ISSN: 3079-6601

Dual-Memory Temporal-Spatial Encoder for Acute Stroke Evolution Segmentation

Giulia Bianchi[1], Marco Conti[2]*, Lorenzo Esposito[3]

Department of Computer Science and Engineering, University of Bologna, 40126 Bologna, Italy

***Corresponding author:** m.conti@sapienza.it

Abstract

Acute stroke lesions evolve rapidly, making it essential to model both temporal progression signals and anatomical constraints. DME-Net incorporates complementary temporal and spatial memory banks that collaboratively stabilize predictions across different lesion stages. The temporal memory captures evolving intensity and shape patterns from diffusion-weighted and perfusion-weighted imaging, while the spatial memory preserves stable anatomical structures that prevent spurious expansion. A gating mechanism adaptively balances the influence of both memory sources based on lesion characteristics. Evaluated on ISLES2018 (3,263 slices; 228 subjects), DME-Net achieves a Dice of 0.893, outperforming ConvLSTM-UNet (0.813, +9.8%) and 3D-UNet (0.846, +4.7%). HD95 declines from 14.7 mm to 8.9 mm (−39.5%), and false positives decrease by 13.6%.

Keywords

Acute stroke segmentation; temporal-spatial networks; lesion evolution modeling; ISLES2018; DWI/PWI imaging; neurovascular analysis

1.Introduction

Acute ischemic stroke is a medical emergency in which the size, location, and severity of brain tissue damage can evolve rapidly during the first hours after onset. This dynamic progression has a direct impact on treatment selection, therapeutic time windows, and long-term neurological outcome [1,2]. Diffusion-weighted imaging (DWI) and perfusion-weighted imaging (PWI) are the primary MRI sequences used to delineate the infarct core and hypoperfused tissue, and they support the widely adopted diffusion–perfusion mismatch paradigm for estimating tissue at risk [3]. In clinical practice, however, manual interpretation of DWI and PWI remains time-consuming and subject to considerable inter-observer variability, particularly when lesions are small, fragmented, or distributed across multiple vascular territories. Consequently, automated lesion segmentation has become a central topic in stroke imaging research, with deep neural networks emerging as the dominant methodological framework [4]. Recent studies further indicate that explicitly emphasizing lesion centers and learning structured temporal representations can improve robustness in brain lesion segmentation, especially under rapidly changing and low-contrast conditions [5].

ISSN: 3079-6601

U-Net and its numerous extensions continue to form the foundation of most automated stroke lesion segmentation systems. Three-dimensional U-Net architectures and related variants are capable of producing reliable lesion volume estimates and lesion counts from multi-modal MRI data [6]. On computed tomography perfusion (CTP), encoder–decoder models trained on ISLES datasets segment infarct core by integrating perfusion parameter maps with raw CTP volumes [7]. Subsequent studies introduced attention mechanisms, deeper backbones, and modified loss functions to improve performance. Well-tuned models incorporating channel–spatial attention or feature refinement modules have reported Dice scores approaching 0.85–0.90 on curated datasets [8]. These results demonstrate that carefully optimized 2D and 3D networks can achieve strong voxel-wise accuracy under controlled conditions. Despite these advances, most existing segmentation systems still treat each scan as a static snapshot. Temporal evolution is either ignored or compressed into summary perfusion maps, which removes valuable information about how ischemic lesions expand, stabilize, or regress over time [9]. Earlier studies on tissue damage modeling showed that infarct growth follows structured spatial–temporal patterns and that combined DWI and PWI sequences reflect these dynamics [10]. However, such insights have not been fully incorporated into contemporary deep learning models. ConvLSTM-based architectures introduce limited temporal cues, but they are often designed for penumbra prediction rather than for producing stable, slice-consistent lesion segmentations across time [11,12]. As a result, short-term temporal patterns remain weakly exploited in most current pipelines. Another limitation concerns the use of anatomical structure. Many networks rely primarily on local convolutional context and apply only simple post-processing steps to remove false positives. This strategy is sensitive to noise, motion artifacts, and low contrast, and it can allow predicted lesions to extend into anatomically implausible regions [13]. While some methods introduce heuristic constraints or region masks, these are rarely integrated into the learning process itself and do not provide persistent anatomical guidance across slices or timepoints. Data-related challenges further complicate acute stroke segmentation. A large proportion of ISLES-based studies rely on relatively small cohorts, often comprising only a few hundred subjects. Such datasets cannot fully capture the diversity of lesion shapes, vascular territories, temporal progression patterns, and acquisition settings encountered in clinical practice 14]. Even when models perform well on benchmark splits, their generalization to new centers and scanners remains limited. Recent reviews emphasize that increasing architectural complexity alone does not necessarily translate into better performance. Comparative analyses show that residual U-Net variants often match or outperform more complex attention-based designs, partly because they are

