Fırat Üniversitesi Müh. Bil. Dergisi
37(1), 387-400, 2025

Araştırma Makalesi
https://doi.org/10.35234/fumbd.1598597

Stroke Classification in Brain Computed Tomography Images Using Vision Transformers and GAN-based Data Augmentation

Erdem YELKEN[1*], Murat CEYLAN[2]
[1,2] Konya Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Konya, Türkiye
[*1] e218121001002@ktun.edu.tr, [2] mceylan@ktun.edu.tr

(Geliş/Received: 09/12/2024; Kabul/Accepted: 21/03/2025)

Abstract: This study presents an innovative approach to stroke classification. The research utilizes brain computed tomography (CT) images to distinguish between three classes: "no stroke" "ischemic stroke" and "hemorrhagic stroke" employing Vision Transformers (ViTs), a deep learning-based method incorporating attention mechanisms. In this work, ViTs were effectively applied as a powerful method for image-based classification. To enhance model performance, various training strategies and data augmentation techniques were implemented. Specifically, GAN-based architectures such as SRGAN (Super-Resolution GAN) and BSRGAN (Blind Super-Resolution GAN) were used to expand the dataset and improve its diversity. These GAN-based augmentation techniques significantly improved the model's overall performance and classification accuracy. The Vision Transformer model was rigorously evaluated through multi-class classification tasks using a range of performance metrics. In the three-class classification task, the model achieved 99.06% accuracy, 98.18% precision, 98.94% recall, and a 98.54% F1-score. For the binary classification of ischemic vs. hemorrhagic stroke, the model reported 99.78% accuracy, 99.02% precision, 99.66% recall, and a 99.26% F1-score. In the binary classification of stroke presence, the model achieved 98.68% accuracy, 97.80% precision, 98.54% recall, and a 98.14% F1-score. These findings demonstrate the potential of Vision Transformers to assist in faster and more reliable stroke diagnosis and highlight their contribution to the development of decision support systems in medical applications.

Key words: Stroke classification, vision transformers, CT imaging, data augmentation, deep learning.

Beyin Bilgisayarlı Tomografi Görüntülerinde Görü Dönüştürücüler ve GAN Tabanlı Veri Artırma Kullanılarak İnme Sınıflandırması

Öz: Bu çalışma, inme sınıflandırması için yenilikçi bir yaklaşım sunmaktadır. Araştırmada, beyin bilgisayarlı tomografi (BT) görüntüleri kullanılarak "inme yok", "iskemik inme" ve "hemorajik inme" olmak üzere üç farklı sınıfı ayırt etmeyi amaçlayan, dikkat mekanizmaları içeren derin öğrenme tabanlı bir yöntem olan Görü Dönüştürücüler (GD) uygulanmıştır. GD modelleri, bu çalışmada görüntü verisi sınıflandırması için güçlü ve etkili bir yöntem olarak kullanılmıştır. Modelin performansını artırmak amacıyla çeşitli eğitim stratejileri ve veri artırma teknikleri uygulanmıştır. Özellikle, GAN tabanlı SRGAN (Süper Çözünürlük GAN) ve BSRGAN (Kör Süper Çözünürlük GAN) mimarileri, veri setini genişletmek ve çeşitliliği artırmak için kullanılmıştır. Bu GAN tabanlı artırma teknikleri, modelin genel başarımını ve sınıflandırma doğruluğunu önemli ölçüde iyileştirmiştir. Görü Dönüştürücü modeli, çok sınıflı sınıflandırma görevleri kapsamında çeşitli performans ölçütleriyle kapsamlı biçimde değerlendirilmiştir. Üç sınıflı sınıflandırma görevinde model, %99,06 doğruluk, %98,18 hassasiyet, %98,94 duyarlılık ve %98,54 F1 skoru elde etmiştir. Hemorajik ve iskemik inme sınıflandırmasında modelin doğruluğu %99,78, hassasiyeti %99,02, duyarlılığı %99,66 ve F1 skoru %99,26 olarak raporlanmıştır. İkili "inme var/yok" sınıflandırmasında ise model, %98,68 doğruluk, %97,80 hassasiyet, %98,54 duyarlılık ve %98,14 F1 skoru elde etmiştir. Bu bulgular, Görü Dönüştürücüler'in hızlı ve güvenilir inme teşhisine katkı sunma potansiyelini ve tıbbi uygulamalarda karar destek sistemlerinin gelişimine önemli ölçüde katkı sağlayabileceğini göstermektedir.

Anahtar kelimeler: İnme sınıflandırması, görü dönüştürücüler, BT görüntüleme, veri artırma, derin öğrenme.

1. Introduction

Stroke is a condition that occurs when blood flow to the brain tissue is suddenly interrupted or reduced and is one of the leading causes of death and disability worldwide [1]. Strokes are divided into two main categories: ischemic stroke (blockage of brain vessels) and hemorrhagic stroke (bleeding due to rupture of brain vessels) [2]. According to the World Health Organization (WHO), stroke affects approximately 15 million people each year, 5

* Corresponding author: e218121001002@ktun.edu.tr. Authors' ORCID Numbers: [1] 0000-0001-9307-2959, [2] 0000-0001-6503-9668

millions of whom are permanently disabled [1]. The recovery process after a stroke is quite complex and long-term, so early diagnosis and accurate treatment planning are vital [3,4].

The main methods used today for stroke diagnosis include imaging techniques such as CT, magnetic resonance imaging (MRI), and cerebral angiography [3]. CT imaging is usually the first choice because it is fast and widely available [4]. However, each of these methods has its advantages and limitations. For example, CT scans may be limited in detecting early blood flow changes in brain tissue, such as ischemic stroke, because these changes usually take some time to become visible on CT [5]. Magnetic resonance imaging can provide higher-resolution images and more clearly reveal differences in brain tissue. However, the high cost of MRI devices and the longer scanning process can disadvantage patients with time constraints in emergencies [6].

The rapid development of artificial intelligence and deep learning techniques has enabled significant advances in medical imaging and has provided new-generation alternatives to traditional analysis methods [7,8]. In this field, convolutional neural networks (CNNs), in particular, are widely used in stroke classification in the medical image field because of their high accuracy and generalization capacity [9,10,11,12]. However, Vision Transformers (VTs) have recently attracted more attention in image processing and have shown superior performance compared to CNNs in some tasks [13]. Vision Transformers are based on the transformer architecture that derives its foundation from the successes in natural language processing (NLP) and optimizes information processing by using attention mechanisms in image processing processes [14,15]. These models can capture prominent features with attention mechanisms, mainly by providing effective learning in large data sets, and can model relationships in a global context more effectively, unlike ESAs [16]. Thus, they offer significant advantages in image analysis in medical imaging [17].

