Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 4, October 2026, pp: 1796-1818 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Segmentation of Acute Ischemic Stroke Infarct using Multi-scale U-Net Architecture on Non-contrast CT Scan Images

Osmalina Nur Rahma[1], Aisyah Widayani[2], Nimas Maharani Putri[1], and Rifai Chai[3]

[1] Biomedical Engineering, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia
[2] Radiology Imaging Technology, Department of Health, Faculty of Vocational Studies. Universitas Airlangga, Surabaya, Indonesia
[3] Department of Engineering Technologies, School of Science, Computing and Engineering Technologies, Swinburne University of Technology, Australia

Corresponding author: Osmalina Nur Rahma (e-mail: osmalina.n.rahma@fst.unair.ac.id), **Author(s) Email**: Aisyah Widayani (e-mail: aisyahwidayani@vokasi.unair.ac.id), Nimas Maharani Putri (e-mail: nimas.maharani.putri-2022@fst.unair.ac.id), and Rifai Chai (e-mail: rchai@swin.edu.au)

Abstract **Acute ischemic stroke must be treated within a few hours, and the first scan is usually a non-contrast CT (NCCT), on which an early infarction is often hard to see. Automatic segmentation assist image interpretation, and multi-scale convolution is a commonly proposed approach. This study measured how much such a block improves U-Net performance for this task, and what it costs. Every convolutional block of U-Net, in the encoder and the decoder, was replaced by a block with three parallel branches using 1×1, 3×3 and 5×5 kernels. Both networks were trained on the public Acute Ischemic Stroke Dataset, on 375 studies with supratentorial infarcts. The data were divided by study; the scans of the dataset's own test split were kept for testing, and augmentation was applied to the training subset alone. Four configurations were compared: each architecture with and without augmentation. On the 51 test patients, Multi-scale U-Net did not improve overlap. Its Dice score was 0.6372 against 0.6632 for U-Net with augmentation, and the paired difference per patient was 0.0233 in favour of U-Net (95 % CI −0.060 to 0.013, p = 0.19). The multi-scale block instead altered the balance between the two types error: sensitivity increased by 0.060 (p = 0.002), and precision decreased by 0.100 (p < 0.001). Augmentation improved both networks (+0.106 and +0.048 Dice per patient, p < 0.001), mainly by cutting false-positive lesions from 15.5 to 5.6 per patient. The multi-scale block costs 2.31 times as many parameters and 3.37 times as much inference time. An ablation over the target definition favoured the dataset authors' labels: adding the label for infarct invisible on NCCT cost 0.040 Dice and tripled false-positive lesions. Both networks segment small infarcts poorly, with Dice below 0.16 under 100 pixels, which remains the main obstacle.**

Keywords **Acute Ischemic Stroke, Non-contrast CT-scan, Augmentation, U-Net, Multi-scale Convolution.**

I. Introduction

Stroke is the second leading cause of death worldwide [1], [2], and its incidence, mortality and disability burden all rose between 1990 and 2021 [3], [4]. In Indonesia, the recorded prevalence was 10.9 per 1,000 population in 2018 and 8.3 per 1,000 in 2023 [5], [6]; the decline may reflect improved primary prevention as much as a change in case detection, and 67 per cent of the national total is recorded as acute ischemic stroke (AIS) [7]. Because interrupted cerebral blood flow destroys neurons within minutes, treatment must begin within a narrow therapeutic window when thrombolysis is most effective [8], [9]. Assessment begins at the bedside with the National Institutes of Health Stroke Scale, a check of vital signs and a short laboratory panel to exclude mimics such as hypoglycaemia [10]; symptoms vary widely between patients, however, and mild or atypical deficits delay recognition and introduce disagreement among clinicians [11].

Imaging then begins with non-contrast CT (NCCT), which is fast and excludes haemorrhage before thrombolysis, although it is less sensitive to early ischemia than diffusion-weighted MRI and carries none of the contrast-related risk of CT perfusion [12], [13]. These properties keep NCCT as the routine first-line modality for AIS [14], [15]. During the first hours after onset, however, the scan often looks unremarkable, because the density shift produced by a fresh infarct is too slight to be seen reliably [16], which leaves interpretation dependent on the reader's experience and open to disagreement between observers [17]. In Indonesia, the difficulty is compounded by the distribution of neurological expertise: most neuro-intervention facilities and specialists are concentrated on Java, and the neurologist-to-population ratio is approximately 1:108,000 [7]. Automated support for reading subtle NCCT findings within the therapeutic window is therefore of practical value [18].

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Deep learning (DL) has emerged as a practical response to this gap, sharpening what can be read from NCCT and flagging infarct signs that a human reader may pass over [19], [20], [21]. Classical learners such as SVM, Random Forest, and KNN handle the complex spatial structure of medical images poorly [22]. Convolutional Neural Networks (CNNs) displaced them, and within that family the U-Net has become the standard choice for segmentation, because its skip connections carry forward the spatial detail that pooling discards and so allow boundaries to be traced accurately even when labelled examples are scarce [23], [24], [25].

However, U-Net's performance deteriorates when lesion size varies substantially, a limitation that is particularly relevant for AIS infarcts, which can range from small lacunar lesions to large territorial infarcts. Several distinct families of remedies have been proposed, and they should be distinguished carefully because they enlarge or exploit the receptive field in different ways. Dilated (atrous) convolution widens the receptive field by inserting gaps between sampled positions, keeping the parameter count of a single kernel but sampling the input sparsely, which can produce gridding artefacts on small low-contrast targets [26]. Attention mechanisms such as Attention U-Net leave the receptive field unchanged and instead reweight skip-connection features so that irrelevant background is suppressed [27]. Nested and densely connected variants such as U-Net++ rework the skip pathway itself to reduce the semantic gap between encoder and decoder [28]. Symmetry-aware designs exploit an anatomical prior specific to the brain, comparing a slice with its mirror image across the midline to expose subtle unilateral hypoattenuation [29]. Transformer-based segmentation networks, including TransUNet and Swin-UNet, replace or supplement convolution with self-attention to model long-range dependencies, at the cost of substantially larger data and compute requirements [30], [31].

Parallel multi-kernel convolution, the approach adopted here, differs from all of these: within a single block it applies kernels of several sizes densely to the same input and concatenates the responses, so that features at several receptive-field sizes are available to the next layer simultaneously. This design descends from the Inception block [32] and has been transferred to segmentation in histopathology [33] and in ultrasound and dermoscopy imaging [34], [35]. Among these, the variants most closely related to the present work are MSU-Net [34], MUNet [35] and the multi-scale U-Net of Li et al. [36]. MSU-Net fuses convolutions of different kernel sizes within a residual block; MUNet builds a multi-scale pyramid by operating on several downsampled copies of the input; Li et al. combine multi-scale processing with dilated convolution and evaluate it on acute ischemic stroke NCCT from 30 patients. The architecture examined here differs in three specific respects, and this study claims no more than that: (i) every convolutional block of both encoder and decoder is replaced, so multi-scale filtering is applied at all five resolution levels rather than at selected stages; (ii) the wider branches are preceded by a 1×1 bottleneck that fixes the channel count entering the 3×3 and 5×5 convolutions, which is what keeps the 5×5 branch affordable; and (iii) the block uses dense kernels only, with no dilation, so the receptive field is enlarged without sparse sampling of the input. The contribution is therefore not the invention of multi-scale convolution, which is well established, but a controlled comparison of this particular dense parallel-kernel configuration against a matched U-Net baseline, with the effect of the architecture separated experimentally from the effect of augmentation.

Three things are consequently missing from the literature on this problem. Dense parallel multi-kernel convolution has been evaluated on acute ischemic stroke NCCT in a single small single-centre collection [36], so its behaviour on a public dataset of several hundred patients is not established. In that evaluation, the parallel kernels were combined with dilated convolution and with data augmentation, so the contribution of the parallel kernels themselves was not isolated from the rest. The computational cost of the additional branches is also rarely reported alongside the accuracy, although it is what decides whether the design is usable in the setting that motivates it, namely a service with few readers and a short time window. This study addresses the three points together.

The objective of this study is therefore to determine how much of the segmentation performance attainable on acute ischemic stroke NCCT is attributable to dense parallel multi-kernel convolution, measured against a matched U-Net baseline under identical conditions, and at what computational cost. Every standard convolutional block of U-Net, in the encoder and the decoder alike, is replaced by a three-branch block carrying 1×1, 3×3 and 5×5 kernels without dilation, and the two networks are trained and evaluated on the publicly available Acute Ischemic Stroke Dataset (AISD), a secondary source providing a substantially larger and independently collected sample. Segmentation is confined to infarcts of the supratentorial territory, meaning the cerebral hemispheres perfused by the anterior (ACA), middle (MCA) and posterior cerebral arteries (PCA), rather than the infratentorial compartment formed by the brainstem and cerebellum, whose appearance on NCCT differs and whose proximity to bone introduces additional artefact.

Three contributions follow from this work: 1) the adaptation of a Multi-scale U-Net architecture using dense parallel multi-kernel convolution blocks, without dilation, for automatic AIS infarct segmentation on

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

NCCT, an application that remains relatively underexplored compared with its use in other medical imaging modalities; 2) a 2×2 comparison between U-Net and the proposed Multi-scale U-Net under both non-augmented and augmented data conditions, which separates the contribution of the architecture from that of augmentation and reports the computational cost of each architecture exactly; and 3) an explicit account of what the resulting numbers do and do not establish, including the partitioning limitation set out in Section II.B, which determines how the absolute values should be read and what must be done before they can be described as evidence of generalization.

Section II sets out the dataset and the proposed Multi-scale U-Net. Section III reports the experiments and the comparison against the baseline. Section IV weighs the findings together with the limitations of the work, and Section V closes the paper.

II. Method

The work proceeded in three stages: assembling the NCCT data, preparing the images, and then training and testing the Multi-scale U-Net on them. The intervention studied here is deliberately narrow, in that the standard convolutional block of U-Net is exchanged for a parallel multi-scale block and the two architectures are then compared directly. Preprocessing steps, including resizing, augmentation, and normalization, were applied prior to training, and U-Net and the proposed Multi-scale U-Net were each evaluated under both non-augmented and augmented data conditions to assess their relative segmentation performance. It is hypothesized that the incorporation of parallel multi-scale convolution improves the model's sensitivity to infarct lesions of varying sizes and shapes, especially the small or irregularly shaped ones that U-Net tends to miss, and that augmentation compensates further for the modest size of the available dataset. The full procedure is traced in Fig. 1.

A. Data Collection

The images used here are secondary data drawn from the Acute Ischemic Stroke Dataset (AISD), released publicly through a GitHub repository maintained by Chang Gung University [29]. It holds 397 non-contrast computed tomography (NCCT) studies, each acquired at Linkou Chang Gung Memorial Hospital in Taiwan from a patient presenting with acute ischemic stroke, and each recorded no more than 24 hours after symptoms began. An experienced physician traced the infarct on every scan by hand, taking the corresponding diffusion-weighted magnetic resonance imaging (DWI-MRI) study as the reference; the labels were then double-reviewed by a senior doctor, and the slice thickness of the NCCT acquisitions was 5 mm. The annotations were projected onto the corresponding NCCT images using ImageJ software. The dataset is distributed for non-commercial academic research through the repository of Liang et al. [29]; it was used under those terms, and the permanent citation for the dataset is reference [29]. All statements regarding ethical approval and participant consent refer to the original collection as documented by those authors and are reproduced in the Declarations section without independent verification by the present authors.

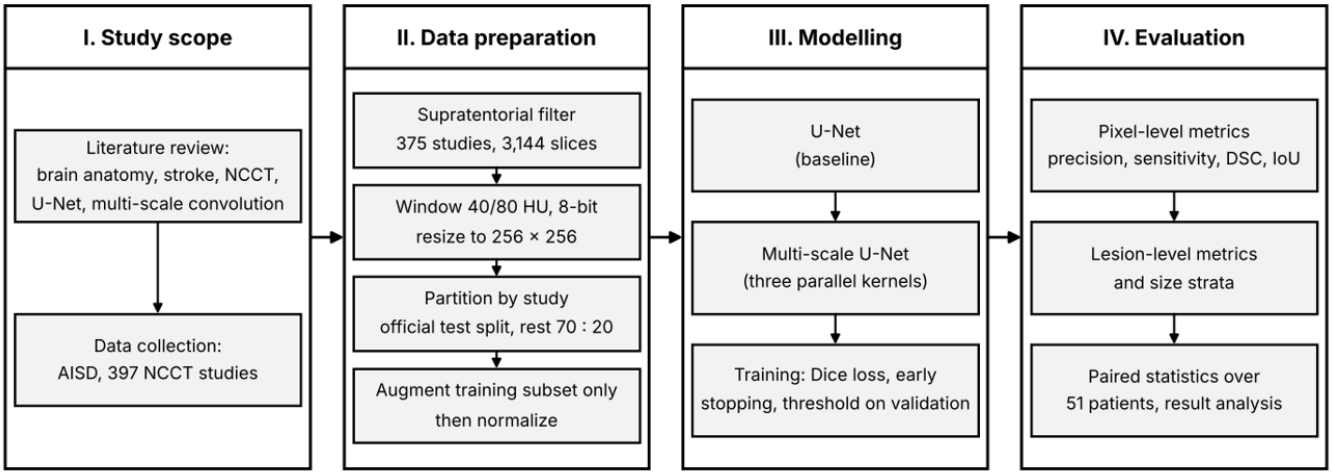


Fig. 1. Overall workflow of the proposed acute ischemic stroke infarct segmentation framework, including NCCT data acquisition, preprocessing, patient-level partitioning, model training, and performance evaluation. The four labelled stages of the diagram are study scope, data preparation, modelling and evaluation.

