AJNR

This information is current as of August 5, 2025.

Automated Midline Shift Detection in Head CTs Using Localization and Symmetry Techniques Based on User Selected Slice

Nooriel E. Banayan, Hrithwik Shalu, Vaios Hatzoglou, Nathaniel Swinburne, Andrei Holodny, Zhigang Zhang and Joseph Stember

AJNR Am J Neuroradiol published online 11 April 2025
http://www.ajnr.org/content/early/2025/04/10/ajnr.A8767

This represents the accepted version of the manuscript and also includes the supplemental material; it differs from the final version of the article.

ORIGINAL RESEARCH

Automated Midline Shift Detection in Head CTs Using Localization and Symmetry Techniques Based on User Selected Slice

Nooriel E. Banayan PhD, Hrithwik Shalu MS, Vaios Hatzoglou MD, Nathaniel Swinburne MD, Andrei Holodny MD, Zhigang Zhang PhD, Joseph Stember MD, PhD

ABSTRACT

BACKGROUND AND PURPOSE: Midline shift is an intracranial pathology characterized by the displacement of brain parenchyma across the skull's midsagittal axis, typically caused by mass effect from space-occupying lesions or traumatic brain injuries. Prompt detection of midline shift is crucial, as delays in identification and intervention can negatively impact patient outcomes. The gap we have addressed in this work is the development of a deep learning algorithm that encompasses the full severity range from mild to severe cases of midline shift. Notably, in more severe cases, the mass effect often effaces the septum pellucidum, rendering it unusable as a fiducial point of reference.

MATERIALS AND METHODS: We sought to enable rapid and accurate detection of midline shift by leveraging advances in artificial intelligence. Using a cohort of 981 patient CT scans with a breadth of cerebral pathologies from our institution, we manually chose an individual slice from each CT scan primarily based on the presence of the lateral ventricles and annotated 400 of these scans for the lateral ventricles and skull-axis midline using Roboflow. Finally, we trained an artificial intelligence model based on the You Only Look Once object detection system to identify midline shifts in the individual slices of the remaining 581 CT scans.

RESULTS: When comparing normal and mild cases to moderate and severe cases of midline shift detection, our model yielded an AUC of 0.79 with a sensitivity of 0.73 and specificity of 0.72 indicating our model is sensitive enough to capture moderate and severe midline shifts and specific enough to differentiate them from mild and normal cases.

CONCLUSIONS: We developed an artificial intelligence model that reliably identifies the lateral ventricles and the cerebral midline across various pathologies in patient CT scans. Most importantly, our model accurately identifies and stratifies clinically significant and emergent midline shifts from non-emergent cases. This could serve as a foundational element for a future clinically integrated approach that flags urgent studies for expedited review, potentially facilitating more timely treatment when necessary.

ABBREVIATIONS: CT =Computed Tomography; AUC = Area Under the Curve; MLS = Midline Shift; IML = Ideal Midline.

Received January 3, 2025; accepted after revision March 17, 2025.
From the Departments of Radiology, Memorial Sloan Kettering Cancer Center, NY, NY, 10065 (N.E.B., V.H., N.S., A.H., Z.Z., J.S.)
and Aerospace Engineering, Indian Institute of Technology Madras, Chennai, India, 600036 (H.S.)
Two authors have obtained provisional patents for a company, Authera Inc, tangentially related to the research presented in this work. Another author is the owner/president of fMRI Consultants, LLC, a purely educational entity, and not related to this work.

Please address correspondence to Nooriel E. Banayan, PhD, Department of Radiology, Memorial Sloan Kettering Cancer Center, NY, NY, 10065; banayan1@mskcc.org

SUMMARY SECTION

PREVIOUS LITERATURE: Previous studies implementing Artificial Intelligence (AI) models to measure Midline Shift (MLS) have utilized either deviations from symmetry or the septum pellucidum as a landmark for downstream measurements. These methods limit generalizability for a range of MLS pathologies that may have complex symmetry patterns or effacement of the septum pellucidum. Additionally, previous work utilized complex in-house AI models raising the barrier to entry for radiologists and scientists trying to incorporate AI into their workflow.

KEY FINDINGS: Our lightweight AI model based on the publicly available You Only Look Once (YOLO) library rapidly measures and detects MLS in single-slice patient CT scans. Our model accurately differentiates between normal-to-mild and moderate-to-severe cases of MLS potentially setting the groundwork for future AI integration in radiology workflows.

KNOWLEDGE ADVANCEMENT: The use of a publicly available AI library as the core of our workflow allows the broader medical and scientific community to rapidly modify and enhance the model's usability and accuracy in detecting MLS in patient scans.

Copyright 2025 by American Society of Neuroradiology.

INTRODUCTION

Midline shift (MLS) is a critical indicator of intracranial pathology, often associated with traumatic brain injury, hemorrhage, and space-occupying lesions. The presence and severity of MLS are crucial prognostic factors that reflect asymmetrically increased intracranial pressure and correlate with poor neurological outcomes and elevated mortality rates [1]. As such, prompt and accurate detection of MLS is paramount in emergency settings, where timely intervention can significantly impact patient outcomes. Like other acute abnormalities on imaging, re-prioritization on busy worklists or automatically notifying clinicians [2] can decrease the time to diagnosis [3, 4]. Having a prompt diagnosis expedites treatment, leading to better average outcomes [5, 6, 7]. Traditionally, MLS assessment relies on manual measurements by radiologists, a process that is subject to interobserver variability and potentially delayed in busy clinical environments. These limitations have spurred research into automated MLS detection methods, with the field evolving from classical machine learning approaches [8, 9, 10, 11] to more sophisticated deep learning [1, 12, 13, 14, 15] techniques. These can fall into one of two common algorithm categories, symmetry-based and localization-based [3, 16, 17, 18, 19, 20, 8, 9].

Symmetry-based methods analyze asymmetries in the intracranial contents, operating under the assumption that deviations from symmetry can indicate MLS. While this approach helps assess mass effect, asymmetry can also be present in a wide range of asymptomatic and normal anatomical variations, as well as numerous pathological states, many of which are not associated with midline shift. Hence, it could result in false-positive MLS predictions [21, 11]. Localization-based methods estimate MLS by analyzing the relative positions of key cerebral landmarks, notably the septum pellucidum. However, this approach has limitations, as the septum pellucidum may be obscured or effaced by adjacent brain parenchyma, lesions, or blood products, particularly in cases of more pronounced MLS. Hence, we were motivated to develop a convolutional neural network (CNN)-based method to identify and quantify MLS in a manner that remains valid across severities, including those that obscure the septum pellucidum. We believe this provides a step toward clinical applicability since the approach maintains fidelity across the degrees of MLS that can present on imaging.

In this study, we have combined the strengths of deep learning, symmetry-based, and landmark-based approaches for predicting midline shift within a single, efficient model. Specifically, we focused on tracking the lateral ventricular bodies to represent the symmetry-based component. From the landmark perspective, we employed the stable anatomical features of the anterior and posterior falx, defined previously as part of the "skull axis" [22]. To decrease false predictions, we chose to forgo the septum pellucidum as a reference point, as it is commonly obscured by brain parenchyma in severe MLS cases. Our localization approach leveraged the You Only Look Once (YOLO) [23] model to identify ventricles and key anatomical landmarks. YOLO, a state-of-the-art convolutional neural network (CNN), offers exceptional speed, accuracy, and flexibility while benefiting from ongoing refinements by the open-source community [24]. Its streamlined architecture and computational efficiency facilitate parameter tuning, allowing for precise localization tailored to specific image datasets. The model's ability to perform real-time object detection with high accuracy and its adaptability to various applications made it particularly suitable for our research objectives, enabling efficient medical image analysis without compromising precision.

