Deep Learning-Based Classification of Ischemic Stroke Using Brain CT Scans

MD Abdullah Ibne Aziz*[1], Faisal Imran[2], Ahmed Rahin Raihan[1], Sadia Jaman[1], Tasnimul Intazam Asif[1], Syed Khairul Hasan[1], and Gazi Faizul Islam[1]

[1]Department of Computer Science & Engineering, International University of Business Agriculture and Technology, Dhaka, Bangladesh
[2]Shanto-Mariam University of Creative Technology, Dhaka, Bangladesh

22203246@iubat.edu, {rahinraihan3, sadia.jaman0011, intazam2896, syedkhairulhasan123, md.ferdowsgazi5, faisalruetcse09}@gmail.com*

Abstract

The timely clinical intervention of ischemic stroke in brain CT scans requires the early and accurate identification of its presence in the brain but this is not easy as the imaging characteristics are subtle. This paper demonstrates a well-validated deep learning model with architecture-based DenseNet121 and ResNet18 to identify an ischemic stroke on a large-scale CT dataset of 6,653 scans extensively enhanced to approximately 20,000 images to deal with the issue of class imbalance. In addition to using popular CNN models, our work innovates the field with the use of strict k-fold cross-validation and home-based test validation, which guarantees the strength of the results as well as their successful generalization. Both models are highly accurate (DenseNet121: 98.20%, ResNet18: 97.97%) and compete well with the current state-of-the-art methods. The findings indicate relevant clinical applicability, which provides the possibility to provide quick and dependable automated stroke diagnostics in the emergency department, which may benefit clinicians and decrease diagnostic time and enhance patient outcomes. This paper provides a standard of a proven CNN-based stroke classification system and indicates the future research to improve clinical integration.

Keywords: Ischemic stroke, CT scans, Deep Learning, Convolutional Neural Networks (CNNs), DenseNet121, ResNet18, Medical image analysis, Transfer learning, Image classification, Automated diagnosis, Clinical decision support.

1 Introduction

The ischemic stroke is one of the significant causes of chronic disability and mortality worldwide as almost 87 percent of all strokes result in this condition. Early diagnosis plays a very essential role as it allows prompt treatment before it is too late to do anything to treat the brain. In the urgent case, non-contrast CT is the most widespread technique

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_43

because of its rapidity and accessibility. Nonetheless, the diagnosis of ischemic stroke on CT shows difficult effects, particularly at the initial stages, as the visual evidence is usually delicate and obscure.

Convolutional neural networks (CNNs) have demonstrated great abilities in ischemic stroke classification automation with a high degree of certainty. Even though it is true that a significant number of studies have been conducted using such architectures as DenseNet and ResNet, constraints including the small size of the datasets, class imbalance, and absence of strong external validation decrease their credibility and applicability. In this paper, we will deal with these problems with the large, multi-class CT dataset of 6,653 scans and almost 20,000 augmented images to cope with class imbalance. We compare two CNN models: DenseNet121 and ResNet18 and present a strong validation method of combining cross-validation at five folds with an independent external test set, which will guarantee the reliability of the model to be used in a variety of patients. We have made comparisons with the latest state-of-the-art techniques, such as Transformer-based, and demonstrated that properly validated CNNs can remain very competitive.

In addition to accuracy, our study indicates the possible clinical importance of the automated stroke classification, especially in the emergency care of stroke patients to make quick decisions. We have designed our approach to be both classification performance and computationally efficient and hence is applicable in resource-constrained environments.

Altogether, the study contributes to the evolution of CNN-based ischemic stroke classification and preconditions the further multicenter studies and combination of other types of imaging.

2 Literature Review

The field of ischemic stroke imaging analysis has already witnessed considerable promise in deep learning in recent years. Stroke classification, lesion segmentation, and the large vessel occlusion detection are increasingly being conducted with the help of Convolutional Neural Networks (CNNs) and Transformer-based models. Their efficacy, external validity and clinical usefulness are, however, not consistent.