Only scans whose infarct lay in the supratentorial region were retained, as shown in Fig. 2, covering the ACA, MCA, and PCA territories. The supratentorial region was selected because it contains a larger proportion of brain parenchyma and is subject to less

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

interference from skull bone structures than the infratentorial region. Subsequently, each NCCT image was matched with its corresponding segmentation mask to exclude images without ground truth annotations. The final dataset comprised 375 patient folders, containing paired 3,144 NCCT images and corresponding masks, with an image resolution of 512 × 512 pixels.

The composition of the subset used here should be stated precisely, because the released dataset and the subset differ. The annotation folder in the release contains 3,757 masks across 384 studies, while the supratentorial image subset assembled for this work contains 3,150 slices across 375 studies; this difference reflects the restriction to supratentorial infarcts described above. Pairing each image with its mask and discarding images without a corresponding annotation removed six more slices, leaving 3,144 paired image-mask slices from 375 studies, which is the set partitioned in Section II.C.

The segmentation target definition requires its own statement because the AISD masks are not binary. The released annotations carry five labels: remote infarct (1), clear acute infarct (2), blurred acute infarct (3), invisible acute infarct (4) and infarct (5). The dataset authors specify that the infarct lesion consists of labels 1, 2, 3 and 5, while label 4 marks acute infarct that the annotator identified on the reference DWI-MRI but judged invisible on the NCCT itself. The target used here follows that specification: the masks were binarised over labels 1, 2, 3, and 5, and label 4 was excluded, so the networks are not asked to reproduce annotated area that carries no visible signal on the input image. Across the 3,144 paired slices, the annotated area is divided as follows: clear acute infarct 51.6 per cent, blurred acute infarct 27.1 per cent, invisible acute infarct 9.1 per cent, remote infarct 7.7 per cent, and infarct 4.6 per cent, with label 4 appearing in 436 of the 3,144 slices. One property of the annotation scheme should also be stated: because label 1 denotes remote rather than acute infarct, the target includes established infarcts alongside acute ones, so the model localises infarcted tissue without distinguishing acute from chronic hypoattenuation; remote infarct accounts for 7.7 per cent of the annotated area and appears in 688 of the 3,144 slices. An ablation over target definitions, comparing this target with the addition of label 4 and with a restriction to the two acute labels, is reported in Section III.D and Table 6.

The AISD studies are distributed as NCCT volumes in Hounsfield Units (HU) together with the corresponding masks. Before the experiments reported here, each axial slice was converted to an 8-bit grayscale image (512×512, uint8) using a fixed narrow brain window applied identically to every slice, with densities outside the window clipped. The intensity distribution of the converted images is consistent with that procedure and with a window close to the conventional brain setting of level 40 HU and width 80 HU: across a random sample of the converted slices the modal parenchymal intensity is stable at 117 (interquartile range 112–119), which under that mapping corresponds to roughly 37 HU, while 8.6 per cent of pixels are saturated at 255, as expected for bone, and 59.5 per cent sit at 0, corresponding to air outside the head. The stability of the parenchymal mode across slices excludes per-slice minimum-maximum rescaling, under which that mode would vary from slice to slice. The exact window parameters cannot be recovered from 8-bit data alone and are reported here as the setting used during conversion. A narrow brain window is the one used routinely for visual assessment of early ischemic change, and it deliberately allocates the full 8-bit range to the narrow density interval that separates infarcted from normal parenchyma, so the small hypoattenuation of an acute infarct, typically a few HU, is preserved and in fact amplified relative to a wider window. The conversion is nevertheless lossy in two respects that should be stated plainly. Densities outside the window, including bone, calcification and haemorrhage, are saturated and cannot be recovered; and the quantitative HU scale is not available to the network, so absolute attenuation cannot be used as a feature. The normalization

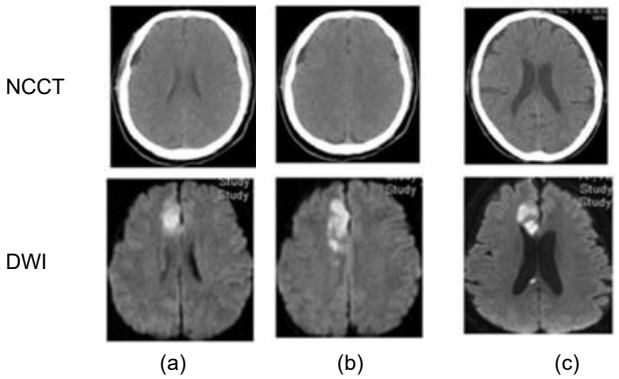


Fig. 2. Representative non-contrast CT images from the Acute Ischemic Stroke Dataset (AISD), illustrating supratentorial infarct regions [29]. Columns (i), (ii) and (iii) are three different studies; the upper row is the NCCT slice used as the network input and the lower row the corresponding DWI-MRI of the same study. The hyperintense region on each DWI image is the infarct from which the reference annotation was traced, here in the middle cerebral artery territory. No colour overlay or marking is applied to either row.

described in Eq. (4), division by 255, therefore acts on windowed 8-bit intensities rather than on raw HU, and is a rescaling step rather than an intensity standardisation. Training directly on windowed HU

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

data, or supplying several windows as separate input channels, is a plausible extension and is noted among the limitations.

The conversion is a windowing operation followed by an 8-bit quantisation, written out in Eq. (1) [14],

$$I_8(x,y) = \text{round}\left(255 \cdot \text{clip}\left(\frac{H(x,y) - L + W/2}{W}, 0, 1\right)\right) \quad (1)$$

where $I_8(x,y)$ is the 8-bit intensity written to the PNG file, $H(x,y)$ the Hounsfield value of that pixel, L the window level (40 HU) and W the window width (80 HU); clip (·, 0, 1) bounds its argument to the unit interval, so densities below $L - W/2$ saturate at 0 and densities above $L + W/2$saturate at 255, and round (·) is rounding to the nearest integer.

B. Data Preprocessing

Three operations were applied in sequence: resizing, augmentation and normalization, after which the resulting set of images was partitioned. The order of these steps is stated explicitly here and discussed in Section II.C, because it governs how the reported values should be interpreted. Every image was first brought to 256×256 so its dimensions matched what the U-Net expects as input. Fixing the geometry in this way lets training and testing proceed on tensors of identical shape. Augmentation was then carried out with the aim of increasing data variation, using a mild affine transformation applied twice with independently drawn random parameters, so each original image produced two new, different images. The affine transformation used is shown in Eq. (2) [37],

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + b \quad (2)$$

where x, y are the pixel coordinates in the original image, x', y' are the pixel coordinates in the transformed image, and *A* is the linear transformation matrix (rotation, scaling, shearing). This can be further expressed in homogeneous form, as shown in Eq. (3) [37],

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

where *b* is the translation vector, ***a*** represents the linear transformation coefficients, and t_x, t_y denote the horizontal and vertical translation. Rotations were drawn at random from −8° to +8°, scaling from 0.97 to 1.03×, and translation along both axes from ±5% of the frame, which is −12.8 to +12.8 pixels at 256×256. These transformations were applied at a small scale to preserve the anatomical structure of the images and avoid excessive distortion [37], [38].

The dataset used in the augmented condition was the union of the original images and all augmented images, three times the size of the original set, giving 9,432 images with their corresponding masks. Intensities were then rescaled onto the interval 0 to 1, each pixel simply divided by 255, following Eq. (4),

$$x_n = \frac{x_i}{255} \quad (4)$$

where x_n is the normalized pixel value and x_i is the original pixel intensity value, ranging from 0 to 255. This stage aims to reduce intensity variation between images and accelerate convergence during the model training process. By performing normalization, each pixel is brought to a comparable scale, so that lesion analysis and segmentation can be carried out in a more stable and consistent manner [39].

With preprocessing complete, the images were partitioned by study. The 52 scans named in the predefined AISD test split were reserved for testing, of which 51 survived the supratentorial filter and contributed 412 slices. The remaining studies were divided 70:20 into training and validation with a fixed seed, giving 252 studies and 2,080 slices for training and 72 studies and 652 slices for validation. No study contributes slices to more than one subset, and the disjointness is asserted in code before training begins. Augmentation was then applied to the training subset alone, producing two transformed copies of each

Algorithm 1. NCCT Image Preprocessing and Patient-Level Dataset Partitioning

```
Input : NCCT volumes V_p and expert annotation masks M_p of the P = 397 AISD studies, p = 1..P
Output: training set T, validation set V and test set S of (image, label) pairs, disjoint by patient
 1: for each study p do
 2:    discard p if its infarct lies outside the supratentorial region
 3:    for each axial slice (x, m) of study p do
 4:       I <- window(x) at level 40 HU and width 80 HU, then quantise to 8 bits       (Eq. (1))
 5:       I <- resize(I, 256 x 256) bilinear;  m <- resize(m, 256 x 256) nearest neighbour
 6:       y <- 1 where m in {1, 2, 3, 5}, else 0     // lesion-label conversion, label 4 excluded
 7:       append (I, y) to the slice list of study p
 8: S <- the slices of the 52 studies named in the official AISD test split; 51 survive step 2
 9: split the remaining studies 70:20 by study, never by slice                    -> T, V
10: assert that T, V and S share no study            // patient-level partition, checked in code
11: for each (I, y) in T do      // augmentation is applied only here, after the partition,
12:    repeat twice:             // so that no transformed copy of a test slice can reach training
13:       draw a random affine A (rotation, scaling, shear, translation)      (Eq. (2), Eq. (3))
14:       append (A(I), A(y)) to T
15: divide every image in T, V and S by 255                                        (Eq. (4))
16: return T, V, S
```

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

training slice and 6,240 training images in the augmented condition; the validation and test subsets are never augmented. Both architectures and both data conditions use exactly this partition, so all four models are trained and evaluated on identical subsets. The complete preprocessing and partitioning procedure is stated as Algorithm 1, so that it can be reproduced without reference to the prose above.

C. Why the partition is made by study

The unit of partition is the study rather than the slice, and augmentation follows the partition rather than preceding it. Both choices are deliberate, and the reason is worth stating because it governs what the reported values measure. Adjacent axial slices of one study share the scanner, the reconstruction kernel, the acquisition protocol and the patient anatomy, and neighbouring positions show almost the same infarct cross-section. If slices were allocated individually, a test slice would almost always have close relatives in the training subset, and the resulting figures would describe interpolation between slices of a known patient rather than generalization to an unknown one. This is a well-documented source of optimistic bias in medical imaging studies [40]. Augmenting before partitioning has the same effect by a second route, because the transformations used here are mild, within ±8° of rotation, 0.97 to 1.03 in scale and ±5 per cent in translation, so a transformed copy of a test image is visually close to the original.

Using the test split distributed with the dataset adds a second property that a partition chosen by the authors of a study cannot have: the test studies are fixed by the authors of the dataset, and they are the same studies on which other work on AISD reports [29]. The values in Section III are therefore estimates of performance on patients the model has not seen, and they can be placed alongside published values with the qualifications set out in Section IV. The comparison between the two architectures is made under matched conditions in either case, since both are trained and evaluated on identical subsets, but only a partition by study licenses the absolute figures.

D. U-Net

U-Net was designed for biomedical segmentation and takes the form of a symmetric encoder-decoder joined by skip connections. Along the contracting path, the encoder builds up spatial features level by level; along the expansive path, the decoder restores resolution until the output mask matches the input pixel for pixel. Each encoder level is tied to its decoder counterpart, so semantic content from the deep layers arrives together with the fine spatial detail held in the shallow ones. That pairing is what keeps boundaries sharp, especially when working with limited medical imaging datasets [24], [41], [42].

Both networks are assembled from four primitive operations, which are stated here so that the blocks described below can be written compactly. A two-dimensional convolution of an input feature map F with a kernel K is given in Eq. (5) [42],

$$(F * K)(i,j) = \sum_m \sum_n F(i+m, j+n) \cdot K(m,n) + b \quad (5)$$

where b is the bias, (i,j) indexes the output position, and m and n index the kernel support, which is 1×1, 3×3 or 5×5 in this work; one such convolution is applied per output channel, and the results are summed over the input channels. Every convolution except the final one is followed by batch normalization, Eq. (6) [39], which standardises the pre-activation values over the mini-batch and is what allows training to proceed stably at the small learning rate used here [39],

$$\mathrm{BN}(z) = \gamma \cdot \frac{z - \mu_B}{\sqrt{{\sigma_B}^2 + \varepsilon}} + \beta \quad (6)$$

where z is a pre-activation value, μ_B and ${\sigma_B}^2$ the mean and variance of z over the mini-batch, ε a small constant that prevents division by zero, and γ and β the learned scale and shift.

The activation is the rectified linear unit of Eq. (7) [22],

$$\mathrm{ReLU}(z) = \max(0, z) \quad (7)$$

which introduces the non-linearity that lets successive convolutions represent more than a single affine map, and which leaves positive responses unchanged so that the weak hypoattenuation of an infarct is not compressed. Downsampling between encoder blocks is 2×2 max pooling, Eq. (8) [22],

$$P(i,j) = \max_{0 \le m,n < 2} F(2i+m, 2j+n) \quad (8)$$

where P is the pooled map and the maximum is taken over each non-overlapping 2×2 window, so both spatial dimensions are halved while the channel count is unchanged.