To address the clinical use case of differentiating normal or near-normal cases from those requiring urgent attention due to moderate or severe midline shift, we hypothesized that our model would consistently and accurately make these distinctions across the range of severities seen at our institution. This approach has the potential to serve as a prototype for tools that highlight pronounced midline shift to radiologists, aiding in the prioritization of critical cases within work lists.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

Anonymized non-contrast head CT scans were obtained from the institutional database using the Extensible Neuroimaging Archive Toolkit (XNAT) system. Scans were categorized based on the severity of midline shift (MLS) as reported in radiological assessments: normal (no appreciable MLS), mild (2-4.9 mm), moderate (5-9.9 mm), and severe (≥ 10 mm). The dataset comprised 215 normal, 244 mild, 349 moderate, and 173 severe cases. For each scan, we selected a single slice at the level of the largest lateral ventricular body volume for analysis.

Image Annotation

In designing our approach, we followed the definition of midline shift based on the ideal midline (IML), a hypothetical line segment that represents the brain's midline in a normal, symmetrical state. [25, 26]. The actual midline shift (MLS) can then be measured or defined as the deviation of brain structures from the IML. Our method used the location of the lateral ventricles as a surrogate for the current location of brain structures, either midline in the normal state or shifted due to mass effect, if present. For each CT exam, we manually selected the 2D slice on which the lateral ventricles were most prominent and of highest volume. We defined the IML primarily via skull axis keypoints at the interface of the calvarial inner table and the falx cerebri. Since this area is rigid, it establishes a midline and one immune to pressure imbalances within the cerebrum. We also included a midsagittal point situated midway in the anteroposterior dimension between the anterior and posterior falx. Figure 1 shows how we measureed MLS as the distance between the centroid of the ventricular bounding box and the nearest point on the IML.

To create the training set, we randomly selected 100 CT scans from each of the four MLS severity categories (normal, mild, moderate, and severe) from the total pool of 981, based on radiology reports. This resulted in a balanced training set of 400 samples, ensuring equal representation across the different severity levels. We annotated the images using Roboflow [27], with a dual-annotation strategy that reflected our definition of MLS:

1. Delineation of the bounding box of the lateral ventricles.
2. Keypoint annotation of the cerebral midline using five anatomical landmarks.

Figure 2 shows examples of the testing set inference predictions for the lateral ventricle bounding boxes and keypoint-based IML. Of note is that the IML (purple line segment) remains valid throughout shifting intracranial structures. As the mass effect increases and the ventricles become effaced, the lateral ventricular bounding boxes (green) shift laterally, causing their centers of mass to deviate from the IML.

Model Architecture and Training

Model development and training were conducted in a Jupyter environment using an in-house modified version of the You Only Look Once (YOLO) Python library [28, 29, 30]. This architecture was chosen for its ability to efficiently detect objects and locate keypoints. We trained two instances of the YOLO model:

1. The first model detected the lateral ventricles using bounding boxes.
2. The second model localized the cerebral midline using keypoints.

Training utilized the AdamW optimizer with automatic hyperparameter tuning. A batch size of 16 and image dimensions of 640×640 pixels were used. Data augmentation included techniques such as random flips, grayscale conversion, blur, median blur, Contrast Limited Adaptive Histogram Equalization (CLAHE), and other transformations. Automatic Mixed Precision (AMP) was enabled to improve computational efficiency. The dataset was split into training and validation sets, with approximately an 80:20 ratio. Both models converged within 150 training epochs. The best-performing models were selected based on validation set performance, with final evaluation conducted on the testing set.

Overview of Midline Shift Quantification

We quantified MLS with a novel approach that combines ventricle-based and midline-based measurements. We calculated the MLS as follows:

1. We computed the centroid of the bounding box encompassing the lateral ventricles, serving as a proxy for the septum pellucidum's position.
2. We defined the IML by the anterior and posterior falcine-inner table attachments.
3. We calculated the MLS as the shortest distance from the ventricular bounding box centroid and the nearest point on the IML.

This method leverages the assumption that in normal anatomy, the centroid of the lateral ventricles closely corresponds to the midpoint of the actual anatomical midline.

Implementation of Midline Shift Quantification

Figure 1 illustrates how we computed MLS values. For each testing set image, once we computed the keypoint-based IML and lateral ventricle bounding box, we computed the MLS as follows. We obtained the bounding box center-of-mass, C_{box}:

$$C_{\text{box}} = (C_x, C_y) = \left(x_{UL} + \frac{w}{2}, y_{UL} + \frac{h}{2}\right) \tag{1}$$

where the upper left corner coordinates (x_{UL}, y_{UL}), width w, and height h define the bounding box. We may write an equation for the line specifying the IML with slope m_{IML} and y-intercept b_{IML}:

$$y_{\text{IML}} = m_{\text{IML}} \times x_{\text{IML}} + b_{\text{IML}} \tag{2}$$

Then, we define the MLS as the distance from C_{box} to the nearest point on the IML, i.e., in the direction perpendicular to that of the IML:

$$MLS = \frac{|m_{\text{IML}} \times C_x - C_y + b_{\text{IML}}|}{\sqrt{m_{\text{IML}}^2 + 1}} \tag{3}$$

.

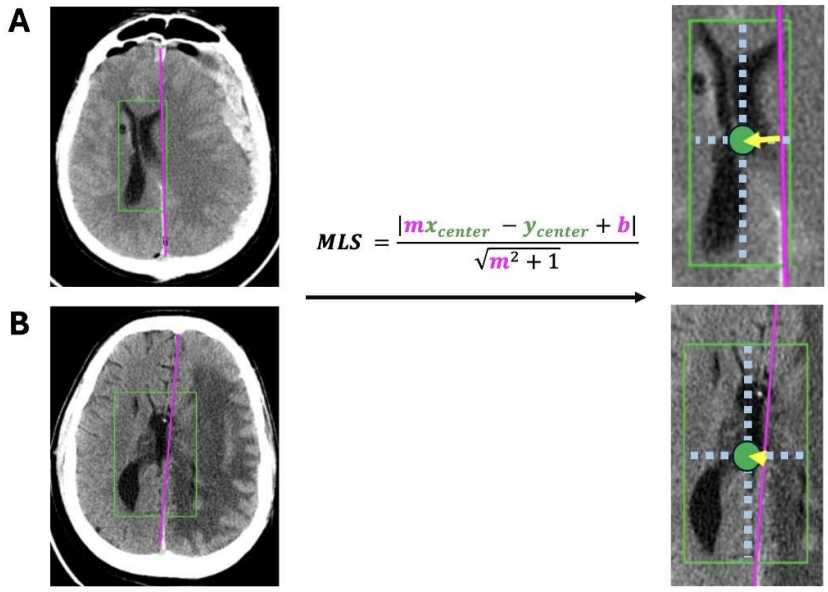


FIG 1. Illustration of midline shift (MLS) estimation based on the centroid (green dot) of the lateral ventricles (green box) and sagittal skull axis midline (purple line). The septum pellucidum is visualized in image A but not in image B. Yet, our algorithm can successfully measure MLS (yellow arrows) and classify MLS severity in both cases, independent of the septum pellucidum.