In recent years, deep learning methods in medical imaging have made significant progress in the diagnosis of serious neurological conditions such as stroke by learning on large data sets [18,19]. Okimoto et al. conducted a study aimed at improving the visualization of acute brain infarcts in CT images using deep learning techniques [20]. Zhu et al. emphasized the potential of artificial intelligence to increase diagnostic support in healthcare services in their study examining how deep learning models were applied to CT and MRI images for ischemic and hemorrhagic stroke classification [21]. Miyamoto et al. showed that an artificial intelligence-supported system could be effective in increasing the diagnosis and classification accuracy in directing stroke patients to the right treatment, especially in hospitals where there is no neurologist, and reported high success rates with 88.7% and 86.1% accuracy rates [22]. Shakunthala et al. classified ischemic and hemorrhagic stroke using magnetic resonance imaging (MRI) data and showed that this approach achieved 94.8% accuracy [23]. Altıntas et al. classified ischemic, hemorrhagic, and normal brain CT images using deep learning models and reported that the AlexNet model achieved 90.89% accuracy with transfer learning approaches [24].

Recent studies have comparatively evaluated different deep learning approaches for stroke classification in brain CT images. Koska et al. performed stroke classification using the dataset provided in the TEKNOFEST 2021 artificial intelligence health competition, which was also used in this study, and reported that it achieved 91% accuracy with MobilNetV2 and 88% accuracy with EfficientNetB0 [25]. Cinar et al. developed a hybrid approach by combining different models such as EfficientNetB0, ResNet50, VGG19 with machine learning algorithms and showed that the EfficientNetB0 + SVM model achieved 95.13% accuracy [26]. On the other hand, there are also studies in the literature that address the effectiveness of ViT models for brain stroke diagnosis. Katar et al. tested the success of stroke classification in unbalanced and balanced data scenarios using the ViT model and showed that 98.75% accuracy was achieved with data augmentation methods [27]. In addition, the Grad-CAM algorithm visualized the regions where the model focused and revealed that it could increase the clinical interpretability of the model. Cinar et al. compared models such as EfficientNetB0, ResNet101, VGG19, MobileNet-V2 and GoogleNet with transfer learning methods and reported that the EfficientNet-B0 model showed the best performance with an accuracy rate of 97.93% [28].

Vision Transformers have shown superior performance in various medical imaging applications and their success has been confirmed by many studies in the literature [29, 30]. For example, in the study by Li et al. it was shown that Vision Transformers can be used as an effective tool in the diagnosis of Alzheimer's disease in brain MRI images [31]. The same study suggested that Vision Transformers can further improve model performance when combined with different data augmentation techniques. In addition, it has been stated that Vision Transformers provide high accuracy in chest X-rays in the diagnosis of COVID-19 [32] and have been successfully applied in two-dimensional CT images in the diagnosis of pancreatic cancer [33]. In addition, studies have been conducted where Vision Transformers are used for medical imaging purposes in stroke assessments and yielded effective results. Ayoub et al. achieved an overall accuracy rate of 87.51% in classifying brain CT scan slices using the ViT architecture and showed high sensitivity in locally identifying stroke patients [34]. Abbaoui et al. The study conducted by [35] demonstrated the superior performance of the ViT-b16 model in ischemic stroke

classification. These findings indicate that Vision Transformer models have a wide potential in the field of medical imaging and their performance can be further improved with various data augmentation techniques.

This study aims to increase the accuracy of stroke diagnosis using the Vision Transformer model. This study, conducted on the Ministry of Health stroke dataset, aims to classify three different classes, namely “no stroke” “ischemia” and “hemorrhagic” In the study, a comparison of traditional data augmentation methods with Generative Adversarial Network (GAN) based data augmentation methods was made. In addition, the performance of the model was evaluated with 5-fold cross-validation. This study aims to contribute significantly to the literature by revealing the effect of using Vision Transformers and GAN-based data augmentation techniques in stroke diagnosis. The findings may contribute to healthcare professionals making more accurate diagnoses and improving treatment plans.

2. Materials and Methods

2.1. Dataset and preprocessing

In this study, a publicly available brain CT dataset belonging to the Ministry of Health of the Republic of Turkey and compiled by Koc et al. was used. The dataset is fully anonymized and does not contain any personal or clinical data. Accordingly, due to the anonymous nature of the dataset, its use in academic research does not constitute any ethical violation. The dataset consists of a total of 6650 computed tomography images. 4427 of these images are non-stroke images and 2223 are stroke images. Images containing stroke were classified according to stroke types; 1130 of them are ischemic type stroke and 1093 of them are hemorrhagic type stroke [36].

The dataset was created with data collected from the e-Pulse and Teleradiology Systems of the Ministry of Health of the Republic of Turkey. The images were reviewed in detail by 7 radiologists over a 6-week period and divided into three categories: “no stroke”, “ischemic stroke” and “hemorrhagic stroke”. Stroke regions in images containing stroke were marked by experts and checked for accuracy. The data provided by the General Directorate of Health Information Systems of the Ministry of Health are presented in DICOM and PNG formats with a size of 512×512 pixels. The dataset contains three different types of images: original images, mask images and superimposed versions of mask images on original images [36].

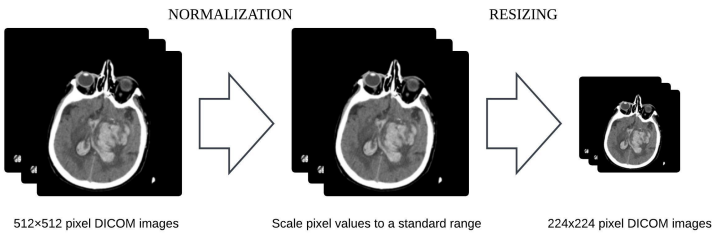

Figure 1. Data preprocessing flowchart.

