Journal of Associated Medical Sciences 2024; 57(2): 49-55

JAMS
Journal of Associated Medical Sciences

Scopus Indexed & Thai-Journal Citation Index Centre (TCI)

Journal of Associated Medical Sciences

Journal homepage: https://www.tci-thaijo.org/index.php/bulletinAMS/index

Comparative analysis of deep learning techniques for accurate stroke detection

Titipong Kaewlek[1,2*] Ketmanee Sitinwan[1] Kunaporn Lueangaroon[1] Wasita Sansuriyawong[1]

[1]*Department of Radiological Technology, Faculty of Allied Health Sciences, Naresuan University, Phitsanulok Province, Thailand.*
[2]*Interdisciplinary Health and Data Sciences Research Unit, Faculty of Allied Health Sciences, Naresuan University, Phitsanulok Province, Thailand.*

ARTICLE INFO

Article history:
Received 20 November 2023
Accepted as revised 9 February 2024
Available online 12 February 2024

Keywords:
Brain CT image, computed tomography image, deep learning, stroke, VGG-16

ABSTRACT

Background: The traditional diagnosis of strokes through computed tomography (CT) heavily relies on radiologists' expertise for accurate interpretation. However, the increasing demand for this critical task exceeds the available radiologist workforce, necessitating innovative solutions. This research addresses this challenge by introducing deep learning techniques to enhance the initial screening of stroke cases, thereby augmenting radiologists' diagnostic capabilities.

Objective: This study aims to compare four techniques for classifying stroke lesions in CT images.

Materials and methods: Four distinct models-CNN-2-Model, LeNet, GoogleNet, and VGG-16-were trained using a dataset comprising 1,636 CT images, including 1,111 normal brain images and 525 stroke images. Seventy percent of the images were used to train the most effective deep learning model, and subsequently, these images were utilized to evaluate the performance of each model. The evaluation involved assessing accuracy, precision, sensitivity, specificity, F1 score, false positive rate, and AUC.

Results: The evaluation process included a comprehensive statistical analysis of the models' prediction results. The findings revealed that VGG-16 emerged as the top-performing deep learning model, achieving an impressive accuracy of 0.969, precision of 0.952, sensitivity of 0.952, specificity of 0.978, F1 score of 0.952, false positive rate of 0.022, and AUC of 0.965.

Conclusion: In conclusion, deep learning techniques, particularly the VGG-16 model, demonstrate significant promise in enhancing the accuracy of stroke lesion classification in CT images. These findings underscore the potential of leveraging advanced technologies to address the growing challenges in stroke diagnosis and pave the way for more efficient and accessible healthcare solutions.

Introduction

A stroke occurs when there is a sudden disruption in the blood supply to the brain, leading to an immediate impairment of brain function. It manifests in two primary forms: ischemic stroke and hemorrhagic stroke.[1] The World Health Organization (WHO) report on World Stroke Day in 2022, observed on October 29, emphasizes stroke as the primary cause of global disability and the second leading cause of death.[2] According to the 2022 Global Stroke Factsheet, strokes contribute to five million annual fatalities worldwide. In Thailand, data from the Ministry of Public Health for 2022 reports 37,802 cases, translating to 58.0 patients per 100,000 individuals in the Thai population. This underscores a rising incidence of stroke cases, aligning with global trends, and notably, stroke

* **Corresponding contributor.**
Author's Address: *Department of Radiological Technology, Faculty of Allied Health Sciences, Naresuan University, Phitsanulok Province, Thailand.*
E-mail address: titipongk@nu.ac.th
doi: 10.12982JAMS.2024.026
E-ISSN: 2539-6056

ranks as the second leading cause of death among the elderly in Thailand.[3]

The traditional diagnostic approach for stroke involves a thorough examination, encompassing medical history assessments, physical examinations, and imaging studies such as CT scans and MRIs.[4,5] Gathering medical history entails obtaining information about the patient's background, including risk factors like hypertension, diabetes, smoking, and prior strokes. Physical examination assesses neurological symptoms, covering aspects such as strength, coordination, reflexes, and sensory function.