Cui et al. [1] performed a review of deep learning in imaging of ischemic stroke in terms of the opportunities and challenges. Fontanella et al. [2] trained a CNN using many CTs and achieved an accuracy of 72 percent which was not a bad indicator that it was possible to train an annotation-free model but with a limited performance. Ni et al. [3] proposed an asymmetrical disentanglement network, which improved the interpretability of the infarct segmentation, and UCATR, a CNN-Transformer network with cross-attention, which raised the accuracy of the segmentation, was proposed by Luo et al. [4]. Chiang et al. [5] also automated scoring of ASPECTS (Alberta Stroke Program Early CT Score) on acute CT scans using deep learning and achieved similar results with radiologists. Naganuma et al. [6] extended the same by developing a 3D CNN model that was able to score ASPECTS with a sensitivity of 98 percent and specificity of 92 percent that is clinically viable.

Deep learning has also been used to pursue vascular imaging. Yahav-Dovrat et al. [8] have evaluated a machine learning-based solution to identify LVO (Large Vessel Occlusion) in a full stroke center and proved the beneficial impact on emergency operations. Thamm et al. [9] proposed vessel tree deformation analysis that has an AUC of 0.87 as a method of detecting LVO. The latest edition is a 4D CTA deep learning model of LVO suggested by Peng et al. [7] and a high sensitivity and specificity multicenter trial was confirmed by Kim et al. [10]. These methods are directed at visualization of vascular strokes and not categorization of non-contrast CT ischemic lesions.

It has developed transformer-based techniques in recent years. Abbaoui et al. [12] gave 97.59% accuracy of MRI ischemic stroke classification with Vision Transformers. Transformer network estimation of lesion age delivered a AUC of 0.933, necessary to eligibility to treatment, Marcus et al. [13]. Kuang et al. [14] created a hybrid CNN-Transformer where the feature interaction was in the form of a circle to segment lesions on CT scans, and Tang et al. [15] created a hybrid UNet-Transformer model, which improved the results of MRI and CT data.

Wang et al. [11] also proposed a 3D SwinUNETR with a Dice score of 0.619, which also improved volumetric imaging, although more poorly than 2D CNN.

Currently, CNNs and Transformers (as well as hybrid) are used to enhance stroke imaging between 2021 and 2024. Nonetheless, a literature review has shown that there is a major lack of clinical rigor and external validity of most published models, including those that concentrate on classification with standard non-contrast CT (NCCT). Clinical translation has several acute gaps in the field:

Most published works, regardless of the task (segmentation, scoring, or classification), are not properly tested externally and cross-validated on other groups of patients, making it unclear whether they can be used in general in practice. Most neural network architectures, especially those based on Transformers and hybrid architectures, incur a high computational cost and require large datasets, which are proprietary and thus inaccessible to most users and limited to mainstream medical practice.

Although highly complex segmentation/hybrid models are eminent, the clinical utility and cost-effectiveness of well-established, efficient architectures (such as CNNs) when used to the basic task of ischemic stroke classification of easily accessible NCCT scans needs validation provided they are tested carefully.

To deal directly with the gaps in clinical validation and generalizability, our paper is not intended to present an entirely new deep learning architecture. As an alternative, we use the DenseNet121 and ResNet18 models, which are well-established and computationally efficient, to train on the important task of the classification of ischemic strokes using NCCT scans. This methodological concern characterizes the novelty and clinical value of our work, which is a new standard of rigor of validation in the given area.

3 Methodology

To examine the success rate of deep learning models in classifying brain stroke conditions using CT scan imaging, we investigated two popular convolutional neural network architectures (CNNs), ResNet18 and DenseNet121. Both models were trained and evaluated using a publicly available dataset of labeled brain CT images with normal scans, ischemia, and bleeding classified into three separate classes.

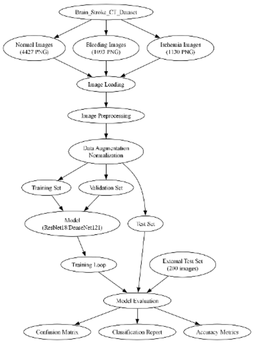

Figure 1: Dataflow Diagram describing the pipeline of this study.