The encoder built here runs five blocks deep, with filter counts rising 32, 64, 128, 256, 512. Within a block, two 3×3 convolutions are applied one after the other, each trailed by batch normalization and a ReLU, as set out in Eq. (9) [42],

$$F_{out} = \sum \left(BN\left(W_2 * \sum \left(BN(W_1 * F_{in} + b_1)\right) + b_2\right)\right) \quad (9)$$

with F_{in} and F_{out} denoting the input and output feature maps, W is a 3×3 convolution, BN batch normalization, and *ReLU* the activation function. Between blocks, a 2×2 max pooling step halves the feature maps. The decoder reverses that path with four 2×2 transposed convolutions, concatenating the matching encoder map after each one through the skip connection, as set out in Eq. (10) [42],

$$D_i = f\left(\left[U\left(D_{(i-1)}\right) \| E_i\right]\right) \quad (10)$$

with D_i and E_i denoting the decoder and corresponding encoder feature maps, UpConv the transposed convolution, and Concat the channel-wise concatenation. A closing 1×1 convolution turns the last decoder output into the segmentation mask. That convolution returns a single channel, and a sigmoid

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

maps each of its outputs to a probability, as shown in Eq. (11) [22],

$$\hat{Y}(i,j) = \frac{1}{1+e^{-z(i,j)}} \quad (11)$$

where $z(i,j)$ is the value produced by the final 1×1 convolution at pixel (i,j), and $\hat{Y}(i,j) \in (0,1)$ is the predicted probability that the pixel belongs to an infarct.

E. Multi-scale Convolution

Multi-scale convolution has gained ground as a way of letting a network read spatial structure at several resolutions at once. In medical image segmentation, anatomical structures often vary in size, shape, and thickness, so a single convolutional kernel size is often insufficient to capture both local detail and global context [23]. This multi-scale concept was originally introduced in the Inception architecture, which employed several kernel sizes in parallel within a single block to produce richer feature representations. The use of 1×1 convolution as a bottleneck has also proven effective in reducing the number of channels, thereby enabling multi-scale operations to be performed with greater computational efficiency [36].

The block used here carries three parallel paths. The first applies a single 1×1 convolution to the input. The second applies a 1×1 convolution followed by a 3×3 convolution, and the third a 1×1 convolution followed by a 5×5 convolution; in both cases, the leading 1×1 convolution acts as a bottleneck that fixes the channel count entering the wider kernel, which is what keeps the 5×5 branch affordable. Every convolution in the three branches is followed by batch normalization and a ReLU activation. The block appears on both the encoder and decoder sides, as illustrated in Fig. 3. After the three branches process the input independently, their outputs, each carrying C channels, are concatenated into a 3C-channel representation. This combination is richer than a single convolution because it captures information from three receptive-field sizes at once: per-pixel intensity from the 1×1 path, local edge and contour structure from the 3×3 path, and broader regional context from the 5×5 path. The concatenated representation is then reduced back to C channels by a final 1×1 convolution, applied without batch normalization or activation, so that the block presents the same interface as the standard U-Net block it replaces [43], [32]. Padding is 1 for the 3×3 kernel and 2 for the 5×5 kernel, so every branch preserves the spatial dimensions of its input. The block can be written out as follows. Let X be its input, C the number of output channels of each branch, and B_1, B_2 and B_3 the three branch outputs. The first branch is a single 1×1 convolution, Eq. (12) [32],

$$B_1 = \mathrm{ReLU}\big(\mathrm{BN}(W_{1\times1} * X)\big) \quad (12)$$

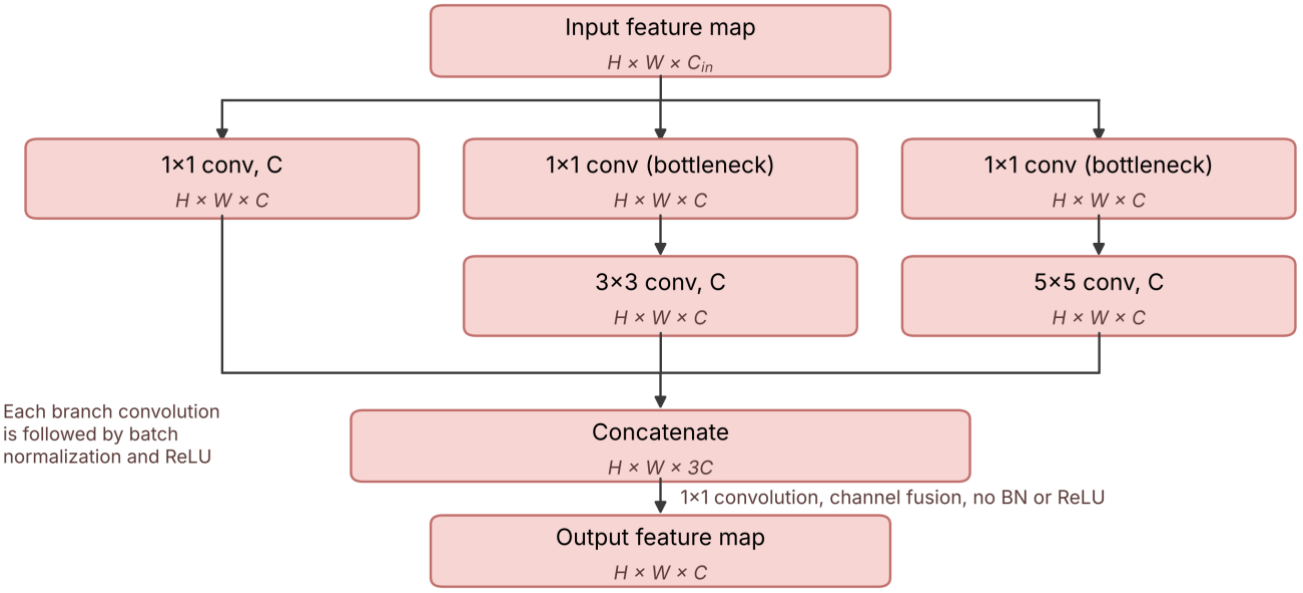


Fig. 3. Architecture of the proposed multi-scale convolution block incorporating parallel 1×1, 3×3, and 5×5 convolutional kernels, feature concatenation, and channel fusion. The input feature map carries C_{in} channels at H×W.

where $W_{1\times1}$ is a 1×1 kernel; this branch mixes channels at a fixed spatial position and therefore preserves per-pixel intensity information.The second and third branches place a wider kernel after a 1×1 bottleneck, Eq. (13) [32],

$$H_k = \mathrm{ReLU}\left(\mathrm{BN}\left({W_{1\times1}}^{(k)} * X\right)\right)$$
$$B_k = \mathrm{ReLU}\big(\mathrm{BN}(W_{k\times k} * H_k)\big), \quad k \in \{3,5\} \quad (13)$$

where H_k is the output of the bottleneck, the leading 1×1 convolution fixes the channel count entering the wider kernel, and $k \times k$ is 3×3 for B_2 and 5×5 for B_3; padding is 1 and 2, respectively, so every branch returns a map of the same spatial size as X. The three

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

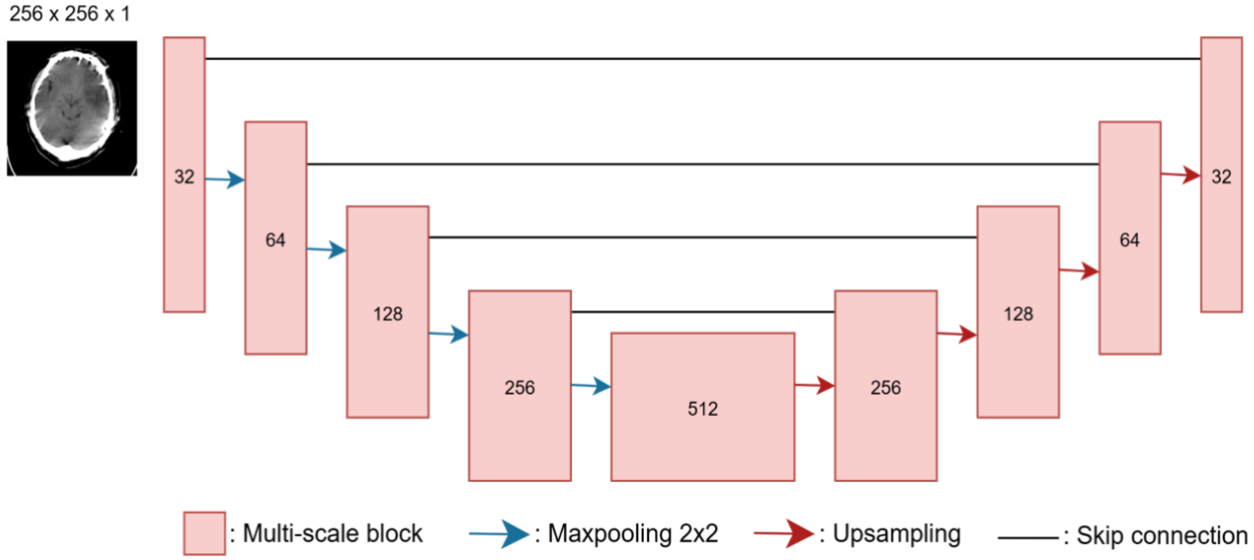


Fig. 4. Detailed architecture of the proposed Multi-scale U-Net for acute ischemic stroke infarct segmentation, showing encoder–decoder stages, skip connections, multi-scale convolution blocks, and output segmentation masks. The input is one 256×256×1 NCCT slice. The five encoder stages carry 32, 64, 128, 256 and 512 channels, each a multi-scale block of Fig. 3 followed by 2×2 max pooling, and the decoder mirrors them with transposed convolutions and skip connections. A final 1×1 convolution with a sigmoid returns a 256×256×1 probability map, which the threshold of Eq. (19) turns into the output segmentation mask.

branch outputs are then concatenated along the channel axis, Eq. (14) [32],

$$Z = [B_1 \oplus B_2 \oplus B_3] \in \mathbb{R}^{H\times W\times 3C} \quad (14)$$

where $\oplus$ denotes concatenation, so the three branches, each carrying C channels, give a 3C-channel representation that holds information from three receptive-field sizes at once and that representation is reduced back to C channels by a single 1×1 convolution, Eq. (15) [43],

$$Y = W_m * Z \in \mathbb{R}^{H\times W\times C} \quad (15)$$

where W_m is a 1×1 kernel applied without batch normalization or activation. The block output Y therefore has the same shape as the input of the standard U-Net block it replaces, which is what allows the substitution to leave the rest of the network untouched.

F. Multi-scale U-Net

The two networks differ in one respect only, namely how many kernel sizes each convolutional block carries. Swapping the standard block for the multi-scale module shown in Fig. 4 leaves everything else untouched: five encoder-decoder levels, max pooling on the way down, transposed convolution on the way up. Because the kernels differ in size, padding was set so that feature maps keep their spatial dimensions through the block. The cost of that substitution can be stated in closed form. For a block with C_{in} input channels and C output channels, the multi-scale module carries the parameter count of Eq. (16) [24],

$$N_{\mathrm{MS}} = 37C^2 + 3C \cdot C_{\mathrm{in}} + 16C \quad (16)$$

where the first term collects the 3×3, 5×5 and merging kernels $(9C^2 + 25C^2 + 3C^2)$, the second the three 1×1 convolutions that read the block input, and the third the six biases together with the five batch-normalization scale-and-shift pairs. Summed over the nine multi-scale blocks, the four transposed convolutions and the 1×1 output layer, this reproduces the 17,905,249 parameters of Table 1 exactly, against 7,765,409 when the standard block is used instead, whose count is $9C^2 + 9C \cdot C_{\mathrm{in}} + 6C$ and the multiply-accumulate count of Eq. (17) [24],

$$MACs_{\mathrm{MS}} = H \cdot W \cdot C \cdot (3C_{\mathrm{in}} + 37C) \quad (17)$$

where H and W are the spatial dimensions at that stage. Summed over the network, this gives 24.91 GMACs per 256×256 slice; the corresponding expression for the standard block, $H \cdot W \cdot C \cdot (9C_{\mathrm{in}} + 9C)$, gives 12.03 GMACs. The multi-scale network therefore carries 2.31 times the parameters of U-Net and costs 2.07 times as many operations, which is the price of the three-branch block.

G. Training Configuration

Both models were implemented in PyTorch and trained on a single GPU, using NVIDIA L4 on Google Colab Pro and NVIDIA T4 on Kaggle Notebooks. Optimization used Adam with a learning rate of 1×10^{-5}, selected after a preliminary comparison with 1×10^{-6} and 1×10^{-4} in which the smaller value had not converged within the epoch budget and the larger produced an oscillating

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 1. Layer-wise architectural specifications of the proposed Multi-scale U-Net with an input image size of 256×256×1. MS denotes the three-branch multi-scale block of Fig. 3, Params the number of trainable parameters of the stage and MACs the multiply-accumulate operations per slice in giga-operations. Channel dimensions are given as input → output. The stage values sum to the 17,905,249 trainable parameters and 24.906 GMACs of the complete network reported in Table 7.

