



Research article



Deep learning-based prediction of mortality using brain midline shift and clinical information

An-Rong Wu ^a, Sun-Yuan Hsieh ^{a,b}, Hsin-Hung Chou ^{b,**}, Cheng-Shih Lai ^c,
Jo-Ying Hung ^d, Bow Wang ^{c,e}, Yi-Shan Tsai ^{c,e,*}

^a Department of Computer Science and Engineering, National Cheng Kung University, Tainan, Taiwan

^b Department of Computer Science and Engineering, National Chi Nan University, Nantou, Taiwan

^c Department of Medical Imaging, National Cheng Kung University Hospital, Tainan, Taiwan

^d Department of Statistics, National Cheng Kung University, Tainan, Taiwan

^e Department of Radiology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

ARTICLE INFO

Keywords:

Artificial intelligence
Medical intelligence
Brain midline shift
Midline detection
Keypoint detection network
Deep learning

ABSTRACT

Brain midline shift (MLS) indicates the severity of mass effect from intracranial lesions such as traumatic brain injury, stroke, brain tumor, or hematoma. Brain MLS can be used to determine whether patients require emergency surgery and to predict patients' prognosis. Since brain MLS is usually emergent, it must be diagnosed immediately. Therefore, this study presents a computer-aided deep-learning method for detecting MLS, aiming to predict mortality in a prognosis-predicting cohort using brain MLS and clinical information. The brain midline is a 3-dimensional structure, but computed tomography (CT) slices are 2-dimensional which limits brain MLS detection. Here we propose a keypoint detection method to detect brain midline on each CT slice, acquiring brain MLS distance and area in each slice. Combined with clinical information, patient mortality can be predicted using the multilayer perceptron (MLP) model. The accuracy, precision, sensitivity, specificity, and F1-score for slice selection with the proposed model are 0.966, 0.952, 0.991, 0.932, and 0.971, respectively. Both MLS distance and volume were precisely predicted at slice-level and case-level with only the slightest error. The detected midlines were clearly separated into left and right brain with a dice coefficient of 0.98. The accuracy and AUC of the MLP model were both above 0.8. The model detected large brain MLS cases well in the prediction of outcomes in the prognosis-predicting cohort. The method performs well on slice selection and brain MLS detection, and predictions of MLS distance and volume combined with clinical information predicts the patient's prognosis well.

1. Introduction

The brain midline shift (MLS) is a critical situation from a traumatic brain with hemorrhage, stroke with brain swelling, tumor with mass effect, and other space-occupying lesions. It can be measured on computed tomography (CT), magnetic resonance imaging (MRI), or ultrasound (US), and CT is preferred for acute neurological conditions or stroke. Therefore, most of the computer-aided methods are

** Corresponding author.

* Corresponding author.

E-mail addresses: p76094046@gs.ncku.edu.tw (A.-R. Wu), hsiehsy@mail.ncku.edu.tw (S.-Y. Hsieh), chouhh@ncn.edu.tw (H.-H. Chou), laics@mail.ncku.edu.tw (C.-S. Lai), belinda19960903@gmail.com (J.-Y. Hung), wangbow1227@gmail.com (B. Wang), n506356@gmail.com (Y.-S. Tsai).

<https://doi.org/10.1016/j.heliyon.2024.e41271>

Received 18 October 2024; Received in revised form 3 December 2024; Accepted 15 December 2024

Available online 8 January 2025

2405-8440/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

based on CT images. If the MLS distance exceeds 5 mm, the patient needs immediate surgery [1]. Traditionally, the clinicians use visual assessment to manually select the slice with measurement of a maximum of MLS or use the typical image slice at the level of the foramen of Monro (FM) [1]. Also, MLS is usually used as a quantitative indicator to predict outcomes in trauma patients [2–6]. Hence, developing deep learning tools to automatically detect and quantify the brain MLS as soon as possible may assist clinicians in diagnosing quickly, treating patients immediately, and lowering the variability in the quantification of MLS. The various etiologies of MLS make the detection of MLS very difficult. So, understanding how to detect and quantify MLS accurately remains challenging. The brain midline is a virtual line that symmetrically separates the left and right brain. Traditionally, the brain midline has been evaluated using typical image slices passing through the five anatomic structures, including the anterior falx (AF), septum pellucidum (SP), third ventricle (V3), pineal region (PR), and posterior falx (PF), as shown in Fig. 1(a). However, not all slices of the brain CT series contain five anatomic structures. The ideal midline for the whole brain is the line that joins the falx, crista galli, and frontal crest at the anterior cranium fossa and cruciform eminence at the posterior cranium fossa, as the blue line shown in Fig. 1(b). The midline is defined by a curve drawn along the AF to the PF to separate the right and left brains. Then, the distance from the farthest point on the deformed midline to the ideal midline is the MLS distance, as shown in Fig. 1(b). We can use the deep learning technique to identify the perfect midline. In that case, we can provide the information on the farthest distance of MLS and calculate with volumetric MLS by summarizing the displaced areas in each slice for the quantification of MLS.

Previous studies can be divided into symmetry-based, landmark-based and deep learning method-based [7]. Liao et al. [8] developed a symmetry-based MLS detection method in which the deformed brain midline is decomposed into three segments: the upper and the lower straight segments separating the two brain hemispheres, and the curved segment formed by a Bezier curve. The average of the absolute value of the MLS in their study was 4.0 mm with a standard deviation of 4.4 mm. Chen et al. [9] used a landmark-based approach based on the Gaussian mixture model and shape matching. The detected midline of these methods is not always smooth, and sometimes it is difficult to find the five anatomical points. For the actual midline, above 80 % has less than 2.25 mm difference. Jain et al. [10] detected the MLS through the segmentation of brain, cistern and acute intracranial lesions. They evaluated their method for MLS detection >5 mm with a classification accuracy of 0.89. Xia et al. [11] proposed a three-stage deep learning-based method to do the brain midline delineation as a segmentation task and explore the relationship between MLS and the Glasgow Coma Scale (GCS) score and Glasgow Outcome Scale (GOS) score. Jiang et al. [12] indicated that measuring MLS on a single slice is not enough for whole brain tissue displacement. They analyzed the correlation between the brain MLS distance/volume and the outcomes of traumatic brain injury patients.

In recent years, deep learning has developed rapidly and has been used for several tasks, such as detection and segmentation, on medical images. Keypoint detection is an application of deep learning in which human pose estimation, such as face and skeleton, can be done. The most important aspect of this method is that the location of each keypoint must be accurate. Keypoint detection can be done in two ways. The first is to predict the coordinates of keypoints directly, the second is to predict the heatmap of keypoints, such as Mask R-CNN [13] and OpenPose [14]. The second way to do keypoint detection can be viewed as a semantic segmentation task. Many deep neural networks were designed for semantic segmentation, such as FCN [15], U-Net [16], SegNet [17]. Olaf et al. [16] proposed a U-shaped neural network known as U-Net, which won the IEEE International Symposium on Biomedical Imaging (ISBI) cell tracking challenge 2015. Until now, U-Net is still the most commonly used model for image segmentation. Keypoint detection and brain midline detection have something in common. The points that the brain midline must pass through can be seen as “keypoints”. Therefore, the brain midline can be detected as long as these keypoints are detected well.