The images in the dataset were subjected to a standard preprocessing process for analysis. As shown in Figure 1, the images were first normalized in this process. Normalization aims to scale the pixel values of the images to a specific range and thus enable the model to learn more consistently and effectively. Then, all the images were resized to the same size; this step facilitates the process by standardizing the input dimensions used in the model’s training. Finally, the image data was converted into tensors. Tensors are multidimensional data structures that deep learning models can process, and representing image data in this format allows the model to process the data more effectively. Data preprocessing steps help the model process the data more effectively and increase stroke classification performance.

2.2. Data augmentation techniques

Data augmentation techniques are widely used to increase the diversity of training data, improving the generalization capabilities of deep learning models and reducing the risk of overfitting. In this study, both traditional data augmentation methods and GAN- based data augmentation techniques are applied.

2.2.1. Traditional data augmentation methods

Traditional data augmentation methods aim to create new images by applying various transformations on existing images. These techniques increase the overall performance of the model by allowing it to see a wider variety of data and reduce the risk of overfitting. Horizontal flipping helps the model learn and recognize symmetrical patterns by flipping the images horizontally with a 50% probability. In a study conducted by Shorten and Khoshgoftaar, it was stated that horizontal flipping is a frequently preferred data augmentation method in the field of medical image processing and increases the generalization ability of the model. Applying random rotations between -30 and 30 degrees to the images allows the model to better understand and process images from different angles. The use of such rotation angles is among the techniques widely recommended in data augmentation strategies [32]. Cropping and padding techniques make it easier for the model to generalize in regions with missing information by randomly removing or adding up to 20% of the image. Brightness settings increase the model's robustness to different lighting conditions by changing the image brightness between 80% and 120% and provide consistent performance in various scenarios. It was also emphasized in the same study that such brightness adjustments are effective in increasing the robustness of medical images to different scanning conditions. These measures play an important role in the processing of medical images obtained under different scanning conditions [37,38].

2.2.2. GAN based data augmentation

Generative Adversarial Networks (GANs) are deep learning models that enable two neural networks, the generator and the discriminator, to compete with each other to produce realistic data [39]. In this study, the GAN-based data augmentation method was evaluated comparatively with traditional data augmentation methods. Although traditional methods (reflection, rotation, brightness change, etc.) can increase the generalization capacity of the model to a certain extent, they are limited to the variations in the existing data set and do not have the ability to produce new examples. In contrast, GANs increase the diversity of the data set by creating synthetic images similar to the original data distribution and contribute to the model's better learning of rare classes. Especially since data limitation is a common problem in the field of medical imaging, GAN-based data augmentation methods offer a significant advantage for data sets with limited sample size.

GAN-based data augmentation techniques increase the generalization performance of the model during the training process, but they also have some disadvantages. First, the training of GANs requires high computational cost compared to traditional data augmentation methods. In addition, GANs may have a problem that can cause the generated synthetic images to get stuck in certain patterns. In addition, it should be evaluated whether the generated images are clinically realistic. Therefore, the GAN-assisted data augmentation process has been meticulously optimized and carefully implemented to increase the generalization ability of the model. In this study, two different GAN architectures were used.

2.2.2.1. SRGAN (Super-Resolution GAN)

SRGAN is an advanced GAN architecture that produces high-resolution images from low-resolution images. This model is trained to obtain high-resolution versions of low-resolution images, and in this process, a 4-fold augmentation is performed using pre-trained VGG19 model weights [40]. SRGAN allows more detailed analyses by increasing the resolution, especially in medical imaging. In Figure 2 and Figure 3, examples produced by SRGAN show the results of ischemic and hemorrhagic data sets, respectively. SRGAN contributes to the model's training process and increases its accuracy with the high-resolution images it produces. In addition, the images produced by SRGAN attract attention with their high PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) values; this reveals the quality of the produced images and their closeness to the original.

2.2.2.2. BSRGAN (Blind Super-Resolution GAN)

BSRGAN is an improved version of the SRGAN model and is capable of producing more realistic and diverse high-resolution images. BSRGAN produces more robust and generalizable images by better modeling the noise and distortions present in the training data. In the training of BSRGAN, a special training strategy was applied to obtain four times higher resolution versions from low resolution images and hyperparameters such as loss function, learning rate and filter sizes were meticulously tuned and optimized to increase the accuracy of the model. The

most obvious advantage of BSRGAN is that it processes the noise and distortions in the images more effectively, making low quality and corrupted data more realistic. This method increases the ability of the model to generalize better under different data conditions and improves its performance. Examples generated by BSRGAN in Figure 2 and Figure 3 show the results for ischemic and hemorrhagic data sets. Images generated by BSRGAN, just like SRGAN, have high PSNR and SSIM values; This supports the quality of the generated images and their similarity with real images [41].

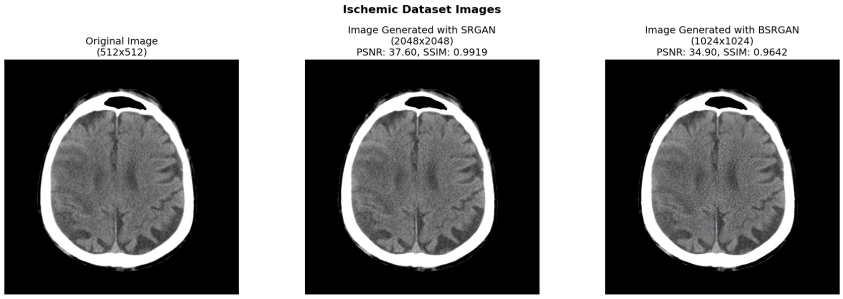

Figure 2. Example images of the ischemic dataset obtained with GAN-Based data augmentation methods.

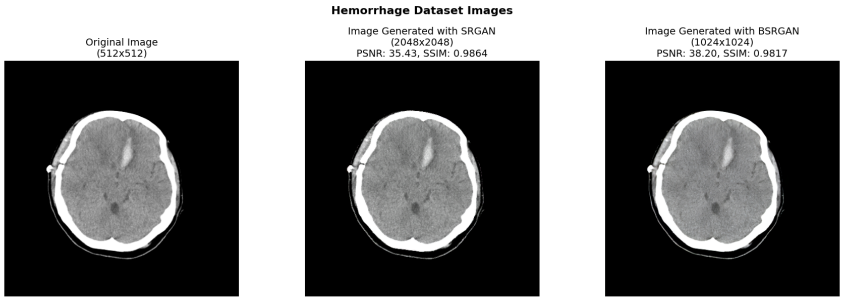

Figure 3. Example images of the hemorrhage dataset obtained with GAN-Based data augmentation methods.