Stage	Operation	Channels	Output size	Params	MACs (G)
Enc1	MS block	1 → 32	256×256	38,496	2.489
Enc2	MaxPool 2×2 + MS block	32 → 64	128×128	158,720	2.584
Enc3	MaxPool 2×2 + MS block	64 → 128	64×64	632,832	2.584
Enc4	MaxPool 2×2 + MS block	128 → 256	32×32	2,527,232	2.584
Enc5	MaxPool 2×2 + MS block	256 → 512	16×16	10,100,736	2.584
Dec4	ConvTranspose + concat + MS	512 → 256	32×32	3,346,688	3.020
Dec3	ConvTranspose + concat + MS	256 → 128	64×64	837,760	3.020
Dec2	ConvTranspose + concat + MS	128 → 64	128×128	209,984	3.020
Dec1	ConvTranspose + concat + MS	64 → 32	256×256	52,768	3.020
Output	Conv 1×1 + sigmoid	32 → 1	256×256	33	0.002
Total				17,905,249	24.906

validation curve. The Dice loss guided training, computed over the flattened pixels of each mini-batch rather than per image, with a smoothing constant of 1. The batch size was 10 and training ran for a fixed 100 epochs with no early stopping and no learning-rate schedule; the reported model is the state of the network at the final epoch, and probability maps were binarised at a fixed threshold of 0.5. Weights were left at the PyTorch defaults, which is Kaiming uniform initialisation for convolutional layers. Each of the four configurations was trained once, with the seed fixed only for the partition; repeated runs with different initialisations were not performed, and the consequences of this are stated in Section IV. Evaluation metrics were accumulated over all pixels of the test subset and computed from those totals, so a large well-segmented lesion contributes more to the reported figures than a small one. Written out over the flattened pixels of a mini-batch, that loss is Eq. (18) [44],

$$L_{\mathrm{Dice}} = 1 - \frac{2\sum_i p_i g_i + s}{\sum_i p_i + \sum_i g_i + s} \quad (18)$$

where $p_i \in (0,1)$ is the predicted probability of pixel i, $g_i \in \{0,1\}$ its ground-truth label, the sums run over every pixel of the mini-batch, and $s = 1$ is a smoothing constant that keeps the expression defined when prediction and ground truth are both empty. The probability map is turned into a binary mask by a fixed threshold, Eq. (19),

$$\hat{\mathrm{Y}}_{\mathrm{bin}}(i,j) = \begin{cases} 1, & \hat{\mathrm{Y}}(i,j) > \tau \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

where τ is the decision threshold and $\hat{\mathrm{Y}}_{\mathrm{bin}}$ the binary mask from which every metric reported below is computed. The threshold was not tuned on the validation subset, which is noted among the limitations.

The Dice loss was chosen because it optimises overlap directly and is robust to the severe foreground–background imbalance of this task, in which infarct pixels form a small fraction of each slice [44]. It should be noted that a pure Dice objective weights false positives and false negatives equally and tends toward conservative predictions when the foreground is small, consistent with sensitivity being the weakest of the four metrics reported here. Objectives that penalise false negatives more heavily, such as the Tversky loss with α < β [45] or its focal variant [46], and compound objectives such as Dice combined with binary cross-entropy, are therefore natural candidates for this task. An ablation over these objectives was not carried out in the present study, so the extent to which the reported performance depends on the choice of objective rather than on the architecture cannot be quantified here; this is stated as a limitation and as planned work in Section IV. The training procedure, including the schedule, the stopping rule, the selection of the reported weights, and the selection of the decision threshold, is stated as Algorithm 2.

H. Metrics Evaluation

The performance was analyzed by evaluating the model using a confusion matrix [47]. The confusion matrix consists of four main components, namely True Positive (TP), representing infarct areas correctly detected by the model; False Negative (FN), representing infarct areas that were not detected by the model; False Positive (FP), representing normal areas incorrectly classified as infarct; and True Negative (TN), representing normal areas correctly recognized by the model, all evaluated at the pixel level. Based on this confusion matrix, several key evaluation metrics were used to assess the model's performance, namely precision, sensitivity, the Dice Similarity Coefficient

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

(DSC), and the Intersection over Union (IoU), with the formulas as follows [48].

a. Precision is used to measure the level of confidence in the infarct predictions generated by the model, namely the proportion of positive predictions that are actually infarct at the pixel level, as shown in Eq. (20) [47],

$$Precision = \frac{TP}{TP+FP} \quad (20)$$

b. Sensitivity describes the system's ability to detect actual positive cases, which is particularly critical in disease diagnosis, and is formulated in Eq. (21) [47],

$$Sensitivity = \frac{TP}{TP+FN} \quad (21)$$

c. Dice Similarity Coefficient (DSC) is defined as twice the area of intersection divided by the sum of the sizes of both areas, as shown in Eq. (22) [48],

$$DSC = \frac{2TP}{2TP+FP+FN} \quad (22)$$

d. Intersection over Union (IoU) is defined as the ratio between the area of intersection and the area of union of the two regions, as shown in Eq. (23) [48].

$$IoU = \frac{TP}{TP+FP+FN} \quad (23)$$

All four metrics are computed at the pixel level and pooled over the test subset. They quantify spatial overlap but do not establish whether each individual infarct was found, and a single large, well-segmented lesion can dominate a pooled pixel count [49]. Lesion-level measures such as lesion detection sensitivity, false-positive lesions per scan and lesion-wise DSC, together with performance stratified by lesion volume, would address this directly and are the natural complement to what is reported here; they were not computed in the present study and are identified in Section IV as required before any claim about small-lesion detection can be supported by evidence rather than inference.

Inference and the evaluation procedure, including the lesion-level measures and the statistical comparison, are stated as Algorithm 3.

Algorithm 2. Training, Validation, Early Stopping, and Decision-Threshold Selection for U-Net and Multi-scale U-Net

```
Input: training set T and validation set V; network f, U-Net or Multi-scale U-Net, with weights w;seed s
Output: selected weights w* and decision threshold t*
 1: seed every random number generator with s              // one seed fixes initialisation and batching
 2: w <- Kaiming uniform initialisation;  optimiser Adam, learning rate 1e-5, batch size 10
 3: best <- infinity;  bad <- 0;  w* <- w
 4: for epoch e = 1 to 100 do
 5:    for each mini-batch B of 10 slices drawn from T do
 6:       P <- sigmoid(f(B; w))                                                    (Eq. (11))
 7:       L <- Dice loss between P and the labels of B                             (Eq. (18))
 8:       w <- Adam update of w from the gradient of L
 9:    Lval <- Dice loss over the whole of V, which is never augmented
10:    halve the learning rate if Lval has not improved for 8 epochs
11:    if Lval < best then best <- Lval;  w* <- w;  bad <- 0   // best-weight selection
12:    else bad <- bad + 1
13:    if bad = 20 then break                                  // early stopping
14: t* <- the t in {0.10, 0.15, ..., 0.90} that maximises DSC on V with weights w*  // threshold search
15: return w*, t*
```

Algorithm 3. Inference, Segmentation Performance Evaluation, and Patient-Level Statistical Analysis

```
Input : test set S of 51 patients; selected weights w*; decision threshold t*
Output: per-patient table R, lesion-level table L and the paired statistical comparisons
 1: for each patient p in S do
 2:    for each slice I of patient p do
 3:       Yhat <- sigmoid(f(I; w*))                       // probability map        (Eq. (11))
 4:       Ybin <- 1 where Yhat > t*, else 0               // binary mask            (Eq. (19))
 5:       remove connected components smaller than 10 pixels
 6:    TP, FP, FN <- the sums of true positives, false positives and false negatives over p
 7:    precision, sensitivity, DSC and IoU of patient p                    (Eq. (20)-(23))
 8:    label the connected components of the annotation and of the prediction
 9:    a true lesion counts as detected if any predicted component overlaps it
10:    record detection sensitivity, false-positive lesions per patient, lesion DSC and size stratum
11: R <- the table of per-patient values;  L <- the lesion-level table
12: report the mean of each column of R with a bootstrap 95% interval from 10,000 resamples
13: compare configurations pairwise over the 51 patients, Wilcoxon signed-rank test, p < 0.05
14: return R, L and the paired test results
```

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 2. Training performance and validation-selected decision thresholds for U-Net and Multi-scale U-Net with and without data augmentation. τ is the probability threshold of Eq. (19), chosen on the validation subset over the grid 0.10 to 0.90 as the value that maximises validation DSC. Epochs is the number completed before the stopping rule halted training, which it did after 20 epochs without an improvement in validation loss; time is given in minutes and peak GPU memory in gigabytes.

Configuration	Epochs	Time (min)	Peak GPU (GB)	τ
U-Net	29	30.5	2.19	0.70
U-Net + augmentation	55	152.0	2.19	0.10
Multi-scale U-Net	40	113.8	5.37	0.90
Multi-scale U-Net + augmentation	41	327.6	5.37	0.90

III. Result

The two architectures were compared on the same partition by study, each with and without augmentation, giving four configurations trained with a single seed. Results are reported at three levels: pooled over the pixels of the test subset, averaged over the 51 test patients, and at the level of individual lesions. The patient-level values carry the statistical tests because the patient is the unit independent across subsets.

A. Training behaviour and threshold selection

Table 2 summarises the four training runs, whose loss curves are shown in Fig. 5 and Fig. 6. All four stopped before the 100-epoch ceiling under the early-stopping rule of Algorithm 2. U-Net without augmentation stopped first, after 29 epochs, having reached its lowest validation loss early and then drifting upward, as expected for 2,080 training slices supporting 7.8 million parameters. Augmentation delayed that point, and the same architecture trained usefully for 55 epochs on the 6,240 augmented slices. The Multi-scale U-Net ran 40 and 41 epochs in the two conditions. Cost per epoch followed the architecture, at roughly 63 and 166 seconds for U-Net without and with augmentation against 171 and 480 seconds for the Multi-scale U-Net, a factor near 2.8, and peak memory during training rose from 2.19 GB to 5.37 GB.

The decision threshold selected on the validation subset, following Algorithm 2, differs sharply between the two architectures and is worth stating before interpreting the metrics. U-Net selected a 0.70 without augmentation and 0.10 with it, while the Multi-scale U-Net selected 0.90 under both conditions. A threshold at the top of the search range means that the multi-scale network assigns high probabilities over a region wider than the lesion, so a strict cut is needed to contain it. That property reappears below as the lowest precision and highest sensitivity of the four configurations, and it is the most consistent effect of the architecture found in this study.

One feature of the curves bears on the comparison that follows. The Multi-scale U-Net reaches the lower validation Dice loss in both conditions, 0.6913 against 0.7497 without augmentation and 0.6326 against 0.6669 with it, although its thresholded overlap on the test subset is the lower of the two. The loss is computed on probabilities while the metrics are computed after binarisation, so the two need not agree; the multi-scale

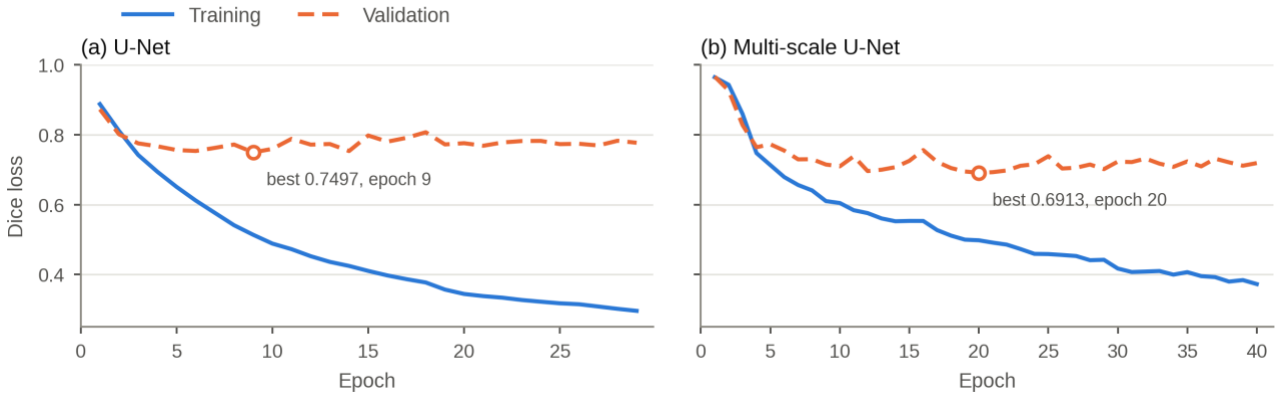


Fig. 5. Training and validation Dice-loss curves of (a) the conventional U-Net and (b) the proposed Multi-scale U-Net without data augmentation. The solid line is the training loss and the dashed line the validation loss. The open circle on each validation curve marks the epoch of lowest validation loss, whose weights were kept for testing. Both panels use the same vertical scale and the same line styles; the horizontal axis spans the epochs each run completed before early stopping.

