatial features to compact outputs, which is advantageous for learning from limited medical datasets; and (iii) the channel-compression bottleneck mitigates overfitting while preserving semantic information. Collectively, these adaptations enable accurate coordinate regression with superior efficiency compared to conventional U-Net decoders.

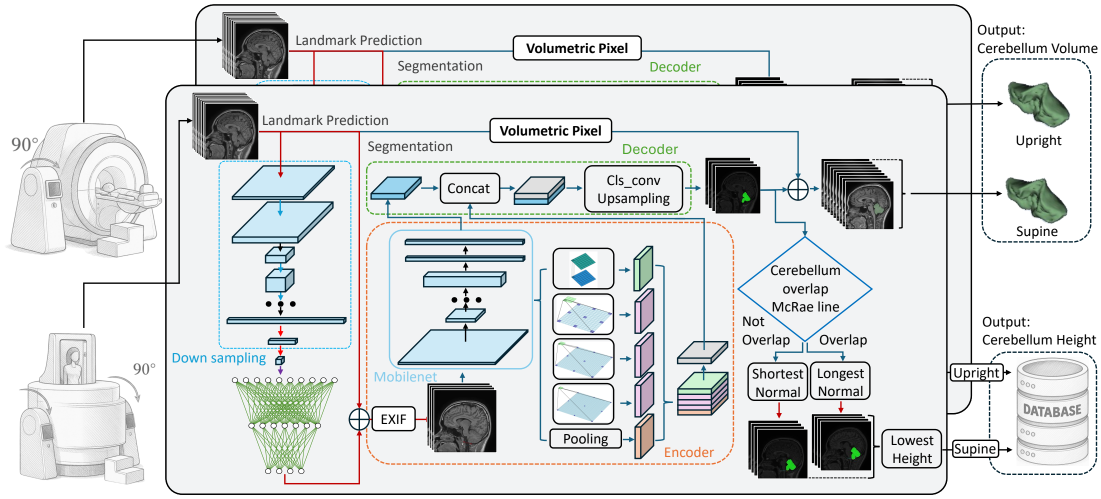


Fig. 2. Overview of the proposed framework. Upright and supine MRI scans are obtained by rotating the scanner 90°. Each scan is processed by a landmark module for osseous references and a MobileNet–DeepLabV3+ network for cerebellar segmentation. Volume is derived from voxel-scaled pixel aggregation, while tonsillar height is measured relative to the McRae line. Outputs include posture-specific 3D reconstructions and height estimates, with upright–supine results synthesized for integrated evaluation.

### C. Cerebellum Segmentation

Standard CNN encoders often struggle to delineate the fuzzy boundaries between the cerebellum and adjacent low-contrast osseous tissues in MRI. While deep networks like ResNet [17] offer high accuracy, their computational cost is often prohibitive for mobile or embedded deployment. Conversely, lightweight models must balance efficiency with the capacity to capture subtle anatomical details.

To address these challenges, we employ a DeepLabV3+ network with a MobileNetV2 backbone, as shown in Fig. 2. This combination was specifically selected to optimize the trade-off between segmentation accuracy and model complexity. The MobileNetV2 backbone utilizes inverted residual structures and linear bottlenecks to significantly reduce the parameter count while preserving representational capacity [18]. Crucially, to handle fuzzy anatomical boundaries, the DeepLabV3+ architecture leverages Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale contextual information [19]. This is further enhanced by the decoder module, which fuses low-level features with high-level semantics to refine boundary delineation [20], ensuring precise segmentation of the cerebellum from surrounding soft tissues.

## III. Experimental Results

### A. Multi-Postural MRI Dataset and Implementation Details

To mitigate scanner-specific variations and standardize intensity distributions, we applied Z-score normalization to all input volumes. Specifically, each MRI scan was normalized by subtracting the mean and dividing by the standard deviation of the brain voxel intensities. This step is critical for stabilizing model convergence given the single-source nature of the dataset. Ground truth annotations for the basion and opisthion were manually performed by two experienced radiologists using LabelMe software, with discrepancies resolved by consensus. Cerebellar masks were manually refined from preliminary segmentations to ensure anatomical precision.

