Mind the Clot: Automated LVO Detection on CTA using Deep Learning

Shubham Kumar[‡], Arjun Agarwal[‡], Satish Golla[‡], Swetha Tanamala[‡], Ujjwal Upadhyay[‡], Subhankar Chattoraj[‡], Preetham Putha[‡], and Sasank Chilamkurthy[‡]

[‡] Qure.ai, Mumbai, Maharashtra, India,

{shubham.kumar, arjun.agarwal, satish.golla, swetha.tanamala, ujjwal.upadhyay, subhankar.chattoraj, preetham.putha, sasank.chilamkurthy}@qure.ai

Abstract

Globally, stroke is a leading cause of death and disability, and accurate and timely diagnosis of large vessel occlusion (LVO) is essential for positive outcomes. We present our robust deep learning-based method for detecting both internal carotid artery (ICA) and middle cerebral artery (MCA) large vessel occlusions (LVO) from computed tomography angiography (CTA) scans. Our proposed two LVO detection models achieved an overall combined accuracy of 90.9% with a sensitivity of 89.8% and specificity of 91.4%. Further, the proposed model is lower in computational complexities and produces the results in less than 40 seconds, validating its adequacy for deployment in the clinical workflow.

1. Introduction

Acute Ischemic Stroke (AIS) is a significant cause of death and disability worldwide, and timely diagnosis and management are essential for better treatment prognosis outcomes [1]. In high-income countries, the thirty-day mortality rate of AIS is around 15% [2][3]. The patients admitted to critical care units generally have poor prognoses while the mortality rate is as high as 50% [4][5]. AIS caused by LVOs requires a prompt response in prehospital triage similar to acute myocardial infarcts detection using electrocardiograms [6]. The reduction in morality and morbidity for suspected LVOs patients depends widely on timely diagnosis and treatment via endovascular thrombectomy (ET) [7][8]. Post 2015, based on several randomized control trials published that endovascular therapy is more efficacious than alteplase alone [9][10][11][12][13]. However, the outcome of using positive endovascular therapy to achieve vessel recanalization and disability-free life for patients is vastly time-dependent. According to a recent study, the medicare population increased by 5% from 2012-2019 while the shortage of diagnostic radiology workforce has grown significantly [14]. Moreover, unfamiliarity among non-neuroradiologists with reading CTA during on-call hours and the comparatively small caliber of the M2 and A2 segments may cause a diagnostic error. A robust automated solution is required to identify LVO from CTA scans to address these issues and reduce inter-reader variability.

In recent years, a deep learning-based computer-aided diagnostic model capable of learning underlying deep discriminative features has proven to outperform hand-crafted based feature techniques in the multidisciplinary domain such as breast cancer detection using histopathological images [15], automated ASPECT scoring [16], and detection of critical finding from head CT scans [17][18]. A recent study published to detect LVO from Multiphase CTA scans using Deep Convolutional Neural Network (CNN) has reported an AUC of 0.74, a sensitivity of 77%, and a specificity of 71% [19] on 63 images. However, the study had relatively small positive cases, and a granular image of the model's capability in detecting the occlusion locations was also not reported. In another work, a commercially available LVO detection algorithm was utilized to detect Middle cerebral artery-M1 (MCA-M1) and Internal Carotid Artery-Terminus (ICA-T) occlusions. Overall analysis on 876 CTA scans was performed, and the AUC of 86.3%, the sensitivity of 90.1%, and the specificity of 82.5% were reported [20]. However, the published study needs to include the identification of LVOs from the source image and the generalizability of the proposed model to be deployed in real clinical settings and the resource constraint environment. In addition, these methods often require significant manual annotation of training data and are limited in performance due to the need for large amounts of training data.

To address these limitations, we present a deep learning-based approach for detecting MCA and ICA LVOs in CTA

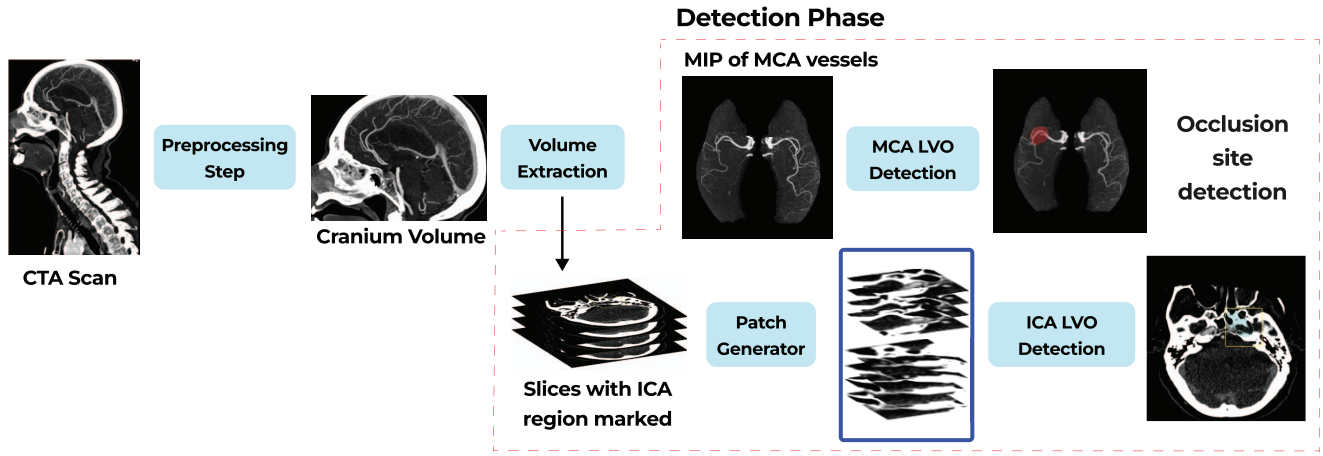

Figure 1: Algorithm Description: First, a CTA scan is cropped to the cranium volume as detailed in section 2.1. As described in section 2.2, we extract the relevant arterial regions to focus on. Finally, it outputs the MCA VT volume, on which the MIP is reconstructed, and the ICA VT volume. The ICA VT volume uses the Patch Generator to generate and stack left and right ICA patches. Later, we feed the MCA MIP to the MCA LVO Detection model, as detailed in section 2.3, to output the occlusion site if present. We also feed the stack of ICA patches to the ICA Detection Model, as detailed in section 2.4, to output per-patch predictions. These predictions are aggregated to give the final label per CTA scan.

scans. The proposed model is extensively validated and demonstrates significant performance in standard evaluation metrics. Moreover, our model eliminates the need for detailed manual annotation and segmentation, enabling radiologists to make faster and more accurate decisions.