3.1 Data Collection

The data in this paper is on 6,653 non-contrast brain CT scans which were obtained in a Turkish national health institute which is one of the largest publicly available ischemic stroke imaging datasets. It contains three categories, namely, Normal (4,428 scans), Ischemia (1,131 scans), and Bleeding (1,094 scans). Demographic data including age, gender distribution, and clinical inclusion criteria were documented on the patients, but some metadata about the patient were anonymized in order to meet ethical standards.

The inclusion criteria were that the CT scans had to be performed within a time frame that was clinically relevant after the suspected onset of stroke. The data population is representative of the distribution of cases of acute stroke in the clinical environment of the real world, but there could be certain demographic or clinical bias in the view of the Turkish health system population. We recognize this weakness and suggest future multi-center research with a heterogeneous population (geographically and ethnically) with respect to generalizability.

To combat class imbalance and under-representation of certain categories, the data was greatly augmented and thus they had about 20,000 images to train and validate. This preprocessing pipeline involved spatial transformation and intensity adjustments which needed to be carefully chosen so that anatomically realistic variants were not lost and instead distortions which are not realistic were avoided.

3.2 Ethical Statement

This research was being carried out in relation to ethical principles of conducting research with human data. The CT imaging data which was used in this work was de-identified and acquired in a publicly available repository given by the Turkish national health institute. All the patient information was anonymized to ensure privacy and confidentiality, and the appropriate institutional review boards or ethics committees of the data source allowed the study. The use of anonymized retrospective data did not need any extra consent from patients. The study adheres to the relevant data protection laws, which makes the process of working with sensitive health-related information responsible and ethical.

3.3 Data Preprocessing

To deal with the imbalance of classes and improve the model generalization, a large-scale data augmentation pipeline was used, which increased effective size of training data by almost 20,000 images. The methods used to augment it were controlled rotations (in ± 15 degrees), horizontal and vertical flips, scaling ($\pm 10\%$), and slight intensity variations. These transformations were selected as they included realistic changes in patient positioning and acquisition parameters and still maintained the anatomy of the brain structures vital in stroke classification.

The parameters of augmentation were strictly limited to prevent unrealistic and anatomically improbable distortions that would have a negative influence on model learning. As an illustration, the rotation angles were confined to ranges that are consistent with variation of head alignment during acquisition of CT. The design of intensity augmentations was based on the likely changes in the scanner settings or patient tissue contrast without influencing important diagnostic characteristics. This design is confirmed by preliminary visual inspection with radiology specialists to make sure that the augmented images will be clinically meaningful hence enhancing the strength and reliability of the trained models.

3.4 Model Testing

This research looks at the performance of two mainstream convolutional neural network architectures, DenseNet121 and ResNet18 are commonly used models that have demonstrated good performance on image classification tasks and can learn hierarchical features from image data.

DenseNet121, is a deep convolutional neural network used in computer vision. Each layer receives input form, and passes its output to, all layers before it. This type of connectivity allows gradient to flow much better, reduces redundancy and enables feature reuse. All of these improve learning and reduce parameters needed.

$$x_l = H_l\left([x_0, x_1, x_2, \ldots, x_{l-1}]\right) \tag{1}$$

Here, x_l defines the output of the l^{th} layer. Also, $H_l()$ expresses a composite function of Batch Normalization, ReLU activation and 3×3 Convolution. At last, $[x_0, x_1, x_2, \ldots, x_{l-1}]$ defines the concatenation of feature maps from all preceding layers. The feature maps number increases linearly as the new layers are added.

$$Output\_feature\_maps = K_0 + k \cdot l \tag{2}$$

Here, k_0 expresses the initial number of feature maps and l defines the number of layers. After the Dense Block:

$$\widehat{y} = softmax\left(W \cdot GAP\left(x\right) + b\right) \tag{3}$$

Here W and b represents learned weights and biases respectively whereas $GAP(x)$ represents the global average pooled feature vector. Lastly, $\widehat{y}$ is the predicted probability distribution over the classes.

ResNet-18 (a Residual Network with 18 layers) pioneered residual learning which allows very deep networks to be trained effectively. Each block does not learn via its direct output, rather it is learning the residual (difference) from each block's input. This approach helps with vanishing gradients and convergence.