For these reasons, we have proposed a keypoint detection method based on a modified U-Net model for directly detecting the brain midline of the whole brain through a deep neural network model. Our method combines the advantages of both the landmark-based method and deep learning method. We performed slice-level detection, then quantified both shift distance and shift volume by summarizing all the slices with MLS. At the same time, we did a prognosis-predicting cohort study using another deep learning model to predict mortality status using MLS and clinical information. The contributions of this paper are summarized as follows.

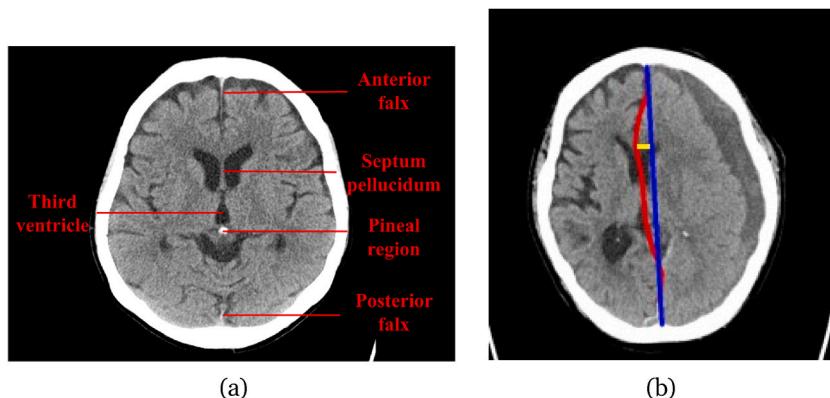


Fig. 1. (a) Example of five anatomic structures. (b) Example of ideal midline (blue line), deformed midline (red line) and MLS distance (yellow line).

- This study proposes the first approach that can select slices with the brain midline and detect the brain midline as keypoints to avoid the shortcomings of previous works. The newly adopted method quantifies the brain MLS distance and volume with only the slightest error.
- To ensure generalizability, a prognosis-predicting cohort study was conducted to validate the connection between brain MLS and mortality.

The rest of this article is organized as follows. Section 2 introduces the materials we used and our proposed method, which is divided into four stages. Section 3 describes the metrics we used, the training details, data preparation, and data preprocessing. Section 4 discusses the experimental results for each stage and the limitations. Finally, Section 5 concludes this article.

2. Materials and methods

The Institutional Review Board (IRB) of our institution reviewed and approved the study protocol, allowing a retrospective review of emergent images on brain CT from July 2015 to May 2022. The IRB number for this study is A-ER-111-185. Because patient data were retrospectively collected and maintained patients' anonymity, the IRB also approved a waiver of informed consent.

2.1. Dataset

Two separate and distinct datasets were used for this retrospective cohort study. First, non-contrast head CT scans were collected from patients aged >20 years who visited the emergency department of National Cheng Kung University Hospital. To build the deep learning model for MLS detection and quantification, data were extracted from the hospital radiological reporting system, including data of 69 head CT series with a diagnosis of brain MLS. In addition, data from 10 cases with normal brains were retrospectively collected from the radiological reporting system for evaluation of the normal range of MLS. Another cohort included patients with the radiological report of brain hemorrhage for prognosis modeling. Included data were demographic data such as sex, age, and clinical data such as emergency surgery history, types of brain hemorrhage, the hospital length of stay (admission duration), and patients' final mortality status (alive or dead). The details of the two datasets are described in Section 2.1.1 and Section 2.1.2, respectively.

2.1.1. MLS detection and quantification dataset

By reviewing the radiological reporting system, 69 head CT series with a diagnosis of MLS were collected. All slices have a 5 mm slice thickness, each head CT series consist of about 30 slices. All slices are 512×512 pixels. The slice-level ground truth, ideal midline, deformed midline and shift distance, were manually labeled by a radiology technician with 30-year-experience (CSL) and re-checked and corrected the labeling by a radiologist (YST) with a 20-year-experience. The ideal midline from head to toe was defined by traceable anatomies such as falx, crista galli, and frontal crest at the anterior cranium fossa and cruciform eminence at the posterior cranium fossa.

Table 1

Summary descriptives of groups by mortality status. The shift distance and volume are the prediction of our midline detection model.

	Alive		p-value
	N = 1915	N = 313	
Age ^a	63.2(17.8) 5.06 (3.24)	67.7 (15.3) 9.52 (6.47)	<0.001 <0.001
Shift distance (mm) ^a	16.69 (8.75)	27.46 (18.32)	<0.001
Shift volume (cm^3) ^a			
Gender ^b	Female Male	873 (45.6 %) 1042 (54.4 %)	111 (35.5 %) 202 (64.5 %)
IPH ^b	No Yes	880 (46.0 %) 1035 (54.0 %)	139 (44.4 %) 174 (55.6 %)
SAH ^b	No Yes	765 (40.0 %) 1150 (60.0 %)	89 (28.4 %) 224 (71.6 %)
SDH ^b	No Yes	1096 (57.2 %) 819 (42.8 %)	173 (55.3 %) 140 (44.7 %)
EDH ^b	No Yes	1811 (94.6 %) 104 (5.4 %)	297 (94.9 %) 16 (5.1 %)
IVH ^b	No Yes	1688 (88.1 %) 227 (11.9 %)	207 (66.1 %) 106 (33.9 %)
Surgery ^b	No Yes	1324 (69.1 %) 591 (30.9 %)	250 (79.9 %) 63 (20.1 %)

Abbreviations: IPH, Intraparenchymal Hemorrhage; SAH, Subarachnoid Hemorrhage; SDH, Subdural Hemorrhage; EDH, Epidural Hemorrhage; IVH, Intra-ventricular Hemorrhage.

^a Continuous variable, descriptive statistics are shown as mean (std), t-test is performed.

^b Categorical variable, descriptive statistics are shown as n (%), Chi-squared test is performed.

2.1.2. Prognosis-predicting dataset

Another dataset includes a total of 2228 patients with brain hemorrhage from the emergency department to validate our method. All CT images from these patients were obtained from Siemens CT Scans with 1780 from SOMATOM Definition Flash and 448 from SOMATOM Definition AS. The prediction of MLS distance and shift volume using our model with demographic and clinical information, including age, sex, previous surgery or not and five brain hemorrhage types, such as IPH, IVH, SAH, EDH, and SDH, to predict patients' prognosis (dead or alive). We excluded the cases in which age is younger than 20 (considering that children's brains are still developing, their brain structure is quite different from that of adults). The demographic and clinical characteristics of the cohort are summarized in Table 1. The age range of the cohort dataset was from 21 to 99 years with mean 63.8 years, and the number of male and females were 1244 (55.8 %) and 984 (44.2 %), the number of dead and alive patients were 313 (14 %) and 1915 (86 %), and the number of patients with and without previous surgery were 654 (29.3 %) and 1574 (70.7 %), respectively.

2.2. Modeling for midline detection

The complete framework of our midline detection method is shown in Fig. 2. Our method is divided into two parts, midline prediction and prognosis prediction.