Given the limited dataset size ($N = 34$), we employed extensive on-the-fly data augmentation to enhance model invariance and prevent overfitting. This included: 1) Geometric Transformations: Random rotations ($\pm 15°$), scaling ($0.8\times$–$1.2\times$), and translations to simulate patient positioning variability; 2) Intensity Perturbations: Random gamma correction and brightness contrast adjustment to mimic different scanner protocols; and 3) Elastic Deformations: Applied to simulate non-rigid anatomical variations.

The Segmentation model was trained using SGD with a momentum of 0.9 and weight decay of $1\times10^{-4}$. We employed a cosine annealing learning rate schedule, decaying from $7 \times 10^{-3}$ to $7 \times 10^{-5}$ over 200 epochs, with a batch size of 16. Crucially, unlike the training-from-scratch paradigm (e.g., nnU-Net), the MobileNetV2 backbone was initialized with ImageNet pre-trained weights and fine-tuned without freezing. This transfer learning strategy provided robust feature priors, enabling effective convergence despite the limited sample size [21]–[23]. In parallel, the Landmark Prediction model utilized the Adam optimizer with a fixed learning rate of $1\times10^{-4}$ and a batch size of 4 for 510 epochs, optimizing the Mean Squared Error (MSE) loss for precise coordinate regression [24]–[26].

The dataset was partitioned into training (80%), validation (10%), and testing (10%) sets. Crucially, to prevent data leakage in our paired dataset, the split was performed strictly at the subject level. This ensures that the upright and supine scans of any given volunteer are exclusive to a single subset (e.g., if Subject A is in the training set, both their positional scans are used solely for training). This rigorous separation guarantees that the test metrics reflect the model's ability to generalize to unseen anatomical structures, rather than memorizing subject-specific features. For nnU-Net, we utilized its standard self-configuring pipeline ("2d_fullres" mode) without manual intervention to reflect its representative "out-of-the-box" performance [27].

### B. Segmentation Comparison

TABLE I
SEGMENTATION PERFORMANCE AND EFFICIENCY COMPARISON

| Method | Params (M) | FLOPs (G) | Precision | IoU | PA |
|---|---|---|---|---|---|
| U-Net | 31.03 | 60.50 | 0.78 | 0.71 | 0.82 |
| nnU-Net | 31.03 | 60.50 | 0.98 | 0.84 | 0.85 |
| **CMInet** | **3.50** | **4.50** | 0.96 | **0.94** | **0.98** |

We compare CMInet with U-Net and nnU-Net, two widely used benchmarks in medical image segmentation [28]. These models represent a lightweight yet boundary-limited design and an automated but computationally intensive framework, respectively [29] [27]. Table I presents the results using three standard metrics: Precision, Intersection over Union (IoU), and Pixel Accuracy (PA). As detailed in Table I, CMInet achieves a 13$\times$ reduction in computational cost (3.5M parameters vs. 31M for baselines) while delivering superior accuracy. U-Net, lacking multi-scale feature extraction, struggles with boundary capture, resulting in IoU and PA values 32% and 20% lower than ours [29], [30]. While nnU-Net yields high precision (0.98), it exhibits markedly lower IoU (-12%) and PA (-15%). This suggests that its heuristic design tends to over-smooth subtle anatomical transitions in low-contrast MRI. In contrast, CMInet's ASPP module effectively captures multi-scale context, reducing computational resources by 13.4x FLOPs and parameters by 8.9x in our setup compared to U-Net/nnU-Net while maintaining high accuracy.

### C. Landmark Comparison

To validate the effectiveness of our approach, we benchmark landmark detection against two representative baselines: YOLOv8 [31] (object detection-based) and MMPose [32] (pose estimation-based). Performance is evaluated using three quantitative metrics calculated in the image coordinate system (pixels): Mean Squared Error (MSE), Defined as the mean of the squared Euclidean distances between predicted and ground truth coordinates, measured in pixels squared ($px^2$). Average Euclidean Distance (Avg Dist), The mean $L_2$ distance between predicted and ground truth landmarks, measured in pixels ($px$). Percentage of Correct Keypoints (PCK@5px) The percentage of predicted keypoints falling within a strict threshold of $\alpha = 5$ pixels from the ground truth.