Data acquisition and preparation

A waiver of informed consent was obtained from our Institutional Review Board for this retrospective study. Patient CT scans of the head were acquired and anonymized using the Extensible Neuroimaging Archive Toolkit (XNAT) system integrated in MSK's database. Table 1 provides an overview of this study's image categories and corresponding counts. We divided CT scans based on the severity of MLS that fellowship-trained neuroradiologists had detailed in their clinical reports:

- **Normal**: No observable MLS
- **Mild**: MLS of $2 - 4.9$ mm
- **Moderate**: MLS of $5 - 9.9$ mm
- **Severe**: MLS ≥ 10 mm

The selection of 10 mm as a cutoff is supported by evidence indicating that functional outcomes worsen when the midline shift surpasses this threshold [5]. The total number of CT scans for each category used in this study was 215 normal scans, 244 mild scans, 349 moderate scans, and 173 severe scans. We manually selected a slice from each scan containing the largest lateral ventricular volume. 100 selected images from each category yielded a total training set of 400 images. Images were uploaded into Roboflow for manual annotation of

- the lateral ventricles using a bounding box, and
- the cerebral midline using keypoint detection, where 5 keypoints were used for annotating the midline.

Downstream training and predictions were done in a Jupyter environment using a modified version of the You Only Look Once (YOLO) Python library. The bounding box and keypoint annotations were downloaded from Roboflow and used as input into our model implemented in YOLO to identify these structures in the testing set. We then calculated the MLS, as described earlier, as the shortest distance between the centroid of the bounding box and the nearest point on the IML.

Table 1: CT scan categories based on radiology reports with midline shift (MLS) criteria, patient counts, and dataset splits for training, validation, and testing. Of note, the MLS values were those measured by radiologist using the septum pellucidum's deviation from the imagined midline.

Category	MLS (mm)	Total	Training/Validation	Testing
Normal	0-2	215	100	115
Mild MLS	2-5	244	100	144
Moderate MLS	5-10	349	100	249
Severe MLS	≥10	173	100	73

RESULTS

Bounding box ventricle localization

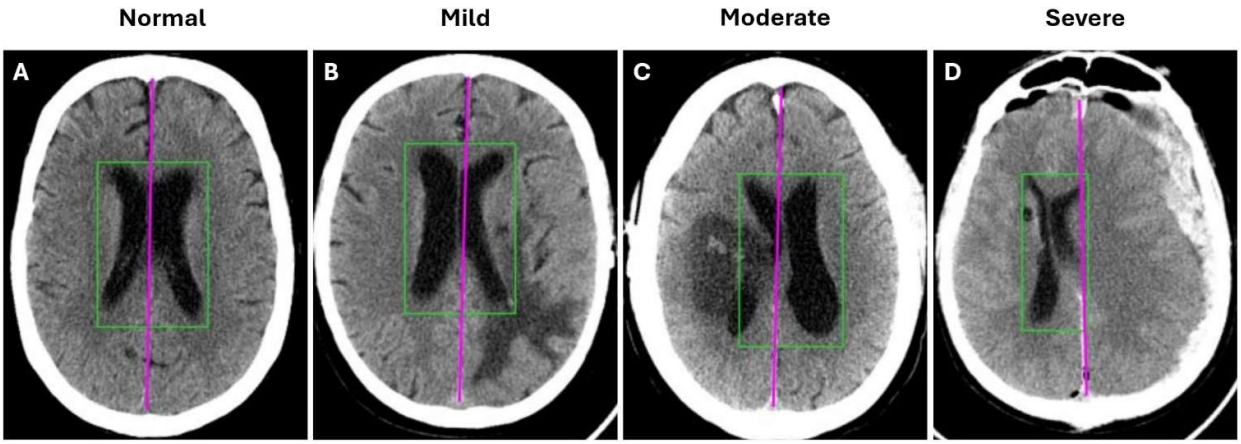


FIG 2. Examples of testing set inference predictions of lateral ventricle bounding box (green) and ideal midline (IML, purple), the latter based on fiducial keypoints related to the skull axis.

After 150 epochs of training, the training and validation losses steadily decreased and eventually plateaued, indicating convergence. The trained ventricle detection network achieved the following performance metrics on the testing set:

- **Mean Average Precision at 50% overlap (mAP@50)** = 0.999: This meant that the model was nearly perfect (99.9%) at drawing bounding boxes around the lateral ventricles when a lenient standard was used. Specifically, if the predicted box overlapped with the true box by at least 50%, the model was credited with being correct. This demonstrated that the model was highly reliable at detecting the general location of the lateral ventricles.

- **Mean Average Precision at 50% to 95% overlap (mAP@50–95)** = 0.841: This score evaluated how well the model performed across stricter overlap thresholds, ranging from 50% to 95%. A score of 0.841 (84.1%) indicated that the model was quite good at matching the actual shape and position of the lateral ventricles, although it showed slightly lower accuracy when requiring a very close match.

- **Precision**: 0.973: This metric reflected the proportion of true positive detections among all positive detections made by the model, indicating high reliability in predicting positive cases.

- **Recall**: 0.987: This score represented the proportion of true positive detections among all actual positive cases, demonstrating the model's effectiveness in capturing as many true instances as possible.

Keypoint detection of imaginary midline

We trained a keypoint detection variant of YOLO to identify five keypoints defining the IML: two near the anterior falx-skull inner table interface, two near the posterior falx-skull interface, and one approximately halfway between the anterior and posterior interfaces.

After 100 epochs of training, the model demonstrated convergence on the training and validation datasets with the following testing set performance metrics:

- mAP@50: 0.889
- mAP@50–95: 0.888
- Precision: 0.933
- Recall: 0.937

Four-category prediction of presence and severity of midline shift

Figure 3 illustrates the average MLS values across the four categories: normal, mild, moderate, and severe MLS. The results demonstrate clear stratification, with progressively higher mean MLS values corresponding to increasing radiologist-measured severity levels. The differences in mean MLS between successive categories surpass the 95% confidence intervals, as indicated by the error bars in the box plot. Furthermore, the mean MLS values align reasonably well with our stratification criteria:

- **Normal**: Mean MLS less than 2 mm
- **Mild**: Mean MLS between 2 and 5 mm
- **Moderate**: Mean MLS between 5 and 10 mm
- **Severe**: Mean MLS approximately 10 mm

The statistical analysis revealed a statistically significant difference between each pair of MLS severity categories, as elaborated in the following subsections.