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

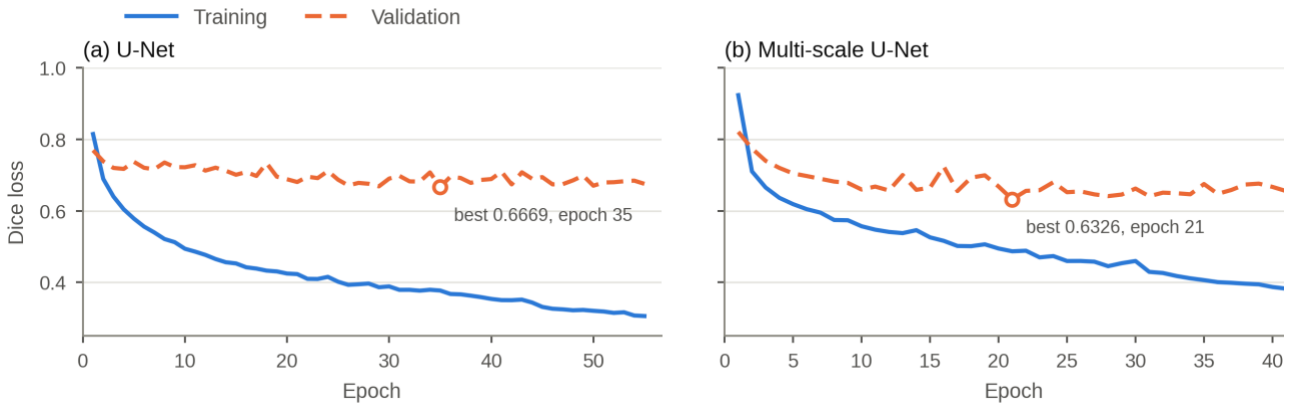


Fig. 6. Training and validation Dice-loss curves of (a) the conventional U-Net and (b) the proposed Multi-scale U-Net with data augmentation. The vertical axis limits, the line styles and the marker convention are those of Fig. 5, so the two training conditions can be compared directly.

network ranks the pixels better and places the boundary worse. This is the same property that the threshold selection exposes, and it explains why a comparison made at a fixed threshold of 0.5 would favour the multi-scale network. Holding the threshold at 0.5 for both architectures, as a fixed-threshold protocol would, penalises them unequally for this reason. Selecting it on the validation subset removes that confound, at the cost of one further quantity estimated from held-out data.

B. Segmentation performance on the test subset

Table 3 reports the four configurations on the 51 test patients. Pooled over the pixels in the test subset, U-Net with augmentation gives the highest overlap of the four, with a DSC of 0.6632 and an IoU of 0.4961, compare with 0.6372 and 0.4676 for the Multi-scale U-Net with augmentation. Without augmentation, the ordering is the same, 0.5963 against 0.5445. Replacing every convolutional block with a three-branch multi-scale block therefore did not improve overlap under a partition by study. The two architectures nevertheless differ in ways the overlap metrics conceal. The Multi-scale U-Net reaches the highest sensitivity of the four configurations, 0.7172 without augmentation and 0.6980 with it, together with the lowest precision, 0.4389 and 0.5862. U-Net is the mirror image, with the highest precision of the four, 0.7511 with augmentation, and the lowest sensitivity, 0.5937. The multi-scale block shifts the operating point toward recall: it labels more infarct and more non-infarct, as seen in the representative cases in Fig. 7 and Fig. 8, where predictions extend beyond the annotated boundary. Because DSC and IoU weight the two error types symmetrically, this kind of shift can leave them unchanged while substantially altering the model's behaviour.

Averaging over patients rather than pooling over pixels changes the comparison in one condition. With augmentation, the ordering is preserved: 0.4938 for U-Net against 0.4705 for the Multi-scale U-Net. Without augmentation, it reverses: 0.3879 against 0.4224. Pooled values are dominated by the patients with the largest infarcts, because every pixel counts once, whereas the patient-level mean gives a patient with a 40-pixel lesion the same weight as one with a 4,000-pixel lesion. The two conventions need not agree, and that they disagree here is a result in itself: the relative standing of the two architectures depends on lesion size.

Table 3. Quantitative comparison of U-Net and Multi-scale U-Net segmentation performance on 51 independent test patients. Precision, sensitivity, the Dice Similarity Coefficient (DSC) and the Intersection over Union (IoU) are defined in Eqs. (20) to (23). The first four columns are pooled over every pixel of the test subset; the last column is the mean ± standard deviation of the DSC computed for each patient separately.

Configuration	Precision	Sensitivity	DSC	IoU	DSC per patient
U-Net	0.5885	0.6043	0.5963	0.4248	0.3879 ± 0.2181
U-Net + aug.	0.7511	0.5937	0.6632	0.4961	0.4938 ± 0.2256
Multi-scale U-Net	0.4389	0.7172	0.5445	0.3741	0.4224 ± 0.2223
Multi-scale U-Net + aug.	0.5862	0.6980	0.6372	0.4676	0.4705 ± 0.2399

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

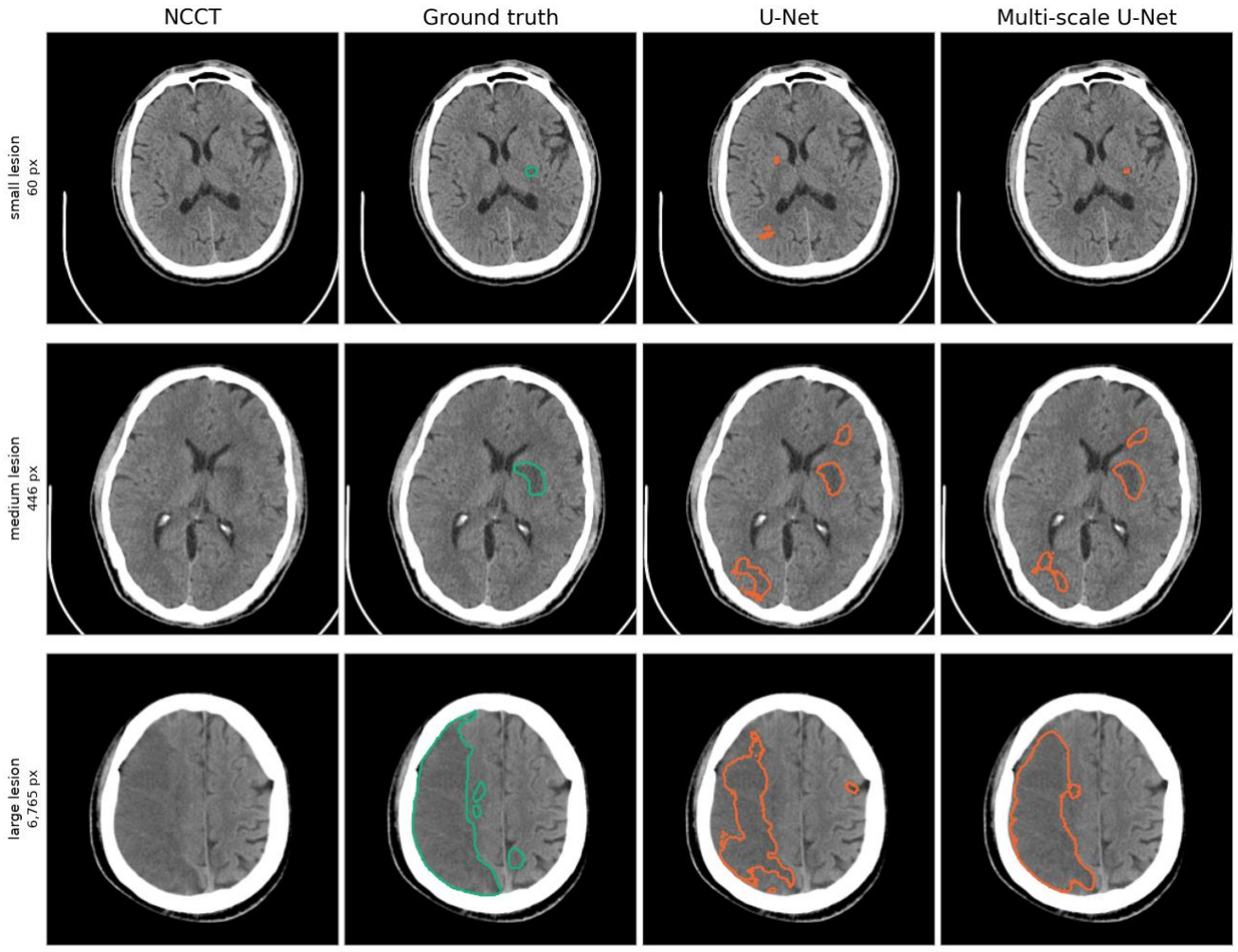


Fig. 7. Qualitative comparison of infarct segmentation results obtained using U-Net and Multi-scale U-Net without data augmentation for small, medium, and large lesions. The columns are the NCCT slice, the reference annotation, the U-Net prediction and the Multi-scale U-Net prediction. The annotation is outlined in green and both predictions in orange, each network at its own validation-selected threshold. The rows are a small lesion of 60 pixels, a medium lesion of 446 pixels and a large lesion of 6,765 pixels from the test subset. The two outline colours differ in lightness as well as in hue, so they remain distinguishable in greyscale print.

Augmentation improved every configuration on every metric except sensitivity, which fell slightly for both architectures. Pooled DSC rose by 0.0669 for U-Net and by 0.0927 for the Multi-scale U-Net, and the per-patient mean by 0.1059 and 0.0481. The gain is concentrated in precision, which rose by 0.1626 for U-Net and 0.1473 for the Multi-scale U-Net, while sensitivity fell by 0.0107 and 0.0193. Augmentation, therefore, did not help the networks find more infarct; it helped them stop labelling tissue that is not infarct.

C. Separating the contribution of the architecture from that of augmentation

The four configurations form a 2×2 design, and reading them as a grid separates the two interventions. Holding the data condition fixed, replacing the standard block with the multi-scale block changes the per-patient DSC by +0.0345 without augmentation and by −0.0233 with it. Holding the architecture fixed, augmentation raises it by 0.1059 for U-Net and by 0.0481 for the Multi-scale U-Net. Augmentation is therefore the larger effect in both rows of the grid, by a factor of three to five, and the only one the statistical test supports. The architecture effect is not merely small; it changes sign between the two data conditions, so the two interventions are not additive here, and the value of the multi-scale block cannot be stated independently of the training data it is given.

Two qualifications belong with this reading. First, each configuration was trained once, with a single seed. The paired tests in Table 4 establish that the

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

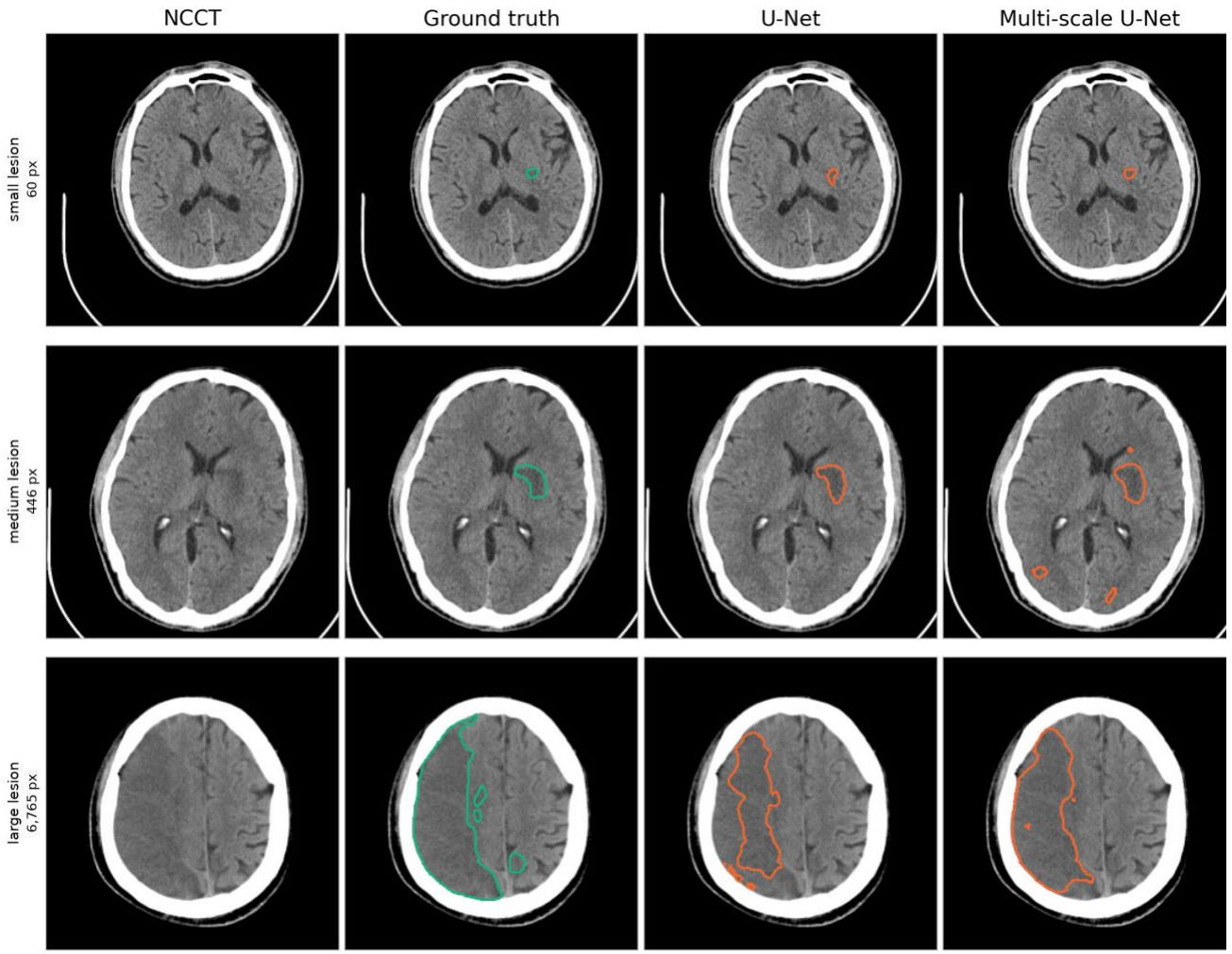


Fig. 8. Qualitative comparison of infarct segmentation results using U-Net and Multi-scale U-Net with data augmentation for small, medium, and large lesions. The columns are the NCCT slice, the reference annotation, the U-Net prediction and the Multi-scale U-Net prediction. The annotation is outlined in green and both predictions in orange, each network at its own validation-selected threshold. The rows are the same small lesion of 60 pixels, medium lesion of 446 pixels and large lesion of 6,765 pixels shown in Fig. 7.

differences between the trained models are unlikely to arise from which patients happen to be in the test subset, but they say nothing about variation between training runs of the same configuration; a second and a third seed were begun and are not complete, and an architecture difference of 0.02 to 0.03 in DSC lies within the range that initialisation alone can produce in segmentation tasks of this size. Second, the comparison between pooled and per-patient values above shows that the ranking of the two architectures depends on the averaging convention. Therefore, neither ordering should be reported without specifying the convention used to obtain it.

