Scientific Reports

https://doi.org/10.1038/s41598-026-40751-9

Article in Press

Classification of dysphagia severity after lateral medullary infarction with deep learning

Received: 30 June 2025

Accepted: 16 February 2026

Published online: 19 February 2026

Cite this article as: Lee T., Kim B.H., Nam K. *et al.* Classification of dysphagia severity after lateral medullary infarction with deep learning. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-40751-9

Taeheon Lee, Bo Hae Kim, Kihwan Nam & Jin-Woo Park

Classification of Dysphagia Severity after Lateral Medullary Infarction with Deep Learning

Taeheon Lee[1], Bo Hae Kim[2], Kihwan Nam[3,*,+], Jin-Woo Park[1,*,+]

1. Department of Physical Medicine and Rehabilitation, Dongguk University Ilsan Hospital, College of Medicine, 27 Dongguk-ro, Ilsandong-gu, Goyang 10326, Republic of Korea
2. Department of Otorhinolaryngology-Head and Neck Surgery, Dongguk University Ilsan Hospital, College of Medicine, 27 Dongguk-ro, Ilsandong-gu, Goyang 10326, Republic of Korea
3. Graduate School of Management of Technology, Korea University, Seoul 02841, Republic of Korea

* Correspondence: namkh@korea.ac.kr (K.H.N); jinwoo.park.md@gmail.com (J.-W.P.)

+ These authors contributed equally to this work as corresponding authors

Abstract

Dysphagia is a common and debilitating complication in patients with lateral medullary infarction (LMI), affecting up to 100% of cases and significantly impairing quality of life. Accurate classification of early dysphagia severity is essential for timely intervention and personalized rehabilitation planning. This study aimed to develop and validate a deep learning algorithm using acute-phase diffusion-weighted MRI to classify dysphagia severity in LMI patients. A retrospective cohort of 163 patients with confirmed acute LMI was analyzed. Dysphagia severity was determined by videofluoroscopic swallowing studies (VFSS), categorizing patients into severe and non-severe groups. Lesion regions were manually labeled and preprocessed for model training. Transformer-based deep learning architecture, the Hierarchical Vision Transformer (Hier-ViT), was employed due to its capacity to model spatial hierarchies and global image context. The model achieved an accuracy of 0.85, with a precision of 0.70, recall of 0.75, F1-score of 0.72, and an area under the ROC curve (AUC) of 0.69. These findings suggest that Hier-ViT can effectively classify dysphagia severity in LMI patients using early MRI, offering a potential tool for early risk stratification. While the model shows a high accuracy, the modest AUC suggests that further refinement and multi-modal integration are necessary to improve its discriminative power in imbalanced clinical datasets.

Keywords:

Lateral medullary infarction, Dysphagia, Classification, Deep learning, Magnetic Resonance Imaging

1. Introduction

Dysphagia is a commonly documented complication after stroke, but its reported prevalence varies widely, ranging between 19% and 81% [1]. Dysphagia can significantly impair patients' quality of life, leading to complications such as malnutrition, dehydration, aspiration pneumonia, social isolation, depression, and anxiety [2,3].

Among stroke types, lateral medullary infarction (LMI) is caused by ischemia in the lateral part of the medulla oblongata [4]. LMI is characterized by a range of neurological symptoms, including gaze-induced nystagmus [5], Horner's syndrome [6], ipsilateral ataxia [7]. Importantly, dysphagia has been reported in 51% to 100% of patients with LMI [8,9], and it tends to be more severe and prolonged in LMI patients compared to those with hemispheric stroke patients [10]. This is attributed to the anatomical characteristics of the medulla, the location of the central pattern generators (CPGs) responsible for swallowing. CPGs are generally defined as neural networks capable of generating central commands that govern stereotyped, rhythmic motor behaviors such as swallowing [11,12]. In the medulla, the CPG is composed primarily of the nucleus of the tractus solitarius (NTS) and the nucleus ambiguus (NA). Neurons in the NTS function as interneurons that coordinate and program the sequential swallowing motor pattern, while neurons in the NA primarily serve as motoneurons innervating the pharyngolaryngeal muscles and the esophagus [13].