Our method divides the brain midline into ten equal parts, not based on the traditional five typical anatomical structure positioning. The polyline connecting the eleven points would make the precision closer to the ground truth midline. One CT slice is inputted into the midline detection model to detect 11 keypoints on slice, producing 11 heatmaps of each keypoint by which coordinates could be calculated in order. Theoretically, using 11 coordinates could be connected into the midline. Due to limitations of the model, these coordinates would not be in order of the slices without midline. Therefore, the series can be filtered to select the slices which contain the brain midline in this way. Then, the maximum of MLS distance and total shift volume can be computed for each patient.

Traditionally, measurement of MLS is done from the axial brain CT image with the midline passing through five points: AP, SP, V3, PR, and PF. The midline in other slices is defined by a curve drawn along the falx/crista galli/frontal crest to cruciform eminence for separation of right and left brains. In addition, the brain midline is segmented into 10 equal parts, producing 11 keypoints on the midline on each slice, as shown in Fig. 3. Therefore, we proposed a method based on keypoint detection to detect the 11 keypoints.

2.2.1. Model architecture

The keypoint detection network is based on U-Net [16] architecture, which is a U-shaped neural network. The structure of U-Net could be divided into "encoder", "bottleneck" and "decoder" part. The encoder extracts the features of the input image and reduces the size of the image after each convolution and pooling. The bottleneck learns the compression of input data. From output by the bottleneck, the decoder can rebuild a new image of the same size as the input image. U-Net uses "skip-connection" to connect the encoder and decoder, so the decoder is able to take over some important information from the encoder. However, some information is lost in the process of training a deep neural network, and this defect is caused by the up-sampling and pooling layer. As a result, our model failed to learn about the keypoints. Therefore, we also adopted dilated convolution [18] and residual learning [19] in our model to solve these defects.

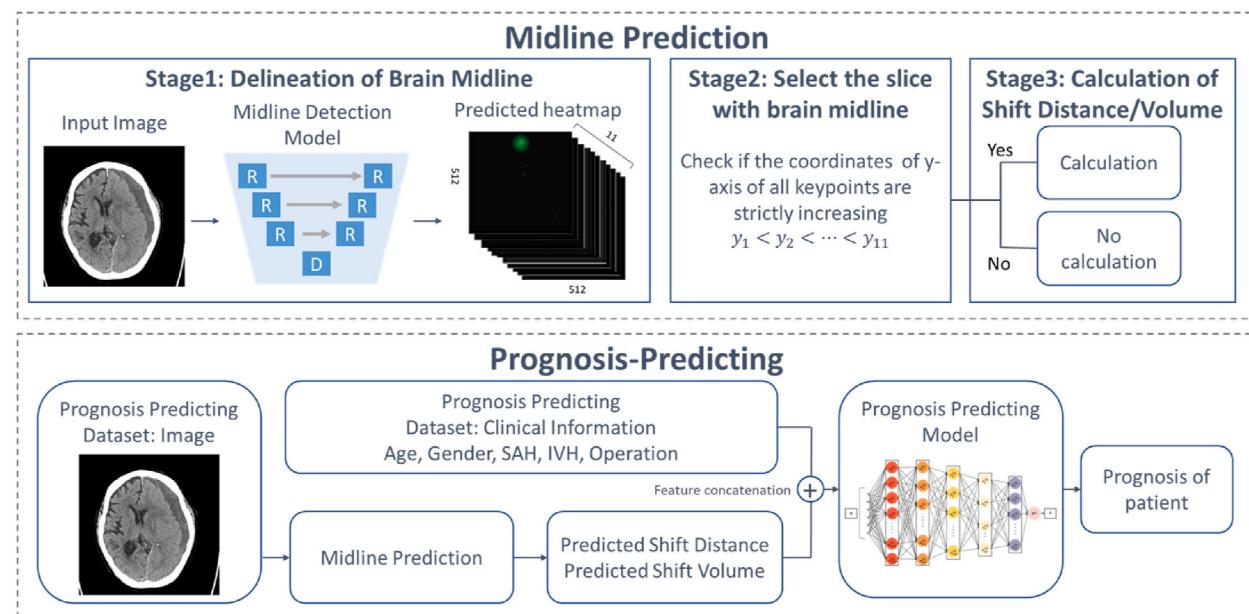


Fig. 2. Flowchart of the proposed method. The plus sign in the circle means feature concatenation.

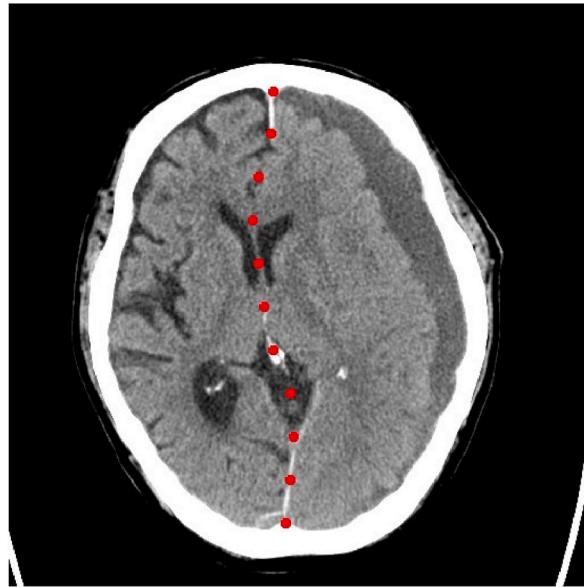


Fig. 3. Brain midline is segmented into 10 equal parts, producing 11 keypoints on the midline.

The difference between convolution and dilated convolution is the filter. Dilated convolution allows for convolutional filters to skip some pixels, as shown in Fig. 4, thereby creating a more spread-out receptive field. This allows the network to capture a larger context of the input data while also preserving the spatial resolution of the feature maps. Hence using dilated convolution lets the model learn keypoints more accurately. We applied the dilated convolution on the bottleneck part of U-Net, since we would like the bottleneck part to learn the location of keypoints accurately.

Residual learning is proposed by He et al. [19]. Those authors pointed out that deep plain networks have the degradation problem, which means when we train a plain network, such as VGG nets [20], the deeper the network, the higher the training error will become. So, He et al. proposed residual learning to solve this problem. Let's consider a building block defined as:

$$y = F(x, \{W_i\}) + x \quad (1)$$

where x and y are the input and output vectors of the layer, and the function $F(x, \{W_i\})$ represents the residual mapping to be learned. He et al. showed that learning residual function is easy than learning unreference functions through experiments. Accordingly, the residual block was employed in our model to prevent degradation.

The details of our model are shown in Fig. 5. We input an image with size of $512 \times 512 \times 1$, then obtain an output with size of $512 \times 512 \times 11$, where each channel of output image corresponds to the heatmap of each keypoint.

2.2.2. Selection and calculation

After midline detection, we must check whether a midline appears on the image or not. Since the 11 keypoints are detected on the midline, when there is a brain midline on the image, these predictive keypoints should be in order. Hence, when the predictive keypoints are not in order, we assume that with no brain midline structure on the image. Our slice selection method checks whether the coordinates of the y-axis of the 11 keypoints are strictly increasing. If they are strictly increasing, the predicted keypoints are in order. If not, the predicted keypoints are not in order.