Imaging studies, crucial to the diagnostic process, include CT scans (Computed Tomography) and MRI (Magnetic Resonance Imaging).[4,5] CT scans are frequently employed for the swift determination of whether a stroke is ischemic or hemorrhagic. Furthermore, they aid in pinpointing the location and extent of the damage. MRI, offering more detailed images compared to CT scans, is particularly valuable for detecting ischemic strokes. The advent of artificial intelligence (AI) in medical imaging has opened avenues for improving the accuracy and efficiency of stroke detection. AI algorithms trained on extensive datasets of medical images, possess the capacity to autonomously analyze intricate patterns, potentially revolutionizing the diagnostic process.[6,7]

In recent studies for stroke classification, Ammar *et al*. compared five deep learning models (ReNet50, VGG-16, Xception, InceptionV3, and InceptionResNetV2), with VGG-16 demonstrating the highest performance accuracy at 96.0% for intracranial hemorrhage.[8] Worachotsueptrakun compared four deep learning models (AlexNet, VGG-16, GoogleNet, and ResNet), with GoogleNet exhibiting the best performance, achieving accuracy, precision, recall, and F1-score of 92.00%, 94.00%, 83.96%, and 88.70%, respectively.[9] Additionally, Vamsi *et al*. achieved an accuracy of 97.81 % using VGG-16 and random forest. [10]

This study aims to compare the efficacy of stroke classification among four models using brain CT images. Specifically, the research seeks to assess the performance of four conventional neural networks-CNN-2-Model, LeNet, GoogleNet, and VGG-16-utilizing deep learning techniques for stroke classification.

Materials and methods

Data collection

The 1,636 CT images were collected from Kaggle.[11,12] This dataset included 1,111 normal brain images and 525 stroke images. Inclusion criteria comprised axial CT image plane, non-ionic contrast media images, and indicated labels for stroke and normal. Exclusion criteria encompassed multi-lesion brain images, unidentified results, and images of the base of the skull region. For image preparation, the dataset was divided into three groups: 70% for training, 20% for validation, and 10% for testing. A total of 328 images were selected for training the brain segmentation.

Stoke classification technique

The stroke detection process in this study has two parts: brain segmentation and stroke classification, as shown in Figure 1. For brain segmentation, we use the U-net architecture to separate the brain tissue from the skull bone in CT images.[13] Subsequently, the brain-segmented images were trained using four convolutional neural network models (CNN-2-Model, LeNet, GoogleNet, and VGG-16), and their performance in stroke classification was compared.

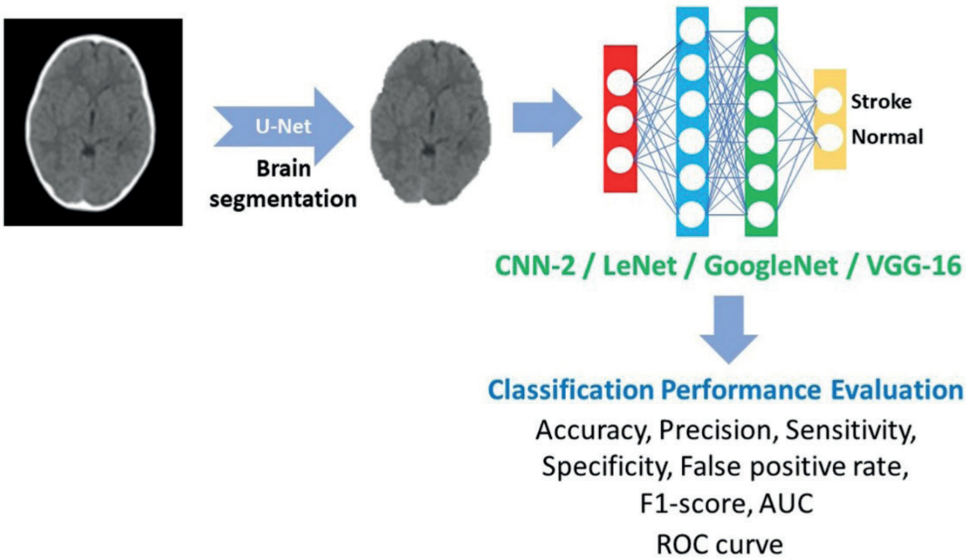

Figure 1. *Flowchart of stoke detection.*

Brain segmentation

In brain segmentation, we apply the U-net architecture modified from Kairess's model as shown in Table 1.[13] The input consists of 328 images (256x256x1 pixels). The contracting path follows the typical architecture of a convolutional network, with the repeated application of 3x3 convolutions, each followed by a rectified linear unit (ReLU), and a 2x2 max-pooling operation with a stride of 2 for down sampling. Every step in the expansive path consists of upsampling the feature map followed by a 3x3 convolution that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and each followed by a sigmoid. The 32x32x128 dense layer was connected to the down-up sampling. At the final layer, 256x256x64 pixels, with a 3x3 convolution and a 1x1 convolution, is followed by a sigmoid. The model was trained using a learning rate of 0.000001. The network training was set to a batch size of 32 for 80 epochs. The Dice similarity coefficient (DSC), Dice loss, and Weighted bce Dice loss were used to evaluate the performance of brain segmentation.