Initial dysphagia is considered a key factor associated with poor outcomes following LMI [14]. Although most patients with dysphagia experience mild symptoms and recover quickly [15,16], some patients with severe dysphagia require tube feeding for several months or even years [12,17,18]. Therefore, early and accurate classification of dysphagia severity is clinically important for developing appropriate therapeutic strategies. Several studies have attempted to identify factors associated with dysphagia prognosis in LMI patients [18,19] A previous lesion-symptom mapping study demonstrated that posterolateral involvement in both the upper and lower medulla, as visualized on diffusion-weighted imaging, was significantly associated with severe dysphagia, typically characterized by decreased pharyngeal constriction and the apparent absence of any evidence of esophageal passage in videofluoroscopic swallowing study (VFSS) findings, as well as the initial requirement for enteral tube feeding [18,20]. These findings suggest that the anatomic factor related to the extent and vertical distribution of the lesion play a critical role in determining swallowing impairment. This underscores the importance of early and accurate lesion localization in predicting dysphagia severity and guiding individualized rehabilitation planning.

Recently, artificial intelligence (AI) based on deep learning has been expanding its applications in various medical fields to enhance the diagnosis and treatment of diseases [21]. Deep learning methods are a type of representation learning that leverage multiple levels of abstraction. These methods consist of simple but non-linear modules that transform the representation at one level (starting with raw input) into a higher, more abstract level of representation [22].

Deep learning techniques have been widely applied across various organ systems, including the kidney, prostate, and spine [23]. Notably, brain image analysis has emerged as one of the most extensively studied areas in the field of medical imaging. A variety of tasks have been successfully addressed, such as the staging and early diagnosis of Alzheimer's disease using multimodal magnetic resonance imaging (MRI) data [24], automated segmentation of brain tumors including glioblastoma and meningioma from heterogeneous clinical MRI scans [25,26], and lesion detection in patients with multiple sclerosis [27]. Additional applications include skull stripping for brain extraction [28], and classification of brain functional connectomes [29].

Given that anatomical lesion patterns in the medulla have been shown to significantly influence the severity of dysphagia in patients with LMI, accurate lesion localization is of particular clinical relevance. At the same time, recent advances in deep learning have led to substantial progress in brain imaging analysis, with applications ranging from disease classification to lesion segmentation and functional mapping. These developments suggest that integrating anatomical insights with data-driven modeling may offer a powerful approach to classifying dysphagia severity.

In this study, we aim to develop and validate a Transformer-based deep learning algorithm to automatically analyze the location of initial MRI lesions in patients with LMI and classify the severity of dysphagia at the subacute phase. By leveraging the ability of Transformers to capture both local and global patterns, we expect to enhance the accuracy of severity classification and facilitate the development of more effective rehabilitation strategies.

2. Materials and methods

2.1. Study design

We performed a retrospective study analyzing clinical and imaging data of patients who were admitted to Dongguk University Ilsan Hospital with acute LMI from September 2005 to the

July 2024. Participants satisfied the following inclusion criteria: (1) first-ever onset of acute stroke, (2) patients aged 20 years or older and (3) diagnosis of lateral medullary infarction confirmed by MRI. Patients with other structural brain disorders, including neurodegenerative or neuromuscular conditions that could independently influence swallowing function, were excluded from the study. The study was approved by the Institutional Review Board of Dongguk University Ilsan Hospital (No. 2024-07-004). All procedures were performed in compliance with the relevant guidelines and regulations.

2.2. Data collection and outcome measures

All stroke patients underwent brain MRI within 24 hours of admission using a 1.5-T MR machine (MAGNETOM-Avanto, Siemens, Erlangen, Germany). Diffusion-weighted MRI (DWI) parameters were as follows: b-values of 0 and 1000 s/mm², repetition time of 5400 ms, echo time of 77 ms, field of view of 220 × 220 mm, slice thickness of 3.0 mm, and an interslice gap of 0.3 mm. Following established protocols [20], three DWI slices were extracted from the lower, middle, and upper medulla levels based on the MNI brain template for analysis.

VFSS was conducted at a mean of 14 days following stroke onset using a fluoroscopic system (Sonialvision-100, Shimadzu Corporation, Kyoto, Japan). All examinations adhered to a standardized protocol, which included the following: (1) patients maintained an upright seated position throughout the procedure; (2) the VFSS videos were recorded in the lateral view; (3) each subject was given 5 mL of diluted barium solution (35% w/v), curd-type yogurt, and mashed boiled pumpkin, each administered twice; and (4) the protocol was modified as needed according to the patient's clinical status and level of cooperation.