After slice selection, the MLS distance, area and volume for each selected slice are computed. The first and the last keypoint are used

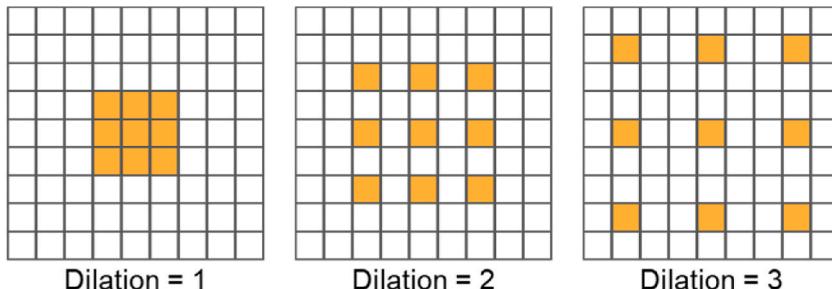


Fig. 4. Dilated convolution.

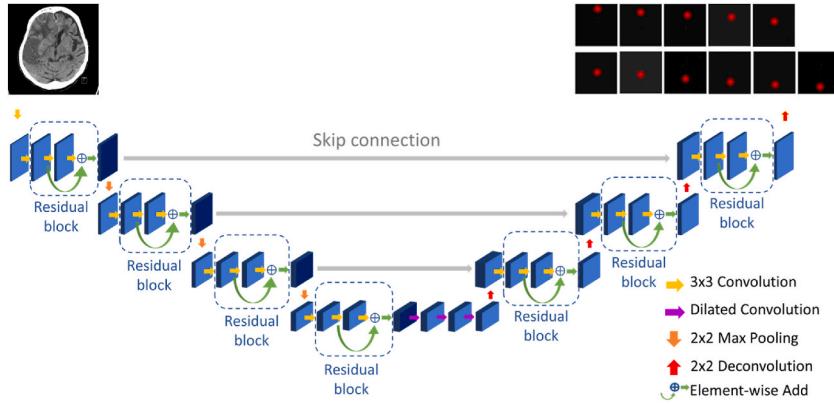


Fig. 5. The structure of the midline detection model.

as the ends of the predicted ideal midline, and the line that connects all 11 keypoints from the first keypoint to the last keypoint is the predicted deformed midline. Then the distance from the farthest point on the predicted deformed midline to the predicted ideal midline is the shift distance on that slice, denoted by d_i . Also, the area bounded by the predicted deformed midline and the predicted ideal midline is the shift area, denoted by a_i . Then the MLS volume can be calculated by multiplying the slice thickness of CT, denoted by v_i . The results are presented at slice-level and case-level. For case-level, we chose the maximum shift distance of these slices, denoted by MLS_D . And summation of all shift volumes of these slices, denoted by MLS_V .

$$MLS_D = \max_i(d_i) \quad (2)$$

$$v_i = a_i \times \text{slice thickness} \quad (3)$$

$$MLS_V = \sum_i v_i \quad (4)$$

2.2.3. Evaluation metric for slice selection

The accuracy, precision, sensitivity, specificity and F1-score are used as our evaluation metrics for slice selection. The definitions of the accuracy, precision, sensitivity, specificity are shown in (5), (6), (7) and (8), where TP , FP , FN and TN denote the number of true positive, false positive, false negative and true negative results, respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Sensitivity} = TPR = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (8)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (9)$$

2.2.4. Evaluation metric for midline detection

Mean absolute error (MAE) and dice coefficient are used as the evaluation metric for midline detection. The definitions are shown as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

where y_i is the ground truth shift distance/volume and \hat{y}_i is the prediction of shift distance/volume.

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (11)$$

It is inadequate to use MAE of shift volume as the evaluation metric, since MAE cannot show the difference in location between ground truth and prediction, so we also used the dice coefficient as the evaluation metric for shift volume. The brain midline separates the brain into the left and right brain, as shown in Fig. 6. Brain tissue was extracted from the skull by using a threshold for pixel value, since a certain difference exists between the values of brain tissue and skull. The morphology method [21] was used to find the maximal connected components of brain tissue. Hence, the dice coefficient was computed for the left and right brain to verify the similarity between the ground truth midline and the predicted midline. The slice-level and case-level dice coefficient for the left and right brain were computed by formula (11). The slice-level dice is the mean of the dice for each CT slice, and the case-level dice is the mean dice after all CT slices of the same patient are concatenated into one.

2.3. Prognosis-predicting cohort study

As stated above, the objective of this cohort study was to predict mortality, that is, patients' with dead or alive classification. Since the data of the cohort study was imbalanced as stated in Section 2.1.2, k-means clustering was applied to make sure that similar data from each class existed in each fold, and to ensure that our model was able to learn similar data from each case. Data balancing methods were used to deal with the problem of imbalanced data. Also, feature selection was done before training the MLP model.

2.3.1. Data balancing

Since the data of the cohort are imbalanced, the model may ignore the minority class when training the model, but the minority class is the main interest. Resampling methods, such as oversampling and undersampling, were used to resolve this problem. Over-sampling duplicate samples from the minority class. In contrast, undersampling delete samples from the majority class. Also, random oversampling and random undersampling were used in the data preprocessing stage to find a better resampling method for developing the model.

2.3.2. Feature selection

In general, many different methods are available for doing feature selection [22], including filter methods, wrapper methods and embedded methods. Filter methods are independent of the classifier, with low computational cost and fast. Hence, we chose the filter method for doing the feature selection. The criteria are based on the p-value of each variable computed by chi-squared test or t-test. The features with p-value greater than 0.05 in Table 1 were filtered out.

2.3.3. Multilayer perceptron

Multilayer Perceptron (MLP) [23] is a class of artificial neural networks (ANN). An MLP comprises at least three layers: an input layer, a hidden layer and an output layer, with a nonlinear activation function. Each layer of MLP is fully connected through a certain weight, and MLP learns from data by changing the weights of perceptron through the technique called backpropagation [24,25]. Backpropagation works by computing the gradient of the loss function with regard to the weights of the network. Our MLP model is shown in Fig. 7. The 6-layer MLP model was used for learning patients' mortality status, where the input layer is size 6, followed by five hidden layers with size of 1024, 512, 256, 64, and 32, respectively. Given that prediction is a binary classification task, an output layer with a single node was used.

2.3.4. Evaluation metric for prognosis-predicting model

Since the cohort study is a classification problem, accuracy, precision, sensitivity, specificity, F1-score and area under the curve (AUC) were used to evaluate the model. AUC is the area under Receiver Operating Characteristic (ROC) curve, where ROC curve is plotted with true positive rate (TPR) against false positive rate (FPR) at different classification thresholds. TPR equals sensitivity, which defined in Eq (7) and FPR is defined as follows:

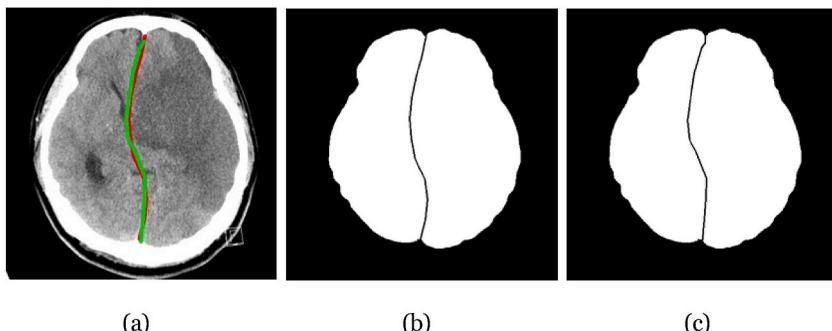


Fig. 6. Brain tissue was extracted and then the ground truth midline and predicted deformed midline were used to separate the left brain and right brain and calculate dice. (a)Predicted midline (green) and ground truth (red). (b)Left and right brain separated by ground truth. (c)Left and right brain separated by predicted midline.