Table 1. U-net architecture parameters.

Layer (type)	Output Shape	Parameter
input_1 (InputLayer)	(None, 256, 256, 1)	0
conv2d (Conv2D)	(None, 256, 256, 32)	320
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)	0
conv2d_1 (Conv2D)	(None, 128, 128, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 64)	0
conv2d_2 (Conv2D)	(None, 64, 64, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 128)	0
dense (Dense)	(None, 32, 32, 128)	16512
up_sampling2d (UpSampling2D)	(None, 64, 64, 128)	0
conv2d_3 (Conv2D)	(None, 64, 64, 128)	147584
up_sampling2d_1 (UpSampling2D)	(None, 128, 128, 128)	0
conv2d_4 (Conv2D)	(None, 128, 128, 64)	73792
up_sampling2d_2 (UpSampling2D)	(None, 256, 256, 64)	0
conv2d_5 (Conv2D)	(None, 256, 256, 1)	577
Total parameters: 331,137, Trainable parameters: 331,137, non-trainable parameters: 0		

Comparative analysis for stroke detection

The 1,636 CT images were collected from Kaggle.[11,12] This dataset included 1,111 normal brain images and 525 stroke images. For image preparation, 70 % of the images were used for training, 20% for validation, and 10% for testing. The brain segmentation images were trained, validated, and tested using four model techniques (CNN-2-Model, LeNet, GoogleNet, and VGG-16). The optimizer used was the Adam method, and the loss function employed was cross-entropy. The structure of the four models is shown in Table 2. The models were trained on Google Colab with Tesla T4.

Table 2. The structure of four models.

Model	CNN2-model	LeNet	GoogleNet	VGG-16
Input layer	256×256×1	256×256×1	256×256×1	256×256×1
1st Blocks	1. A Conv 2. Relu 3. a MaxPooling	1. A Conv 2. Relu 3. a MaxPooling	1. A Conv 2. Relu 3. a MaxPooling	1. Two Conv 2. Relu 3. a MaxPooling
2nd Block	1. A Conv 2. Relu 3. a MaxPooling	1. A Conv 2. Relu	1. Two Conv 2. Relu 3. a MaxPooling	1. Two Conv 2. Relu 3. a MaxPooling
3rd Block	-	-	1. Two inceptions 2. MaxPooling	1. Three Conv 2. Relu 3. a MaxPooling
4th Block	-	-	1. Five inceptions 2. MaxPooling	1. Three Conv 2. Relu 3. a MaxPooling
5th Block	-	-	1. Two inceptions 2. AvePooling (0.4 dropout)	1. Three Conv 2. Relu 3. a MaxPooling
Output layer	1. A flattened layer 2. Two dense layers 3. Softmax (2 class)	1. A flattened layer 2. Two dense layers 3. Softmax (2 class)	1. A flattened layer 2. Two dense layers 3. Softmax (2 class)	1. A flattened layer 2. Two dense layers 3. Softmax (2 class)

Note: *Conv: convolutional, Relu: rectified linear unit*

Statistical evaluation

Table 3 represents the confusion matrix for evaluating the efficiency of the four models in classifying stroke and normal brain images on CT images where:

True Positives (TP): The model correctly detects and classifies strokes.

True Negatives (TN): The model correctly identifies normal images without detecting a stroke.

False Positives (FP): The model incorrectly detects a stroke on a normal image or an incorrect lesion.

False Negatives (FN): The model fails to detect a stroke on an actual stroke image.

Table 3. Confusion matrix of stoke detection.