Dysphagia severity was determined according to VFSS findings [20]. Severe dysphagia was defined by the presence of diminished pharyngeal contraction and a clear lack of observable esophageal passage. Patients who did not exhibit these features were categorized as having non-severe dysphagia. This specific dichotomization, while distinct from standardized scales like the Penetration-Aspiration Scale (PAS), was chosen because pharyngeal paralysis and the resulting inability to achieve esophageal passage are the primary clinical hallmarks of severe dysphagia in LMI patients, directly necessitating enteral tube feeding. This criteria-based approach ensures that the classification is grounded in the functional pathophysiology most relevant to early clinical management in the acute phase. The classification of severe versus non-severe dysphagia based on VFSS was performed by a single physiatrist (J.W.P.) experienced in stroke rehabilitation and dysphagia assessment, following the predefined

criteria (diminished pharyngeal contraction and lack of esophageal passage).

2.3. Data preprocessing

Three brain MRI images were selected based on consistent anatomical landmarks to ensure comparability across subjects and were input into the model as independent samples. These three slices (lower, middle, and upper medulla levels based on the MNI template) were chosen to capture the vertical extent of the lesion, which is a key anatomical factor associated with dysphagia severity, as suggested by prior lesion-symptom mapping studies [20]. While these three slices do not guarantee capturing the entire extent of every lesion, the selection strategy prioritized slices that align with known functionally critical regions for swallowing CPGs and allowed for a standardized input across subjects. Each patient was labeled as having either severe or non-severe dysphagia based on VFSS results, creating a binary classification task. All images were resized to a standardized resolution and normalized for pixel intensity. The dataset was split into training (70%) and testing (30%) sets to evaluate model performance. No additional clinical variables were included at this stage, and the analysis focused solely on the spatial imaging features from early post-stroke MRI scans. This preprocessing ensured that the model learned discriminative patterns directly from medullary lesion characteristics while controlling for variations in image size and intensity.

2.4. Deep learning models

This study employed a Hierarchical Vision Transformer (Hier-ViT) model to classify dysphagia severity based on early brain MRI data. Hier-ViT is a Transformer-based vision model that constructs multi-level representations by hierarchically merging non-overlapping image patches. Each input image was divided into fixed-size patches and linearly embedded, with positional encoding added to preserve spatial information. These embeddings were then passed through a series of Transformer encoder layers that progressively reduced spatial resolution while increasing feature abstraction, allowing for both local and global contextual learning. The model was trained using the preprocessed medulla MRI slices and labeled dysphagia severity, with performance evaluated based on classification metrics such as accuracy, precision, recall, and AUC. A schematic overview of the entire deep learning pipeline—from diffusion-weighted image preprocessing to patch embedding, hierarchical feature extraction, and severity classification—is presented in Figure 1.

2.5. Statistical analysis

Statistical analysis was performed using Python 3.8 on an Ubuntu 22.04 operating system,

utilizing the Scikit-learn (version 1.5) and SciPy (version 1.11.4) libraries. Model performance was evaluated using receiver operating characteristic (ROC) curve analysis, and the area under the curve (AUC) was calculated as a primary indicator of discriminative ability. Confidence intervals for the mean AUC were estimated using the bias-corrected and accelerated bootstrap method.

3. Results

3.1. Patient characteristics

A total of 163 patients met the inclusion criteria for this study. The mean age of the participants was 60.5 ± 13.0 years. Among them, 44 patients (27.0%) were classified as having severe dysphagia, including 33 males (75.0%) and 11 females (25.0%), with a mean age of 60.0 ± 14.5 years. The remaining 119 patients (73.0%) were classified as having non-severe dysphagia, comprising 71 males (59.7%) and 48 females (40.3%), with a mean age of 60.7 ± 12.6 years. VFSS was conducted at a mean of 12.4 ± 3.4 days from stroke onset. Table 1 summarizes the baseline demographic characteristics stratified by dysphagia severity.

3.2. Model performance

The Hier-ViT model achieved an accuracy of 0.85, precision of 0.70, recall of 0.75, and an F1-score of 0.72 in classifying dysphagia severity. The area under the receiver operating characteristic curve (AUC) was 0.69, indicating fair overall classification performance. The performance metrics are summarized in Table 2, and the classifier's ability to distinguish between severity groups is demonstrated in Fig. 2 (Confusion matrix and ROC curve).