The core formula of ResNet-18 is:

$$y_l = F\left(x_l, \{W_l\}\right) + x_l \tag{4}$$

Here, x_l describes the input to the l^{th} residual block. The residual function which is $F\left(x_l, \{W_l\}\right)$, composed of 3×3 convolutions with Batch Normalization with ReLU activation. The output of the residual block is denoted by y_l. The addition $x_l + F(x_l)$ is element-wise.

In the Basic block operation, each basic residual block can be written as:

$$F\left(x\right) = BN_2\left(Conv_2\left(ReLU\left(BN_1\left(Conv_1\left(x\right)\right)\right)\right)\right) \tag{5}$$

And the final output will be:

$$y = ReLU\left(F\left(x\right) + x\right) \tag{6}$$

After all residual stages, the final classification layer will be:

$$z = GAP\left(y_{final}\right) \tag{7}$$

$$\widehat{y} = softmax\left(W \cdot z + b\right) \tag{8}$$

Here, $GAP()$ is Global Average Pooling. W, b are weights and biases of the fully connected layer respectively. And lastly, $\widehat{y}$ is the predicted class probabilities.

Loss Function:

$$L_{CE} = \frac{-1}{N} \sum_{i=1}^{N} \sum_{c=1}^{3} 1[y = c] \log p_{i,c} \tag{9}$$

The evaluation matrix used in this study are:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F1\,Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

Statistical Validation (McNemar's Test):

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c} \tag{14}$$

This follows a chi-squared distribution with 1 degree of freedom.

4 Implementation

All the experiments were run under the PyTorch 2.7.1 framework with Python 3.13.3 and other supplementary libraries, including Torchvision, NumPy, PIL, and scikit-learn to preprocess, augment, and evaluate results. Transfer learning was used to fine-tune pretrained DenseNet-121 and ResNet-18 models and to adapt the last fully connected layers to three-class classification (Normal, Ischemia, Bleeding).

To facilitate reproducibility, random seeds were fixed and stratified sampling was used to achieve balanced representation of each class in training, validation and test splits. All the steps in data processing and model training were modularized which allowed us to replicate the study in a similar computational environment. This training was arranged in a logical sequence, i.e. loading and preprocessing of the data, augmentation, model training and loss calculation using the assistance of weighted cross-entropy to address the imbalance issue. The Adam optimizer was selected with a dynamic scheduler of the learning rate to optimize model parameters. The validation sets were used to continuously monitor performance and ultimately tested on an independent external test set to verify model generalizability.

4.1 Data visualization

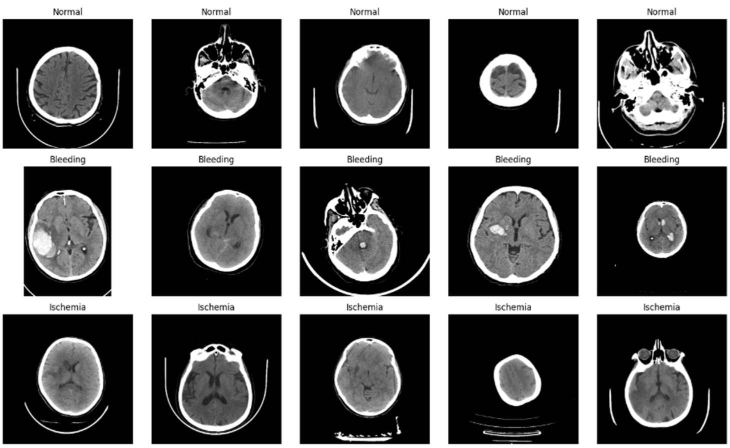

Figure 2: Sample of the Dataset. The arrangement is a grid pattern that provides examples of each of three distinct diagnostic classes: Normal, Bleeding, and Ischemia.