The main contributions of this work are summarized as follows:

- We propose a novel method to generate high-quality Maximum Intensity Projections (MIP) by mitigating venous phase contamination caused by delayed acquisition. Even if our detection model may fail for distal MCA occlusions, these MIPs can still be valuable for visual inspection.
- For ICA detection, we introduce a novel generalized framework that utilizes patch-level predictions and aggregates them to obtain a final decision.
- The proposed model classifies LVOs in under 40 seconds, making it suitable for implementation in time-sensitive stroke management settings, crucial for positive stroke outcomes.

1.1. Related Work

Several existing solutions describe automated CTA LVO detection. In [21], they use elastic registration to remove the skull and then compare contralateral hemispheres to detect vessel density asymmetry, determining normal blood flow or occlusion. The median turn-around time was 158 seconds (IQR: 140–176 seconds), with elastic registration being the most time-consuming step (130 seconds).

However, this method has limitations. It may miss distal M1-MCA occlusions if there is reconstitution of proximal M2 segments [22]. Delayed CTA acquisition can cause venous phase contamination, affecting prediction results when comparing contralateral hemispheres. Similarly, there are many FPs noticed because due to stenosis in ICA causes asymmetrical vascular density in contralateral hemisphere resulting in false prediction. Moreover, many petrous and cervical ICA occlusions are missed as the algorithm's focus is not on these regions. The results of contralateral comparison heavily depend on skull stripping, and elastic registration does not guarantee intact ICA vessels due to its vicinity.

In [23], a deep learning CNN solution was presented to detect occlusions. It provides a binary output (LVO detected: 'Yes' or 'No'). For positive LVO cases, it displays an occlusion box centered around the proximal occlusion site in axial, coronal, and sagittal views. The focus is on ICA Terminus occlusion.

However, one drawback is the relatively long processing time, taking approximately 4 minutes and 59 seconds (SD±1 minute and 12 seconds) as reported in the paper. In their supplementary paper, they evaluated the algorithm's diagnostic performance on the MR CLEAN Registry [24] and PRESTO [25] datasets, considering the impact of CTA acquisition timing on imaging quality. They observed a significant drop in performance metrics in the venous phase compared to the peak arterial or equilibrium phase.

In [26], they propose a 3D segmentation CNN based on the U-Net [27] architecture to segment ICA-T or M1 segment vessels. This CNN is trained on numerous manual

segmentations of the ICA-T and M1 regions. To identify cases where ICA-T and M1 segments are not visible due to an ICA occlusion and no retrograde filling, they compare the lengths of the left and right segmentations. If one side is significantly shorter than the other, it indicates an LVO, and the system triggers an alert.

Additionally, another 3D CNN is trained to segment all vessels (not limited to ICA-T and M1 vessels). The outputs of both networks are combined to refine the MCA vessel tree. Finally, they identify the end points of the MCA vessels and check if the total distance between the ICA-T and the end point falls below a predefined threshold. If so, an LVO is detected, and the system triggers an alert.

2. Methodology

2.1. Data Pre-processing

In stroke evaluation, CTA scans are commonly used to assess the head and neck region. Our primarily focus on to detect LVO in the head, including the cervical segment of ICA.

Let $\mathbf{X} \in \mathbb{R}^{N \times H \times W}$ represent the input CTA scan, where N is the total number of slices. We denote each slice of $\mathbf{X}$ as $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_N$, forming a sequence of DICOM slices in $\mathbb{R}^{H \times W}$.

The **Cranium Slice Selector**, denoted as $f_{\text{CSS}} : \mathbb{R}^{H \times W} \rightarrow [0, 1]$, is a deep learning model that classifies CTA slices based on the presence of the cranium.

We feed all the slices $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_N$ into the Cranium Slice Selector model, resulting in predicted probabilities $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_N$, where $\hat{p}i = f_{\text{CSS}}(\mathbf{I}_i)$.

Using a threshold value t, we define a binary sequence $\mathbf{B} = [b_1, b_2, \ldots, b_N]$ where $b_i = \begin{cases} 1, & \text{if } \hat{p}_i > t \\ 0, & \text{otherwise} \end{cases}$.

The estimated cranium location range is determined by finding the length ℓ of the longest consecutive subsequence of slices with a value of 1 in $\mathbf{B}$, given by:

$$\ell = \max_{i,j} \sum_{k=i}^{j} b_k, \quad \text{where } 1 \leq i \leq j \leq N$$

By obtaining the maximum length ℓ of the consecutive subsequence, we estimate the cranium location with a high probability of containing the cranium.

Let $\mathbf{X}_{\text{cranium}} \in \mathbb{R}^{L \times H \times W}$ represent the set of slices obtained from the Cranium Slice Selector model.

The **Cranium Localization model** , denoted as $f_{\text{loc}} : \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^4$, is developed using bounding box regression. The model is trained by minimizing the mean-square-error (MSE) between the predicted coordinates and the ground truth coordinates. Here, H and W represent the height and width of the CT slice.