Model	Pathology		Total
	Stroke	**Normal**	
Stroke (positive)	TP	FP	TP+FP
Normal (negative)	FN	TN	FN+TN
Total	TP+FN	FP+TN	TP+FP+FN+TN

The performance of the model was analyzed. The accuracy, positive predictive value (precision), sensitivity (recall), specificity, F-1 score, and false positive rate were calculated using Equations (1) to (6). The receiver operating characteristic curve (ROC) and area under the curve (AUC) were also evaluated.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4)$$

$$\text{F1-score} = 2*\frac{Precision*Sensitivity}{Presision+Sensitivity} \quad (5)$$

$$\text{False Posotive Rate} = \text{1-Specificity} \quad (6)$$

Results

All brain CT images were segmented, as demonstrated in Figure 2, showing the result that the U-net deep learning model can segment the brain CT images in this study. The predicted area indicates a higher performance of brain segmentation. The average Dice similarity coefficient (DSC) is 0.963, with a Dice loss of 0.037 and a Weighted BCE Dice loss of 0.185.

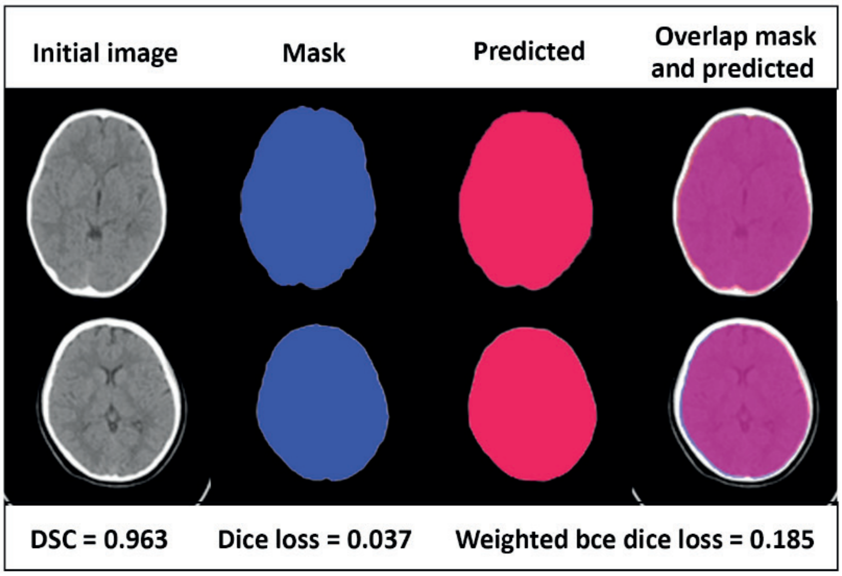

Figure 2. *Segmented brain CT image.*

Table 4 presents the results of evaluating the four models for detecting stroke lesions in CT images. The highest accuracy was achieved by VGG-16 (0.969), while the best precision was observed with GoogleNet (0.976), with VGG-16 as the second-best (0.952). VGG-16 (0.952) also demonstrated the highest sensitivity, whereas GoogleNet had the lowest sensitivity. The specificity of VGG-19 (0.978) was slightly lower than that of GoogleNet (0.988). The F1-score of VGG-16 (0.952) was higher than that of GoogleNet (0.901), CNN-2 model (0.886), and LeNet (0.886). Additionally, the false positive rate of VGG-16 (0.022) was lower than that of the CNN-2 model (0.074) and LeNet (0.074), and comparable to GoogleNet (0.012).

Table 4. The performance of four models

Model	Accuracy	Precision	Sensitivity	Specificity	F-1 score	False positive Rate
CNN-2-model	0.932	0.833	0.946	0.926	0.886	0.074
LeNet	0.932	0.833	0.946	0.926	0.886	0.074
GoogleNet	0.932	0.976	0.837	0.988	0.901	0.012
VGG-16	0.969	0.952	0.952	0.978	0.952	0.022