TABLE II
LANDMARK PERFORMANCE AND EFFICIENCY COMPARISON

| Method | Params (M) | FLOPs (G) | MSE | PCK | Avg Euc Dist |
|---|---|---|---|---|---|
| YOLO v8 | $\sim$3.20 | $\sim$8.10 | 59.53 | 0.15 | 9.89 |
| MMPose | 5.36 | 9.12 | 49.94 | **0.37** | **6.35** |
| **CMInet** | **1.82** | **3.47** | **26.02** | 0.34 | 6.44 |

The results highlight distinct architectural trade-offs. YOLOv8's bounding-box regression inherently limits precision for fine anatomical points, resulting in the highest MSE (59.53). MMPose maximizes PCK but incurs high computational cost (9.12 GFLOPs). CMInet balances these extremes by replacing the decoder with a flattening and FC regression head. As shown in Table II, CMInet achieves the lowest regression error (MSE 26.02), reducing error by 48–56% compared to baselines, while requiring only 1.82M parameters.

### D. Ablation Study

To validate the effectiveness of our specific architectural components, we conducted an ablation study focusing on the Flattening operation in the landmark branch and the ASPP module in the segmentation branch.

TABLE III
ABLATION STUDY ON ARCHITECTURAL COMPONENTS

| Configuration | Landmark MSE ↓ | Seg IoU ↑ |
|---|---|---|
| Baseline (Global Avg Pooling) | 45.12 | - |
| Baseline (w/o ASPP) | - | 0.89 |
| Baseline (w/o Pre-training) | 52.30 | 0.82 |
| **CMInet (Full)** | **26.02** | **0.94** |

As presented in Table III, replacing our Flattening operation with standard Global Average Pooling (GAP) resulted in a sharp increase in MSE (from 26.02 to 45.12). This confirms that GAP discards critical spatial information required for precise coordinate regression, whereas flattening preserves the spatial arrangement of features. Similarly, removing the ASPP module caused a 5% drop in Segmentation IoU, validating its necessity for capturing multi-scale context along fuzzy osseous boundaries. Finally, training from scratch (w/o Pre-training) significantly degraded performance across both tasks, underscoring the importance of transfer learning for this limited dataset.

### E. Visual Results Comparison

Fig. 3 visually corroborates the quantitative findings. CMInet yields landmarks precisely aligned with anatomical references (B), whereas YOLOv8 (A) shows displacement due to its bounding-box limitations. For segmentation, CMInet (D) accurately delineates the cerebellar margin while excluding adjacent osseous tissue. In contrast, U-Net (C) mislabels portions of the occipital bone as cerebellum, reflecting its limited feature hierarchy.

### F. Statistical Analysis of Cerebellar Tonsil Height

To evaluate the significance of posture-dependent changes, we performed a paired t-test between supine and upright measurements (Fig. 4). Among 34 healthy subjects, the number of cases with tonsillar descent below the McRae line increased from 6 in the supine position (mean height: 4.16 $\pm$ 6.08 mm) to 9 in the upright position (2.74 $\pm$ 5.72 mm). Concurrently, cerebellar volume showed a statistically significant decline from 131.49 $\pm$ 12.87 $cm^3$ to 129.66 $\pm$ 13.03 $cm^3$ ($p < 0.001$). Notably, despite overlapping distributions caused by high inter-subject variability, paired analysis reveals a highly consistent within-subject trend ($p < 0.001$). This confirms that posture-dependent shifts are masked in unpaired population

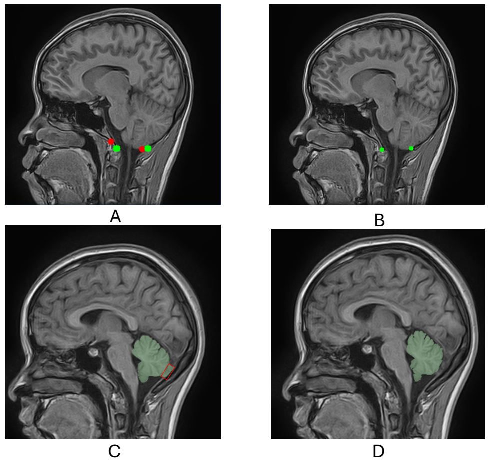


Fig. 3. Visual comparison of landmark detection and segmentation: (A, C) Results from YOLOv8 and U-Net; (B, D) Results from CMInet. Ground truth annotations are shown in green, model predictions in red.

studies but are statistically significant when tracking individual morphology under gravity load.