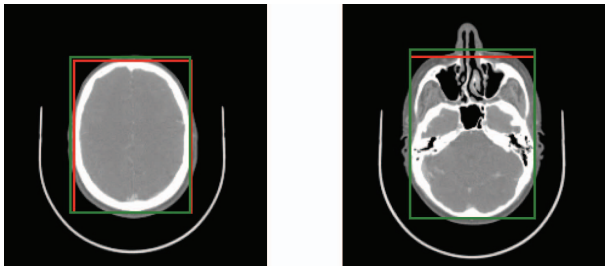

Figure 2: Bounding box regressor output (green one) vs ground truth (red one)

For each slice $\mathbf{D}_i \in \mathbf{X}_{\text{cranium}}$, the object localization model f_{loc} predicts the bbox coordinates as follows:

$$f_{\text{loc}}(\mathbf{D}_i) = (x_{\text{min},i}, y_{\text{min},i}, x_{\text{max},i}, y_{\text{max},i}) \tag{1}$$

We take the minimum value of $x_{\text{min},i}$ and $y_{\text{min},i}$ among all slices, and the maximum value of $x_{\text{max},i}$ and $y_{\text{max},i}$ among all slices. Now final bounding box we get be this $(x_{\text{min}}, y_{\text{min}}, x_{\text{max}}, y_{\text{max}})$, applied on $\mathbf{X}_{cranium}$ to crop out unnecessary background and we get $\mathbf{X}_{cropped} \in \mathbb{R}^{L \times H' \times W'}$.

$$\mathbf{X}_{cropped} = \mathbf{X}_{cranium}[:, y_{\text{min}} : y_{\text{max}}, x_{\text{min}} : x_{\text{max}}] \tag{2}$$

We introduced **Tilt-Correction** to standardized the input for our proposed LVO Detection models.

Let $\mathbf{I}_{\text{fixed}}$ be the fixed template, which is a symmetric CTA Head scan or tilt-corrected CTA scan. Let $\mathbf{I}_{\text{moving}}$ be the moving image, which is $\mathbf{X}_{cropped}$ obtained from the object localization model.

The goal is to register the moving image $\mathbf{I}_{\text{moving}}$ with respect to the fixed template $\mathbf{I}_{\text{fixed}}$ resulting in a symmetric or tilt-corrected CTA scan using the antspy library, which utilizes the affine-fast transform for image registration.

The registration process can be represented as a function $f_{\text{registration}} : \mathbb{R}^{D_{\text{moving}} \times H_{\text{moving}} \times W_{\text{moving}}} \rightarrow \mathbb{R}^{D_{\text{fixed}} \times H_{\text{fixed}} \times W_{\text{fixed}}}$, where H_{moving}, W_{moving}, and D_{moving} are the height, width, and depth of moving image, and H_{fixed}, W_{fixed}, and D_{fixed} are the height, width, and depth of the fixed template.

The registered image $\mathbf{I}_{\text{registered}}$ is obtained by applying the image registration function to the moving image:

$$\mathbf{I}_{\text{registered}} = f_{\text{registration}}(\mathbf{I}_{\text{moving}}) \tag{3}$$

2.2. Volume Extraction

In this section, we present two semantic segmentation models developed for the extraction of the intracranial region and vascular territory of interest from CTA scans. These volumes are essential for the proposed Large Vessel Occlusion (LVO) detection models.

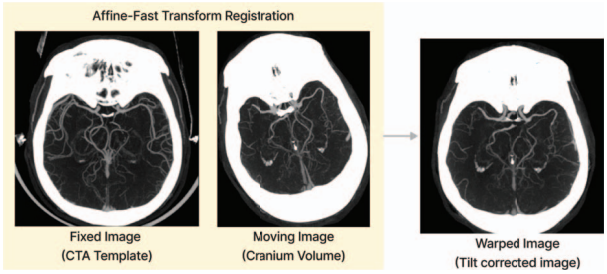

Figure 3: Tilt correction using affine fast transform registration using ANTSpy [28] .

Our motivation for undertaking this to overcome the challenges posed by bones and overlapping large vessels, as well as the presence of visible venous phases caused by delayed acquisition. These factors often obstruct the vessels of interest and introduce noise into the MIP image. Some example of noisy MIP shown in Fig 4.

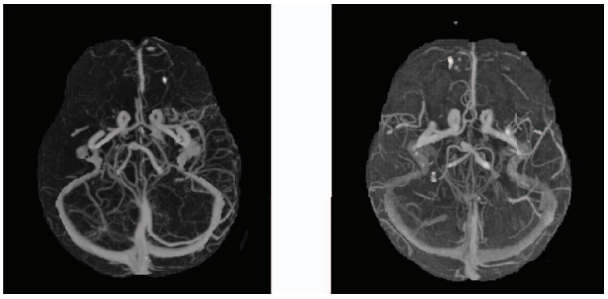

Figure 4: MIP of ICV showing venous phase and vessel overlap

Some example of clear MIP only showing vessel of interest, in our case is MCA is shown in Fig 5.

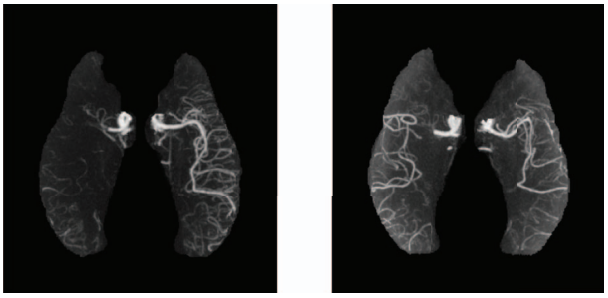

Figure 5: MIP showing MCA vessels only

The **Intracranial Volume semantic segmentation model** aims to accurately separate the vasculature from the surrounding bone structures within the cranium. This segmentation model was trained and validated using a dataset consisting of 485 CTA scans. To annotate the Intracranial Volume (ICV) mask on CTA slices, corresponding non-contrast CT scans (NCCT) with matching dimensions were used as a reference.

The **Vascular Territory semantic segmentation model** is specifically designed to identify and segment arterial territories in the brain from CTA scans. For training and validation, a dataset of 942 CTA scans was meticulously annotated by an expert neurologist.