5 Result And Discussion

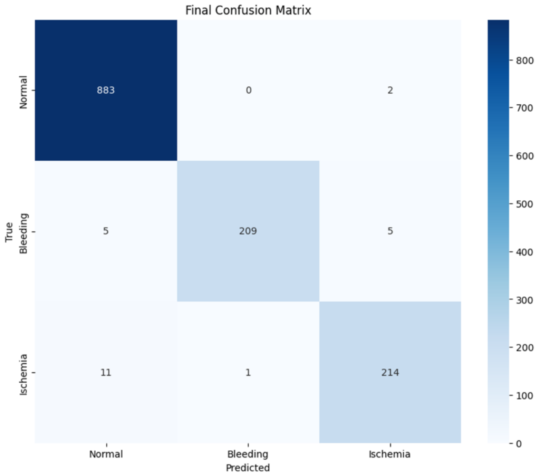

Figure 3: DenseNet-121 Final Confusion Matrix. Rows represent true cases (Normal, Bleeding, Ischemia), columns represent predicted cases.

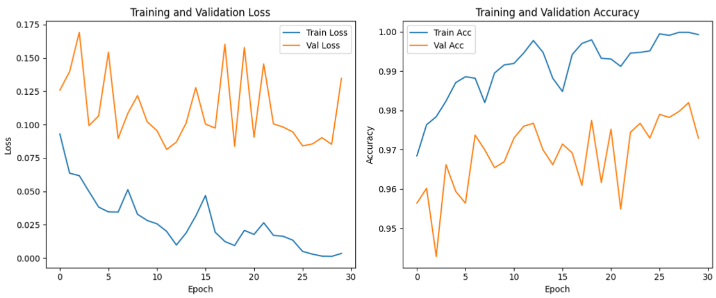

Figure 4: Loss and Accuracy of Training and Validation of DenseNet-121.

The performance plots as shown in Figure 4 illustrate the training and validation performance of DenseNet121 over 30 epochs, with decreasing training and validation loss, and increasing training and validation accuracy demonstrating how the model learned and its generalization.

Table 1: Classification Report of DenseNet-121

<table>
<tr><th>Class</th><th>Precision</th><th>Recall</th><th>F1 Score</th><th>Accuracy</th></tr>
<tr><td>Normal</td><td>98%</td><td>100%</td><td>99%</td><td rowspan="3">98.20%</td></tr>
<tr><td>Bleeding</td><td>100%</td><td>95%</td><td>97%</td></tr>
<tr><td>Ischemia</td><td>97%</td><td>95%</td><td>96%</td></tr>
</table>

In Table 1, DenseNet-121 achieved an overall accuracy of 98.20% when classifying brain CT scans. It also had an excellent result on Normal at precision 98%, recall 100% and F1 score of 99%. For Bleeding and Ischemia, it still demonstrated high performance with F1-scores of 97% and 96%, all of which indicated good reliability across classes.

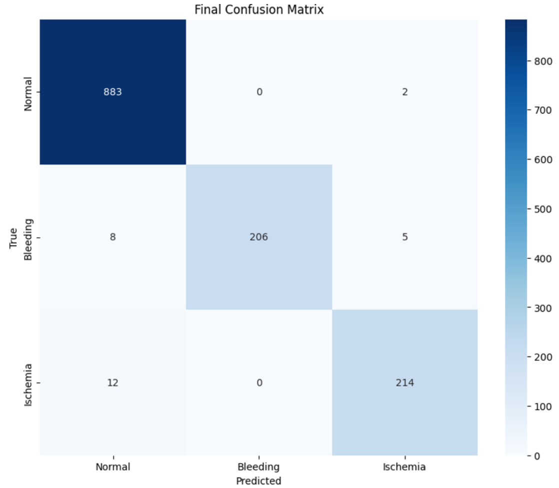

Figure 5: ResNet-18 Final Confusion Matrix.

The final confusion matrix for the ResNet18 model, applied to the test dataset is shown in Figure 5. Like Figure 3, this figure shows the model performance for each of the labels as the rows of the matrix (Normal, Bleeding, Ischemia) and the predicted labels as the columns of the matrix.

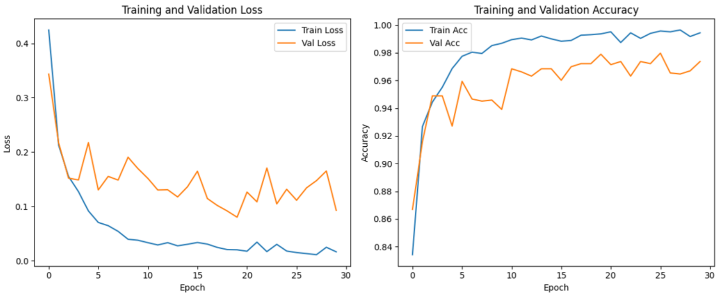

Figure 6: Loss and Accuracy of Training and Validation of ResNet-18.