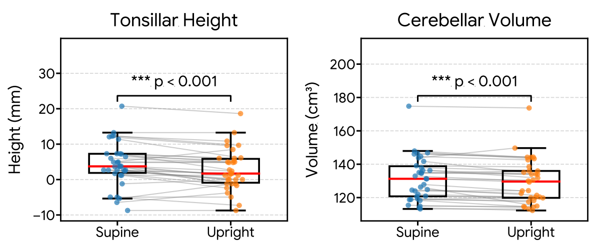


Fig. 4. Boxplots of tonsillar height and cerebellar volume in supine versus upright MRI positions. Individual data points are linked to visualize paired within-subject shifts. While inter-subject variability is high, the paired analysis reveals a consistent, significant trend ($p < 0.001$).

### G. Clinical Validation

To validate the clinical utility of our framework, we assessed the agreement between automated tonsillar height measurements and radiologist-derived ground truth on the test set using Mean Absolute Error (MAE) and Pearson correlation ($r$). CMInet achieved an MAE of 1.14 mm and a strong correlation of $r = 0.95$ ($p < 0.001$). Notably, this error margin is commensurate with the reported inter-observer variability of manual measurements (0.5–3.1 mm), indicating that CMInet achieves clinical-grade precision. Unlike intermediate component metrics such as Segmentation IoU or Landmark MSE, this validation directly confirms the reliability of the integrated diagnostic output for clinical assessment.

## IV. Discussion

### A. Physiological Implications of Postural Dynamics

The most clinically significant finding of this study is the quantifiable impact of gravity on cerebellar morphology. Our paired analysis revealed that transitioning from supine to upright posture caused a statistically significant caudal displacement of the cerebellar tonsils ($p < 0.001$), increasing the prevalence of tonsillar descent below the McRae line (from 6 to 9 cases) within the same healthy cohort. This suggests that standard supine MRI systematically underestimates the degree of tonsillar descent. Consequently, diagnoses based solely on static supine imaging may miss "occult" Chiari malformations that only manifest symptomatically and radiologically under gravitational load [13].

Furthermore, the observed reduction in cerebellar volume in the upright position, while small ($\sim$1.4%), aligns with the Monro-Kellie doctrine regarding intracranial fluid dynamics. We hypothesize that as gravity pulls the neuraxis caudally, cerebrospinal fluid (CSF) is displaced from the cranial subarachnoid space to the spinal canal, resulting in a measurable decrease in intracranial tissue volume.

### B. Bridging the Modality Gap

Historically, precise McRae line definition required CT imaging for osseous contrast, while soft-tissue assessment required MRI. CMInet effectively bridges this gap by demonstrating that deep learning can extract osseous landmarks from MRI with precision comparable to inter-observer variability. By integrating this landmarking capability with segmentation, our framework offers a "one-stop" radiation-free assessment tool. This is particularly valuable for pediatric populations requiring serial monitoring, where cumulative radiation dose from repeated CT scans is a major concern.

### C. Clinical Implications and Limitations

While our results demonstrate robust segmentation and landmark regression, we acknowledge constraints inherent to the experimental scope.

First, the reliance on a single scanner reflects the global scarcity of specialized rotatable MRI hardware [12], [13]. While this currently limits multi-vendor validation for postural analysis, it highlights the uniqueness of the dataset. To minimize potential domain bias, we rigorously employed Z-score normalization and intensity perturbation augmentations, ensuring the model learns anatomical features rather than scanner-specific noise.

Second, the cohort focused exclusively on healthy subjects. This was a deliberate design choice to establish a strictly normative baseline for physiological tonsillar mobility. Quantifying the natural range of gravity-dependent variance is a critical scientific prerequisite before AI can reliably distinguish pathological herniation from normal physiological shifts. Consequently, this study provides the necessary control benchmark for future clinical applications.