During the annotation process, two distinct arterial regions were delineated: the Middle Cerebral Arterial (MCA) region and the Internal Carotid Arterial (ICA) region. Furthermore, the arterial regions were subdivided based on the contra-lateral hemispheres, resulting in four binary masks per CTA scan. These masks enable precise localization and differentiation of arterial territories within the brain, providing valuable information for subsequent analysis and detection algorithms.

The input to both models is a CT slice resized to 224×224, with four channels generated by manipulating the window width and window level parameters. The optimal values for these parameters are (250,500), (900,300), (600,300), and (max, min). The 4-channel input image is generated using the following equation:

$$\mathbf{I_{4ch}} = f(\mathbf{I_{CT}}, ww, wl) \tag{4}$$

where $\mathbf{I_{CT}}$ represents the original CT slice, ww and wl are the window width and window level, respectively, and f is a function that manipulates the Hounsfield Unit (HU) values of the CT slice to generate the four channels.

Both segmentation models utilize the U-Net [27] architecture with ResNet-18 [29] backbones similar to as shown in figure 8. The U-Net architecture is depicted as follows:

$$\mathbf{M}_{\text{pred}} = \text{U-Net}(\mathbf{I}_{4ch}) \tag{5}$$

Here, $\mathbf{M}_{\text{pred}}$ denotes the predicted mask

During the ICV segmentation, differentiating the Internal Carotid Artery (ICA) from other structures poses a challenge for the U-Net model [27], as the ICA passes in close proximity to the cranium, and its HU values fall within a similar range. To address this issue, additional weightage is given to the loss function derived from the ICA territory region and the inferior slices of the cranium.

$$L(z) = \begin{cases} L_1, & \text{if } z > 0.4n \\ \alpha * L_1 + \beta * L_{ICA}, & \text{otherwise} \end{cases}$$

Here, n represents the number of slices present in the cranium volume. The function L calculates the loss based on the slice location z within the cranium. L_1 denotes the focal loss [30] calculated on the segmentation output, and L_{ICA} represents the focal loss calculated specifically for the ICA territory region. Through experimentation, we determined that the values $\alpha = 2$ and $\beta = 5$ yielded the best performance for our task.

In the case of the *Vascular Territory Semantic Segmentation Model*, the output is a 4-channel output with the same height and width as the original image. Each channel corresponds to one class of output, representing the left MCA territory, right MCA territory, left ICA territory, and right ICA territory. This can be represented as:

$$\mathbf{M}_{\text{pred}} = \begin{bmatrix} \mathbf{M}_{\text{leftMCA}} \\ \mathbf{M}_{\text{rightMCA}} \\ \mathbf{M}_{\text{leftICA}} \\ \mathbf{M}_{\text{rightICA}} \end{bmatrix} \tag{6}$$

Where $\mathbf{M}_{\text{leftMCA}}$, $\mathbf{M}_{\text{rightMCA}}$, $\mathbf{M}_{\text{leftICA}}$, and $\mathbf{M}_{\text{rightICA}}$ are the binary masks representing the respective territories.

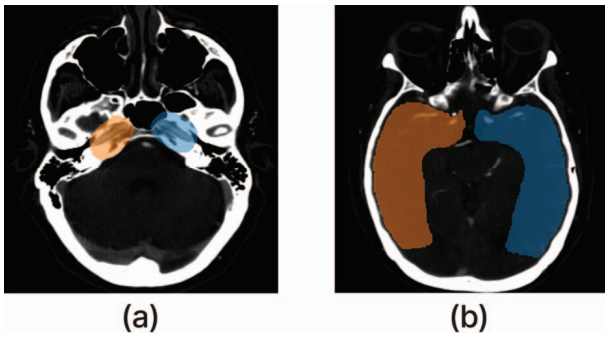

Figure 6: Vascular territory segmentation output (a) ICA Territory (b) MCA Territory

On the other hand, the *Intracranial Volume Semantic Segmentation Model* produces a single-channel output, which is the ICV mask. This can be represented as:

$$\mathbf{M}_{\text{pred}} = \mathbf{M}_{\text{ICV}} \tag{7}$$

Where $\mathbf{M}_{\text{ICV}}$ is the binary mask representing the Intracranial Volume.

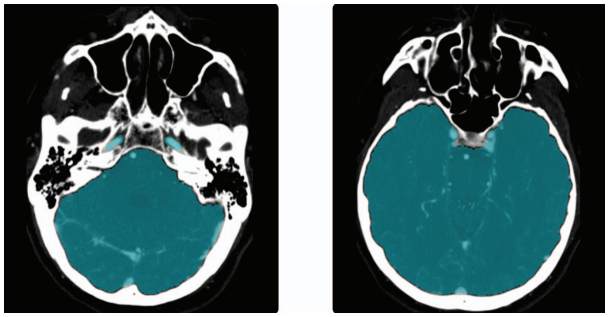

Figure 7: Intracranial volume Segmentation model output

Volume required for LVO Detection models,

$$\mathbf{X}_{\text{ICV}} = \mathbf{X}_{cropped} \odot \mathbf{M}_{\text{ICV}} \tag{8}$$

$$\mathbf{M}_{\text{MCA}} = \mathbf{M}_{\text{leftMCA}} \vee \mathbf{M}_{\text{rightMCA}} \tag{9}$$

$$\mathbf{X}_{\text{MCA}} = \mathbf{X}_{\text{ICV}} \odot \mathbf{M}_{\text{MCA}} \tag{10}$$

$$\mathbf{X}_{\text{leftICA}} = \mathbf{X}_{cropped} \odot \mathbf{M}_{\text{leftICA}} \tag{11}$$

$$\mathbf{X}_{\text{rightICA}} = \mathbf{X}_{cropped} \odot \mathbf{M}_{\text{rightICA}} \tag{12}$$

Finally on $\mathbf{X}_{\text{MCA}}$ and $\mathbf{X}_{\text{ICV}}$ volume, CT windowing are applied with a window width of 600 and window level of 300 to enhance the visibility of contrast-enhanced vessels.