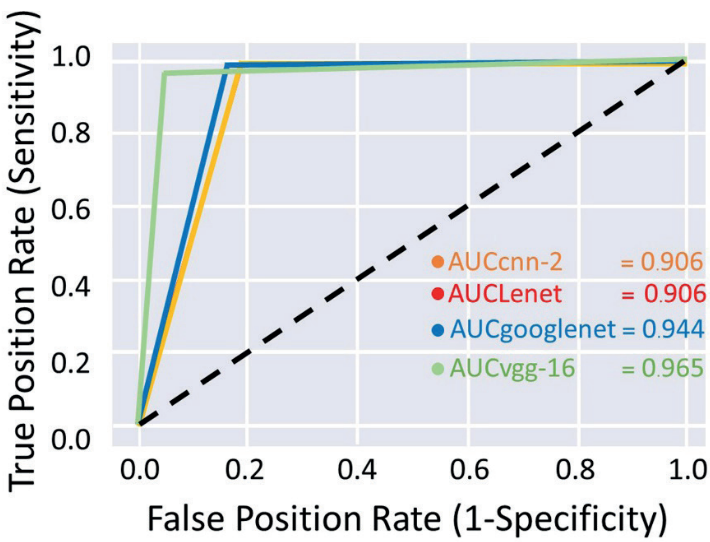

Figure 3. *The receiver operating characteristic curves and areas under the curve of the CNN-2, LeNet, GoogleNet, and VGG-16 models.*
ROC curve and AUC of the VGG-16 (Green Line) model, presented in Figure 3, confirm the highest performance in the detection of strokes in brain CT images. The second-best performance is observed in GoogleNet (Blue Line). The ROC curve and AUC of the CNN-2 (Yellow Line) model are equal to those of LeNet (Red Line), and they exhibit the lowest performance.

Discussion

The primary goal of this study was to address the challenges in stroke diagnosis by introducing and comparing four deep learning techniques for classifying stroke lesions in CT images. The overarching objective was to enhance the diagnostic capabilities of radiologists and contribute to the development of innovative solutions to meet the increasing demand for accurate stroke interpretation. Our evaluation process involved training and assessing four distinct models-CNN-2-Model, LeNet, GoogleNet, and VGG-16-using a dataset comprising 1,636 CT images. Notably, the dataset included a balanced representation of 1,111 normal brain images and 525 stroke images, providing a comprehensive basis for model training and evaluation.

Our study revealed that among the evaluated models, VGG-16 emerged as the top-performing deep learning model. With an accuracy of 0.969, an F-1 score of 0.952, and an AUC of 0.965, VGG-16 demonstrated superior capabilities in accurately classifying stroke lesions in CT images. The robust performance of VGG-16 can be attributed to its deep architecture and its ability to capture intricate patterns within the image data. The model's capacity to learn hierarchical features is particularly advantageous in the nuanced task of stroke lesion classification, contributing to its superior performance compared to other models.

In previous studies, Ammar *et al*. compared five deep learning models (ReNet50, VGG-16, Xception, InceptionV3, and InceptionResNetV2), with VGG-16 demonstrating the highest accuracy performance at 96.0 % for intracranial hemorrhage.[8] Worachotsueptrakun compared four deep learning models (AlexNet, VGG-16, GoogleNet, and ResNet), with GoogleNet exhibiting the best performance, achieving accuracy, precision, recall, and F1-score of 92.00%, 94.00%, 83.96%, and 88.70%, respectively.[9]

Chen *et al*. studied four methods (CNN-2, VGG-16, ResNet-50, and ResNet-50 without dropout) and varied the batch size of training models to 8, 16, 32, 64, and 128, respectively.[14] The best performance was observed in CNN-2 and ResNet-50 without dropout, using a batch size of 128 (98.72% accuracy), followed by CNN-2 with a batch size of 32. Chen's work indicates the effect of batch size on the accuracy of training model performance. Additionally, hybrid algorithms have shown high performance. Ozaltin *et al*.[15] used the hybrid technique of OzNet (Deep Learning combined with Decision tree (Machine learning)) and achieved 98.42% accuracy, compared with only OzNet, where the accuracy of the hybrid method surpassed that of the sole deep learning method (87.74% accuracy). Vamsi *et al*. achieved an accuracy of 97.81% using VGG-16 and random forest, representing another successful hybrid approach.[10]

Implications and future directions

The success of VGG-16 in this study underscores the potential of leveraging deep learning techniques to enhance the accuracy and efficiency of stroke diagnosis. The application of such models in the initial screening of stroke cases holds promise for reducing the burden on radiologists and addressing workforce shortages. However, it is crucial to acknowledge certain limitations in our study, such as the reliance on a specific dataset and the need for further validation on diverse datasets to ensure the generalizability of the findings. Additionally, the interpretability of deep learning models remains an ongoing challenge, and efforts to enhance model explainability are essential for fostering trust in clinical applications.