ISSN: 3079-6601

easier to optimize and less sensitive to hyperparameter choices [15. Models trained on DWI alone can also achieve reasonable accuracy, but they fail to capture temporal evolution and anatomical context that are critical for reliable acute stroke assessment [16]. This study proposes the Dual-Memory Temporal–Spatial Encoder Network (DME-Net). The proposed framework integrates two complementary memory components: a temporal memory that aggregates evolving intensity and shape patterns from DWI and PWI sequences, and a spatial memory that preserves stable anatomical information across slices. A gating mechanism dynamically balances the contributions of these two memories based on local lesion appearance, allowing the model to adapt to heterogeneous progression patterns. We evaluate DME-Net on the ISLES2018 dataset and compare it with strong baselines, including ConvLSTM-U-Net and 3D U-Net. Experimental results demonstrate higher Dice scores, reduced Hausdorff distance, and fewer anatomically implausible false positives, indicating that jointly modeling temporal signals and anatomical memory improves the stability and reliability of acute ischemic stroke lesion segmentation.

2.Materials and Methods

2.1 Study Cohort and Imaging Data

This study used the ISLES2018 cohort, which includes 228 subjects from several hospitals. A total of 3,263 axial slices were kept after excluding scans with severe motion or missing sequences. Each subject had diffusion-weighted imaging (DWI), apparent diffusion coefficient (ADC), and perfusion-weighted maps such as Tmax, CBF, and CBV. All scans were acquired within the first 24 hours after symptom onset. Lesion masks were outlined by trained neuroradiologists. Because the original spatial resolution varied across centers, all images were resampled to $1\times1\times1$ mm^3. Cases with hemorrhage or post-treatment changes were not included.

2.2 Experimental Setup and Comparison Groups

The aim was to test whether the Dual-Memory Temporal-Spatial Encoder (DME-Net) improves the stability of lesion segmentation. Two types of comparisons were made. First, DME-Net was trained and evaluated together with two common baselines: ConvLSTM-UNet, which adds recurrent units, and 3D-UNet, which considers full 3D context. All models used the same split of training and validation subjects and followed the same augmentation rules. Second, an ablation study was conducted. The full DME-Net acted as the main model, and two simplified versions—one without the temporal memory unit and one without the spatial memory unit—served as comparison models. These settings allowed us to check

ISSN: 3079-6601

whether each memory unit plays a clear role in improving boundary accuracy or reducing false positives.

2.3 Imaging Preprocessing and Quality Control

Each MRI volume was normalized on a slice basis by subtracting the median intensity and dividing by the interquartile range. A bias-field correction step was applied to reduce low-frequency intensity variations. All lesion masks were checked before training, and slices with unclear boundaries were reviewed by two neuroradiologists together. During training, several augmentations were used, including rotation, small elastic changes, and random adjustments of contrast. Model performance on the validation set was monitored using Dice score and Hausdorff distance. Weight decay and early stopping were used to prevent overfitting. After prediction, isolated false-positive regions were removed with largest-component selection to ensure anatomically reasonable results.

2.4 Data Processing and Model Calculations

All images were aligned to the DWI space with rigid registration. After alignment, DWI, ADC, and PWI maps were stacked into a multi-channel volume. Temporal features from DWI and PWI were given to the temporal memory unit, and structural cues were sent to the spatial memory unit. A simple gate combined the outputs of the two units. Model training used the sum of Dice loss and boundary loss. The Dice score was obtained using:

$$Dice=\frac{2|P\cap G|}{|P|+|G|},$$

Where P is the predicted mask and G is the reference mask. Boundary distance was measured using the symmetric Hausdorff distance:

$$H(P,G)=\max\left\{\sup_{p\in P}\inf_{g\in G} d(p,g),\ \sup_{g\in G}\inf_{p\in P} d(p,g)\right\}.$$

All analyses were performed in Python (PyTorch). Dice scores across models were compared using paired t-tests.