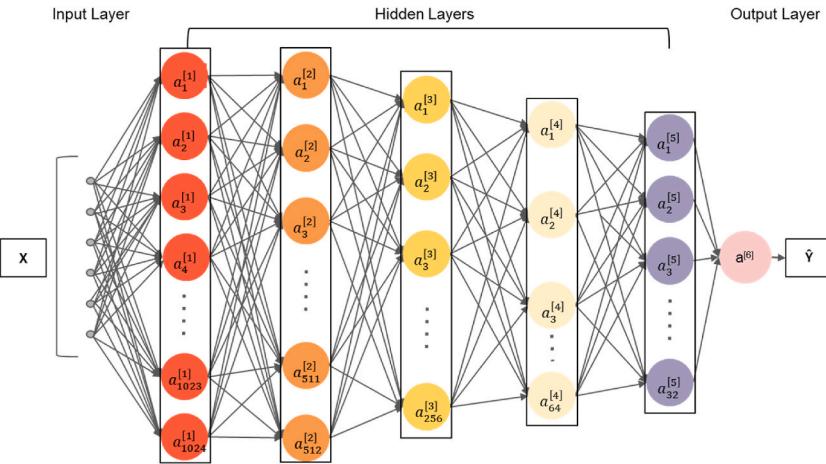


Fig. 7. The MLP model used for cohort study.

$$FPR = \frac{FP}{FP + TN} \quad (12)$$

AUC near to 1 means it has a good measure of separability, by contrast AUC close to 0 means it has a poor measure of separability.

2.4. Experimental environment and parameters

The developed model was implemented using with PyTorch package (version 2.2.0) in Python. The model was trained using a single NVIDIA GeForce RTX 2070 SUPER GPU with 8 GB memory for 100 epochs with a batch size of 4. We use Adam optimizer [26] with a learning rate of 0.001, and employed the mean square error (MSE) as the loss function. The deformed midline labeled by a radiologist was segmented into 10 equal parts, producing 11 keypoints; then the ground truth heatmap was constructed for each keypoints by Gaussian distribution with $\sigma = 10$.

For the cohort study, 6 layers MLP was implemented with Keras package in Python. ReLU was used as the activation function for hidden layers, and Sigmoid for the output layer. Binary cross entropy was used as the loss function. MLP was trained on the same device for 100 epochs with a batch size of 64.

For the midline detection dataset, the data were randomly split into 45 training cases, 5 validation cases and 19 testing cases. The numbers of slices for training, validation and testing were 1336, 155 and 592, respectively. For the prognosis-predicting dataset, a 5-fold cross-validation was adopted for a reliable result. The data were split into five folds, of which three folds had 446 data, and the other two folds had 445 data.

2.5. Statistical analyses

Continuous variables are shown as number (n) and percentage (%) and performed by *t*-test; and categorical values are shown as mean \pm standard deviation (SD) and performed by Chi-squared test, using *scipy* package in Python. Probability (p) values of <0.05 were established as statistical significance.

3. Results

3.1. Results of midline detection model

The results of slice selection are shown in Table 2. All evaluation metrics reached above 0.9. Only 3 false negative slices, and 17 false positive slices were found. Fig. 8 shows three false negative slices with traceable midline drawing along crista galli and cruciform eminence. The causes of false positives are six from the sagittal sinus and seven from the slices near the skull base without clear anatomic structure by partial volume effect from crista galli with bony calvarium, and the other four by subarachnoid hemorrhage (SAH) at anterior cranium fossa. Fig. 9 shows false positive slices with bright linear midline caused by small SDH along the superior sagittal sinus.

Table 2

Results of slice selection.

Accuracy	Precision	Sensitivity	Specificity	F1-score
0.966	0.952	0.991	0.932	0.971

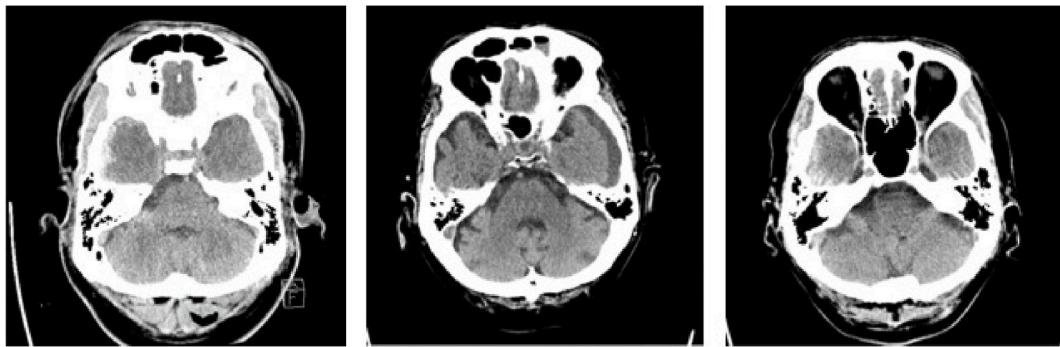


Fig. 8. Three false negative slices in slice selection.

For midline detection, the model's performance was evaluated at slice-level and case-level. For case-level results, MLS_D and MLS_V were used, which are defined in Section 2.2.2, representing one patient. The results of MLS distance and MLS volume are shown in Table 3, respectively. The newly developed method detects the MLS distance with an average error of about 2 mm, and the MLS volume within 1 cm^3 on slice-level, within 5 cm^3 at case-level on average. Fig. 10 shows the successfully detected case.

The model was compared with U-Net and U-Net with residual block. As shown in Table 4, adding residual block and dilated convolution improves the performance greatly. The midline detection model was also tested on ten regular brain CT series (Fig. 11).

The average MLS distances of the brains with MLS and those without MLS are 12.340 ± 5.618 and 2.289 ± 0.765 , respectively. It shows that the brain midline is slightly shifted even in the normal brain. Fig. 12 shows that the model detects large brain MLS cases well.

The results of dice coefficient are shown in Table 5. Both slice-level and case-level dice achieved 0.98. As the results are shown above, our midline detection model performed well for both slice selection and midline detection.

3.2. Results for prognosis-predicting model

For feature selection, p-values for all variables are shown in Table 1. As mentioned in Section 2.3.2, we chose the variables with p-value less than 0.05. That is, age, sex, SAH, IVH, shift distance, shift volume and surgery were shown as features for training the MLP model. Results of the MLP model using undersampling and oversampling are shown in Table 6. The MLP model with undersampling had poor accuracy about 0.5. In contrast, the results of the MLP model with oversampling had accuracy about 0.8. Results of several MLP using different hidden layers are shown in Table 8. The results of MLP using 6 and 4 hidden layers are slightly worse than others. After finding oversampling is the better data balancing method, the model with higher sensitivity, F1 score and AUC was chosen. Different thresholds were used on the output of the MLP model to get the best performance of the model. Outcomes greater than threshold is regarded as dead and the output less than threshold is regarded as not dead, these results are shown in Table 7. Along with the increment of threshold, the accuracy, precision and specificity also increased, while the sensitivity decreased. Fig. 13 shows the ROC curves for three MLP models with different training features.