4. Discussion

This study investigated the performance of a deep learning model based on the Hier-ViT in classifying the severity of dysphagia following LMI using early brain MRI findings.

While the Hier-ViT model achieved a high classification accuracy of 0.85; however, the AUC was relatively modest at 0.69. This discrepancy is likely attributable to the class imbalance within our cohort, where 73% of patients were in the non-severe group. In such instances, while the model correctly identifies a high percentage of overall cases, the AUC provides a more

nuanced reflection of its ability to distinguish between the two classes across all possible thresholds. The interpretation of "fair" performance is based on established general guidelines for AUC interpretation where values between 0.7 and 0.8 are typically considered "fair" discriminative ability. The precision of 0.70 means that when the model predicts severe dysphagia, it is correct 70% of the time—an important factor in minimizing false-positive predictions and avoiding unnecessary interventions. The recall of 0.75 shows that the model successfully identifies 76% of patients who truly have severe dysphagia, which is essential for ensuring that patients at high risk are not overlooked. The balance between precision and recall is demonstrated by the F1-score of 0.72, which indicates moderate yet clinically meaningful performance in terms of both identifying true cases and minimizing misclassification.

From an anatomical perspective, this performance is particularly meaningful in light of previous findings indicating that anatomical lesion patterns within the medulla significantly influence the severity of dysphagia in patients with LMI. Specifically, lesion-symptom mapping studies have demonstrated that posterolateral involvement of both the upper and lower medulla, as identified on diffusion-weighted imaging, is strongly associated with severe dysphagia requiring enteral feeding [20]. Given that infarctions in the medullary region typically involve extremely small anatomical territories that are often indistinguishable by visual inspection alone, the model's ability to achieve this level of classification accuracy suggests that it may capture subtle spatial patterns in the lesion that are correlated with dysphagia severity.

Recent advances in deep learning have enabled automated analysis of brain imaging data across various domains, including the segmentation of brain tumors, the staging of Alzheimer's disease, and the classification of neurological disorders [23]. In this context, our study applies a Hierarchical Vision Transformer model to early brain MRI data to classify dysphagia severity, integrating anatomical lesion characteristics with data-driven categorization of dysphagia severity.

Existing clinical tools, such as VFSS or FEES (fiberoptic endoscopic evaluation of swallowing), are the current gold standards for stratifying dysphagia severity. However, access to these studies can be delayed or unavailable in acute settings, especially in centers with limited resources. Looking ahead, this model is proposed as a complementary triage tool rather than a replacement for gold-standard assessments such as VFSS or FEES. In clinical practice, the model could provide immediate risk stratification upon the completion of an initial MRI, allowing clinicians to prioritize high-risk patients for early intervention and appropriate

nutritional support when specialized swallowing studies are delayed or unavailable.

The present study applies a Hier-ViT architecture to early brain MRI data to classify dysphagia severity. In this approach, the input image is divided into patches, which are then linearly projected and processed by the Transformer encoder [30]. Although CNN-based methods have demonstrated commendable performance, their ability to model long-range dependencies is inherently limited due to the localized nature of convolution operations. In contrast, Hier-ViT effectively addresses this limitation by capturing both local and global contextual relationships through its hierarchical self-attention mechanism [31]. Hier-ViT has demonstrated effectiveness in various medical imaging tasks and has shown superior performance compared to conventional CNN-based approaches [32]. The implementation of a Hier-ViT model in this study highlights its potential utility in medical imaging domains that involve anatomically small but functionally critical regions. This architecture is particularly advantageous when analyzing the small, anatomically complex medulla, as it incorporates global contextual information while preserving spatial hierarchies. Despite inherent limitations, including modest dataset size, the model demonstrated stable classification performance, suggesting that transformer architectures may offer robustness in complex neuroanatomical classification tasks. This reinforces the potential of early imaging-based models in supporting clinical decisions.