Although SRGAN and BSRGAN models perform similar tasks, they offer different data augmentation strategies. While SRGAN produces more detailed and clear images by increasing the resolution, BSRGAN provides more accurate and realistic data by considering the noise and distortions in the images. While the images obtained with SRGAN enable the model to learn fine details, the data produced by BSRGAN increases the model's classification accuracy in low-quality and distorted images. In addition, the high PSNR and SSIM values of the images produced with these methods increase the image quality and model accuracy [42]. The use of these two models together expands the generalization capacity of the model by representing the various variations in the datasets more comprehensively and improves the classification performance. This allows more reliable and sensitive results in medical imaging studies.

Table 1. Number of Samples of the Dataset According to Data Augmentation Methods.

Data Augmentation Method	No Stroke	Ischemic	Hemorrhagic	Total
None	4427	1130	1092	6649
Traditional	4427	4427	4427	13281
GANs	4427	4369	4369	13165

The dataset has three different classes: no-stroke, ischemic, and hemorrhagic, and the number of samples for each class is given in Table 1. When data augmentation methods are not used, the imbalance between the classes is noticeable. To reduce the class imbalance, classical data augmentation methods and GAN architecture were applied. This shows that different data augmentation techniques play an important role in ensuring the balance of the dataset.

2.3. Vision transformers

Vision Transformers have emerged as a significant innovation in the field of image processing in recent years. This model was developed by adapting the transformer architecture, which has achieved great success in the field of natural language processing (NLP), to image processing tasks. Unlike traditional CNNs, the VIT model processes images in small patches and allows these patches to be analyzed sequentially with attention mechanisms. This method allows the model to preserve the sequential information of these patches by taking each image patch as a separate input and adding positional coding. Attention layers learn the relationships of these patches with each other and use multi-head attention mechanisms to capture important features. In this way, ViT models can effectively learn on large datasets and better represent complex features [43,44].

In the input layer of the model, images are divided into patches of a certain size (e.g., 16x16 pixels) and these patches are converted to a vector by linear projection [43]. Positional coding is added to preserve the sequential information of these patches, as in the transformer architecture. Then, attention layers process these vectors with multi-header attention mechanisms. Each attention header is used to learn different types of information and capture important features. The information from the attention layers is represented at a higher level through a feedforward neural network (MLP) and the features obtained in the final stage are transmitted to an output layer for classification. Each of these stages, the steps of the model from input to output and the operations performed in each step are given in Figure 4.

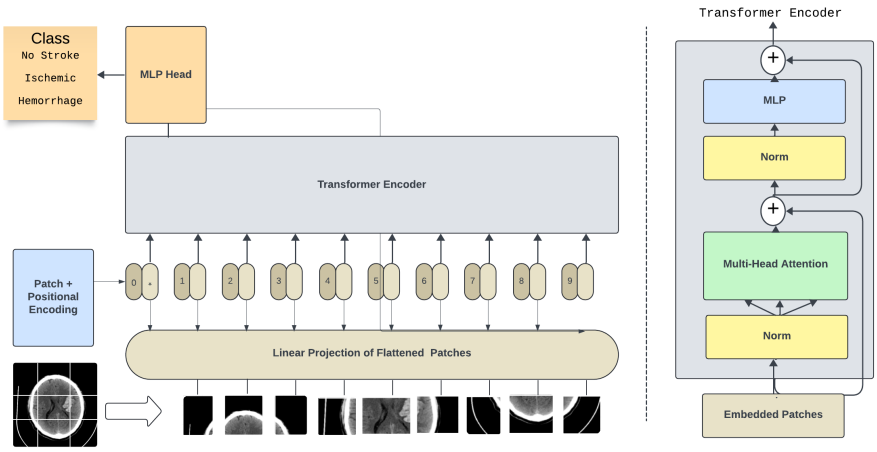

Figure 4. Vision Transformers model overview.

This study applies various strategies and configurations to train the VIT model. A pre-trained VIT model on a large-scale dataset is utilized in the first stage. In particular, the google/vit-base-patch16-224-in21k model, which is trained on the ImageNet-21k dataset, is taken as a basis (https://huggingface.co/google/vit-base-patch16-224-in21k, accessed on May 25, 2024). ImageNet-21k is a large-scale dataset containing 14 million images with a resolution of 224×224 pixels and approximately 22,000 classes, and the training performed on this dataset increases the general feature extraction ability of the model [40]. This approach is built on the transfer learning methodology, which enables the use of the knowledge acquired in the source domain to improve the performance of the target domain [45]. During the training process, the input images were brought to a standard resolution of 224×224 pixels, aiming for the model to work on a more homogeneous and consistent input set. This standardization was considered an important step towards increasing the model's classification accuracy and overall performance.

2.4 Training processes

The training processes of the ViTs model are an important stage used to increase the success of deep learning models. In this process, it is aimed to carefully determine the parameters of the model, to apply strategies to prevent overfitting during training, and to optimize the generalization ability of the model. The model was trained on the CentOS 7 operating system with three NVIDIA GeForce RTX 2080 (total 24 GB VRAM) and 24 GB RAM.

In order to evaluate the performance of the model and prevent overfitting, 5-fold cross-validation was used. This method allows the model to be trained on each part by dividing the dataset into five equal parts. In this way, it is a useful approach to understand how well the model generalizes on different data parts.

The model parameters used in the training process were carefully selected to optimize the performance and generalization ability of the model. The parameters in Table 2 were determined by considering both literature studies and previous experimental results. The learning rate was selected as 2×10^{5}, and this value was optimized to ensure the stability of the model during the learning process. Parameters such as the number of hidden layers and the number of attention heads were determined to prevent overfitting while increasing the capacity of the model. Different parameter combinations were tested in the trial-and-error process, and the structure that gave the best results was adopted. In this process, the accuracy values on the training sets were monitored.

Table 2. Vision Transformer Model Parameters for Training.