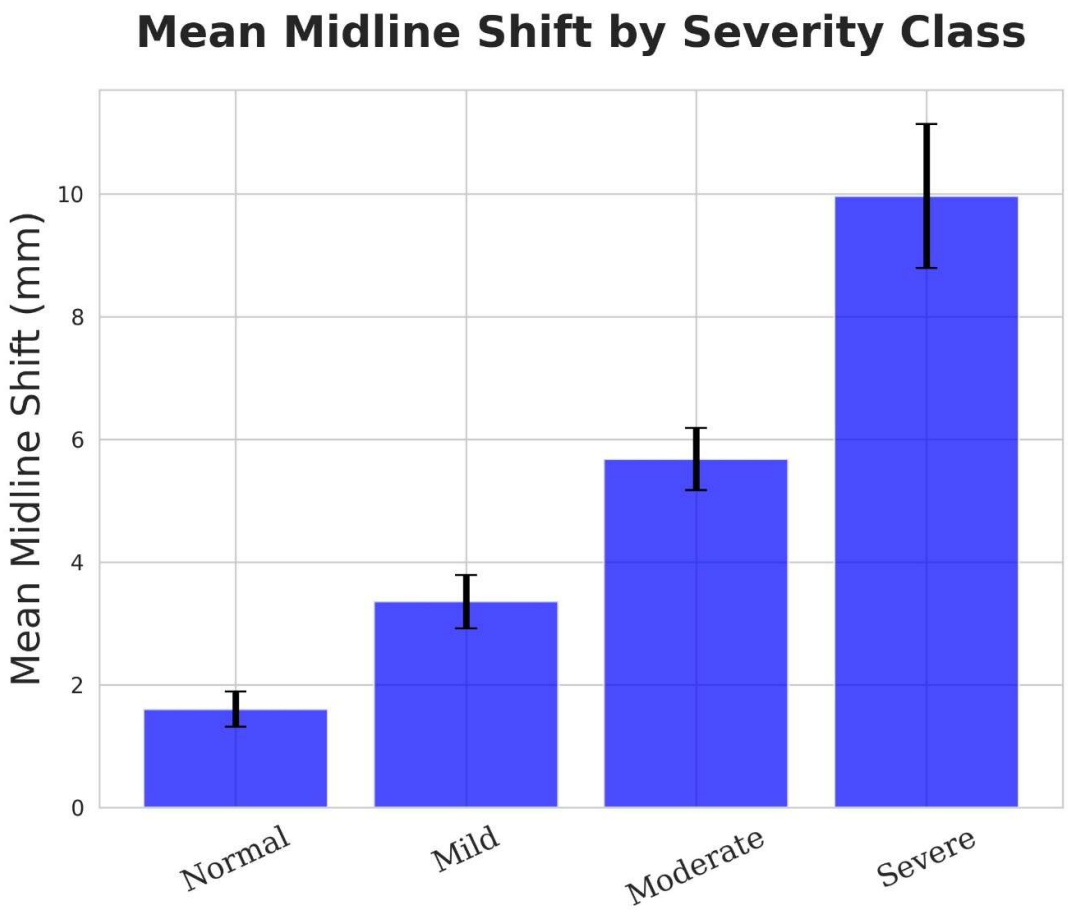


FIG 3: Mean predicted midline shift values for the testing set across four clinical categories. The bar labels denote MLS severity as determined by radiologists during clinical interpretation, while the bar heights indicate the mean MLS values predicted by our automated method. Error bars represent the 95% confidence intervals.

STATISTICAL ANALYSIS

Coarse-grained stratification into two classes of the degree of midline shift for clinical utility

As outlined earlier, our statistical analysis primarily focused on a binary classification of normal-to-mild MLS versus moderate-to-severe MLS. This approach is driven by the clinical importance of moderate-to-severe cases, which often necessitate urgent review and warrant higher prioritization on radiologist worklists. By systematically varying the threshold MLS value to distinguish between normal-to-mild and moderate-to-severe cases, we generated the Receiver Operating Characteristic (ROC) curve depicted in Figure 4.

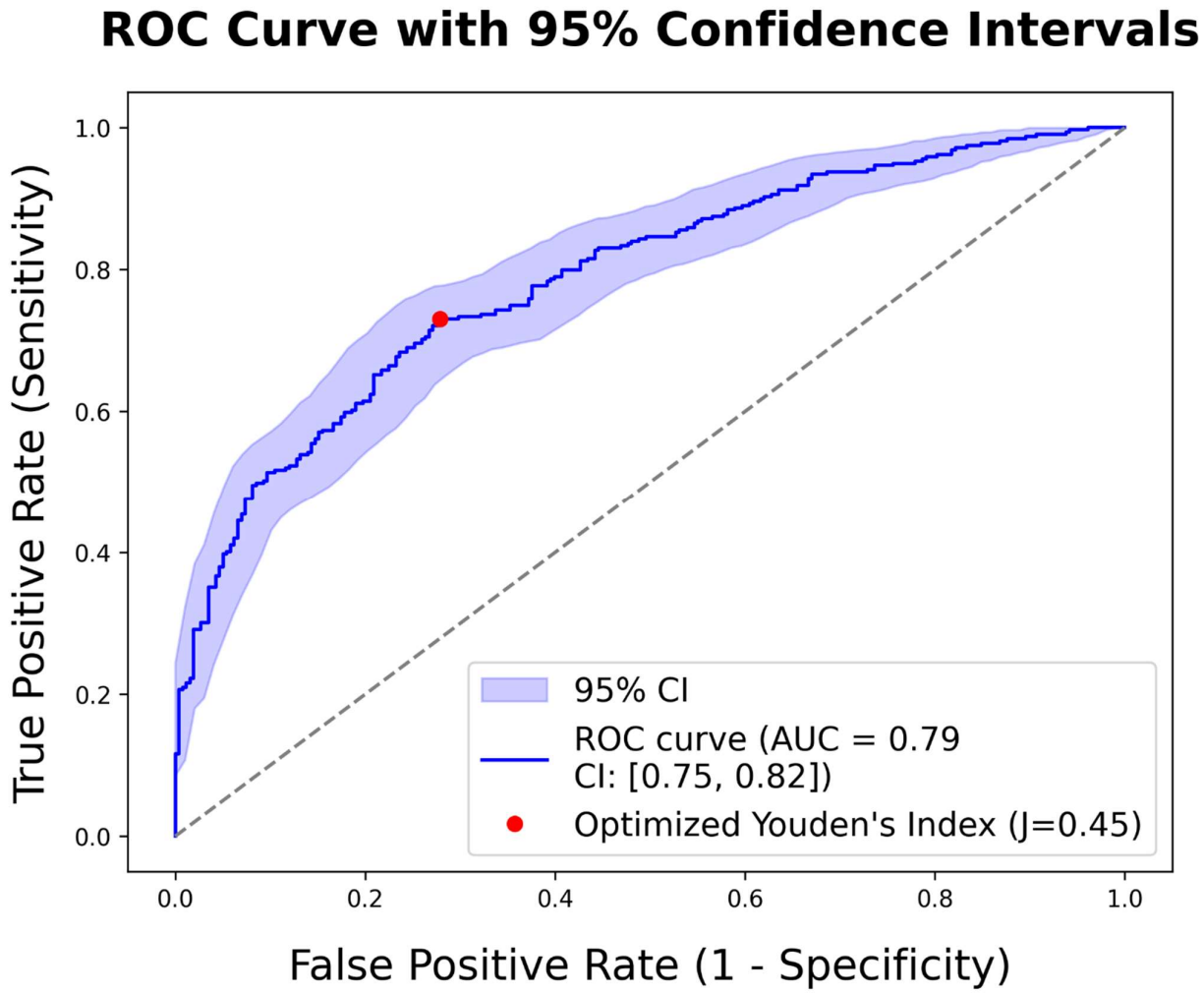


FIG 4: Receiver Operating Characteristic (ROC) Curve illustrating the model's performance in distinguishing between normal-to-mild and moderate-to-severe midline shift categories by varying the decision threshold. The blue curve represents the algorithm's True Positive Rate, while the dashed line indicates the performance of a random classifier. The blue region represents 95% confidence intervals, and the red dot displays the Youden optimal point