This study has several limitations that should be acknowledged. First, the sample was drawn from a single-center cohort, which may limit the generalizability of the model's performance to broader populations with different demographic or clinical characteristics. Second, external validation was not performed, restricting the applicability of the model to other clinical settings and raising concerns about potential overfitting to institutional-specific data. Third, the reliability and consistency of the dysphagia severity labeling were not assessed through inter-rater agreement, as only a single rater was involved in the classification, which is a potential limitation of the ground truth data used for training. Fourth, the follow-up period was confined to the early phase of dysphagia recovery, thereby limiting insight into long-term functional outcomes or true prognosis. Lastly, the model was developed using only MRI and VFSS data, without incorporating additional clinical variables such as medical history or neurological assessments, which may have further enhanced the model's predictive capacity through multi-modal integration.

To further enhance translational relevance, future research should focus on developing multi-modal models that integrate imaging data with key clinical variables. Furthermore, external validation using multi-center datasets will be essential to ensure the model's

generalizability across different institutional protocols and patient populations. Incorporating additional clinical variables—such as neurological assessments, comorbidities, or laboratory findings—as well as multi-modal imaging may further improve predictive performance. Moreover, extending the follow-up period to capture long-term swallowing outcomes will be essential to fully evaluate the clinical relevance of such models for true prognostic applications. As deep learning approaches continue to evolve, efforts should also be made to streamline model architecture for real-time clinical implementation and to assess their utility in guiding personalized rehabilitation strategies.

5. Conclusion

This study demonstrated the feasibility of utilizing a Hier-ViT model to classify the severity of dysphagia in patients with LMI based on early brain MRI findings. The model showed moderate but clinically meaningful classification performance, indicating that Transformer-based architectures are capable of effectively capturing spatial features associated with swallowing dysfunction. Given the clinical importance of early severity stratification in formulating individualized rehabilitation strategies, these findings support the potential application of deep learning models as adjunctive tools that complement traditional swallowing assessments. Although further refinement is needed to address class imbalance, these results offer a promising avenue for improving early risk stratification in individualized rehabilitation planning.

Acknowledgements

We thank the Department of Medical Information at Dongguk University Ilsan Hospital for their support in accessing radiologic data.

Author contributions

T.L. collected clinical and imaging data, performed data preprocessing and annotation, and drafted the manuscript. K.N. developed and implemented the deep learning model and contributed to data analysis. B.H.K. critically reviewed the manuscript and contributed to overall interpretation. J.-W.P. conceptualized and supervised the study, provided critical revisions, and served as corresponding author. All authors read and approved the final version

of the manuscript.

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare no competing interests.

Ethics approval

This study was approved by the Institutional Review Board of Dongguk University Ilsan Hospital (IRB No. 2024-07-004). All methods were carried out in accordance with relevant guidelines and regulations.

Consent to participate/consent to publish

Informed consent was waived by the IRB due to the retrospective nature of the study using de-identified patient data.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT). (No. RS-2023-00252208)

References

1 Barer, D. The natural history and functional consequences of dysphagia after hemispheric stroke. *Journal of Neurology, Neurosurgery & Psychiatry* **52**, 236–241 (1989).

2 Clavé, P. & Shaker, R. Dysphagia: current reality and scope of the problem. *Nature Reviews Gastroenterology & Hepatology* **12**, 259–270 (2015).

3 Ekberg, O., Hamdy, S., Woisard, V., Wuttge–Hannig, A. & Ortega, P. Social and psychological burden of dysphagia: its impact on diagnosis and treatment. *Dysphagia* **17**, 139–146 (2002).

4 Vuilleumier, P., Bogousslavsky, J. & Regli, F. Infarction of the lower brainstem: Clinical, aetiological and MRI-topographical correlations. *Brain* **118**, 1013–1025 (1995).

5 Lee, H. & Sohn, C. H. Axial lateropulsion as a sole manifestation of lateral medullary infarction: a clinical variant related to rostral–dorsolateral lesion. *Neurological research* **24**, 773–774 (2002).

6 Dieterich, M. & Brandt, T. Wallenberg's syndrome: lateropulsion, cyclorotation, and subjective visual vertical in thirty-six patients. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **31**, 399–408 (1992).

7 Nowak, D. A. & Topka, H. R. The clinical variability of Wallenberg's syndrome: The anatomical correlate of ipsilateral axial lateropulsion. *Journal of neurology* **253**, 507–511 (2006).

8 Norrving, B. & Cronqvist, S. Lateral medullary infarction: prognosis in an unselected series. *Neurology* **41**, 244–244 (1991).

9 Sacco, R. L. *et al.* Wallenberg's lateral medullary syndrome: Clinical-magnetic resonance imaging correlations. *Archives of neurology* **50**, 609–614 (1993).