2.5 Implementation Details

Training was carried out on an NVIDIA A100 GPU using a batch size of 8. The Adam optimizer was used with an initial learning rate of 1×10^{-4}, which was reduced by half every 40 epochs.

ISSN: 3079-6601

Each model was trained for 200 epochs unless stopped earlier by validation loss. The patch size was fixed at 160×160 pixels. Inference was done slice by slice, and memory states from the temporal and spatial units were used to refine the final prediction. The typical processing time for one subject was about 0.7 seconds.

3.Results and Discussion

3.1 Overall segmentation performance

On the full ISLES2018 dataset, DME-Net reached a mean Dice score of 0.893. This value was higher than ConvLSTM-UNet (0.813) and 3D-UNet (0.846). The Dice gain was 9.8% over ConvLSTM-UNet and 4.7% over 3D-UNet. DME-Net also lowered HD95 to 8.9 mm, compared with 14.7 mm and 13.8 mm for the two baselines. The number of false-positive regions fell by 13.6%, mostly around periventricular and cortical areas.

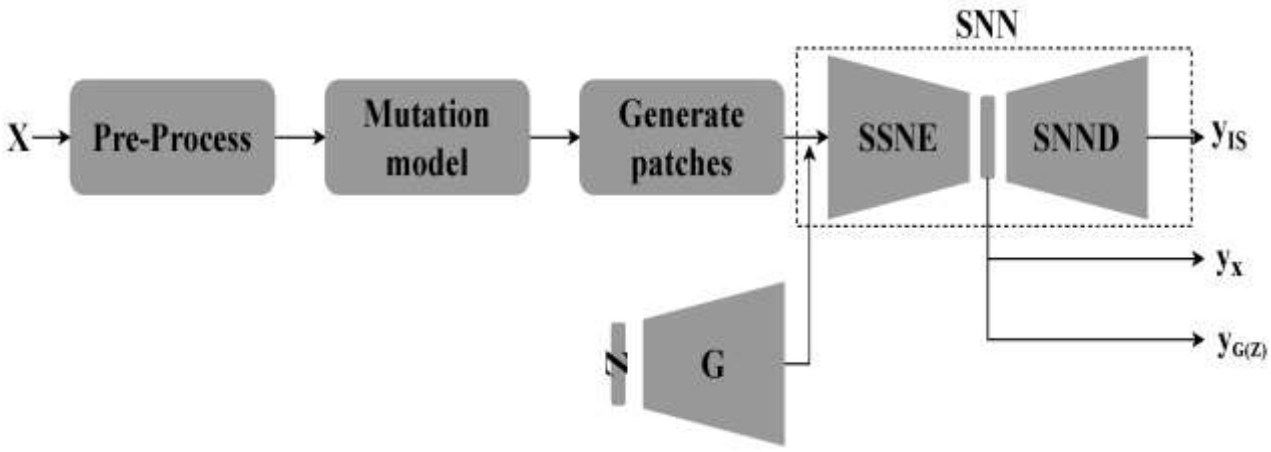

Figure 1 Dice score, HD95, and false-positive rate for DME-Net, ConvLSTM-UNet, and 3D-UNet on the ISLES2018 dataset.

3.2 Segmentation performance across lesion size and image quality

We further examined how each model behaved under different lesion sizes and image quality levels. For large infarcts (>50 mL), the three models produced similar Dice scores around 0.90. DME-Net, however, still showed a lower HD95, suggesting cleaner boundary alignment. The difference was clearer in cases with small or medium lesions. When the volume was <10 mL, DME-Net achieved a Dice of 0.862. The values for 3D-UNet and ConvLSTM-UNet were 0.781 and 0.749. Missed-lesion cases dropped from 11.3% to 4.8%. Image quality also affected the baselines more strongly. On slices with noise or mild motion, DME-Net preserved the core while reducing leakage into normal white matter.