The resulting Area Under the Curve (AUC) was 0.79, indicating fair discriminative ability, with 1.0 representing perfect classification and 0.5 reflecting random chance (dashed line). Additionally, the optimized Youden's index point is marked on the figure, corresponding to the optimal threshold value of 3.2 mm. Using this threshold, the following confusion matrix was obtained:

$$\text{Confusion matrix} = \begin{pmatrix} 186 & 72 \\ 87 & 232 \end{pmatrix} \quad (4)$$

Based on the confusion matrix, the primary metrics for binary classification are presented in Table 2.

Table 2: Performance metrics for binary classification, normal-to-mild versus moderate-to-severe, for the testing set. PPV: Positive Predictive Value; NPV: Negative Predictive Value; F1 Score: Harmonic mean of precision and recall. These values are based on a threshold value of 3.24 mm, that was obtained by optimization of Youden's index, as shown in Figure 4.

Metric	Value
Accuracy	0.72
Sensitivity (Recall)	0.73
Specificity	0.72
PPV (Precision)	0.76
NPV	0.68
F1 Score	0.74

Four categories of midline shift severity

We conducted a one-way analysis of variance (ANOVA) to compare midline shift values across four severity categories: normal, mild, moderate, and severe. The ANOVA revealed a significant effect of severity on midline shift values, with an F-statistic of 99.9 and a corresponding p-value 1.3×10^{-5} . These results indicate a highly significant difference in midline shift across the four severity groups, suggesting that the degree of severity is strongly associated with variations in midline shift values. We performed a post-hoc Tukey's Honest Significant Difference (HSD) test following the one-way ANOVA to further investigate pairwise differences between the midline shift values across the four severity categories. The results are summarized in Table 3.

Table 3: Tukey's Honestly Significant Difference (HSD) test results for pairwise comparisons of group means, specifically midline shift (MLS) values across severity categories. 'p-adj' represents the adjusted p-values calculated using the Tukey's HSD method, which controls the family-wise error rate. ⟨Δ⟩ denotes the mean pairwise distance between MLS values.

Group 1	Group 2	⟨Δ⟩	*p*-adj	Lower CI	Upper CI
Mild	Moderate	2.3	$< 5x10^{-4}$	1.4	3.3
Mild	Normal	-1.8	$5x10^{-4}$	-2.9	-0.6
Mild	Severe	6.6	$< 5x10^{-4}$	5.3	7.9
Moderate	Normal	-4.1	$< 5x10^{-4}$	-5.1	-3.1
Moderate	Severe	4.3	$< 5x10^{-4}$	3.1	5.5
Normal	Severe	8.4	$< 5x10^{-4}$	7.0	9.7

In all pairwise comparisons, the differences between groups were statistically significant, as indicated by:

- p-values below the threshold of 0.05, and
- 95% confidence intervals for the mean differences not including zero

Our analysis supports the conclusion that predicted midline shift values varied significantly across the different severity categories.

DISCUSSION

In this study, we developed a fast and efficient model for detecting and quantifying MLS in patient CT scans using bounding box and keypoint localization techniques. By combining elements of symmetry-and localization-based approaches, our method leverages the lateral ventricles to evaluate mass effect on intracranial structures and the inner table-falx junctions to establish a stable reference for estimating the ideal midline (IML). This approach enabled the model to reliably differentiate moderate and severe cases from normal and mild MLS. Such differentiation is crucial, as moderate-to-severe MLS often requires prompt medical attention and possible intervention. YOLO is particularly advantageous due to its rapid inference speed, enabling swift deployment in clinical settings. This efficiency makes it well-suited for simultaneous use across multiple instances, allowing numerous radiologists to leverage the model in diverse clinical environments.

One previous study has shown that AI can capture different severities of MLS across thresholds of 2 mm, 5 mm and 10 mm with high sensitivity (83.3%-91.7%) and specificity (72.7%-98.5%) [13]. Their model used the distance between the cerebral midline and the septum pellucidum to measure MLS and correlate MLS severity with patient outcome. Although we do not correlate our model's performance with patient outcomes directly, our model builds on this approach in multiple ways. First, the integration of bounding boxes and keypoint detection within a single YOLO-based model can improve computational efficiency and accuracy. Second, using the lateral ventricle bounding box centroid as a robust proxy for septum pellucidum position may be advantageous in cases of severe deformation or poor visibility of traditional landmarks. Finally, combining ventricle and "skull axis"-based definitions of midline and MLS may improve reliability across a spectrum of pathological conditions. Finally, we report consistent performance of our model across 581 scans that were acquired over the past 10 years with various scanners and settings, and annotated by numerous different radiologists across our institution. Despite the variability present in our dataset, our model was able to accurately differentiate non-emergent from emergent MLS cases.

Along with the improvements our model introduces, several limitations warrant consideration. One drawback is that the input is a single lateral ventricle-containing slice from the full CT scan z-stack. The radiologist must manually select this slice, which introduces lag between the radiologist receiving the scan and identifying MLS. Although our model aims to streamline the detection of MLS, this manual step could be a source of delay in busy hospital settings where time-sensitive decisions are critical. Second, manual slice selection

introduces the potential for human error. Given numerous slices in a typical CT scan z-stack, even experienced radiologists might face difficulty selecting the exact slice that best represents the lateral ventricles, especially in cases where the ventricles are effaced due to pathologies such as swelling or mass effect. This variability in slice selection could lead to slight discrepancies in model input, which can affect the accuracy of MLS measurements. Lastly, a misaligned or suboptimal slice could result in an inaccurate estimation of MLS, undermining the model's performance. To this end, relying on a single slice to represent the entire three-dimensional (3D) structure of the brain means that important contextual information from adjacent slices is disregarded. MLS is not necessarily uniform across all axial slices and in some cases, the degree of MLS may vary between slices, making it difficult to capture an accurate representation of the entire shift from a single, isolated slice. This limitation could be particularly problematic in cases where MLS is subtle or inconsistent across different z-stack slices.

To address these limitations, future pipeline iterations could incorporate automated slice selection. By leveraging deep learning techniques trained to identify the most relevant slice based on the presence of the lateral ventricles, the system could automatically select the appropriate slice, reducing the need for manual intervention and the potential for human error in slice selection. Moreover, extending the model to consider a series of slices rather than just a single slice could enhance its ability to capture the full 3D context of MLS. This approach would allow the model to analyze the relationship between adjacent slices, potentially leading to a more accurate estimation of MLS across the entire z-stack. Combining 2D and 3D analysis could improve the robustness of the model, especially when tilt artifact is present, and in cases where MLS is subtle or variable across different brain regions.