10 Ertekin, C., Aydogdu, I., Tarlaci, S., Turman, A. B. & Kiylioglu, N. Mechanisms of dysphagia in suprabulbar palsy with lacunar infarct. *Stroke* **31**, 1370–1376 (2000).

11 Steuer, I. & Guertin, P. A. Central pattern generators in the brainstem and spinal cord: an overview of basic principles, similarities and differences. *Reviews in the Neurosciences* **30**, 107–164 (2019).

12 Vigderman, A. M., Chavin, J. M., Kososky, C. & Tahmoush, A. J. Aphagia due to pharyngeal constrictor paresis from acute lateral medullary infarction. *Journal of the neurological sciences* **155**, 208–210 (1998).

13 Car, A. Inputs to the swallowing medullary neurons from the peripheral afferent fibers and the swallowing cortical area. *Brain research* **178**, 567–572 (1979).

14 Kim, T. J. *et al.* Dysphagia may be an independent marker of poor outcome in acute lateral medullary infarction. *Journal of Clinical Neurology* **11**, 349–357 (2015).

15 Kim, H., Chung, C.-S., Lee, K.-H. & Robbins, J. Aspiration subsequent to a pure medullary infarction: lesion sites, clinical variables, and outcome. *Archives of*

neurology **57**, 478–483 (2000).

16 Chun, M. H., Kim, D. & Chang, M. C. Comparison of dysphagia outcomes between rostral and caudal lateral medullary infarct patients. *International Journal of Neuroscience* **127**, 965–970 (2017).

17 Gupta, H. & Banerjee, A. Recovery of Dysphagia in lateral medullary stroke. *Case reports in neurological medicine* **2014**, 404871 (2014).

18 Kim, H., Lee, H. J. & Park, J.-W. Clinical course and outcome in patients with severe dysphagia after lateral medullary syndrome. *Therapeutic Advances in Neurological Disorders* **11**, 1756286418759864 (2018).

19 Jang, S. H. & Kim, M. S. Dysphagia in lateral medullary syndrome: a narrative review. *Dysphagia* **36**, 329–338 (2021).

20 Cho, Y.-J., Ryu, W.-S., Lee, H., Kim, D.-E. & Park, J.-W. Which factors affect the severity of dysphagia in lateral medullary infarction? *Dysphagia* **35**, 414–418 (2020).

21 Miller, D. D. & Brown, E. W. Artificial intelligence in medical practice: the question to the answer? *The American journal of medicine* **131**, 129–133 (2018).

22 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444 (2015).

23 Lundervold, A. S. & Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift fuer medizinische Physik* **29**, 102–127 (2019).

24 Islam, J. & Zhang, Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain informatics* **5**, 1–14 (2018).

25 Laukamp, K. R. *et al.* Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. *European radiology* **29**, 124–132 (2019).

26 Perkuhn, M. *et al.* Clinical evaluation of a multiparametric deep learning model for glioblastoma segmentation using heterogeneous magnetic resonance imaging data from clinical routine. *Investigative radiology* **53**, 647–654 (2018).

27 Yoo, Y. *et al.* Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *NeuroImage: Clinical* **17**, 169–178 (2018).

28 Kleesiek, J. *et al.* Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *NeuroImage* **129**, 460–469 (2016).

29 Li, H., Parikh, N. A. & He, L. A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. *Frontiers in neuroscience* **12**,