The comparisons above are differences between single numbers. Table 4 tests them on the 51 test patients, treating each patient as one paired observation and using the Wilcoxon signed-rank test with a bootstrap interval on the mean difference. The architecture comparison, made in the augmented condition in which both networks perform best, does not reach significance on either overlap metric: the Multi-scale U-Net is lower by 0.0233 in DSC, with a 95 per cent interval from −0.0601 to 0.0126 and p = 0.19, and lower by 0.0169 in IoU with p = 0.18. Both intervals contain zero. The same comparison is strongly significant on the two-component metrics, in opposite directions: sensitivity is higher by 0.0602, interval 0.0251 to 0.0947, p = 0.0015, and precision is lower by 0.0999, interval −0.1469 to −0.0567, p = 0.000026. The effect of the architecture is therefore real and

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 4. Patient-level paired statistical comparison of segmentation performance between experimental configurations using the Wilcoxon signed-rank test. Each comparison rests on n = 51 paired observations, one per test patient. A difference is the second configuration minus the first, so a positive value favours the second; the interval is a bootstrap 95 per cent confidence interval on the mean difference from 10,000 resamples, and p < 0.05 is treated as significant.

Comparison	Metric	Difference	95 % CI	*p*
MS-U-Net (aug) − U-Net (aug)	DSC	−0.0233	−0.0601 to 0.0126	0.19
	IoU	−0.0169	−0.0473 to 0.0138	0.18
	Sensitivity	+0.0602	0.0251 to 0.0947	0.0015
	Precision	−0.0999	−0.1469 to −0.0567	< 0.001
U-Net (aug) − U-Net (no aug)	DSC	+0.1059	0.0683 to 0.1441	< 0.001
MS-U-Net (aug) − MS-U-Net (no aug)	DSC	+0.0481	0.0169 to 0.0770	< 0.001

measurable, but it falls on the balance between the two kinds of error rather than on their total.

The augmentation comparisons are significant for both architectures. For U-Net, the per-patient DSC rises by 0.1059, interval 0.0683 to 0.1441, p = 0.000002; for the Multi-scale U-Net, by 0.0481, interval 0.0169 to 0.0770, p = 0.00034. Augmentation is thus the only intervention examined here that improves overlap by an amount these data support, and its effect on U-Net is roughly twice its effect on the Multi-scale U-Net.

Pixel overlap does not say whether an individual infarct was found. The upper part of Table 5 reports the proportion of annotated lesions that received an overlapping prediction, the number of false-positive lesions per patient, and the mean DSC of detected lesions. The ordering again differs from the DSC ordering. The Multi-scale U-Net detects more lesions than U-Net in both conditions, 0.7023 against 0.6654 without augmentation and 0.6145 against 0.6056 with it. The price appears in the false-positive count: with augmentation, the Multi-scale U-Net leaves 8.6 spurious lesions per patient against 5.6 for U-Net. Augmentation is what controls that number, cutting it from 15.5 to 5.6 for U-Net and from 14.6 to 8.6 for the Multi-scale U-Net while slightly reducing detection. Five to nine false lesions per patient is far above what reading support would tolerate, and it is the clearest single statement of how far this task remains from application.

The lower part of Table 5 stratifies the test slices by annotated lesion area, and the three strata are very unequal in difficulty. With augmentation, U-Net reaches a DSC of 0.6983 on the 193 slices whose lesion exceeds 500 pixels, 0.4240 on the 161 slices between 100 and 500 pixels, and 0.1508 on the 55 slices below 100 pixels, whose median lesion is 49 pixels. The architecture comparison across the strata does not support the hypothesis that motivated the multi-scale block. U-Net has the higher DSC in every stratum and in both data conditions; on the smallest lesions, it reaches 0.1508 against 0.0753 with augmentation and 0.0779 against 0.0513 without. The multi-scale network does detect more of the small lesions in the non-augmented condition, 0.4655 of them against 0.3621, but it delineates them worse. Parallel multi-kernel filtering as implemented here therefore improves the chance that a small infarct is noticed, not the quality with which it is outlined.

D. Effect of the target definition

The target used throughout this paper is the one the dataset authors specify, labels 1, 2, 3 and 5. Two alternatives are defensible: adding label 4, which marks acute infarct identified on the reference DWI-MRI but judged invisible on the NCCT, and restricting the target to the two purely acute labels. Table 6 reports U-Net with augmentation trained and evaluated under each of the three definitions, evaluated on the test subset of 51

Table 5. Lesion-level detection performance and segmentation accuracy stratified by annotated infarct area. Lesion detection is the proportion of annotated lesions that receive at least one overlapping predicted component, FP lesions per patient the number of predicted components that overlap no annotated lesion, and lesion DSC the mean Dice coefficient of the detected lesions.

Configuration	Lesion detection	FP lesions per patient	Lesion DSC	DSC small	DSC medium	DSC large
U-Net	0.6654	15.5	0.3460	0.0779	0.3034	0.6487
U-Net + aug.	0.6056	5.6	0.3864	0.1508	0.4240	0.6983
Multi-scale U-Net	0.7023	14.6	0.3707	0.0513	0.2617	0.6146
MS-U-Net + aug.	0.6145	8.6	0.3734	0.0753	0.3369	0.6905

Note: The strata hold slices whose annotated lesion area is below 100 pixels (small, 55 slices, median 49 pixels), between 100 and 500 pixels (medium, 161 slices, median 268 pixels), and above 500 pixels (large, 193 slices, median 1,508 pixels).

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 6. Influence of infarct annotation-label definitions on U-Net segmentation performance with data augmentation. The AISD annotations carry five labels: 1 remote infarct, 2 clear acute infarct, 3 blurred acute infarct, 4 acute infarct invisible on NCCT, and 5 infarcts. The first row binaries the labels 1, 2, 3 and 5 specified by the dataset authors, the second adds label 4, and the third keeps only the acute labels 2 and 3. Each row is trained and evaluated against its own ground truth, so the comparison is between definitions of the target rather than between models. Overlap measures are given to four decimal places and false-positive lesions per patient to one.

Target	Labels	Precision	Sensitivity	DSC	IoU	FP lesions
Dataset definition	1, 2, 3, 5	0.7511	0.5937	0.6632	0.4961	5.6
Label 4 added	1, 2, 3, 4, 5	0.6025	0.6460	0.6235	0.4530	15.3
Acute labels only	2, 3	0.4556	0.6128	0.5226	0.3537	6.8

patients. Each row is trained and evaluated against its own ground truth, so the comparison is between definitions of the target rather than between models. FP lesions are false-positive lesions per patient. The three rows are not three models measured on one task, because each is scored against its own ground truth; the comparison is between definitions of the target, and what it answers is which definition yields the most learnable task. The first row was retrained from scratch for this experiment and reproduced every value of Table 3 to four decimal places, at the same stopping epoch of 55 and the same threshold of 0.10, which establishes that the pipeline is deterministic given its seed.

Adding label 4 costs 0.0396 in DSC and 0.1485 in precision, and raises false-positive lesions from 5.6 to 15.3 per patient. Sensitivity rises by 0.0524 and lesion detection by 0.0425, which is exactly what the change asks for: a network trained to mark tissue that carries no visible signal on the input image learns to mark more tissue everywhere. Label 4 appears in 436 of the 3,144 annotated slices and accounts for 9.1 per cent of the annotated area, so the cost in precision is out of proportion to the area added. This is the mechanism by which a target that includes label 4 inflates sensitivity at the expense of precision, and it is the reason the specification given by the dataset authors was adopted here.

Restricting the target to labels 2 and 3 is worse again, and the reason is instructive. DSC falls to 0.5226 and precision to 0.4556, while false-positive lesions stay near the baseline at 6.8 per patient: the false positives are few but large. Labels 1 and 5, remote and unspecified infarct, account for 12.3 per cent of the annotated area and are visible on the image; removing them from the target does not remove them from the scan, so the network marks them and is scored as wrong for doing so. A target restricted to the acute labels is therefore not the cleaner choice it appears to be, unless studies carrying remote infarct are excluded as well. The definition given by the dataset authors performs best on every overlap measure, which is a result in its own right, and it is also the definition that keeps published work on AISD comparable.

E. Computational cost

Replacing every convolutional block with a three-branch block is not free, and Table 7 quantifies the price. Timings are measured on an NVIDIA T4 at batch size 1. The Multi-scale U-Net holds 17,905,249 trainable parameters against 7,765,409 for the baseline, a factor of 2.31, and requires 24.91 GMACs, equivalently 49.81 GFLOPs, per 256×256 slice against 12.03 GMACs or 24.07 GFLOPs, a factor of 2.07; in 32-bit precision the stored model grows from 29.6 MB to 68.3 MB. These counts are exact and follow from Eq. (16) and Eq. (17) together with Table 1. Measured on

Table 7. Computational complexity and inference efficiency of conventional U-Net and the proposed Multi-scale U-Net for 256×256 input images. Trainable parameters are the weights and biases the network learns, model size the storage they occupy in 32-bit precision, MACs the multiply-accumulate operations per slice and FLOPs the floating-point operations, one MAC counting as two FLOPs; G denotes giga-operations.

Property	U-Net	Multi-scale U-Net	Ratio
Trainable parameters	7,765,409	17,905,249	2.31×
Model size (FP32)	29.6 MB	68.3 MB	2.31×
MACs per slice	12.03 G	24.91 G	2.07×
FLOPs per slice	24.07 G	49.81 G	2.07×
GPU inference per slice	6.8 ms	23.0 ms	3.37×
Peak GPU memory in training	2.19 GB	5.37 GB	2.45×

Note: Parameter, MAC and FLOP counts are exact and follow from Eq. (16) and Eq. (17) together with Table 1. Inference time and peak training memory were measured on an NVIDIA T4 GPU at batch size 1 with a 256×256×1 input.

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 8. Comparison of the proposed segmentation approach with previously published U-Net-based methods for acute ischemic stroke infarct segmentation. For each entry the table gives the dataset and the partitioning protocol, the architecture, and the overlap and error metrics that the cited work reports. A dash marks a value the cited work does not state and n.s. a partitioning protocol that is not stated. Studies that use different datasets or different partitions are not directly comparable; only the AISD entries evaluated on the official 52-scan test split share the protocol of this work.

Study	Dataset and partition	Architecture	DSC	IoU	Sens.	Prec.
Liang et al. [29]	AISD 397; patient, official 345/52	U-Net (baseline)	0.4588	–	0.5019	0.5196
Liang et al. [29]	AISD 397; patient, official 345/52	SEAN	0.5784	–	0.5880	0.6597
Sun et al. [50]	AISD 397; patient	U-Net	0.4578	–	–	–
Sun et al. [50]	AISD 397; patient	Attention U-Net	0.4710	–	–	–
Sun et al. [50]	AISD 397; patient	U-Net++	0.4631	–	–	–
Sun et al. [51]	AISD 397; patient, official 345/52	3D nnU-Net with symmetry prior	0.6720	–	–	–
Li et al. [36]	private 30; n.s.	U-Net, augmented	0.81	–	–	–
Li et al. [36]	private 30; n.s.	Multi-scale U-Net, augmented	0.86	–	–	–
Cherikbayeva et al. [52]	private 50; n.s.	ensemble of three 3D U-Nets	0.7008	–	0.6050	0.7425
This work	AISD 375; patient, 51 of 52 official	U-Net	0.5963	0.4248	0.6043	0.5885
	AISD 375; patient, 51 of 52 official	U-Net, augmented	0.6632	0.4961	0.5937	0.7511
	AISD 375; patient, 51 of 52 official	Multi-scale U-Net, augmented	0.6372	0.4676	0.6980	0.5862

the training GPU at batch size 1, inference took 6.8 ms and 23.0 ms per slice, a ratio of 3.37 that exceeds the ratio of operations because the three branches are executed one after another rather than concurrently, and peak memory during training rose by a factor of 2.45. At 23 ms per slice, a 30-slice study is processed in under a second, so latency is not the obstacle. The obstacle is that the additional cost buys a shift in the balance of errors rather than an improvement in overlap.