4. Discussion

Considering the results of slice selection in the present study, false negative and false positive cases are of concern. Since both are these cases could affect the results of MLS_D and MLS_V . Fig. 8 shows three false negative slices. The midline on these slices is obvious but



Fig. 9. Three false positive slices in slice selection.

Table 3

The left side of the table shows mean and standard deviation of the ground truth, prediction and MAE between ground truth and prediction for shift distance. The right side of the table shows shift volume. The values are shown as mean and standard deviation.

	Shift distance (mm)			Shift volume (cm ³)		
	Ground Truth	Prediction	MAE	Ground Truth	Prediction	MAE
Slice-level	9.479 ± 5.214 13.344 ± 5.283	8.382 ± 4.815 12.340 ± 5.618	2.175 ± 2.386 2.496 ± 2.581	3.258 ± 1.802 34.218 ± 19.409	2.816 ± 1.411 29.576 ± 15.306	0.613 ± 0.742 4.539 ± 5.714
Case-level						

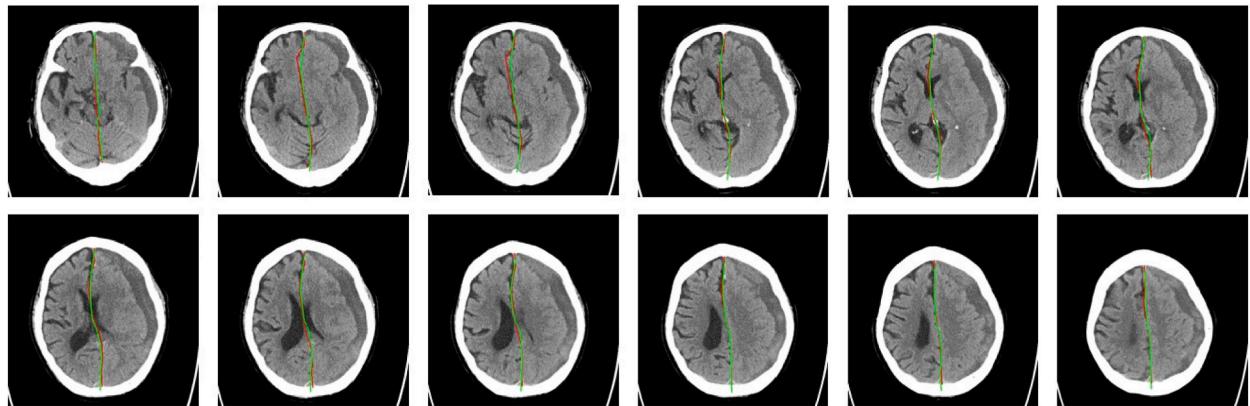


Fig. 10. Prediction of the proposed method on brain MLS. The ground truth is shown as red line and the prediction is shown as green line.

Table 4

Slice-level MAE results of different model structures.

	Shift distance (mm)	Shift volume (cm ³)
U-Net	9.286 ± 10.859	2.479 ± 2.820
	6.028 ± 4.587	1.087 ± 1.012
U-Net + RB	2.175 ± 2.386	0.613 ± 0.742
U-Net + RB + DC ^a		

Abbreviations: RB, Residual block; DC, Dilated Convolution.

^a U-Net + RB + DC is the developed model.

near the skull base. The midline can be drawn by crista galli and cruciform eminence. The crista galli is a bony midline structure with wider and denser than the falx cerebri and with fewer training slices. Fig. 9 represents false positive slices. The image shows bright line density, which is not the brain midline. It is a venous structure called a sagittal sinus with a left-sided small SDH, and its direction is similar to the brain midline. The developed method could not identify the difference between the brain midline and vessels at the present stage. It's just as well that these slices have almost no MLS. Hence these slices were selected not to affect the MLS quantification for further determination in patients' treatment. The human brain is not perfectly symmetrical, hence a slightly MLS exists on the normal brain. The results in Fig. 11 demonstrate this fact.

In the prognosis-predicting cohort study, the performance of MLP with different data preprocessing strategies is shown in Tables 6, 7 and 9. Table 6, depicts the results of undersampling, which were quite poor, since the remaining data after random undersampling became so small that MLP could not learn from these data. Even though k-means clustering and the oversampling method were used to solve the problem of imbalanced data, satisfying results for precision, sensitivity, and F1-score could not be achieved. The results in Table 7 show that using different threshold resulted in different performance. The models with balanced sensitivity and specificity are preferred since we don't want too many false negative and false positive cases. The MLS_D and MLS_V of moderate and severe GCS score patients in Ref. [11] is similar in our cases with mortality status. MLS_V was found to better predict the prognosis of patient than MLS_D . Comparing the present results of using only MLS_D with results of using both MLS_D and MLS_V in Table 9, adding MLS_V as a training feature was shown to improve the performance slightly in the present study. MLS_V performance appears not to stand out significantly when compared to the findings of the study [11]. Further, the results in Table 7 and the green ROC curve in Fig. 13 show that adding clinical information such as training features make the results much better indeed.

We have proposed a three-stage method that is the first to automatically detect the midline of the whole brain. At first, the brain midline was detected for each slice in a CT series. Second, slices were selected that already contain a brain midline. Finally, the MLS distance and volume were quantified for the whole brain. The whole process is automatic and uses only one model, without needing an

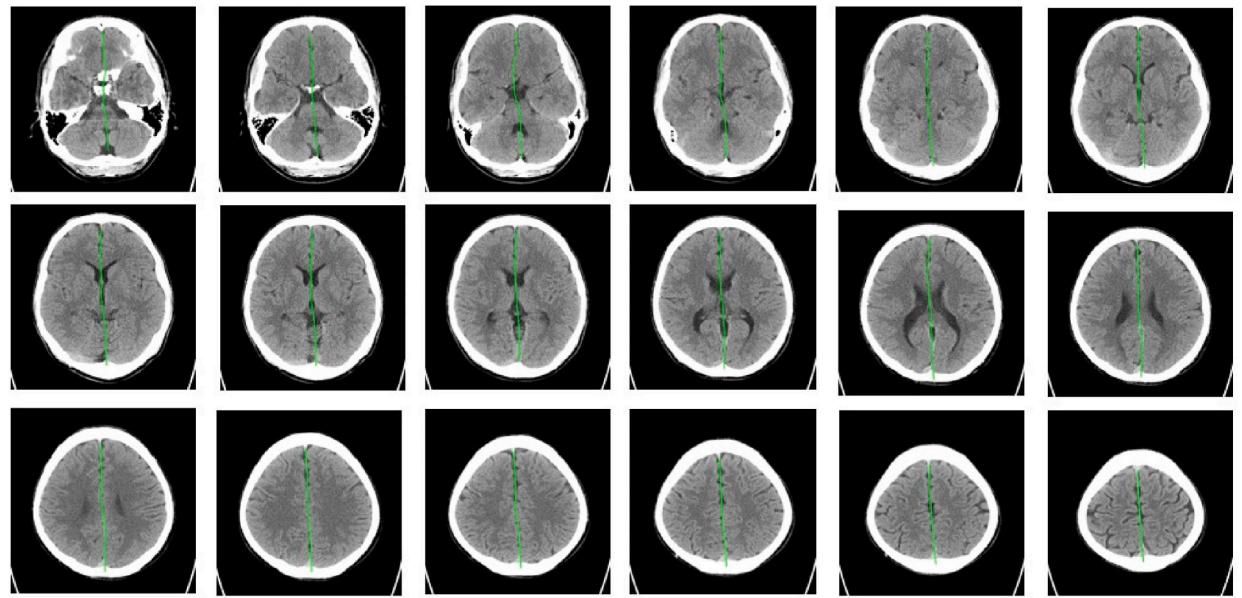


Fig. 11. Prediction of brain without MLS. The prediction is shown as green line.