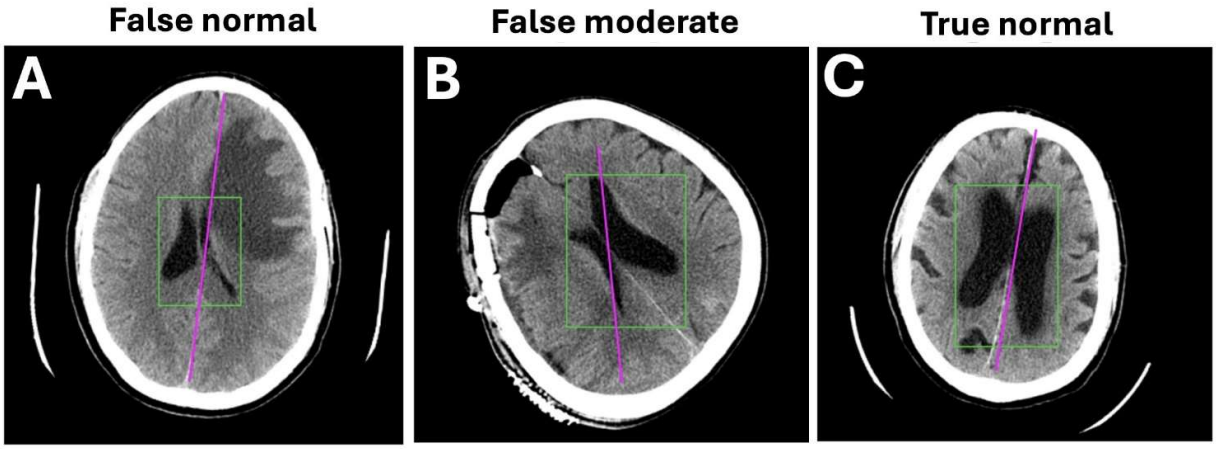


FIG 5: Examples of Failure Modes. (A) False normal and (B) false moderate classification predictions. Relative bounding box rotation is chiefly responsible for these errors, noting that uncorrected tilt artifact and extensive craniotomy changes in (B) likely contribute to an erroneous IML estimation (purple). (C) A true normal case is shown, demonstrating that despite tilting, the ventricular bounding box's center of mass remains relatively stable and near the IML, due to its resistance to rotational displacement.

Examining additional limitations of the current approach, it is important to analyze how the model failed during inference on the testing set (Figure 5). Understanding these failure modes provides valuable insights into the model's shortcomings and areas for improvement. Figure 5A illustrates a false normal ("false negative") prediction, where the ventricular bounding box's center of mass is erroneously close to the IML due to rotational effects. In contrast, Figure 5B presents a false prediction of moderate MLS ("false positive"), where the model incorrectly identifies MLS in a case that should be classified as normal, with no MLS present. In this case, the primary issue lies in excessive patient tilting in the axial plane. Although tilt correction successfully improved head CT orientations in most instances, it failed in this scenario. Likely influenced by the tilt and compounded by extensive craniotomy changes, the model predicted an incorrect IML. In the training set, nearly all IML predictions accurately connected the anterior and posterior falx-inner table junctions, making this failure an uncommon occurrence. Additionally, in Figure 5B, the ventricular bounding box's center of mass is deflected due to rotation, further contributing to the false prediction of moderate MLS. Potential solutions to address these limitations include enhancing tilt correction algorithms and increasing the representation of training set images featuring recent post craniotomy changes, as observed in this case. Additionally, the latest version of YOLO (v11) offers a built-in capability to orient bounding boxes along the dominant direction or major axis of the objects being bounded, which could help mitigate rotational artifacts and improve model performance. Finally, it is important to note that, despite rotation-related failure modes, the model frequently produced results consistent with radiologist assessments. For instance, Figure 5C illustrates a normal case without significant MLS. Although the ventricular bounding box is rotated, its center of mass remains relatively stable and still aligns closely with the nearest point on the IML, ensuring an accurate prediction.

Although our primary focus was distinguishing between normal-to-mild MLS and the more critical moderate-to-severe MLS for worklist prioritization, further stratification between no MLS and mild MLS could enhance the model's granularity. In cases that do not fall under moderate or severe MLS, the septum pellucidum is generally unobstructed, allowing for precise calculation of its deviation from the IML to differentiate between normal and mild MLS. However, edge cases, such as the presence of a cavum septum pellucidum, would

need to be addressed through their inclusion in the training dataset to improve the model's accuracy and robustness. Another area for future inquiry would be assessing changes in midline shift over time. This could be particularly useful in cases of long-standing midline shifts related to architectural brain changes, such as large resection cavities. We envision a future scenario where such a case is flagged for expedited review only if significant worsening occurs between two subsequent scans.

Finally, there is ongoing debate about the efficacy of artificial intelligence for worklist prioritization and its role in improving radiologists' diagnostic performance or turnaround time. The literature presents mixed evidence on whether AI-based worklist prioritization for intracranial hemorrhage in CT head scans enhances diagnostic speed or accuracy [31, 32]. Ultimately, its utility may vary depending on the clinical setting.

CONCLUSIONS

In conclusion, our study demonstrates the potential of combining fast, lightweight localization techniques like YOLO with a hybrid approach integrating symmetry-based and landmark localization methods for automated MLS detection and severity assessment. If the approach is further developed, particularly to include automated slice selection from the full 3D image stack, we can envision the following clinical scenario: cases with significant midline shift (MLS) could trigger automated alerts, prioritizing them to the top of radiologists' worklists to ensure timely review and intervention.