2.3. MCA LVO Detection Model

As mentioned in section 1.1, some solutions focus on comparing MCA vessel voxels between contralateral hemispheres, but they may not perform well in cases of vessel reconstitution or venous phase contamination. Another solution detects the occlusion site, but its performance significantly drops in venous phase scan. Moreover, their algorithm takes 4-5 minutes, which is impractical for triage.

To address these issues, our algorithm detects abrupt changes in MCA blood flow even their is reconstitution. We also mitigate the effects of venous phase contamination by extracting only the arterial region, leaving behind contrast-enhanced veins as described in section 2.2. Furthermore, we reconstruct a 2D MIP from 3D CTA volume, which reduces time complexity compared to 3D models using the whole 3D volume as input.

Let's define the MIP of a given 3D volume as follows:

$$\mathbf{MIP}(\mathbf{X}) = \max_{n=1}^{N} \mathbf{X}(n, i, j)$$

Here, $\mathbf{X} \in \mathbb{R}^{N \times H \times W}$ represents the 3D volume, and $\mathbf{MIP}(\mathbf{X}) \in \mathbb{R}^{H \times W}$ denotes the resulting 2D image.

The equation calculates the maximum intensity value at each pixel location (i, j) across the entire depth dimension (N) of the 3D volume $\mathbf{X}$. These values are then stored in the corresponding locations within the 2D image $\mathbf{MIP}(\mathbf{X})$.

Our proposed model aims to classify whether an MIP image exhibits LVO on MCA sites. Our objective was to test the hypothesis that factors like venous phase and other vasculature overlap could significantly affect the performance of AI systems in detecting LVO. To investigate this, we explored two distinct sets of inputs: $\mathbf{MIP}(\mathbf{X}_{\text{ICV}})$ and $\mathbf{MIP}(\mathbf{X}_{\text{MCA}})$ along with their tilt-corrected counter-parts, as described in section 2.1. The summarized results of these experiments are available in Table 2. MIPs are rescaled to 224×224 to be used as inputs. Data augmentations are applied such as random rotation, shift and flipping.

Initially, we utilized a ResNet-18 architecture for binary classification on both input sets to determine the presence of MCA LVO. To incorporate pixel-level annotations, provided by an experienced neuro-radiologist, around the occlusion site, we extended the ResNet-18 encoder by adding

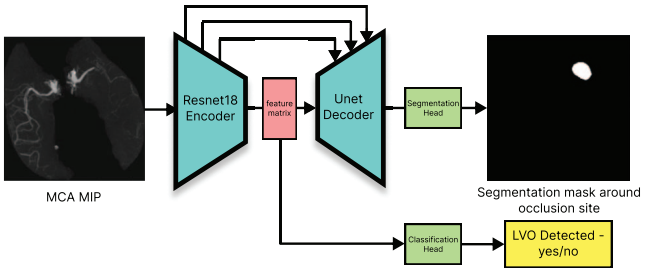

Figure 8: Used architecture in MCA LVO Detection model

a decoder, resulting in a U-Net-like architecture. This modification allowed us to train the model with additional information. Consequently, the model produces two outputs: the classification probability and a segmentation mask as shown in figure 8.

These two outputs contributed to the total loss for our model, as shown in Equation 13.

$$L = \alpha * L_1 + \beta * L_2 \tag{13}$$

Here, L_1 represents cross-entropy loss on the classification output, and L_2 represents the focal loss [30] on the segmentation output. We performed a grid search on a log-linear scale between 1 and 100 to determine the optimal values for the hyperparameters α and β. Ultimately, we found that $\alpha = 1$ and $\beta = 0.5$ worked best for our task.

2.4. ICA LVO Detection Model

As highlighted in section 1.1, the current solutions primarily focus on proximal ICA, which might encounter challenges when dealing with situations involving clots at the ICA origin and reconstitution at the proximal level. In order to overcome this drawback, we introduce a novel approach that expands the analysis from ICA-T to the origin of ICA.

Furthermore, existing solutions commonly rely on a skull stripping method, which poses a challenge in preserving intact ICA vessels. To overcome this limitation, our proposed method eliminates this step and instead adopts the use of the entire ICA territory volume, as thoroughly explained in section 2.2.

Our proposed model, as discussed in section 2.1, utilizes tilt-corrected CTA scans obtained through registration with a CTA Template. By using the CTA Template as a reference, we precisely identify a slice range of 65 slices containing the ICA.

In section 2.2, we describe the application of ICA territory masks on the cranium volume, as shown in equations 11 and 12. This application ensures that regions outside the ICA territory are effectively masked out, resulting in empty regions with voxel values set to zero.

By cropping out these empty regions from $\mathbf{X}_{\text{leftICA}}$ and $\mathbf{X}_{\text{rightICA}}$ and selecting the 65 slices containing the ICA region, as mentioned earlier, we successfully obtain the desired patches. This process is visually depicted in figure 9, illustrating the extraction of the ICA patch from the CT volume.

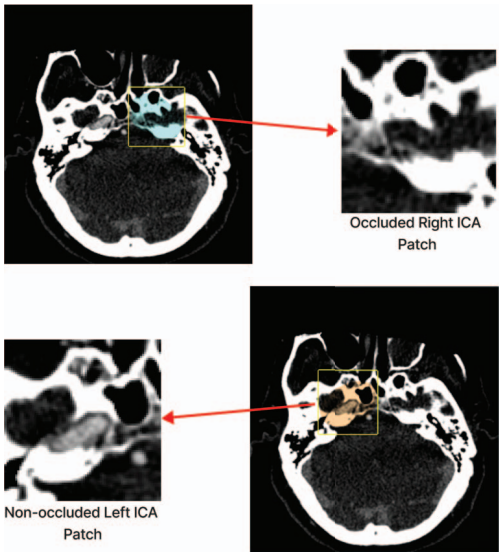

Figure 9: A total of 130 patches containing the ICA region are obtained from both sides, with 65 patches extracted from each contra-lateral side.