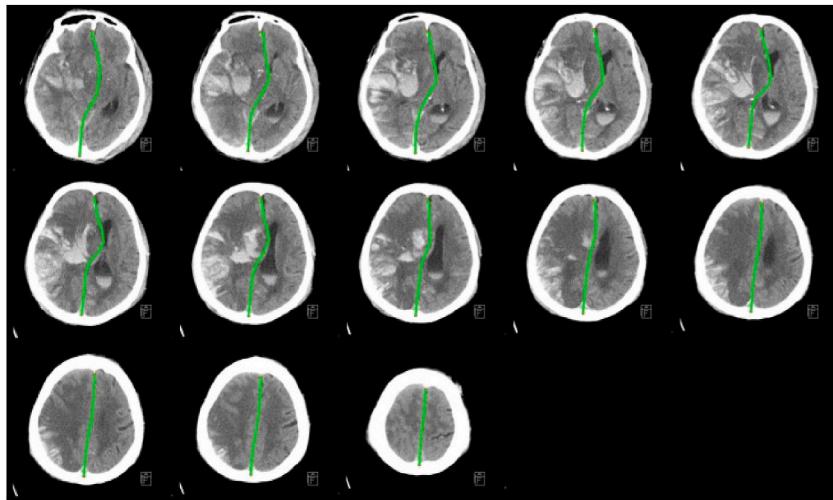


Fig. 12. The developed method performed well on heavy bleeding and cases with large MLS.

Table 5
Dice coefficient for left and right brain, shown at slice-level and case-level respectively.

	Left brain	Right brain
Slice-level	0.983 0.983	0.984 0.983
Case-level		

additional algorithm to find the midline. This method is able to detect and quantify the MLS in 0.42 s, avoiding subjective evaluation by clinicians and providing objective quantitative data quickly. Additionally, with the integration of clinical data, the developed method also demonstrated good performance in predicting patients' mortality. The present study has several limitations, including first that the data only include the brain CT series from one medical center with limited data. Although the slice selection method in this study achieved favorable results, the slices at the top of the skull and the bottom of skull were suboptimal, but without affecting the quantification. When lesions occur on both sides of the brain midline, such as interhemispheric SDH, although the brain parenchyma is

Table 6

Results of MLP for cohort dataset with clinical information using undersampling and oversampling.

	D.B.	F.S.	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
MLP	US	No	0.532	0.202	0.723	0.498	0.315	0.677
	OS	No	0.812	0.416	0.570	0.856	0.480	0.780
MLP	OS	Yes	0.805	0.418	0.683	0.827	0.512	0.819
MLP								

D.B.: Data balancing.

F.S.: Feature selection.

US: Undersampling.

OS: Oversampling.

Table 7

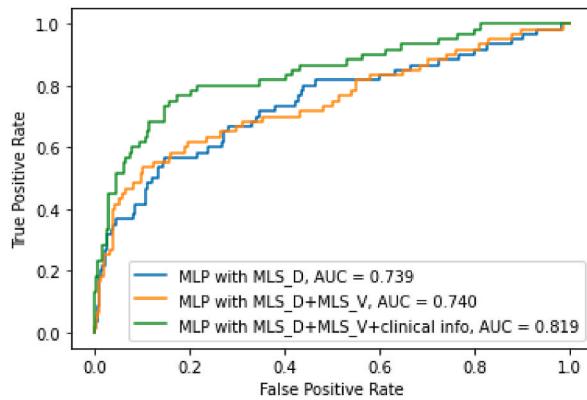
Results of MLP with feature selection using oversampling and different thresholds.

	Threshold	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
MLP	0.4	0.788	0.387	0.713	0.801	0.501	0.819
	0.5	0.805	0.418	0.683	0.827	0.512	0.819
MLP	0.6	0.865	0.556	0.536	0.923	0.543	0.819
MLP	0.7	0.882	0.664	0.443	0.959	0.529	0.819
MLP							

Table 8

Results of MLP using different hidden layers.

Hidden Layers	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
1024, 512, 256, 128, 64, 32	0.796	0.392	0.676	0.817	0.496	0.806
1024, 512, 256, 128, 64	0.802	0.407	0.646	0.829	0.492	0.813
1024, 512, 256, 64, 32	0.805	0.418	0.683	0.827	0.512	0.819
1024, 512, 256, 64	0.802	0.405	0.669	0.826	0.504	0.805
512, 256, 128, 64, 32	0.809	0.419	0.673	0.833	0.511	0.806
256, 128, 64, 32, 16	0.809	0.415	0.656	0.836	0.508	0.813

**Fig. 13.** The ROC curves of the MLP model using different training features.

displaced, there is no MLS. The real medical data often exhibit data imbalance problems, which require the use of data preprocessing techniques or modified model to resolve the problem. The prognosis-predicting dataset of this study was imbalanced, which required data balancing. In the prognosis-predicting cohort, the cases of intracerebral hemorrhage were analyzed solely based on medical history without excluding possibilities of brain tissue displacement, such as tumors, stroke with brain swelling, and quantitative interference caused by midline diseases, such as midline meningiomas.

Our study has several limitations. In this study, we have only 69 cases for building an MLS prediction model. We need to collect more study cases to make our model more robust. Secondly, our model's generalizability is limited by its reliance on a single-institution dataset. Future studies should incorporate external validation using datasets from multiple institutions or open-source databases to

Table 9

Results of MLP using brain hemorrhage types and further only MLS_D , $MLS_D + MLS_V$, and $MLS_D + MLS_V +$ clinical information. The values are shown as mean and standard deviation.

	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
MLP + MLS_D	0.809 ± 0.032	0.415 ± 0.060	0.578 ± 0.028	0.851 ± 0.039	0.477 ± 0.038	0.767 ± 0.017
	0.808 ± 0.016	0.406 ± 0.023	0.561 ± 0.077	0.852 ± 0.030	0.468 ± 0.027	0.766 ± 0.025
MLP + $MLS_D + MLS_V$						
MLP + $MLS_D + MLS_V$ +clinical info.	0.813 ± 0.015	0.417 ± 0.026	0.597 ± 0.086	0.851 ± 0.026	0.488 ± 0.040	0.781 ± 0.041

enhance its applicability. Thirdly, our simplified representation of the midline using only 11 keypoints can introduce quantitative errors in the simulated shift curve. To improve accuracy, future research should explore increasing the number of keypoints and investigating lightweight model architectures for efficient clinical deployment. Fourthly, it's important to note that even in a normal brain, the midline can exhibit slight variations. A clinically significant midline shift is generally considered to be greater than 5 mm. Finally, our model's performance may be limited in emergency settings due to the diverse range of acute neurological conditions, including tumors, stroke, brain edema, and encephalomalacia. Additionally, the model may struggle to differentiate between true midline shifts and false positives caused by midline lesions like interhemispheric tumors or bihemispheric subdural hematomas. To address these limitations, future research should focus on expanding the dataset to include a broader spectrum of acute neurological CT scans.