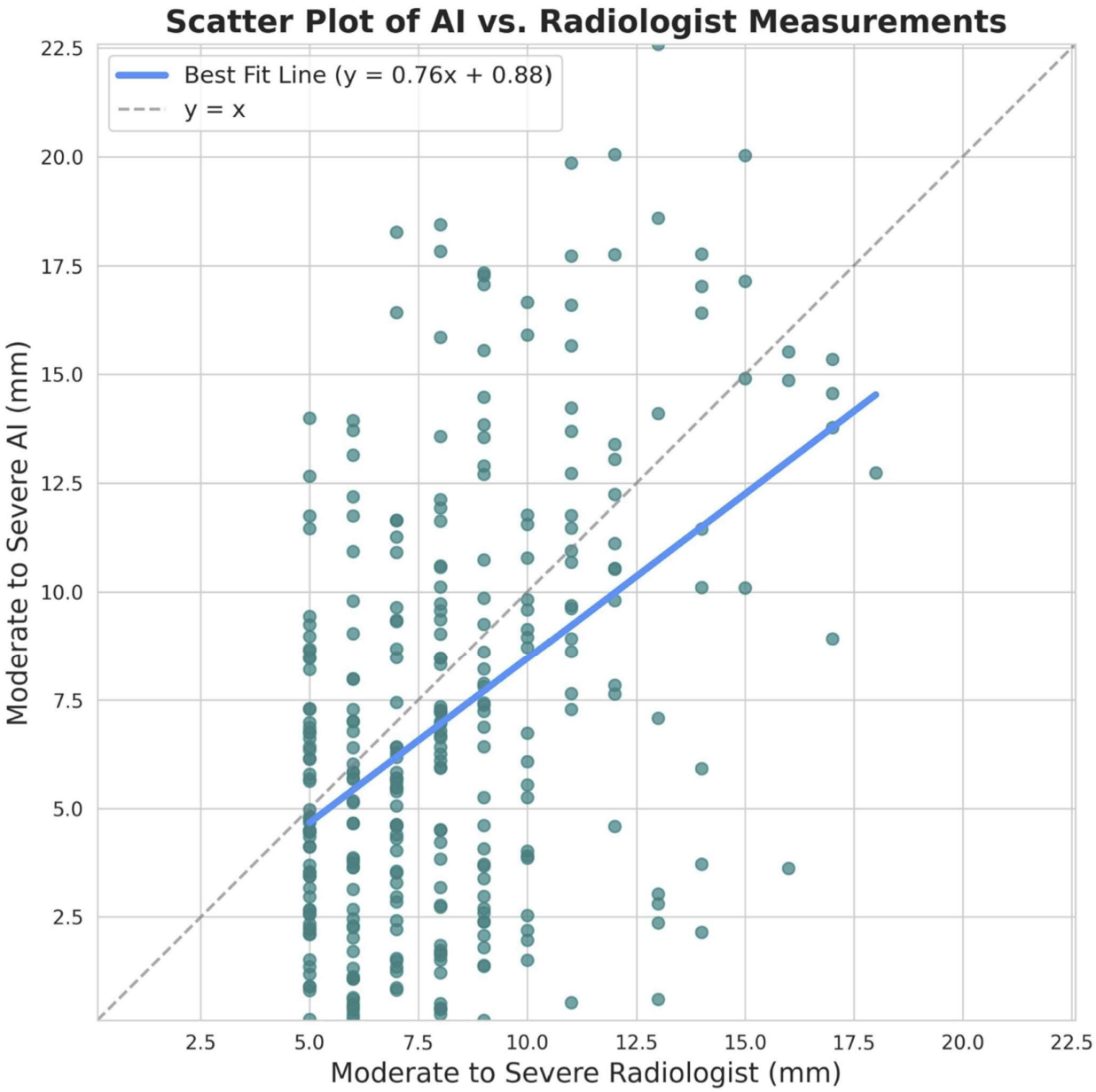
ACKNOWLEDGMENTS

We gratefully acknowledge external support from the Radiological Society of North America (RSNA), the American Society of Neuroradiology (ASNR), Memorial Sloan Kettering Cancer Center, NIH Grant 5R25CA020449, and NIH/NCI Cancer Center Support Grant P30 CA008748.