In collaboration with experienced neuro-radiologists, we annotated ICA patches and classify them into several sub-classes. Considering our focus on determining LVO, we combine these sub-classes into two main classes: LVO or non-LVO. The sub-classes we consider are as follows:

- Normal blood flow in ICA: Non-LVO
- Calcified ICA with blood flow: Non-LVO
- Narrowing in ICA: Non-LVO
- Occluded ICA: LVO
- Occluded ICA with calcification: LVO

Model Development We first used ResNet-18 [29] as encoder and performs binary classification on single patches converted into 6 channel inputs as described in figure 10, outputting the classification label- LVO/Non-LVO.

After experimenting with different configurations, we achieved better sensitivity and specificity of patch classfication model.

To achieve our ultimate goal of identifying LVOs per CTA scan, we utilize a method that aggregates predictions from all the patches. This involves training another model using the pre-trained weights from the patch model, with a batch size of 130. By employing this strategy, we seamlessly combine patch-wise and scan-wise learning within

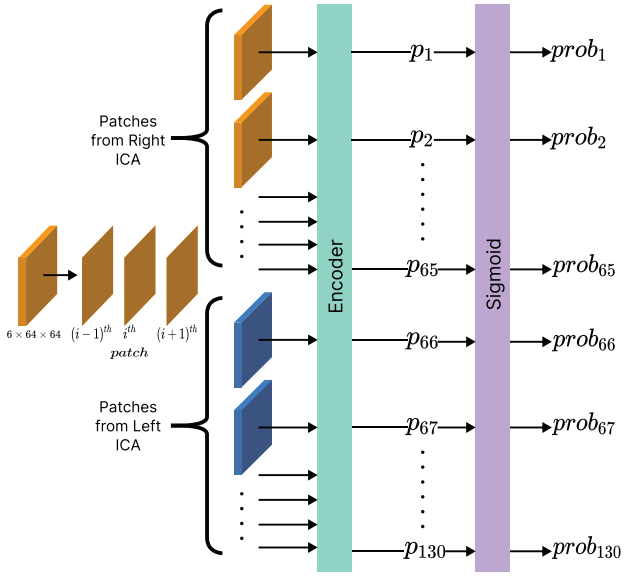

Figure 10: Each i^{th} patch is concatenated with its adjacent $(i-1)^{th}$ and $(i+1)^{th}$ patches to incorporate nearby context. Consequently, each patch is transformed into a two-channel input using CT windowing, resulting in a model input to be dimension of $6 \times 64 \times 64$.

the same model, allowing for comprehensive analysis of the entire scan.

To ensure a balanced representation from both contralateral sides we choose tilt-corrected CTA scans to obtain an equal number of patches from both sides.

The equations below describe the process of combining these losses:

$$prob_i = \text{sigmoid}(p_i) \tag{14}$$

$$loss_{patch} = \text{loss}([prob_1, ..prob_{130}], [gt_1, ..gt_{130}]) \tag{15}$$

$$logits_{scan} = f(p_1, p_2, ..., p_{130}) \tag{16}$$

$$loss_{scan} = \text{loss}(logits_{scan}, gt_{scan}) \tag{17}$$

$$prob_{scan} = \text{sigmoid}(logits_{scan}) \tag{18}$$

$$L = \alpha \cdot loss_{patch} + \beta \cdot loss_{scan} \tag{19}$$

Here, $prob_i$ represents the patch probability calculated using the sigmoid activation function applied to the logits output p_i from the patch classification model, gt_i denotes the corresponding ground truth labels for the patches. The log-sum-exponential function f aggregates the patch logits to obtain $logits_{scan}$, gt_{scan} refers to the ground truth for LVO per CTA scan.

The combined loss L is determined by adding the cross-entropy loss $loss_{patch}$, which takes into account the predictions and ground truth labels for all the patches in a scan, and the cross-entropy loss $loss_{scan}$, which is based on the predictions and ground truth for the entire scan. To optimize the model's performance, we found that setting $\alpha = 1$ and $\beta = 1.5$ as hyperparameters worked best for our task.

By leveraging the information from both individual patches and the complete scan, our model effectively integrates local and global context for accurate LVO identification in the CTA scans.

Finally, the second model takes all the patches present in a CTA scan as inputs and outputs the probability $prob_{scan}$ of LVO for the entire scan. By applying an optimal threshold to the output probabilities, we obtain the final LVO detection result.

3. Experimental Result and Discussion

3.1. Dataset Description

The dataset used in this study consists of cases with positive and negative instances of MCA and ICA LVOs. All the CTA scan in our dataset have a resolution of 512×512 pixels and spacing between slices $\leq 1.25mm$. The proposed model was validated using this dataset, and it's worth noting that the datasets used in the ICV and VT segmentation models are independent of the dataset used for the development of the LVO models.

The detailed distribution of the dataset is presented in Table 1:

Table 1: Dataset Distribution

Dataset	MCA		ICA	
	Positive	Negative	Positive	Negative
Train	831	3837	250	507
Val	170	1119	93	64
Test	372	538	256	538
Total	1373	5494	599	1209

3.2. Experimental Settings and Performance Evaluation Metrics

In section 2, when employing U-Net for the segmentation task, we conducted experiments with different decoder depths. After thorough evaluation, we opted to use a decoder depth of 3 for our final model. This choice was driven by the finding that increasing the decoder depth did not lead to substantial performance improvements and instead introduced higher computation time as a trade-off.

Our chosen architecture consisted of an ResNet-18 encoder depth of 3 with decoder channels set to [64, 32, 16]. We incorporated batch normalization [31] for improved training stability, and for decoder attention, we utilized scSE (Spatial and Channel Squeeze & Excitation) [32] block.