Future work will focus on adapting this pipeline to standard clinical MRI via domain adaptation techniques and extending the analysis to symptomatic Chiari populations to validate diagnostic generalizability against this established baseline.

## V. Conclusion

In this work, we presented CMInet, an automated deep-learning framework for the precise quantification of tonsillar height and cerebellar volume directly from MRI. By successfully addressing the challenge of osseous landmark localization in low-contrast MRI, our approach eliminates the reliance on CT, thereby overcoming the modality gap that has historically hindered automated Chiari assessment. Validated on a unique dataset acquired via a rotatable MRI system, the model accurately captured gravity-induced tonsillar descent and volumetric deformation, aligning with biomechanical expectations. This integration of landmark regression and segmentation not only ensures high reproducibility by mitigating inter-observer variability but also provides a normative baseline for posture-dependent cerebellar morphology. Ultimately, CMInet establishes a standardized, objective foundation for investigating Chiari I malformation, paving the way for more robust, computer-aided neuroanatomical analysis in dynamic imaging environments.

## REFERENCES


[1] A. H. Aiken, J. A. Hoots, and A. M. Saindane, "Incidence of cerebellar tonsillar ectopia in idiopathic intracranial hypertension: A mimic of the chiari i malformation," *AJNR. American Journal of Neuroradiology*, vol. 33, no. 10, pp. 1901–1906, 2012.

[2] B. Sadler, T. Kuensting, J. Strahle, T. S. Park, M. Smyth, D. D. Limbrick, M. B. Dobbs, G. Haller, and C. A. Gurnett, "Prevalence and impact of underlying diagnosis and comorbidities on chiari 1 malformation," *Pediatric Neurology*, vol. 106, pp. 32–37, 2020.

[3] A. J. Barkovich, F. J. Wippold, J. L. Sherman, and C. M. Citrin, "Significance of cerebellar tonsillar position on mr," *AJNR. American Journal of Neuroradiology*, vol. 7, no. 5, pp. 795–799, 1986.

[4] K. W. Tanaka, C. Russo, S. Liu, M. A. Stoodley, and A. Di Ieva, "Use of deep learning in the mri diagnosis of chiari malformation type i," *Neuroradiology*, vol. 64, no. 8, pp. 1585–1592, 2022.

[5] Y. Li, L. Jiayi, X. Yang, C. Li, X. Xiong, Y. Fang, S. Ding, and T. Cui, "Low-cost automated visual screw inspection system," in *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2023, pp. 1735–1740.

[6] L. Xing, M. Chen, J. Yao, Y. Zhang, and Y. Wang, "Pre-post interaction learning for brain tumor segmentation with missing mri modalities," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 1711–1715.

[7] L. Morbée, E. Vereecke, F. Laloo, M. Chen, N. Herregods, and L. Jans, "Mr imaging of the pelvic bones: The current and cutting-edge techniques," *Journal of the Belgian Society of Radiology*, vol. 106, no. 1, p. 123, 2022.

[8] E. Chai, S. Chen, T. Zhang, X. Li, and T. Cui, "Ssmamba: A self-supervised hybrid state space model for pathological image classification," *Medical Image Analysis*, vol. 111, p. 104080, 2026.

[9] F. B. Tesema, A. Guerra-Manzanares, T. Cui, Q. Zhang, M. M. Solomon, and X. He, "Lgps: A lightweight gan-based approach for polyp segmentation in colonoscopy images," *Biomedical Signal Processing and Control*, vol. 118, p. 109777, 2026.

[10] E. Chai, X. Gu, Z. Lu, and T. Cui, "Cellmixer: Pathological image classification using dual-branch vmamba with randomly mixing gradient features data augmentation," *Expert Systems with Applications*, p. 131298, 2026.

[11] B. S. T. Nwotchouang, M. S. Eppelheimer, A. Ibrahimy, J. R. Houston, D. Biswas, R. Labuda, J. R. Bapuraj, P. A. Allen, D. Frim, and F. Loth, "Clivus length distinguishes between asymptomatic healthy controls and symptomatic adult women with chiari malformation type i," *Neuroradiology*, vol. 62, no. 11, pp. 1389–1400, 2020.