Model Parameters	Parameter Values
Learning Rate	2×10^{-5}
Number of Epochs	100
Number of Hidden Layers	12
Number of Attention Heads per Encoder Layer	12
Patch Size	16
Weight Decay	0.01

The hyperparameters specified in Table 2 were carefully determined to provide an optimum balance between the complexity of the model, the learning process, and the generalization ability. The learning rate was manually adjusted to ensure the stability of the model in the optimization process, and the number of hidden layers and the number of attention heads were determined to increase the learning capacity of the model. Parameters such as the learning rate, the number of epochs, and the number of hidden layers were optimized to prevent the model from overfitting and to obtain the best generalization ability. Different hyperparameter combinations were tested experimentally, and the structure that gave the best results was selected using the trial-and-error method. In this process, the accuracy rate of the model, loss function values, and generalization performance were analyzed to determine the most suitable structure.

The training process began by dividing the dataset into two as training and validation. The training data was used for the learning process of the model, and the validation data was separated to evaluate the generalization performance of the model. During the training of the model, the performance of the model was measured via the loss function at the end of each epoch. The loss function used in this process was optimized to minimize the accuracy rate of the model's predictions. The early stopping technique was also used during training. This technique aimed to prevent overfitting by stopping training when the performance on the validation set did not improve. In this process, the epoch that gave the best performance was recorded and training was completed.

The results obtained during training clearly demonstrate the success of the model in the multi-classification task and the effects of different data augmentation methods on the model performance. In the graphs presented in Figure 5, the learning curve of the model is shown depending on the accuracy value. In the graphs, three different data augmentation methods for multi-classification were examined: No Data Augmentation, Traditional Data Augmentation and GAN-Based Data Augmentation. This comparison clearly demonstrates the effects of different data augmentation techniques on the training accuracy of the model.

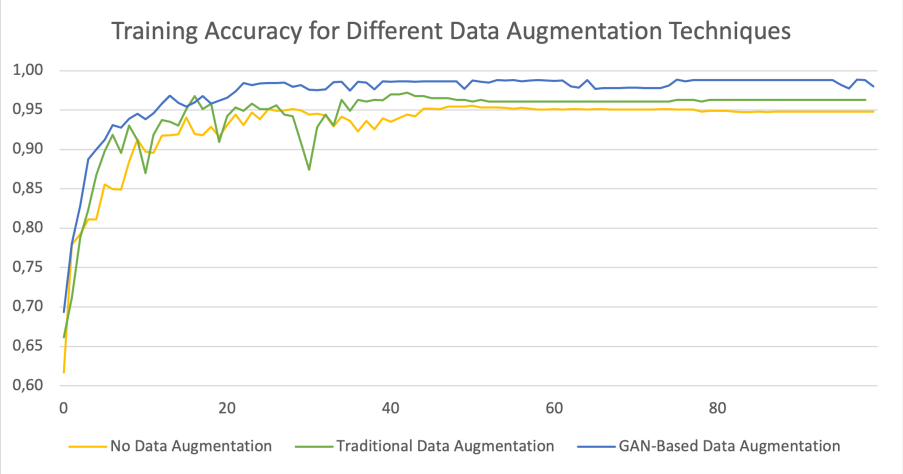

Figure 5. Training accuracy for different data augmentation techniques across epochs.

3. Results

This section presents in detail the results of binary and multi-class classification experiments, the effects of data augmentation techniques on model performance, and the findings.

Key metrics such as accuracy, precision, recall, and F1 score were utilized to evaluate model performance. True Positives (TP) indicate correctly predicted positive instances, whereas True Negatives (TN) correspond to correctly predicted negative instances. Conversely, False Positives (FP) represent incorrectly predicted negative instances as positive, and False Negatives (FN) indicate incorrectly predicted positive instances as negative. These metrics are mathematically defined as given in Equation 1, Equation 2, Equation 3, and Equation 4:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1\ Score = 2\ x\ \frac{Precision\ x\ Recall}{Precision+Recall} \tag{4}$$

3.1. Binary classification results

The binary classification results were evaluated on two distinct classification problems. The first problem involved distinguishing between ischemic and hemorrhagic stroke cases, while the second problem focused on classifying whether a stroke was present or not (stroke vs. no stroke). This evaluation was conducted to assess the model's classification performance across both categories comprehensively. The results for each classification category are summarized in Table 3 and Table 4.

Table 3 shows the binary classification results performed on hemorrhagic and ischemic stroke categories. The accuracy, precision, recall and F1 score values obtained using data augmentation methods were calculated by 5-fold cross-validation method, and the mean values and standard deviations of the relevant metrics are presented in the table. The results reveal that GAN-based data augmentation method has significantly higher performance than other methods. In particular, the accuracy, precision, recall and F1 score values obtained with GAN-based data augmentation were calculated as 99.78 ± 0.3, 99.02 ± 0.4, 99.66 ± 0.3 and 99.26 ± 0.4, respectively. Standard deviations (±) show the variations of the model in different validation layers.

Table 3. Binary classification results (Hemorrhagic and İschemic Stroke).

Data Augmentation Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
None	94.48 ± 0.1	93.66 ± 0.3	94.28 ± 0.2	94.00 ± 0.1
Traditional	95.96 ± 0.2	95.12 ± 0.3	95.82 ± 0.3	95.38 ± 0.2
GAN	99.78 ± 0.3	99.02 ± 0.4	99.66 ± 0.3	99.26 ± 0.4

Figure 6 is constructed using the average values of the confusion matrices obtained from the 5-fold cross-validation processes and represents the generalization performance of the model.

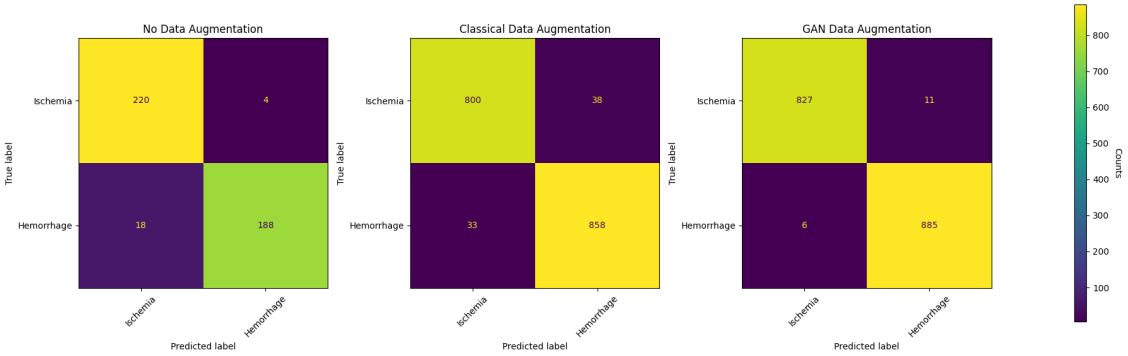

Figure 6. Binary classification confusion matrices (Hemorrhagic and Ischemic Stroke).