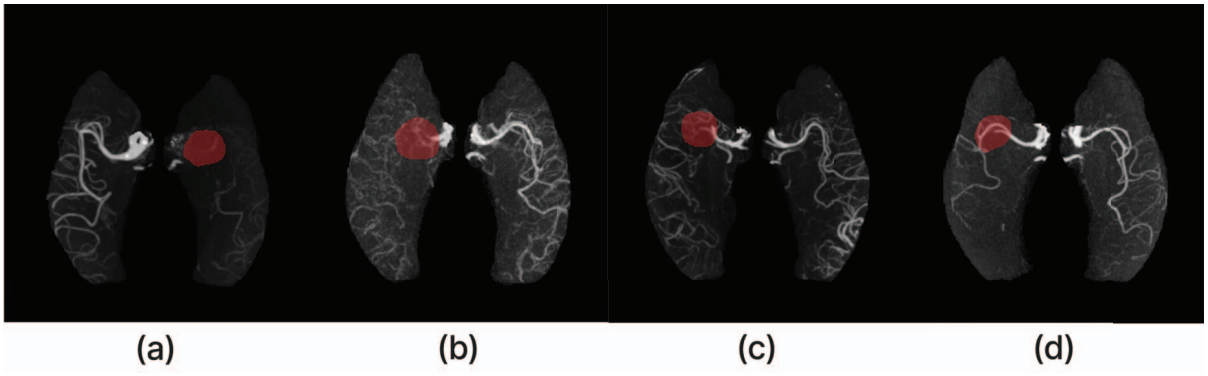

Figure 11: Some examples of MCA LVO detection model output (a) complete occlusion on the left side of the brain, (b) occlusion at the ICA terminus, (c) occlusion in M1 segment of MCA branch, (d) proximal M2 occlusion

Table 2: Performance Evaluation based on Test Set

MCA LVO Detection			
Task	**Sensitivity**	**Specificity**	**DSC**
ICV MIP Classification	0.65	0.67	-
MCA MIP Classification	0.76	0.77	-
MCA MIP Classification + Segmentation	0.85	0.86	0.86
MCA MIP Classification + Segmentation (Tilt corrected)	0.88	0.91	0.90
ICA LVO Detection			
ICA Per Scan Classification	0.93	0.94	-

In the context of MCA LVO Detection, we applied average pooling and introduced a dropout rate of 0.3 at the classification head to enhance generalization and prevent overfitting during training. To optimize the model during the training process, we employed Stochastic Gradient Descent (SGD) with a momentum value set to 0.9.

In this section, we outline our iterative approach to enhance the MCA LVO detection model. The outcomes of each iteration, along with corresponding quantitative results, are summarized in Table 2.

Initially, we encountered subpar performance using ICV MIP as input for the binary classification model due to interference from other vasculature, as depicted in figure 4. To overcome this limitation, we trained the model on MCA VT MIPs, resulting in notable improvements in sensitivity (from 0.65 to 0.76) and specificity (from 0.67 to 0.77), as shown in Table 2. We also introduced pixel-wise ground truth for enhanced model supervision, leading to an additional 9% boost in sensitivity and specificity.

Furthermore, applying a preprocessing step to correct tilt contributed to an extra 4% enhancement in sensitivity and specificity. To validate our segmentation output throughout all iterations, we employed the Dice Similarity Coefficient (DSC).

Our iterative approach effectively improved the performance of the MCA LVO detection model. Some of the model outputs on the tilt-corrected MCA MIPs are shown in figure 11. Additionally, we expanded the test set by including 63 distal M2 LVO positive scans, from which our model accurately identified 40 cases while missing 23.

For the Vascular Territory Segmentation model, we achieved an overall DSC of 0.90 on 232 CTAs, while the Intracranial Volume Segmentation model achieved a DSC of 0.98 on 98 CTAs.

Table 3: Performance Evaluation based on Validation Set; SN: Sensitivity, SP: Specificity.

Valdation Set	**SN**	**SP**	**AUC**	**Threshold**
MCA LVO	0.9059	0.8892	0.9483	0.5649
ICA LVO	0.914	0.9024	0.9731	0.5464

Overall, our MCA LVO detection model achieved an accuracy of 89.8% with a sensitivity of 87.4% and specificity of 91.4%. The ICA LVO detection model achieved an accuracy of 92.6% with a sensitivity of 93.4% and specificity of 92.2%.

4. Conclusion

In conclusion, our proposed solution utilizes deep-learning models to automate the identification of large vessel occlusions (LVOs) in acute ischemic stroke (AIS)

patients from CTA scans. This approach reduces inter-observer variability and enhances stroke care efficiency.

To strengthen our models, we aim to validate and improve their generalizability by testing them on larger and more diverse datasets. This will refine their performance and reliability. Additionally, we plan to extend their capabilities to detect distal occlusions and differentiate and detect stenosis, further enhancing their usefulness in stroke diagnosis and treatment.

An exciting aspect of our research is the potential integration of these models into clinical decision support systems and assist radiologists and clinicians in diagnosing and treating stroke patients, ultimately leading to better patient outcomes and quality of life.

References

[1] Ujjwal Upadhyay, Jemin Webster, Swetha Tanamala, Sasank Chilamkurthy, Satish Golla, and Justy Antony Chiramal. Ai improves stroke diagnosis and care at a low resource hospital in india. *Journal of the Neurological Sciences*, 429, 2021.

[2] Valery L. Feigin, Carlene M M Lawes, Derrick A. Bennett, Suzanne Lyn Barker-Collo, and Varsha Parag. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *The Lancet Neurology*, 8:355–369, 2009.

[3] Gustavo Saposnik, Moira K. Kapral, Ying Yu Liu, Ruth E. Hall, Martin O'Donnell, Stavroula R Raptis, Jack V. Tu, Muhammad M Mamdani, and Peter C. Austin. Iscore: A risk score to predict death early after hospitalization for an acute ischemic stroke. *Circulation*, 123:739–749, 2011.

[4] Wei-Min Ho, Jr-Rung Lin, Hui-Hsuan Wang, Chia-Wei Liou, Ku-Chou Chang, Jiann-Der Lee, Tsung-Yi Peng, Jen-Tsung Yang, Yeu-Jhy Chang, Chien-Hung Chang, and Tsong-Hai Lee. Prediction of in-hospital stroke mortality in critical care unit. *SpringerPlus*, 5, 2016.