[12] Y. Wang, S. Ke, J. Li, J. Zeng, J. Zhang, S. Niu, C. Yao, J. Zheng, T. Meersmann, G. Liu, H. Zhu, L. Zhao, and C. Wang, "Assessing morphological changes in the choroid plexus between standing and supine positions using a rotatable mri system," *Scientific Reports*, vol. 15, no. 1, p. 22329, 2025.

[13] S. Ke, Y. Wang, J. Zhang, J. Zeng, S. Niu, J. Zheng, T. Meersmann, and C. Wang, "Assessing the impact of posture on brain volume in healthy subjects with a rotatable cryogen-free 1.5t superconducting mri," *Frontiers in Neuroscience*, vol. 19, p. 1644236, 2025.

[14] S. K. P. Tam, J. Chia, A. Brodbelt, and M. Foroughi, "Assessment of patients with a chiari malformation type i," *Brain and Spine*, vol. 2, no. 2, p. 100850, 2021.

[15] E. Chai, L. Chen, L. Wei, and T. Cui, "Harmonizing classification and localization in small object detection," in *International Conference on Neural Information Processing*. Springer Nature Singapore Singapore, 2025, pp. 426–440.

[16] F. Jia, X. Yang, J. Wang, N. Xue, J. Li, and T. Cui, "Marker-free multi-modal motion capture for 6-dof object position and orientation estimation," in *2025 IEEE Symposium on Computational Intelligence in Image, Signal Processing and Synthetic Media (CISM)*. IEEE, 2025, pp. 1–7.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

[19] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[20] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

[21] G. E. Humpire-Mamani, C. Jacobs, B. van Ginneken, and N. Lessmann, "Transfer learning from a sparsely annotated dataset of 3D medical images," *arXiv preprint arXiv:2311.05032*, Nov. 2023.

[22] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27. Curran Associates, Inc., Dec. 2014, pp. 3320–3328.

[23] F. Jia, K. Yang, J. Xiao, T. Hu, X. Yang, A. Rushworth, H. Yu, and T. Cui, "Safety-driven amr end-to-end navigation framework based on sparse sensor human behavior prediction," in *2025 IEEE Symposium on Computational Intelligence on Engineering/Cyber Physical Systems (CIES)*. IEEE, 2025, pp. 1–7.

[24] E. Chai, X. Li, T. Cui, Z. Lu, and F. B. Tesema, "Accelerating convergence in bounding box regression with a refined iou loss function," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[25] E. Chai, W. Yu, T. Cui, J. Ren, and S. Ding, "An efficient asymmetric nonlinear activation function for deep neural networks," *Symmetry*, vol. 14, no. 5, p. 1027, 2022.

[26] X. Gu, C. Gao, Z. Lu, and T. Cui, "You get what you focus on: A weighting factor for iou-based regression loss," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[27] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[28] B. Zhao, Y. Ge, G. Yang, B. Dong, X. Luo, and et al., "The state-of-the-art in cardiac mri reconstruction: Results of the cmrxrecon challenge," *Medical Image Analysis*, vol. 97, p. 103156, 2024.

[29] J. Long *et al.*, "Learning u-net based multi-scale features in encoding-decoding framework for brain tissue segmentation of mri," *Sensors*, vol. 21, no. 9, p. 3232, 2021.

[30] C. Wang, H. Ning, X. Chen, and S. Li, "Db-unet: MLP based dual branch UNet for accurate vessel segmentation in OCTA images," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[31] M. Kang, C.-M. Ting, F. F. Ting, and R. C.-W. Phan, "BGF-YOLO: Enhanced YOLOv8 with multiscale attentional feature fusion for brain tumor detection," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2024*, ser. Lecture Notes in Computer Science (LNCS), vol. 15008. Springer Nature Switzerland, Oct. 2024, pp. 35–45.

[32] E. Stansfield, J. A. Mitterer, A. Altahhan *et al.*, "Landmark detection for medical images using a general-purpose segmentation model," *arXiv preprint arXiv:2507.11551*, 2025.