Table 4 summarizes the binary classification results performed in the categories of stroke and no stroke. The results obtained with the 5-fold cross-validation method again show that the GAN-based data augmentation method provides the highest performance. In this category, the accuracy, precision, recall and F1 score values obtained with GAN-based data augmentation are determined as 98.68 ± 0.2, 97.80 ± 0.3, 98.54 ± 0.3 and 98.14 ± 0.3, respectively. Standard deviations (±) show the variations of the model in different validation layers.

Table 4. Binary classification results (No-stroke and Stroke).

Data Augmentation Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
None	94.28 ± 0.3	93.38 ± 0.2	94.08 ± 0.5	93.68 ± 0.4
Traditional	95.80 ± 0.4	94.88 ± 0.2	95.60 ± 0.3	95.18 ± 0.2
GAN	98.68 ± 0.2	97.80 ± 0.3	98.54 ± 0.3	98.14 ± 0.3

Figure 7 is constructed using the mean values of the confusion matrices obtained from the 5-fold cross-validation processes and represents the generalization performance of the model.

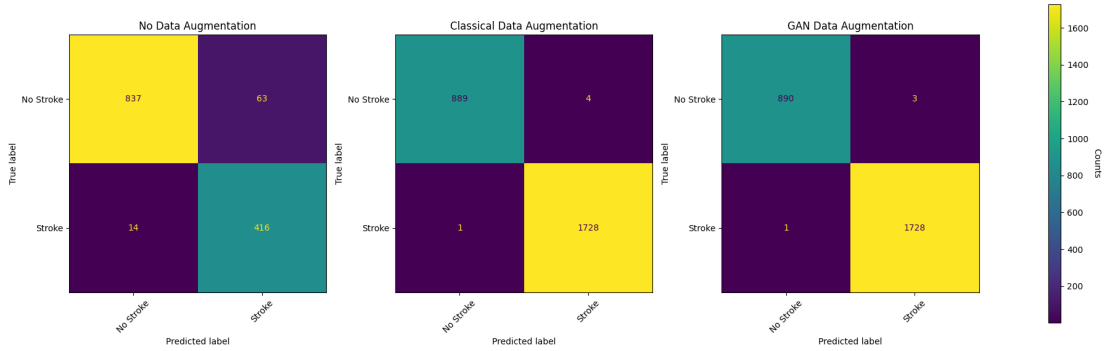

Figure 7. Binary classification confusion matrices (Stroke and No Stroke).

3.2. Multiple classification results

Table 5 shows the multiple classification results obtained using different data augmentation methods. The results are calculated with 5-fold cross-validation, and the accuracy, precision, recall and F1 score metrics are presented separately for each class. In order to better analyze the effect of data imbalance on model performance, weighted metrics are also evaluated in the case without data augmentation. The results in the table show that the GAN-based data augmentation method significantly improves the model performance and reduces the performance differences between classes, especially in imbalanced datasets.

Table 5. Multiple classification results: No Stroke, Ischemic, Hemorrhagic.

Data Augmentation Method	Class	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
None	No Stroke	95.70 ± 0.3	97.73 ± 0.5	95.84 ± 0.4	96.86 ± 0.3
	Hemorrhagic	98.04 ± 0.4	95.41 ± 0.4	94.11 ± 0.2	94.11 ± 0.4
	Ischemic	95.86 ± 0.2	82.96 ± 0.4	92.23 ± 0.4	87.35 ± 0.4
Traditional	No Stroke	99.33 ± 0.4	99.44 ± 0.6	99.88 ± 0.5	99.66 ± 0.2
	Hemorrhagic	86.28 ± 0.2	90.99 ± 0.2	94.33 ± 0.2	92.63 ± 0.2
	Ischemic	86.16 ± 0.2	94.55 ± 0.2	90.66 ± 0.2	92.56 ± 0.2
GAN	No Stroke	99.89 ± 0.2	99.89 ± 0.2	99.78 ± 0.2	99.83 ± 0.5
	Hemorrhagic	99.16 ± 0.3	98.57 ± 0.3	98.81 ± 0.4	98.69 ± 0.3
	Ischemic	99.12 ± 0.3	98.76 ± 0.3	98.65 ± 0.3	98.79 ± 0.3

Figure 8 is constructed using the average values of the confusion matrices obtained from the 5-fold cross-validation processes and represents the generalization performance of the model.

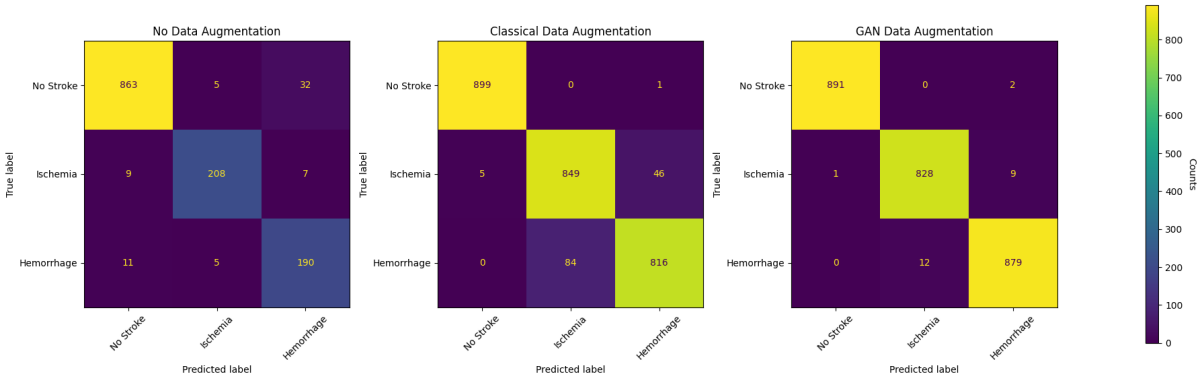

Figure 8. Multiple classification confusion matrices.