[5] Angelika Alonso, Anne Ebert, Rolf Kern, Simone Rapp, Michael Hennerici, and Marc Fatar. Outcome predictors of acute stroke patients in need of intensive care treatment. *Cerebrovascular Diseases*, 40:10 – 17, 2015.

[6] Alexander G. Chartrain, Christopher Paul Kellner, and J. Mocco. Pre-hospital detection of acute ischemic stroke secondary to emergent large vessel occlusion: lessons learned from electrocardiogram and acute myocardial infarction. *Journal of NeuroInterventional Surgery*, 10:549 – 553, 2018.

[7] Raul G. Nogueira and Ashutosh P. Jadhav et.al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *The New England Journal of Medicine*, 378:11–21, 2018.

[8] Mayank Goyal and Bijoy K. Menon et.al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *The Lancet*, 387:1723–1731, 2016.

[9] Gregory W. Albers and M.P. Marks et.al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *The New England Journal of Medicine*, 378:708–718, 2018.

[10] Bruce C.V. Campbell and Peter J. Mitchell et.al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *The New England journal of medicine*, 372 11:1009–18, 2015.

[11] Mayank Goyal and Andrew M. Demchuk et.al. Randomized assessment of rapid endovascular treatment of ischemic stroke. 372 11:1019–30, 2015.

[12] Tudor G. Jovin and Ángel Chamorro et.al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *The New England journal of medicine*, 372 24:2296–306, 2015.

[13] Jeffrey L. Saver and Mayank Goyal et.al. Stent-retriever thrombectomy after intravenous t-pa vs. t-pa alone in stroke. *The New England journal of medicine*, 372 24:2285–95, 2015.

[14] Aditya Khurana, Bhavik Patel, and Richard J. Sharpe. Geographic variations in growth of radiologists and medicare enrollees from 2012 to 2019. *Journal of the American College of Radiology : JACR*, 2022.

[15] Sawon Pratiher and Subhankar Chattoraj. Diving deep onto discriminative ensemble of histological hashing & class-specific manifold learning for multi-class breast carcinoma taxonomy. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1025–1029, 2018.

[16] Ujjwal Upadhyay, Mukul Ranjan, Satish Golla, Swetha Tanamala, Preetham Sreenivas, Sasank Chilamkurthy, J. Arun Pandian, and Jason W Tarpley. Deep-aspects: A segmentation-assisted model for stroke severity measurement. In *ECCV Workshops*, 2022.

[17] 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*, 392:2388–2396, 2018.

[18] U Upadhyay, S Golla, S Kumar, K Szweda, R Shahripour, and J Tarpley. O-064 ai based gaze deviation detection to aid lvo diagnosis in ncct. *Journal of NeuroInterventional Surgery*, 14(Suppl 1):A41–A42, 2022.

[19] Matthew T Stib, Justin Vasquez, Mary P Dong, Yun Ho Kim, Sumera S Subzwari, Harold J Triedman, Amy Wang, Hsin-Lei Charlene Wang, Anthony D Yao, Mahesh V. Jayaraman, Jerrold L. Boxerman, Carsten Eickhoff, Ugur Çetintemel, Grayson L. Baird, and Ryan A. McTaggart. Detecting large vessel occlusion at multiphase ct angiography by using a deep convolutional neural network. *Radiology*, page 200334, 2020.

[20] Clara Monteiro Antunes Barreira, Mehdi Bouslama, J. W. Lim, Alhamza R. Al-Bayati, Y. Saleem, Thomas G Devlin, Diogo C. Haussen, Michael T Froehler, Jonathan A. Grossberg, Blaise Baxter, Michael Frankel, and Raul G. Nogueira.

E-108aladin study: automated large artery occlusion detection in stroke imaging study – a multicenter analysis. *Electronic Poster Abstracts*, 2018.

[21] Shalini A. Amukotuwa, Matús Straka, Seena Dehkharghani, and Roland Bammer. Fast automatic detection of large vessel occlusions on ct angiography. *Stroke*, 2019.

[22] Shalini A Amukotuwa, Matus Straka, Heather Smith, Ronil V Chandra, Seena Dehkharghani, Nancy J Fischbein, and Roland Bammer. Automated detection of intracranial large vessel occlusions on computed tomography angiography: a single center experience. *Stroke*, 50(10):2790–2798, 2019.

[23] Sven PR Luijten, Lennard Wolff, Martijne HC Duvekot, Pieter-Jan van Doormaal, Walid Moudrous, Henk Kerkhoff, Geert J Lycklama a Nijeholt, Reinoud PH Bokkers, SF Lonneke, Jeannette Hofmeijer, et al. Diagnostic performance of an algorithm for automated large vessel occlusion detection on ct angiography. *Journal of neurointerventional surgery*, 14(8):794–798, 2022.

[24] Ivo GH Jansen, Maxim JHL Mulder, and Robert-Jan B Goldhoorn. Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (mr clean registry). *bmj*, 360, 2018.

[25] Esmee Venema, Martijne HC Duvekot, Hester F Lingsma, Anouk D Rozeman, Walid Moudrous, Frederique H Vermeij, Marileen Biekart, Aad van der Lugt, Henk Kerkhoff, Diederik WJ Dippel, et al. Prehospital triage of patients with suspected stroke symptoms (presto): protocol of a prospective observational study. *BMJ open*, 9(7):e028810, 2019.

[26] A Yahav-Dovrat, M Saban, G Merhav, I Lankri, E Abergel, A Eran, D Tanne, RG Nogueira, and R Sivan-Hoffmann. Evaluation of artificial intelligence–powered identification of large-vessel occlusions in a comprehensive stroke center. *American Journal of Neuroradiology*, 42(2):247–254, 2021.

[27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597, 2015.

[28] Antspy. `https://github.com/ANTsX/ANTsPy`. Accessed: 2022-01-26.

[29] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015.

[30] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[31] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[32] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.