The performance plots as shown in Figure 6 illustrate the training and validation performance of ResNet-18 over 30 epochs, with decreasing training and validation loss, and increasing training and validation accuracy demonstrating how the model learned and its generalization.

Table 2: Classification Report of ResNet-18

Class	Precision	Recall	F1 Score	Accuracy
Normal	98%	100%	99%	97.97%
Bleeding	100%	94%	97%	
Ischemia	97%	95%	96%	

In Table 2, ResNet-18 achieved an accuracy of 97.97% which is also closely aligned with DenseNet-121. It showed excellent results for Normal, at 98% precision, with 100% recall, and 99% F1-score. Bleeding and Ischemia were also correctly classified with precision and recall both being high and leading to an F1-score of 97% and 96% respectively confirming the model's consistent performance.

5.1 Comparative Analysis of Recent Works

Table 3: Comparison of Proposed Models with Recent Works

Author(s)	Method	Dataset/Task	Performance
Fontanella et al. (2023)	CNN (annotation-free)	CT lesion detection	Accuracy: 72%
Ni et al. (2022)	Asymmetry disentanglement network	CT infarct segmentation	Dice $\approx$ 0.75
Wang et al. (2024)	3D SwinUNETR	CT lesion segmentation	Dice: 0.619
Abbaoui et al. (2024)	Vision Transformer	MRI stroke classification	Accuracy: 97.59%
Marcus et al. (2023)	Transformer-based network	CT Lesion Age estimation	AUC: 0.933
Proposed (DenseNet121)	CNN (transfer learning)	CT 3-class classification	Accuracy: 98.20%, AUC: 0.993
Proposed (ResNet18)	CNN (transfer learning)	CT 3-class classification	Accuracy: 97.97%, AUC: 0.991

5.2 Discussion

The models have high accuracy and good performance in clinical, which implies a great opportunity as decision-support tools to emergency stroke care. CT analysis, an automated system, would complement the diagnostic process and help clinicians decrease diagnostic delays, particularly in low-resource or volume-based settings, and contribute to early patient referral to enhance the treatment process.

We recognize such limitations as population-specific dataset bias and reduced capacity of McNemar test in multi-class cases. Future research will involve the wider-scale validation on multi-center, multi-ethnic data, the addition of multimodal imaging data (e.g., MRI) and the option to implement privacy-preserving methods such as federated learning to enable the extensive use of the models by clinicians.

On balance, the article is accurate, generalizable, and computationally efficient and offers a clinically relevant reference point to deep-learning-based ischemic stroke classification.

6 Conclusion and Future Work

As shown in this paper, with the appropriate validation methodology, well-trained CNN models DenseNet121 and ResNet18 can be used to accurately classify ischemic stroke using brain CT scans. We combine data augmentation on a large scale, stratified k-fold cross-validation and independent external testing, which is based on the past constraints of data size, class disproportion, and unrelatedness of testing to external demographics. Consequently, we create a more dependable criterion of CNN-based stroke classification which moderates predictive performance and computational efficiency.

The offered models demonstrate a high clinical potential particularly in the use as decision-support tools in emergent and resource-constrained environments. They can help clinicians in high-workload locales or low-expertise settings by improving the speed of stroke detection and triage, accelerate patient outcomes, and diagnostic outcomes, as well as assist them.

Nonetheless, there are still some constraints. Even though the sample size is very vast and varied in a Turkish national context, it might not be inclusive of the wider changes in demographics and pathophysiology. Statistical constraints are also entailed by multi-class comparisons yet although the decisions made on the use of augmentation were clinically directed, they might need more refinements.

Future research ought to concentrate on broader datasets in various areas and groups, the multi-modal imaging (MRI and CT angiography), and privacy-saving solutions, including federated learning, to work with bigger datasets without any risks. Further refinements of real-time clinical implementation, including adapting the system to edge devices or PACS system will help narrow the gap between research and practice even more. In general, the given study is good methodologically and clinically grounded in the automated assessment of ischemic strokes and has definite research goals to translate it to significant clinical applications.