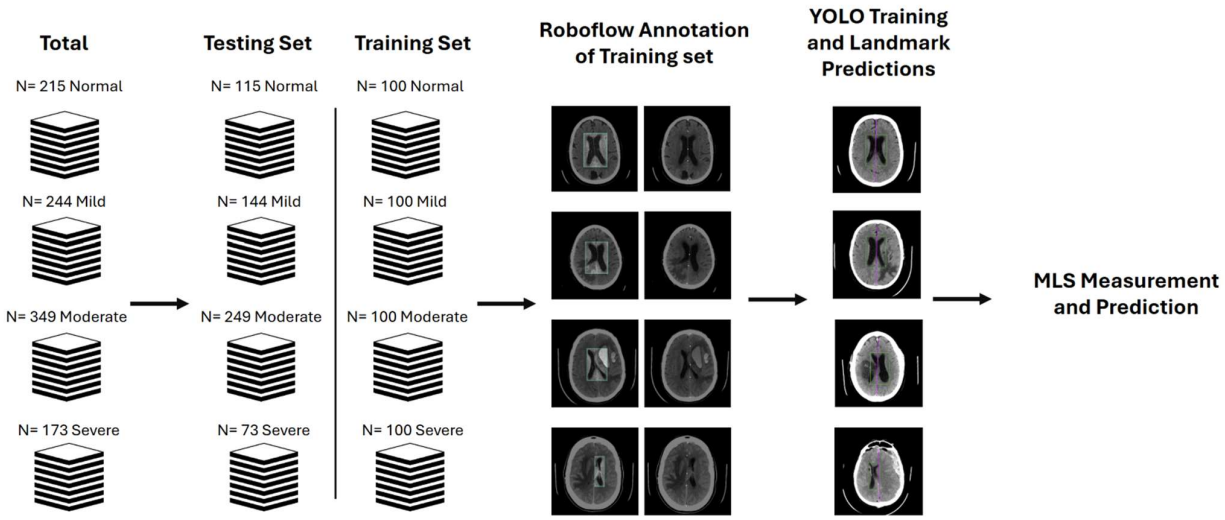
REFERENCES

1. Xiaona Xia, Xiaoqian Zhang, Zhaodi Huang, Qingguo Ren, Jie Gao, and Yanqiu Feng. Automated detection of 3d midline shift in spontaneous supratentorial intracerebral haemorrhage with non-contrast computed tomography using deep convolutional neural networks. European Radiology, 31(12):9285–9294, 2021.
2. Muhannad Seyam, Thomas Weikert, Alexander Sauter, Alex Brehm, Marios-Nikos Psychogios, and Kristine A Blackham. Utilization of artificial intelligence–based intracranial hemorrhage detection on emergent noncontrast ct images in clinical workflow. Radiology: Artificial Intelligence, 4(2):e210168, 2022.
3. Thomas Weikert, Lucas A Noordtzij, Jens Bremerich, Bram Stieltjes, Vikas Parmar, Joshy Cyriac, Alexander W Sauter, and Gregor Sommer. Smart chest x-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. European Radiology, 31(5):3837–3845, 2020.
4. Thomas J O'Neill, Yin Xi, Edward Stehel, Travis Browning, Karthik Rajendran, Nishant Jain, Daniel Lamus, Anuj Bhatia, and William Erdman. Radiologist worklist reprioritization using artificial intelligence: A retrospective study evaluating the effect on pulmonary embolism detection on ct pulmonary angiograms. American Journal of Roentgenology, 217(1):45– 54, 2021.
5. Ross C Puffer, John K Yue, Matthew Mesley, Julia B Billigen, Jane Sharpless, Anita L Fetzick, Ava Puccio, Ramon Diaz-Arrastia, and David O Okonkwo. Long-term outcome in traumatic brain injury patients with midline shift: a secondary analysis of the phase 3 cobrit clinical trial. Journal of Neurosurgery, 131(2):596–603, 2018.
6. Marta Penas-Prado, Terri S Armstrong, and Mark R Gilbert. Proposed additions to the nccn guidelines for adult medulloblastoma. Journal of the National Comprehensive Cancer Network, 18(11):1579–1584, 2020.
7. Robert H Press, Jim Zhong, Saumya S Gurbani, Brent D Weinberg, Bree R Eaton, Hyunsuk Shim, and Hui-Kuo G Shu. The role of standard and advanced imaging for the management of brain malignancies from a radiation oncology standpoint. Neurosurgery, 85(2):165–179, 2019.
8. Furen Xiao, I-Jen Chiang, Jau-Min Wong, Yi-Hsin Tsai, Ke-Chun Huang, and Chun-Chih Liao. Automatic measurement of midline shift on deformed brains using multiresolution binary level set method and hough transform. Computers in Biology and Medicine, 41(9):756–762, 2011. Epub 2011 Jun 30.
9. Furen Xiao, Chun-Chih Liao, Ke-Chun Huang, I-Jen Chiang, and Jau-Min Wong. Automated assessment of midline shift in head injury patients. Clinical Neurology and Neurosurgery, 112(9):721–726, 2010.
10. Wenan Chen, Ashwin Belle, Charles Cockrell, Kevin R Ward, and Kayvan Najarian. Automated midline shift and intracranial pressure estimation based on brain ct images. Journal of Visualized Experiments, 74:3871, 2013.
11. Chun-Chih Liao, Furen Xiao, Jau-Min Wong, and I-Jen Chiang. Automatic recognition of midline shift on brain ct images. Computers in Biology and Medicine, 40(3):331–339, 2010. Epub 2010 Feb 4.
12. Deepak Agrawal, Sharwari Joshi, Vaibhav Bahel, and Latha Poonamallee. Three dimensional convolutional neural network-based automated detection of midline shift in traumatic brain injury cases from head computed tomography scans. Brain Circulation, 9(3):191, 2023.
13. Jiun-Lin Yan, Yao-Lian Chen, Moa-Yu Chen, Bo-An Chen, Jiung-Xian Chang, Ching-Chung Kao, Meng-Chi Hsieh, Yi-Ting Peng, Kuan-Chieh Huang, and Pin-Yuan Chen. A robust, fully automatic detection method and calculation technique of midline shift in intracranial hemorrhage and its clinical application. Diagnostics, 12(3):693, 2022.
14. Maxim Pisov, Mikhail Goncharov, Nadezhda Kurochkina, Sergey Morozov, Victor Gombolevskiy, Valeria Chernina, Anton Vladzymyrskyy, Ksenia Zamyatina, Anna Chesnokova, Igor Pronin, Michael Shifrin, and Mikhail Belyaev. Incorporating task-specific structural knowledge into cnns for brain midline shift detection. In Kenji Suzuki, Mauricio Reyes, Tanveer Syeda-Mahmood, Ender Konukoglu, Ben Glocker, Roland Wiest, Yaniv Gur, Hayit Greenspan, and Anant Madabhushi, editors, Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support, pages 30–38, Cham, 2019. Springer International Publishing.
15. Hao Wei, Xiangyu Tang, Minqing Zhang, Qingfeng Li, Xiaodan Xing, Xiang Sean Zhou, Zhong Xue, Wenzhen Zhu, Zailiang Chen, and Feng Shi. The delineation of largely deformed brain midline using regression-based line detection network. Medical Physics, 47(11):5531–5542, 2020. Epub 2020 Oct 15.
16. Saurabh Jain, Thijs Vande Vyvere, Vasilis Terzopoulos, Diana Maria Sima, Eloy Roura, Andrew Maas, Guido Wilms, and Jan Verheyden. Automatic quantification of computed tomography features in acute traumatic brain injury. Journal of Neurotrauma, 36(11):1794–1803, 2019. Epub 2019 Feb 1.

17. Mohsen Hooshmand, S M Reza Soroushmehr, Craig Williamson, Jonathan Gryak, and Kayvan Najarian. Automatic midline shift detection in traumatic brain injury. In 2018 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 131–134, 2018.
18. Nguyen P. Nguyen, Youngjin Yoo, Andrei Chekkoury, Eva Eibenberger, Thomas J. Re, Jyotipriya Das, Abishek Balachandran, Yvonne W. Lui, Pina C. Sanelli, Thomas J. Schroeppel, Uttam Bodanapally, Savvas Nicolaou, Tommi A. White, Filiz Bunyak, Dorin Comaniciu, and Eli Gibson. Brain midline shift detection and quantification by a cascaded deep network pipeline on non-contrast computed tomography scans. In 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pages 487–495, 2021.
19. Sasank Chilamkurthy, Rohit Ghosh, Swetha Tanamala, Mustafa Biviji, Norbert G Campeau, Vasantha Kumar Venugopal, Vidur Mahajan, Pooja Rao, and Prashant Warier. Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. The Lancet, 392(10162):2388–2396, 2018.
20. Ruizhe Liu, Shimiao Li, Bolan Su, Chew Lim Tan, Tze-Yun Leong, Boon Chuan Pang, C C Tchoyoson Lim, and Cheng Kiang Lee. Automatic detection and quantification of brain midline shift using anatomical marker model. Computerized Medical Imaging and Graphics, 38(1):1–14, 2014. Epub 2013 Nov 26.
21. Alejandro Zuluaga Santamar´ıa, Sergio L. Vargas, Sandra Arango, and Ricardo Uribe. Brain asymmetry : Diagnostic approach asimetr´ıa cerebral : enfoque diagn´ostico. 2017.
22. Yang Liao, Gordon K Smyth, and Wei Shi. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Research, 41(10):e108, 2013.
23. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 779–788, 2016.
24. Juan Terven and Diana-Margarita C´ordova-Esparza. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. arXiv preprint arXiv:2304.00501, 2023.
25. Chun-Chih Liao, Ya-Fang Chen, and Furen Xiao. Brain midline shift measurement and its automation: a review of techniques and algorithms. International journal of biomedical imaging, 2018(1):4303161, 2018.
26. Deepak Agrawal, Sharwari Joshi, and Latha Poonamallee. Automated midline shift detection and quantification in traumatic brain injury: A comprehensive review. Indian Journal of Neurotrauma, 2024.
27. B. Dwyer, J. Nelson, T. Hansen, et al. Roboflow (version 1.0), 2024. Available from https://roboflow.com.
28. Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 779–788. IEEE, 2016.
29. Jiang Peiyuan, Ergu Daji, Liu Fangyao, Cai Ying, and Ma Bo. A review of yolo algorithm developments. Procedia Computer Science, 200:1234–1241, 2022.
30. Glenn Jocher et al. Ultralytics yolov5, 2023.
31. Savage CH, Tanwar M, Elkassem AA, Sturdivant A, Hamki O, Sotoudeh H, Sirineni G, Singhal A, Milner D, Jones J, Rehder D, Li M, Li Y, Junck K, Tridandapani S, Rothenberg SA, Smith AD. Prospective Evaluation of Artificial Intelligence Triage of Intracranial Hemorrhage on Noncontrast Head CT Examinations. AJR Am J Roentgenol. 2024 Nov;223(5):e2431639. doi: 10.2214/AJR.24.31639. Epub 2024 Sep 4. PMID: 39230402.
[32] O'neill TJ, Xi Y, Stehel E, Browning T, Ng YS, Baker C, Peshock RM. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. Radiology: Artificial Intelligence. 2020 Nov 18;3(2):e200024.



SI Figure 1: Head-to-head scatterplot of MLS measured automatically by our model and manually by radiologists. The gray dashed line is the y=x line, the blue line is the line of best fit with a correlation coefficient of 0.76.



SI Figure 2: Flowchart of the AI training and testing methodology presented in this work.

Package	Version
Python	3.11.11
Ultralytics	v8
scikit-learn	1.6.1
seaborn	0.13.2
statsmodels	0.14.4
pandas	2.2.2

SI Table 1: Package versions used in the analysis presented in this work.