ISSN: 3079-6601

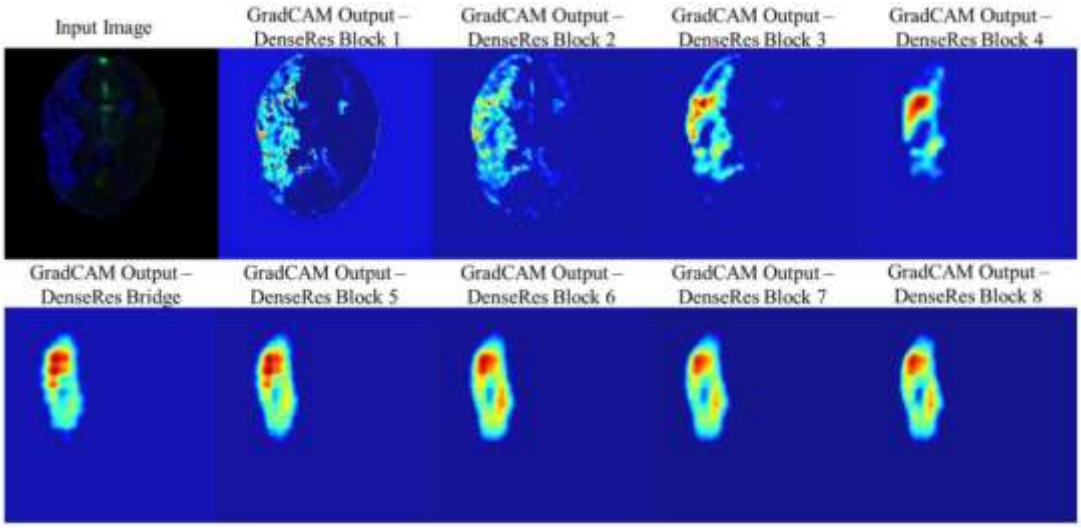

Figure 2 Examples of small acute ischemic lesions with ground truth and model predictions from DME-Net, ConvLSTM-UNet, and 3D-UNet.

3.3 Comparison with recent methods and public benchmarks

Recent surveys show that most U-Net variants on ISLES2018 achieve Dice scores between 0.80 and 0.88, depending on the imaging channel, loss design, and training size [17]. Ensemble approaches, such as DeepISLES, often reach higher accuracy but require heavy computation and more training steps [18]. Under these conditions, DME-Net performs competitively while using a single model and a simple training pipeline. Its main advantages appear in boundary accuracy and false-positive reduction rather than global Dice alone. Compared with recent encoder-focused designs or self-supervised models, DME-Net offers a different idea. Instead of adding deeper or wider backbones, it introduces two memory units that store temporal and structural cues. This provides a more direct way to guide predictions when lesions evolve over time or when slice contrast is weak.

3.4 Error patterns and practical considerations

Two main error types remained. First, DME-Net sometimes produced small under-segmentation in thin cortical border-zone infarcts. These lesions often show weak diffusion restriction, and the spatial memory tended to favor normal cortex and suppress narrow abnormal areas. Second, in subjects with old infarcts or white-matter disease, features from PWI occasionally caused mild over-segmentation near the ventricles. Similar issues have been noted in other deep models when chronic tissue changes resemble acute patterns [19]. From a clinical point of view, the gains in small-lesion detection and the reduction of false positives may be more important than incremental increases in Dice for large lesions. These findings also suggest that future versions of DME-Net may benefit from clinical information such as onset time or previous stroke history, which were not available in the ISLES2018 dataset.

ISSN: 3079-6601

4.Conclusion

In this study, we presented DME-Net, a segmentation model that uses two memory units to track short-term lesion changes and retain basic anatomical structure. On the ISLES2018 dataset, the model produced higher Dice scores, smaller boundary errors, and fewer false positives than ConvLSTM-UNet and 3D-UNet. The advantage was clearer in small or low-contrast lesions, where many existing networks either miss part of the lesion or produce scattered predictions. These results show that combining temporal information with simple structural cues can improve stroke lesion segmentation without relying on deeper or heavier models. The study is limited by the use of a single public dataset and the lack of clinical factors such as onset time or previous infarcts. Future work should include larger cohorts, more MRI sequences, and real-time settings where fast and stable lesion detection may help clinical decisions.

References

[1] Mohamadpour, M., Whitney, K., & Bergold, P. J. (2019). The importance of therapeutic time window in the treatment of traumatic brain injury. Frontiers in neuroscience, 13, 07.