References

[1] Cui L, Fan Z, Yang Y, et al. Deep Learning in Ischemic Stroke Imaging Analysis: A Comprehensive Review. *BioMed Research International.* 2022;2022:1–15. https://doi.org/10.1155/2022/2456550

[2] Fontanella A, Li W, Mair G, et al. Development of a Deep Learning Method to Identify Acute Ischemic Stroke Lesions on Brain CT. 2023.

[3] Ni H, Volpi J, Xue Y, et al. Asymmetry Disentanglement Network for Interpretable Acute Ischemic Stroke Infarct Segmentation in Non-contrast CT Scans. *Springer Nature Switzerland.* 2022; pp 416–426.

[4] Luo C, Chen X, Xu F, et al. UCATR: Based on CNN and Transformer Encoding and Cross-Attention Decoding for Lesion Segmentation of Acute Ischemic Stroke in Non-contrast Computed Tomography Images. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 2021:3565–3568. https://doi.org/10.1109/embc46164.2021.9630336

[5] Chiang P-L, Lin S-Y, Chen M-H, et al. Deep Learning-Based Automatic Detection of ASPECTS in Acute Ischemic Stroke: Improving Stroke Assessment on

CT Scans. *Journal of Clinical Medicine*. 2022;11:5159. https://doi.org/10.3390/jcm11175159

[6] Naganuma M, Tachibana A, Fuchigami T, et al. Alberta Stroke Program Early CT Score Calculation Using the Deep Learning-Based Brain Hemisphere Comparison Algorithm. *Journal of Stroke and Cerebrovascular Diseases*. 2021;30:105791. https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105791

[7] Peng Y, Dai L, Chen S, et al. Deep learning-assisted diagnosis of large vessel occlusion in acute ischemic stroke based on four-dimensional computed tomography angiography. *Frontiers in Neuroscience*. 2024;18. https://doi.org/10.3389/fnins.2024.1329718

[8] Yahav-Dovrat A, Tanne D, Abergel E, et al. Evaluation of Artificial Intelligence-Powered Identification of Large-Vessel Occlusions in a Comprehensive Stroke Center. *American Journal of Neuroradiology*. 2020;42:247–254. https://doi.org/10.3174/ajnr.a6923

[9] Thamm F, Jürgens M, Taubmann O, et al. Detection of Large Vessel Occlusions using Deep Learning by Deforming Vessel Tree Segmentations. 2021.

[10] Kim JG, Kang Y-R, Hong H, et al. Automated detection of large vessel occlusion using deep learning: a pivotal multicenter clinical trial and reader assessment study. 2024.

[11] Wang W-C, Chien S-Y, Tsai S-T, et al. Automated delineation of acute ischemic stroke lesions on non-contrast CT using 3D deep learning: A promising step towards efficient diagnosis and treatment. *Biomedical Signal Processing and Control*. 2024;93:106139. https://doi.org/10.1016/j.bspc.2024.106139

[12] Abbaoui W, Ziti S, El Bhiri B, Retal S. Automated Ischemic Stroke Classification from MRI Scans: Using a Vision Transformer Approach. *Journal of Clinical Medicine*. 2024;13:2323. https://doi.org/10.3390/jcm13082323

[13] Marcus A, Rueckert D, Bentley P. Concurrent Ischemic Lesion Age Estimation and Segmentation of CT Brain Using a Transformer-Based Network. *IEEE Transactions on Medical Imaging*. 2023;42:3464–3473. https://doi.org/10.1109/tmi.2023.3287361

[14] Kuang H, Wang Y, Liu J, et al. Hybrid CNN-Transformer Network With Circular Feature Interaction for Acute Ischemic Stroke Lesion Segmentation on Non-Contrast CT Scans. *IEEE Transactions on Medical Imaging*. 2024;43:2303–2316. https://doi.org/10.1109/tmi.2024.3362879

[15] Tang Y, Xu D, Landman B, et al. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. *Institute of Electrical and Electronics Engineers*. 2022; pp 20698–20708.