IV. Discussion

Two findings follow from the experiments, and they point in different directions. Under a partition by study using the dataset's own predefined test split, replacing every convolutional block of U-Net with a three-branch multi-scale block did not improve overlap: the per-patient DSC fell by 0.0233, with an interval containing zero and p = 0.19. What the block did change, significantly and in opposite directions, is the balance between the two kinds of error, raising sensitivity by 0.0602 and lowering precision by 0.0999. Augmentation improved both architectures by 0.1059 and 0.0481 in per-patient DSC, and did so chiefly by suppressing false positives rather than by finding more infarct: false-positive lesions fell from 15.5 to 5.6 per patient for U-Net. The mechanistic explanation offered in Section II, that filtering at several receptive-field sizes should help where lesion size varies, is not supported by the stratified results. U-Net segments better at every lesion size, and the advantage of the multi-scale block is confined to noticing small lesions rather than outlining them. The architectural change examined here is therefore best described as a change in operating point, which a decision threshold can also produce, and not as an increase in the information the network extracts.

Table 8 lists the values reported by five previous studies that applied U-Net-based architectures to acute ischemic stroke. Sensitivity is recall, as the cited work uses that name. A dash marks a value that the cited work does not state beside its Dice score, and n.s. indicates that the partitioning protocol is not stated. None of these studies reports a parameter count or an inference time, so computational cost cannot be compared across them. The present study can be placed alongside the four entries that use AISD, because it uses the same dataset and the same predefined test split of 52 scans, of which 51 remain after the supratentorial filter. Two independent groups report a plain U-Net on AISD at 0.4588 [29] and 0.4578 [50], values that agree to within 0.001, and the

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

strongest published result on the official test split is 0.6720 for a three-dimensional nnU-Net with a symmetry prior and a generated lesion-probability channel [51]. The U-Net trained here reaches 0.5963 without augmentation and 0.6632 with it; the Multi-scale U-Net reaches 0.5445 and 0.6372. The augmented U-Net therefore sits 0.0088 below the best published value on this dataset and well above the plain U-Net baselines, although the protocols are not identical.

Three differences keep even this comparison approximate, and all three work in the direction of raising the figures reported here. The present evaluation is two-dimensional and slice-based, while [51] is volumetric. It is restricted to studies whose infarct is supratentorial, which removes the infratentorial cases that are hardest to read on NCCT. And 409 of the 412 test slices carry an annotated lesion, because the subset was assembled from annotated slices, so the evaluation does not measure false positives on lesion-free slices in the way a volume-level evaluation does. The comparison is best read as showing that these results now sit inside the published range for AISD, not that they exceed it.

Two further properties of Table 8 deserve comment, because both bear on how the comparison should be read. The first is that most of these studies report Dice alone. Where a second measure is given beside Dice it is a boundary distance in one case [51], and where the error types are separated they are reported as recall and precision [29], [52]; the remaining entries give no basis for judging whether a model errs towards over- or under-segmentation, which Section III shows to be the dimension on which the two architectures examined here actually differ. The second is that none of the studies reports a parameter count, a multiply-accumulate count or an inference time. The cost of an architectural change is therefore invisible in the published record for this task, which is why Table 7 of this paper states it explicitly.

The two studies on private collections stand apart from all of this. Li et al. [36] report 0.81 and 0.86 on 30 patients from a single hospital, and Cherikbayeva et al. [52] report 0.7008 for an ensemble of three three-dimensional U-Nets on 50 patients from a single centre; both sit above every value published on AISD, including the ones reported here. The homogeneity-versus-generalizability explanation that might be offered for that pattern, namely that a small single-centre collection varies little in scanner protocol, patient profile or lesion appearance and therefore poses an easier task, is consistent with the figures but is not supported by anything measured here. Testing it would require evaluating the same architecture on both datasets under a common protocol. Several alternative explanations fit the differences equally well, among them the partitioning regimes discussed above, differences in slice inclusion criteria, in the size distribution of the annotated lesions, in the intensity window applied before training, in annotation procedure, and in whether metrics were pooled over pixels or averaged per case. The stratified results in Table 5 give that last point quantitative force: on this dataset, DSC varies from 0.15 to 0.70 across lesion-size strata within a single model, a range wider than the entire spread between published studies.

Several limitations qualify these results, and the first governs the rest. Each configuration was trained once, with a single seed. A second and a third seed were started and are not complete, so variation between training runs is unmeasured, and an architecture difference of 0.02 to 0.03 in DSC lies within the range that initialisation alone can produce. The paired tests establish that the observed differences are stable across patients, not that they are stable across training runs; these are different claims, and only the first is supported here. Until repeated runs are available, the finding that the multi-scale block does not improve overlap should be read as a failure to demonstrate an improvement rather than as a demonstration of equivalence.

Further limitations follow. No ablation over the loss function was performed, so the contribution of the Dice objective cannot be separated from that of the architecture, and objectives that penalise false negatives more heavily, such as the Tversky loss [45] or its focal variant [46], remain untested on this partition. Evaluation was performed slice by slice in two dimensions, without enforcing consistency between adjacent slices. The decision threshold was selected on the validation subset, which is the correct procedure but adds one further quantity estimated from held-out data.

The remaining limitations concern scope rather than method. Segmentation was confined to infarcts in the supratentorial region, covering the ACA, MCA and PCA distributions, while infratentorial regions such as the brainstem and cerebellum were excluded; those infarcts present with distinct imaging characteristics and greater artifact interference on NCCT owing to their proximity to bone, so nothing demonstrated here transfers to them [14]. All images derive from a single public source acquired at one institution, so the patient-level figures describe generalization within AISD and say nothing about scanners, reconstruction kernels, institutions or acquisition protocols elsewhere; no external test set and no prospectively collected cases were examined. A further consequence of how the subset was assembled is that 409 of the 412 test slices carry an annotated lesion, so the reported metrics do not measure the behaviour of either network on normal slices, which a deployed system would encounter constantly. Conversion of the volumes to 8-bit images through a single brain window discards the quantitative

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

HU scale and saturates densities outside that window. And the annotations, derived from DWI-MRI by one reader and double-reviewed, carry their own uncertainty, which sets an upper bound on the agreement any model can reach.

These findings carry implications for future work, and none yet for clinical practice. Methodologically, the most useful conclusion is negative and therefore worth stating plainly: adding parallel kernels at every resolution level is not, by itself, a productive direction for this task, and the 2.31-fold increase in parameters it costs would be better spent elsewhere. The two quantities that did move, false-positive lesions per patient and performance on lesions below 100 pixels, are where the remaining distance lies, and both are addressed more directly by the loss function, by the decision rule and by three-dimensional context than by the width of the convolutional kernels. Repeated runs across seeds and an ablation over objectives are the immediate next steps.

On the clinical side, the distinction between what has been shown and what would be required is wide. A model that leaves five to nine false lesions in every study, and that reaches a DSC near 0.15 on infarcts below 100 pixels, cannot support reading of early ischemic change, where the lesions that matter most for treatment decisions are precisely the small and faint ones. All images come from a single institution, so even the patient-level figures describe generalization within AISD and say nothing about other scanners, reconstruction kernels or acquisition protocols. External validation on an independent cohort, and a reader study, would both be required before any claim of clinical usefulness could be considered.

V. Conclusion

This work asked how much a dense parallel multi-kernel block improves U-Net for acute ischemic stroke infarct segmentation on NCCT, and at what cost, under an evaluation in which no patient appears in more than one subset. On the 51 patients of the predefined AISD test split, the answer is that it does not improve overlap: the Multi-scale U-Net reached a DSC of 0.6372 against 0.6632 for U-Net with augmentation, and the paired difference of 0.0233 per patient in favour of U-Net was not significant ($p = 0.19$). The block does have a significant and consistent effect, but on the balance of errors rather than their total, raising sensitivity by 0.0602 and lowering precision by 0.0999, and it costs 2.31 times the parameters and 3.37 times the inference time. Augmentation, applied to the training subset alone, improved both architectures significantly, by 0.1059 and 0.0481 in per-patient DSC, mainly by reducing false-positive lesions from 15.5 to 5.6 and from 14.6 to 8.6 per patient. Performance on infarcts below 100 pixels remains poor for both networks, with DSC between 0.05 and 0.15. For this task, therefore, the training data and the decision rule matter more than this architectural change, and the remaining distance lies in small-lesion detection and false-positive suppression. An ablation over the target definition confirmed the labelling specified by the dataset authors as the best of the three defensible choices, the addition of label 4 costing 0.0396 in DSC and tripling false-positive lesions. Repeated runs across seeds, an ablation over the loss function, and external validation are the next steps.

Declarations

Ethics Approval and Consent to Participate

This study used publicly available, de-identified non-contrast computed tomography (NCCT) images obtained from the Acute Ischemic Stroke Dataset (AISD). According to the original dataset publication, the patient imaging data were collected with appropriate ethical approval and informed consent. The present study involved secondary analysis of existing anonymized data and did not include direct participant recruitment, clinical interventions, or collection of additional patient information. No additional ethical approval was sought for this secondary analysis, in accordance with the applicable institutional policy. The original ethical approval and consent procedures are described in the AISD dataset publication [29].

Consent for Publication

The study used previously collected, de-identified medical imaging data made available through a public research dataset. No personally identifiable patient information, facial photographs, or identifying clinical records are presented in this manuscript. Therefore, no additional consent for publication was obtained by the present authors. The original dataset creators were responsible for obtaining applicable participant consent during data acquisition.

Availability of Data and Materials

The dataset analyzed in this study is publicly available through the Acute Ischemic Stroke Dataset (AISD) repository at https://github.com/GriffinLiang/AISD. The dataset includes non-contrast CT images and corresponding infarct annotations used for developing and evaluating the segmentation models. The data were used for non-commercial academic research in accordance with the conditions specified by the dataset providers. No new patient data were collected. The experimental implementation and associated code are available from the corresponding author upon reasonable request, subject to applicable institutional and intellectual property restrictions.

Funding

This research was supported by the Airlangga Research Fund under the Research Group Scheme, Faculty of Science and Technology, Universitas Airlangga, in 2025, under Grant Number 7551/B/UN3.FST/PT.01.03/2025.

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Competing Interests

The authors declare that they have no known financial or non-financial competing interests that could have influenced the research, analysis, interpretation of results, or preparation of this manuscript.

Authors' Contributions

Osmalina Nur Rahma contributed to research conceptualization, supervision, and manuscript review and editing. Aisyah Widayani contributed to methodological review and data analysis. Nimas Maharani Putri contributed to data collection, data processing, and preparation of the experimental dataset. Rifai Chai contributed to critical review and intellectual revision of the manuscript. All authors reviewed and approved the final manuscript and agree to be accountable for the integrity and accuracy of the published work.

Acknowledgments

The authors gratefully acknowledge the Biomedical Engineering Study Program, Department of Physics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia, for providing institutional support and research facilities. The authors also thank colleagues at Rumah Sakit Umum Daerah Haji Jawa Timur for their constructive comments during manuscript preparation. The authors acknowledge the developers and contributors of the Acute Ischemic Stroke Dataset (AISD) for making the medical imaging data available for scientific research.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors used generative AI and AI-assisted technologies to support language editing, to help organise and draft parts of the text, and to assist in preparing analysis scripts and figures. The study design, the experiments, the analysis of the data, the interpretation of the results and the conclusions are the work of the authors. Every AI-assisted output was checked, verified and edited by the authors, who reviewed and approved the final manuscript and take full responsibility for its content.

Declaration of Research Integrity and Originality

The authors affirm that the results presented in this manuscript were obtained through the research procedures described in the Methods section. The study reports an original analysis of an existing publicly available dataset. The authors accept responsibility for the accuracy, integrity, and reproducibility of the reported methodology, experimental results, and conclusions. All external research materials, datasets, and previously published studies have been acknowledged and cited appropriately.

Clinical Applicability Statement

This study was conducted for methodological research purposes using retrospective, publicly available medical imaging data. The proposed segmentation framework has not undergone prospective clinical evaluation or independent external validation and is not intended for direct clinical diagnosis or treatment decision-making. Additional multicenter validation, clinical assessment, and regulatory evaluation would be necessary before considering clinical deployment.

References

[1] N. A. Hilkens, B. Casolla, T. W. Leung, and F.-E. de Leeuw, “Stroke,” Lancet, vol. 403, no. 10446, pp. 2820–2836, Jun. 2024, doi: 10.1016/S0140-6736(24)00642-1.

[2] World Health Organization, “Stroke,” WHO Fact Sheets, Geneva, Switzerland, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/stroke (accessed Oct. 28, 2026).

[3] GBD 2021 Stroke Risk Factor Collaborators, “Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021,” Lancet Neurol., vol. 23, no. 10, pp. 973–1003, Oct. 2024, doi: 10.1016/S1474-4422(24)00369-7.

[4] V. L. Feigin et al., “World Stroke Organization: Global Stroke Fact Sheet 2025,” Int. J. Stroke, vol. 20, no. 2, pp. 132–144, 2025, doi: 10.1177/17474930241308142.

[5] Kementerian Kesehatan Republik Indonesia, “Tekan risiko stroke, masyarakat dianjurkan teratur cek kesehatan,” Jakarta, Indonesia, 2022. [Online]. Available: https://kemkes.go.id/id/tekan-risiko-stroke-masyarakat-dianjurkan-teratur-cek-kesehatan (accessed Oct. 28, 2026).

[6] Kementerian Kesehatan Republik Indonesia, “Cegah stroke dengan aktivitas fisik,” Jakarta, Indonesia, 2022. [Online]. Available: https://kemkes.go.id/eng/cegah-stroke-dengan-aktivitas-fisik (accessed Oct. 28, 2026).

[7] Kementerian Kesehatan Republik Indonesia, “Profil Kesehatan Indonesia 2021,” Jakarta, Indonesia, 2023. [Online]. Available: https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2021 (accessed Oct. 28, 2026).