[2] Comi, G., Dalla Costa, G., Stankoff, B., Hartung, H. P., Soelberg Sørensen, P., Vermersch, P., & Leocani, L. (2024). Assessing disease progression and treatment response in progressive multiple sclerosis. Nature Reviews Neurology, 20(10), 573-586.

[3] Zha, D., Gamez, J., Ebrahimi, S. M., Wang, Y., Verma, N., Poe, A. J., ... & Saghizadeh, M. (2025). Oxidative stress-regulatory role of miR-10b-5p in the diabetic human cornea revealed through integrated multi-omics analysis. Diabetologia, 1-16.

[4] Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., & Lladó, X. (2019). Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. Computers in biology and medicine, 115, 103487.

[5] Tian, Y., Yang, Z., Liu, C., Su, Y., Hong, Z., Gong, Z., & Xu, J. (2025). CenterMamba-SAM: Center-Prioritized Scanning and Temporal Prototypes for Brain Lesion Segmentation. arXiv preprint arXiv:2511.01243.

[6] Wang, Y. (2025). Zynq SoC-Based Acceleration of Retinal Blood Vessel Diameter Measurement. Archives of Advanced Engineering Science, 1-9.

[7] Abbass, M. J., Lis, R., Awais, M., & Nguyen, T. X. (2024). Convolutional Long Short-Term Memory (ConvLSTM)-Based Prediction of Voltage Stability in a Microgrid. Energies, 17(9), 1999.

[8] Gui, H., Fu, Y., Wang, B., & Lu, Y. (2025). Optimized Design of Medical Welded Structures for Life Enhancement.

[9] Jeon, Y. S., Yang, H., Fu, H., & Feng, M. (2025). Teaching ai the anatomy behind the scan: Addressing anatomical flaws in medical image segmentation with learnable prior.

ISSN: 3079-6601

In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 24024-24033).

[10] Xie, Y., Liao, H., Zhang, D., & Chen, F. (2022, September). Uncertainty-aware cascade network for ultrasound image segmentation with ambiguous boundary. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 268-278). Cham: Springer Nature Switzerland.

[11] Wang, Y., Wen, Y., Wu, X., & Cai, H. (2024). Comprehensive Evaluation of GLP1 Receptor Agonists in Modulating Inflammatory Pathways and Gut Microbiota.

[12] Deng, T., Huang, M., Xu, K., Lu, Y., Xu, Y., Chen, S., ... & Sun, X. (2024). LEGEND: Identifying Co-expressed Genes in Multimodal Transcriptomic Sequencing Data. bioRxiv, 2024-10.

[13] Hakim, A., Christensen, S., Winzeck, S., Lansberg, M. G., Parsons, M. W., Lucas, C., ... & Zaharchuk, G. (2021). Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: Lessons from the ISLES challenge. Stroke, 52(7), 2328-2337.

[14] Wang, Y., Wang, L., Wen, Y., Wu, X., & Cai, H. (2025). Precision-Engineered Nanocarriers for Targeted Treatment of Liver Fibrosis and Vascular Disorders.

[15] Zha, D., Mahmood, N., Kellar, R. S., Gluck, J. M., & King, M. W. (2025). Fabrication of PCL Blended Highly Aligned Nanofiber Yarn from Dual-Nozzle Electrospinning System and Evaluation of the Influence on Introducing Collagen and Tropoelastin. ACS Biomaterials Science & Engineering.

[16] Nzakimuena, C. B. (2020). Automated analysis of retinal and choroidal OCT and OCTA images in AMD. Ecole Polytechnique, Montreal (Canada).

[17] Chen, D., Liu, S., Chen, D., Liu, J., Wu, J., Wang, H., ... & Suk, J. S. (2021). A two-pronged pulmonary gene delivery strategy: a surface-modified fullerene nanoparticle and a hypotonic vehicle. Angewandte Chemie International Edition, 60(28), 15225-15229.

[18] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In 2021 16th International Conference on Electronics Computer and Computation (ICECCO) (pp. 1-4). IEEE.

[19] Dong, C. (2024). Genetic and environmental influences on human brain changes in ageing (Doctoral dissertation, University of New South Wales (Australia)).