Conclusion

In conclusion, our research provides valuable insights into the comparative analysis of deep learning techniques for stroke lesion classification in CT images. The superior performance of VGG-16 highlights its potential as a valuable tool in the initial screening of stroke cases. As technology continues to advance, further research and collaboration between clinicians and technologists will be essential to harness the full potential of deep learning in improving stroke diagnosis and patient outcomes.

Conflict of interest

None

Funding

This study received support from the Faculty of Allied Health Sciences, with a grant for general researchers (Grant No. AH-66-01-005) and a research grant for the development of students into researchers in 2023, both provided by the Faculty of Allied Health Sciences, Naresuan University.

Ethical approval

This study was approved by the Ethics Committee of Naresuan University, Thailand (IRB No. P1-0165/2565).

References

[1] Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG. Classification of Stroke Subtypes. Cerebrovasc Dis. 2009; 27 (5): 493-501. doi.org/10.1159/000210432

[2] World Health Organization (WHO). World stroke day 2022 [cited 2022 November 17]. Available from: https://www.who.int/srilanka/news/detail/29-10-2022-world-stroke-day-2022

[3] The Ministry of Public Health of Thailand, Public Health Statistics A.D. 2022. [cited 2022 May 14]. Available from: https://spd.moph.go.th/wp-content/uploads/2023/11/Hstatistic65.pdf

[4] González RG. Current State of Acute Stroke Imaging. Stroke. 2013; 44(11): 3260-4. doi.org/10.1161/STROKEAHA.113.003229

[5] Birenbaum D, Bancroft LW, Felsberg GJ. Imaging in Acute Stroke. West J Emerg Med. 2011; 12(1): 67-76. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3088377/

[6] Raghavendra U, Gudigar A, Paul Aritra, Goutham TS, Inamdar MA, Hegde A, *et al*. Brain tumor detection and screening using artificial intelligence techniques: Current trends and future perspectives. Comput Biol Med. 2023; 163: 107063. doi.org/10.1016/j.compbiomed.2023.107063

[7] Yousaf F, Iqbal S, Fatima N, Kousar T, Rahim MSM. Multi-class disease detection using deep learning and human brain medical imaging. Biomed Signal Process Control. 2023; 85: 104875. doi.org/10.1016/j.bspc.2023.104875

[8] Ammar M, Lamria MA, Mahmoudib S, Laidia A. Deep Learning Models for Intracranial Hemorrhage Recognition: A comparative study. Procedia Comput Sci. 2022; 196: 418-25. doi.org/10.1016/j.procs.2021.12.031

[9] Worachotsueptrakun C. A comparison of 3D Convolutional neural network for brain stroke classification with CT scan images. Chulalongkorn University theses and dissertations. 2022. https://digital.car.chula.ac.th/chulaetd/6664

[10] Vamsi B, Bhattacharyya D, Midhunchakkravarthy D, Kim JY. Early Detection of Hemorrhagic Stroke Using a Lightweight Deep Learning Neural Network Model. Trait du signal. 2021; 38(6): 1727-36. doi.org/10.18280/ts.380616

[11] Tasnia N. brain-stroke-prediction-ct-scan-image-dataset. [cited 2022 May 1]. Available from: https://www.kaggle.com/datasets/noshintasnia/brain-stroke-prediction-ct-scan-image-dataset

[12] Vbookshelf. Brain CT Images with Intracranial Hemorrhage Masks. [cited 2022 May 1]. Available from: https://www.kaggle.com/datasets/vbookshelf/computed-tomography-ct-images

[13] Pandey N. Lung segmentation from Chest X-Ray dataset. Kaggle Inc [cited 2022 April 28]. Available from: https://www.kaggle.com/code/nikhilpandey360/lung-segmentation-from-chest-x-ray-dataset/notebook

[14] Chen YT, Chen YL, Chen YY, Huang YT, Wong HF, Yan

JL, Wang JJ. Deep Learning–Based Brain Computed Tomography Image Classification with Hyperparameter Optimization through Transfer Learning for Stroke. Diagnostics 2022; 12, 807. doi.org/10.3390/diagnostics12040807

[15] Ozaltin O, Coskun O, Yeniay O, Subasi A. A Deep Learning Approach for Detecting Stroke from Brain CT Images Using OzNet. Bioeng. 2022; 9(12): 783. doi.org/10.3390/bioengineering9120783