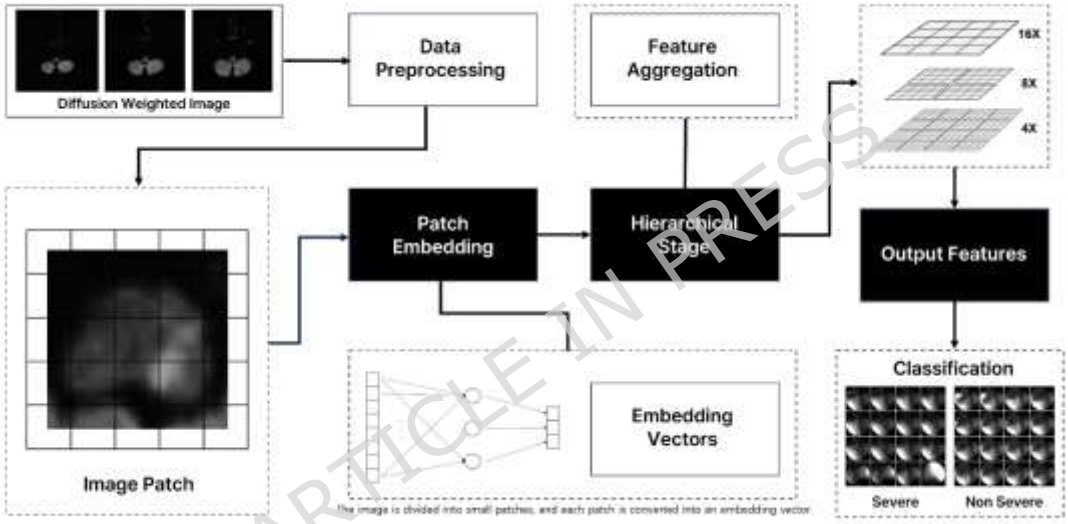
491 (2018).

30 Liu, Z. *et al.* in *Proceedings of the IEEE/CVF international conference on computer vision.* 10012–10022.

31 Cao, H. *et al.* in *European conference on computer vision.* 205–218 (Springer).

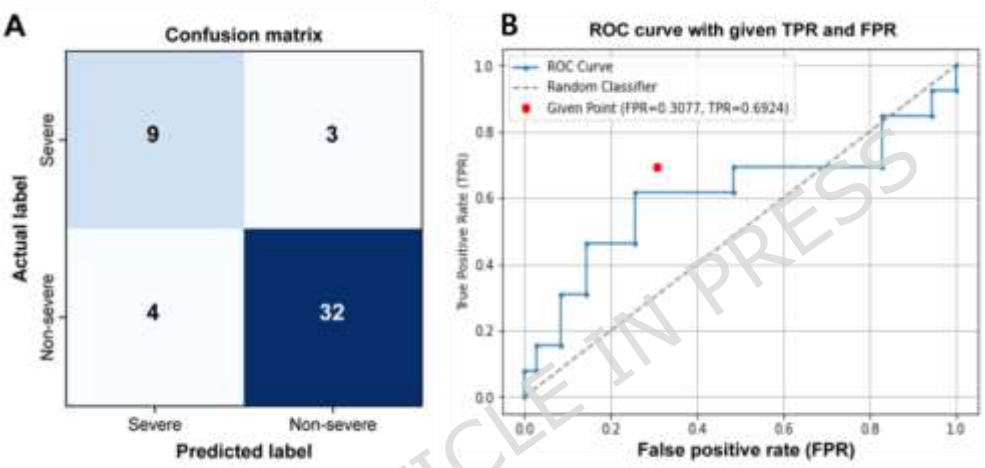
32 Cantone, M., Marrocco, C., Tortorella, F. & Bria, A. Convolutional networks and transformers for mammography classification: an experimental study. *Sensors* **23**, 1229 (2023).

Fig. 1. Deep learning pipeline using Hier-ViT for dysphagia severity classification



Diffusion-weighted images (DWIs) from axial medullary slices were preprocessed and segmented into fixed-size patches. Each patch was converted into an embedding vector and input into the hierarchical vision transformer (Hier-ViT) architecture, enabling multiscale feature aggregation. The output features were then used to classify patients into severe or non-severe dysphagia groups, based on videofluoroscopic swallowing study (VFSS) labels.

Fig. 2. Confusion matrix (A) and ROC curve (B) of Hier-ViT model.



(A) Confusion matrix showing model predictions of dysphagia severity. (B) Receiver operating characteristic (ROC) curve with area under the curve (AUC) of 0.69, indicating fair discriminative performance.

Table 1. Patient characteristics (n = 163).

Characteristic	Severe (n = 44)	Non-severe (n = 119)
Age (yr)	60.0 ± 14.5	60.7 ± 12.6
Gender (M/F)	33 / 11 (75.0 / 25.0%)	71 / 48 (59.7 / 40.3%)
Days from onset to VFSS	12.4 ± 3.4 (overall mean)	

Values are mean ± SD.

Table 2. Performance of prediction models.

	Accuracy	**Precision**	**Recall**	**F1-score**	**AUC (95% CI)**
Hier-ViT	0.85	0.70	0.75	0.72	0.69 (0.61-0.78)