The impact of traditional and GAN-based data augmentation methods on model performance is clearly demonstrated by the tables and graphs presented above. While traditional data augmentation methods improve the generalization ability of the model by increasing the diversity of the dataset, GAN-based data augmentation methods provide even higher performance thanks to the generation of high-quality and diverse data. The overall performance of the VITs model has been significantly increased by the effect of data augmentation techniques. While the results obtained without data augmentation show how the model performs on a limited dataset, traditional and GAN-based data augmentation methods have proven to be effective tools to maximize the potential of the model. In particular, GAN-based data augmentation provided significant improvements in the accuracy, precision, recall and F1 scores of the model. In particular, GAN-based data augmentation significantly increases the model performance on imbalanced datasets, indicating that it can be a reliable approach in clinical applications.

4. Discussion

In this study, integrating ViTs with GAN-based data augmentation methods for stroke classification have provided a significant performance improvement compared to existing approaches in the literature. GAN-based data augmentation techniques, used to address the lack of diversity in datasets with limited size, have enriched the model's learning capacity and increased the representational power of the dataset, enabling superior classification results. The attention mechanisms within the ViT model have captured fine details, especially in low-contrast brain CT images, elevating the model's accuracy to higher levels.

The originality of this study lies in the ViT model's ability to classify ischemic, hemorrhagic, and normal conditions effectively and reliably detect the presence of a stroke. When compared to studies on the same dataset in the literature, it was found that the GAN-supported data augmentation method significantly increased the model's accuracy and introduced an innovative approach to classification performance. Remarkably, the synthetic data generated by the GAN were observed to enhance the model's generalization capacity despite limited data sources, making a meaningful contribution to the learning process.

The comparative results presented in Table 6 demonstrate that the proposed model exhibits competitive and successful performance compared to previous studies utilizing the same dataset in the literature. Specifically, the ViT-based model, enhanced by effective attention mechanisms and GAN-supported data augmentation strategies, has achieved higher classification accuracy than existing methods. The Advanced D-UNet model developed by Yalcin and Vural attained accuracy rates of 98.9% and 98.5% for stroke/no-stroke and hemorrhagic/ischemic classification tasks, respectively. In contrast, the proposed ViT model, when integrated with GAN-based data augmentation, achieved accuracy rates of 99.7% and 98.6% for these tasks, indicating comparable or superior performance to previous approaches. Similarly, Karatas et al. conducted a two-step classification process. In the first step, Stroke "(CVD) / No Stroke (Normal)" classification was performed, achieving 99.72% accuracy. In the second step, hemorrhagic and ischemic stroke classification was conducted, reaching 99.51% accuracy. The proposed model attained results close to this level, demonstrating its competitive performance in stroke classification. The model developed by Cinar et al. using an EfficientNetB0 and SVM combination reported an accuracy of 95.13%, while the study by Koska et al. reported accuracy rates of 91.0% and 88.0% for MobileNetV2 and EfficientNetB0 models, respectively. Given these comparisons, the proposed model demonstrates superior performance over these studies.

The results obtained from the proposed model indicate that, compared to existing deep learning-based methods in the literature, the effective integration of attention mechanisms and GAN-supported data augmentation contributes to higher generalization capability. Particularly in imbalanced datasets, the GAN-based data augmentation method has been shown to improve class balance and enhance model performance. Additionally, the use of attention mechanisms not only increases accuracy rates but also ensures more stable and reliable classification results. These findings highlight the significant contribution of the proposed approach to improving the accuracy and reliability of deep learning-based models for stroke diagnosis. Future research directions should focus on facilitating the integration of this model into clinical applications and evaluating its generalization capacity on large-scale datasets.

In this study, comparisons were made with traditional data augmentation methods in order to examine the effect of synthetic data generated by GANs on model performance. The obtained results show that GAN-based data augmentation improves the class balance especially in imbalanced data sets and increases the accuracy rate of the model. However, there are some limitations of the proposed method. GAN-based data augmentation requires high computational costs; this requires more processing power and time especially during the training process and may create operational difficulties when integrating the model into clinical applications. In addition, since the data set used represents a limited patient population, the generalization capacity of the model should be evaluated on different clinical settings. This reveals the need to test the model on larger data sets and emphasizes the importance of future studies to examine its performance on different patient groups. In the future, optimization techniques can be applied to run the model at lower costs and comparative analyses can be performed with different data sets.

The main objective of this study is to evaluate the effect of the ViT based model and GAN supported data augmentation methods on stroke classification performance. Therefore, Explainable Artificial Intelligence (XAI) methods were not directly applied in the study. However, in future studies, the decision mechanism of the model can be analyzed in more detail by using XAI techniques such as Grad-CAM and SHAP. Thus, it can be visualized which regions the model takes into account more and contribute to the interpretation of clinical experts.

Table 6. Classification results obtained in the literature and in this study using the stroke dataset.

Study	Proposed Model	Problem	Accuracy (%)
Yalcin and Vural (2022) [47]	Advanced D-UNet	Binary Classification (Stroke vs. No Stroke)	98.9
		Binary Classification (Hemorrhagic vs. Ischemic)	98.5
Karatas et al. (2022) [48]	ResNet50	Binary Classification (Stroke vs. No Stroke)	99.7
	Inception-v3	Binary Classification (Hemorrhagic vs. Ischemic)	99.5
Cinar et al. (2023) [26]	EfficientNetB0 + SVM	Binary Classification (Hemorrhagic vs. Ischemic)	95.1
Cinar et al. (2023) [28]	EfficientNet-B0	Binary Classification (Hemorrhagic vs. Ischemic)	98.7
Katar et al. (2023) [27]	Vision Transformer (ViT)	Binary Classification (Stroke vs. No Stroke)	98.7
Koska et al. (2024) [25]	MobilNetV2 / EfficientNetB0	Binary Classification (Hemorrhagic vs. Ischemic)	91.0 / 88.0
This Study	Vision Transformer with Traditional Data Augmentation	Binary Classification (Hemorrhagic vs. Ischemic)	96.4
		Binary Classification (Stroke vs. No Stroke)	96.3
		Multi-class Classification (No Stroke, Hemorrhagic, Ischemic)	96.3
	Vision Transformer with GAN-based Data Augmentation	Binary Classification (Hemorrhagic vs. Ischemic)	98.6
		Binary Classification (Stroke vs. No Stroke)	99.7
		Multi-class Classification (No Stroke, Hemorrhagic, Ischemic)	99.0

As a result, the proposed ViT model offers an innovative approach in the field of stroke classification, demonstrating a performance that can compete with powerful models in the existing literature such as Advanced D-UNet and ResNet50. The high accuracy rates achieved by the model support fast and accurate stroke diagnosis and are promising for early intervention in clinical applications. The proposed method, thanks to the ViT model combined with GAN-supported data augmentation, has the potential to obtain reliable and stable results even under limited data conditions, creating a strong foundation for further research in the field of medical imaging.