[8] R. Rigual, B. Fuentes, and E. Díez-Tejedor, “Management of acute ischemic stroke,” Med. Clínica (English Ed., vol. 161, no. 11, pp. 485–492, 2023, doi: https://doi.org/10.1016/j.medcle.2023.06.032.

[9] N. Yang and Y. B. Hartanto, “Characteristics and reasons for delayed presentation in acute ischemic stroke: single-centered study in Indonesia,” Egypt. J. Neurol. Psychiatry Neurosurg., vol. 60, no. 1, p. 67, 2024, doi: 10.1186/s41983-024-00838-4.

[10] S. Patil, R. Rossi, D. Jabrah, and K. Doyle, “Detection, diagnosis and treatment of acute ischemic stroke: current and future perspectives,” Front. Med. Technol., vol. 4, art. no. 748949, Jun.

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

2022, doi: 10.3389/fmedt.2022.748949.
[11] T. F. Hasan, H. Hasan, and R. E. Kelley, "Overview of Acute Ischemic Stroke Evaluation and Management," Biomedicines, vol. 9, no. 10, 2021, doi: 10.3390/biomedicines9101486.
[12] A. Alsalahat et al., "Current state and advancements of imaging in acute ischemic stroke: a practical review," Neurol. Sci., vol. 47, 2025, doi: 10.1007/s10072-025-08685-8.
[13] K. Muir et al., "Randomised, controlled Trial of CT perfusion and angiography compared to CT alone in thrombolysis-eligible acute ischaemic stroke patients: The penumbra and recanalisation acute computed tomography in ischaemic stroke evaluation (PRACTISE) trial," Eur. Stroke J., vol. 11, 2026, doi: 10.1093/esj/23969873251372348.
[14] H. Luecking, P. Hoelter, S. Lang, M. Schmidt, F. Eisenhut, and A. Doerfler, "Change your Angle of View," Clin. Neuroradiol., vol. 32, pp. 1109–1115, 2022, doi: 10.1007/s00062-022-01172-z.
[15] A. Sarraj, D. Pujara, and B. Campbell, "Current State of Evidence for Neuroimaging Paradigms in Management of Acute Ischemic Stroke," Ann. Neurol., vol. 95, 2024, doi: 10.1002/ana.26925.
[16] L. M. van Poppel, C. B. L. M. Majoie, H. A. Marquering, and B. J. Emmer, "Associations between early ischemic signs on non-contrast CT and time since acute ischemic stroke onset: A scoping review.," Eur. J. Radiol., vol. 155, p. 110455, Oct. 2022, doi: 10.1016/j.ejrad.2022.110455.
[17] N. Venketasubramanian, F. L. Yudiarto, and D. Tugasworo, "Stroke Burden and Stroke Services in Indonesia," Cerebrovasc. Dis. Extra, vol. 12, no. 1, pp. 53–57, 2022, doi: 10.1159/000524161.
[18] C. Capirossi, A. Laiso, L. Renieri, F. Capasso, and N. Limbucci, "Epidemiology, organization, diagnosis and treatment of acute ischemic stroke," Eur. J. Radiol. Open, vol. 11, p. 100527, 2023, doi: https://doi.org/10.1016/j.ejro.2023.100527.
[19] R. Gauriau et al., "Head CT deep learning model is highly accurate for early infarct estimation," Sci. Rep., vol. 13, no. 1, p. 189, 2023, doi: 10.1038/s41598-023-27496-5.
[20] N. Kaothanthong, K. Atsavasirilert, S. Sarampakhul, P. Chantangphol, D. Songsaeng, and S. Makhanov, "Artificial intelligence for localization of the acute ischemic stroke by non-contrast computed tomography," PLoS One, vol. 17, no. 12 December, pp. 1–20, 2022, doi: 10.1371/journal.pone.0277573.
[21] J. Lu et al., "Identification of early invisible acute ischemic stroke in non-contrast computed tomography using two-stage deep-learning model," Theranostics, vol. 12, no. 13, pp. 5564–5573, 2022, doi: 10.7150/thno.74125.
[22] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J. Big Data, vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.
[23] Q. Xia et al., "A comprehensive review of deep learning for medical image segmentation," Neurocomputing, vol. 613, p. 128740, 2024, doi: 10.1016/j.neucom.2024.128740.
[24] R. Azad et al., "Medical image segmentation review: the success of U-Net," IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 12, pp. 10076–10095, Dec. 2024, doi: 10.1109/TPAMI.2024.3435571.
[25] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," Front. Public Heal., vol. Volume 11, 2023, doi: 10.3389/fpubh.2023.1273253.
[26] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, 2018, doi: 10.1109/TPAMI.2017.2699184.
[27] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: learning to leverage salient regions in medical images," Med. Image Anal., vol. 53, pp. 197–207, Apr. 2019, doi: 10.1016/j.media.2019.01.012.
[28] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, LNCS vol. 11045, 2018, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.
[29] K. Liang et al., "Symmetry-enhanced attention network for acute ischemic infarct segmentation with non-contrast CT images," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, Lecture Notes in Computer Science, vol. 12907, Cham, Switzerland: Springer, 2021, pp. 432–441, doi: 10.1007/978-3-030-87234-2_41.
[30] J. Chen et al., "TransUNet: rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," Med. Image Anal., vol. 97, art. no. 103280, Oct. 2024, doi: 10.1016/j.media.2024.103280.
[31] H. Cao et al., "Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation," in Computer Vision – ECCV 2022 Workshops, LNCS vol. 13803, 2023, pp. 205–218, doi: 10.1007/978-3-031-25066-8_9.
[32] C. Szegedy et al., "Going deeper with convolutions," in 2015 IEEE Conference on

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[33] J. Li, K. V. Sarma, K. C. Ho, A. Gertych, B. S. Knudsen, and C. W. Arnold, "A multi-scale U-Net for semantic segmentation of histological images from radical prostatectomies," AMIA Annu. Symp. Proc., vol. 2017, pp. 1140–1148, Apr. 2018, PMID: 29854182, PMCID: PMC5977596. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5977596 (accessed Oct. 28, 2026).

[34] R. Su, D. Zhang, J. Liu, and C. Cheng, "MSU-Net: Multi-Scale U-Net for 2D Medical Image Segmentation," Front. Genet., vol. Volume 12, 2021, doi: 10.3389/fgene.2021.639930.

[35] W. Zhang, H. Cheng, and J. Gan, "MUNet: a multi-scale U-Net framework for medical image segmentation," in Proc. 2020 Int. Joint Conf. Neural Networks (IJCNN), Glasgow, U.K., Jul. 19–24, 2020, pp. 1–7, doi: 10.1109/IJCNN48605.2020.9206703.

[36] S. Li, J. Zheng, and D. Li, "Precise segmentation of non-enhanced computed tomography in patients with ischemic stroke based on multi-scale U-Net deep network model.," Comput. Methods Programs Biomed., vol. 208, p. 106278, Sep. 2021, doi: 10.1016/j.cmpb.2021.106278.

[37] A. Sandru, M.-I. Georgescu, and R. T. Ionescu, "Feature-level augmentation to improve robustness of deep neural networks to affine transformations," in Computer Vision – ECCV 2022 Workshops, Lecture Notes in Computer Science, vol. 13801, Cham, Switzerland: Springer, 2023, pp. 332–341, doi: 10.1007/978-3-031-25056-9_22.

[38] R. Zhang, B. Zhou, C. Lu, and M. Ma, "The Performance Research of the Data Augmentation Method for Image Classification," Math. Probl. Eng., vol. 2022, no. 1, p. 2964829, Jan. 2022, doi: https://doi.org/10.1155/2022/2964829.

[39] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 8, pp. 10173–10196, 2023, doi: 10.1109/TPAMI.2023.3250241.

[40] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," npj Digit. Med., vol. 5, no. 1, p. 48, 2022, doi: 10.1038/s41746-022-00592-y.

[41] B. Murugesakumar, A. S., A. Arunmanikandan, M. Chithra, and M. Dhasaranjan, "U-Net Convolutional Networks for Real-Time Biomedical Image Segmentation and Anomaly Detection," in 2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), 2025, pp. 1–8. doi: 10.1109/iconstem65670.2025.11374860.

[42] X.-C. Tai, H. Liu, R. H. Chan, and L. Li, "A mathematical explanation of UNet," Math. Found. Comput., vol. 8, no. 5, pp. 874–889, 2025, doi: 10.3934/mfc.2024040.

[43] P. Pan, C. Zhang, J. Sun, and L. Guo, "Multi-scale conv-attention U-Net for medical image segmentation," Sci. Rep., vol. 15, 2025, doi: 10.1038/s41598-025-96101-8.

[44] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," Comput. Med. Imaging Graph., vol. 95, p. 102026, 2022, doi: https://doi.org/10.1016/j.compmedimag.2021.102026.

[45] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks," in Machine Learning in Medical Imaging, LNCS vol. 10541, 2017, pp. 379–387, doi: 10.1007/978-3-319-67389-9_44.

[46] N. Abraham and N. M. Khan, "A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 683–687, doi: 10.1109/ISBI.2019.8759329.

[47] S. Yang and G. Berdine, "Confusion matrix," Southwest Respir. Crit. Care Chronicles, vol. 12, no. 53, pp. 75–79, 2024, doi: 10.12746/swrccc.v12i53.1391.

[48] G. V. Vlăsceanu, N. Tarbă, M.-L. Voncilă, and C. A. Boiangiu, "Selecting the right metric: a detailed study on image segmentation evaluation," BRAIN. Broad Res. Artif. Intell. Neurosci., vol. 15, no. 4, pp. 295–318, 2024, doi: 10.70594/brain/15.4/20.

[49] L. Maier-Hein et al., "Metrics reloaded: recommendations for image analysis validation," Nat. Methods, vol. 21, no. 2, pp. 195–212, 2024, doi: 10.1038/s41592-023-02151-z.

[50] J. Sun et al., "Pathological Asymmetry-Guided Progressive Learning for Acute Ischemic Stroke Infarct Segmentation.," IEEE Trans. Med. Imaging, vol. 43, no. 12, pp. 4146–4160, Dec. 2024, doi: 10.1109/TMI.2024.3414842.

[51] J. Sun et al., "Deep learning for segmenting ischemic stroke infarction in non-contrast CT scans by utilizing asymmetry," Clin. Neuroradiol., vol. 36, no. 1, pp. 129–142, Mar. 2026, doi: 10.1007/s00062-025-01559-8.

[52] L. Cherikbayeva, Z. Melis, V. Berikov, A. Yeleussinov, and A. Ataniyazova, "Deep Ensemble for Ischemic Stroke Segmentation on Non-Contrast CT Images," Mater. Proc., vol. 33,

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

no. 1, p. 4, 2026, doi: 10.3390/materproc2026033004.

Author Biography



Osmalina Nur Rahma completed a B.Eng. in biomedical engineering at Universitas Airlangga in 2013, followed by an M.Sc. in the same field at the University of Indonesia in 2016. She now teaches in the Biomedical Engineering Study Program, Department of Physics, Faculty of Science and Technology, Universitas Airlangga, where she supervises student research and takes part in projects supported by the Airlangga Research Fund under the Research Group Scheme. Her work centres on biomedical instrumentation, signal processing, and rehabilitation engineering, and more recently on the use of deep learning for medical image analysis. For the present study, she was responsible for the conceptualization of the research, for the supervision of the experimental work, and for the review and editing of the manuscript.



Aisyah Widayani is a lecturer in the Radiological Imaging Technology program at the Faculty of Vocational Studies at Universitas Airlangga (UNAIR) in Surabaya, Indonesia. She received her Bachelor of Technology (S.T.) in Biomedical Engineering from Universitas Airlangga. During her time as a student, she co-developed assistive medical prototypes, including an exoskeleton for patients with Brachial Plexus Injury (BPI). She received her Master of Engineering (M.T.) in Electrical and Electronics Engineering from Institut Teknologi Sepuluh Nopember (ITS), graduating Cum Laude from the Laboratory of Intelligent Electronics. Her research focus & expertise are computer engineering and healthcare, Deep Learning & Computer Vision, Radiology Innovations and medical Image Processing.



Nimas Maharani Putri received the S.T. degree in biomedical engineering from Universitas Airlangga, Indonesia, in 2026. Her research interests are biomedical image processing, signal processing, and artificial intelligence, particularly the use of convolutional neural networks for the analysis of medical images. For the present study, she collected and curated the non-contrast CT data, carried out the preprocessing and the patient-level partitioning of the studies, prepared the experimental dataset, and ran the training and evaluation experiments reported here. She intends to continue working on deep learning methods for medical image segmentation and on their evaluation under clinically realistic protocols, with an emphasis on reproducibility and on honest reporting of generalization.



Rifai Chai received the B. Eng. degree from Krida Wacana Christian University (UKRIDA), Jakarta, Indonesia, in 2000 and the Ph.D. in Biomedical Engineering from the University of Technology Sydney (UTS), Sydney, Australia, in 2014. From 2000 to 2011, he worked as an electronic product development engineer with companies in Indonesia and Australia. He is a senior lecturer at the Faculty of Science, Engineering & Technology, Swinburne University of Technology, Melbourne, Australia. His research interests are brain-computer interfaces, rehabilitation, medical technologies, and artificial intelligence. He is a senior member of IEEE and serves as Associate Editor for Electronics Letters. For the present study, he contributed the critical review and the intellectual revision of the manuscript.

Manuscript received 5 August 2026; Revised 14 October 2026; Accepted 22 October 2026; Available online 31 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1980
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).