5. Conclusions

Here we have proposed a deep learning method developed using a keypoint detection model to detect the brain MLS and quantify the brain MLS distance and volume. The method selected slices with high accuracy, precision and sensitivity and demonstrated the ability to detect and quantify the brain MLS well. The results of midline detection model were applied to predict patients' prognosis, suggesting that this method can be used in computer-aided diagnosis. Our ultimate goal for the future is to develop a system to guide decision making for clinicians and patients' families. This system will enable them to make objective assessments based on images, clinical data, and the probability of surgical mortality. To the best of our knowledge, this paper is the first to explore the prediction of brain surgery survival rate from brain offset.

CRediT authorship contribution statement

An-Rong Wu: Writing – original draft, Software, Methodology. **Sun-Yuan Hsieh:** Writing – review & editing, Resources, Project administration. **Hsin-Hung Chou:** Writing – review & editing, Validation, Supervision, Software, Project administration, Methodology. **Cheng-Shih Lai:** Writing – review & editing, Data curation. **Jo-Ying Hung:** Writing – review & editing, Validation. **Bow Wang:** Writing – review & editing, Data curation. **Yi-Shan Tsai:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Data availability

Data will be made available for research purposes on request.

Code availability

Code will be made available for research purposes on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Ministry of Science and Technology (MOST) (MOST 109-2634-F-006-023, MOST 110-2634-F-006-020, MOST 111-2634-F-006-007).

References

- [1] C.-C. Liao, Y.-F. Chen, F. Xiao, Brain midline shift measurement and its automation: a review of techniques and algorithms. *International Journal of Biomedical Imaging* 2018, 2018.

- [2] X. Chen, T. Józsa, D. Cardim, C. Robba, M. Czosnyka, S. Payne, Modelling midline shift and ventricle collapse in cerebral oedema following acute ischaemic stroke, *PLoS Comput. Biol.* 20 (5) (2024) e1012145, <https://doi.org/10.1371/journal.pcbi.1012145>.
- [3] S. Gong, C. Chen, Y. Gong, N.Y. Chan, W. Ma, C.H.-K. Mak, J. Abrigo, Q. Dou, Diffusion model based semi-supervised learning on brain hemorrhage images for efficient midline shift quantification, in: A. Frangi, M. de Bruijne, D. Wassermann, N. Navab (Eds.), *Information Processing in Medical Imaging*, 2023, pp. 69–81.
- [4] N.P. Nguyen, Y. Yoo, A. Chekkouri, E. Eibenberger, et al., Brain midline shift detection and quantification by a cascaded deep network pipeline on non-contrast computed tomography scans, in: *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 487–495.
- [5] C. Qin, H. Li, Y. Liu, H. Shang, H. Pei, X. Wang, Y. Chen, J. Chang, M. Feng, R. Wang, J. Yao, 3d brain midline delineation for hematoma patients, in: M. de Bruijne, P.C. Cattin, S. Cotin, N. Padov, S. Speidel, Y. Zheng, C. Essert (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, 2021, pp. 510–518.
- [6] S. Yu, H.J. Choi, B.C. Kim, M. Ha, K. Kim, J.H. Lee, K. Investigators, et al., Prognosis prediction in severe traumatic brain injury according to initial time of brain computed tomography scan using the Rotterdam scoring system, *Korean J. Nutr.* 18 (2) (2022) 161–168.
- [7] A. Gudigar, U. Raghavendra, A. Hegde, G.R. Menon, F. Molinari, E.J. Ciaccio, U.R. Acharya, et al., Automated detection and screening of traumatic brain injury (tbi) using computed tomography images: a comprehensive review and future perspectives, *Int. J. Environ. Res. Publ. Health* 18 (12) (2021) 6499.
- [8] C.-C. Liao, F. Xiao, J.-M. Wong, I.-J. Chiang, Automatic recognition of midline shift on brain ct images, *Comput. Biol. Med.* 40 (3) (2010) 331–339.
- [9] W. Chen, A. Belle, C. Cockrell, K.R. Ward, K. Najarian, Automated midline shift and intracranial pressure estimation based on brain CT images, *JOVE* (74) (2013) 3871.
- [10] S. Jain, T.V. Vyvere, V. Terzopoulos, D.M. Sima, E. Roura, A. Maas, G. Wilms, J. Verheyden, Automatic quantification of computed tomography features in acute traumatic brain injury, *J. Neurotrauma* 36 (11) (2019) 1794–1803.
- [11] X. Xia, X. Zhang, Z. Huang, Q. Ren, H. Li, Y. Li, K. Liang, H. Wang, K. Han, X. Meng, Automated detection of 3d midline shift in spontaneous supratentorial intracerebral haemorrhage with non-contrast computed tomography using deep convolutional neural networks, *Am. J. Tourism Res.* 13 (10) (2021) 11513.
- [12] C. Jiang, J. Cao, C. Williamson, N. Farzaneh, V. Rajajee, J. Gryak, K. Najarian, S.R. Soroushmehr, Midline shift vs. mid-surface shift: correlation with outcome of traumatic brain injuries, in: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2019, pp. 1083–1086.
- [13] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.
- [14] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2d pose estimation using part affinity fields, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7291–7299.
- [15] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [16] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer Assisted Intervention*, Springer, 2015, pp. 234–241.
- [17] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12) (2017) 2481–2495.
- [18] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, *arXiv Preprint arXiv:1511.07122*, 2015.
- [19] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [20] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 *arXiv preprint arXiv:1409.1556*.
- [21] R.M. Haralick, S.R. Sternberg, X. Zhuang, Image analysis using mathematical morphology, *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-9 (4) (1987) 532–550, <https://doi.org/10.1109/TPAMI.1987.4767941>.
- [22] G. Chandrashekhar, F. Sahin, A survey on feature selection methods, *Comput. Electr. Eng.* 40 (1) (2014) 16–28, <https://doi.org/10.1016/j.compeleceng.2013.11.024>.
- [23] R. Lippmann, An introduction to computing with neural nets, *IEEE ASSP Mag.* 4 (2) (1987) 4–22, <https://doi.org/10.1109/MASSP.1987.1165576>.
- [24] F. Rosenblatt, Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Tech. rep., Cornell Aeronautical Lab Inc Buffalo NY, doi:10.1007/978-3-642-70911-1_20, 1961.
- [25] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representation by error propagation, Tech 2, rep., California Univ San Diego La Jolla Inst for Cognitive Science, 1985, p. 50035, <https://doi.org/10.1016/B978-1-4832-1446-7>.
- [26] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 2014 *arXiv:arXiv:1412.6980*.