5. Conclusion

This study presents an innovative approach in stroke classification by combining the VT model with GAN based data augmentation techniques. The experimental results show that the proposed method achieves high accuracy rates in both binary and multi-class classification tasks and exhibits a competitive performance compared to other deep learning-based approaches in the existing literature. In particular, the GAN-supported data augmentation method has improved the class balance in imbalanced data sets and increased the generalization ability of the model.

One of the most important contributions of the proposed model is that it can better capture low-contrast stroke lesions using the attention mechanism of VIT compared to traditional CNN based approaches. In addition, successful results have been obtained even with limited and imbalanced data sets thanks to GAN-based data augmentation. Compared to similar studies in the literature, the proposed method stands out with its high accuracy rates and offers a potential use in clinical decision support systems.

The findings of this study show that by increasing the effectiveness of deep learning techniques in the field of medical imaging, a more reliable and faster approach can be provided in early diagnosis and clinical decision processes. In particular, since timely diagnosis of stroke is of vital importance, the classification success of the developed model can provide a critical advantage in terms of patient management. In future studies, it is planned to test the model with data sets containing larger and more diverse patient populations. In addition, by integrating Explainable Artificial Intelligence methods, the decision mechanism of the model can be better understood and made interpretable for clinical experts. In addition, it is recommended to conduct comprehensive tests with hospital collaborations to evaluate the applicability of the model in real clinical settings. In conclusion, this study shows that the combined use of Vision Transformers and GAN-based data augmentation methods provides an effective

and reliable method for stroke diagnosis. The findings provide an important basis for the future potential use of VIT-based models in medical artificial intelligence applications.

References

[1] World Health Organization. The top 10 causes of death. Retrieved from https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, 2018.

[2] Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R et al. Heart disease and stroke statistics—2017 update: a report from the American Heart Association. Circulation 2017; 135(10): e146-e603.

[3] Donnan GA, Fisher M, Macleod M, Davis SM. Stroke. Lancet 2008; 371(9624): 1612-1623.

[4] González RG. Clinical MRI of acute ischemic stroke. J Magn Reson Imaging 2012; 36(2): 259-271.

[5] Akbarzadeh MA, Sanaie S, Kuchaki Rafsanjani M, Hosseini MS. Role of imaging in early diagnosis of acute ischemic stroke: a literature review. Egypt J Neurol Psychiatr Neurosurg 2021; 57: 1-8.

[6] Moonis M, Fisher M. Imaging of acute stroke. Cerebrovasc Dis 2001; 11(3): 143-150.

[7] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M et al. A survey on deep learning in medical image analysis. Med Image Anal 2017; 42: 60-88.

[8] Suganyadevi S, Seethalakshmi V, Balasamy K. A review on deep learning in medical image analysis. Int J Multimed Inf Retr 2022; 11(1): 19-38.

[9] Gautam A, Raman B. Towards effective classification of brain hemorrhagic and ischemic stroke using CNN. Biomed Signal Process Control 2021; 63: 102178.

[10] Neethi AS, Niyas S, Kannath SK, Mathew J, Anzar AM, Rajan J. Stroke classification from computed tomography scans using 3D convolutional neural network. Biomed Signal Process Control 2022; 76: 103720.

[11] Shakunthala M, HelenPrabha K. Classification of ischemic and hemorrhagic stroke using Enhanced-CNN deep learning technique. J Intell Fuzzy Syst 2023; Preprint: 1-16.

[12] Chen R, Cai Y, Wu J, Liu H, Peng Z, Xie Y et al. Artificial intelligence-based identification of brain CT medical images. In: AOPC 2022: Biomedical Optics; January 2023. Vol. 12560. pp. 53-58.

[13] Wieser M, Siegismund D, Heyse S, Steigele S. Vision transformers show improved robustness in high-content image analysis. In: 2022 9th Swiss Conference on Data Science (SDS); June 2022. pp. 71-72.

[14] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al. Attention is all you need. In: Advances in Neural Information Processing Systems; 2017. Vol. 30.

[15] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

[16] Takahashi S, Sakaguchi Y, Kouno N, Takasawa K, Ishizu K, Akagi Y et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. J Med Syst 2024; 48(1): 1-22.

[17] Azad R, Kazerouni A, Heidari M, Aghdam EK, Molaei A, Jia Y et al. Advances in medical image analysis with vision transformers: a comprehensive review. Med Image Anal 2024; 91: 103000.

[18] Simo AMD, Kouanou AT, Monthe V, Nana MK, Lonla BM. Introducing a deep learning method for brain tumor classification using MRI data towards better performance. Inform Med Unlocked 2024; 44: 101423.

[19] Chavva IR, Crawford AL, Mazurek MH, Yuen MM, Prabhat AM, Payabvash S et al. Deep learning applications for acute stroke management. Ann Neurol 2022; 92(4): 574-587.

[20] Okimoto N, Yasaka K, Fujita N, Watanabe Y, Kanzawa J, Abe O. Deep learning reconstruction for improving the visualization of acute brain infarct on computed tomography. Neuroradiology 2024; 66(1): 63-71.

[21] Zhu G, Chen H, Jiang B, Chen F, Xie Y, Wintermark M. Application of deep learning to ischemic and hemorrhagic stroke computed tomography and magnetic resonance imaging. Semin Ultrasound CT MR 2022; 43(2): 147-152.

[22] Miyamoto N, Ueno Y, Yamashiro K, Hira K, Kijima C, Kitora N et al. Stroke classification and treatment support system artificial intelligence for usefulness of stroke diagnosis. Front Neurol 2023; 14: 1295642.

[23] Shakunthala M, HelenPrabha K. Classification of ischemic and hemorrhagic stroke using Enhanced-CNN deep learning technique. J Intell Fuzzy Syst 2023; Preprint: 1-16.

[24] Altıntaş M, Öziç MÜ. Performance evaluation of different deep learning models for classifying ischemic, hemorrhagic, and normal computed tomography images: transfer learning approaches. Konya J Eng Sci 2024; 12(2): 465-477.

[25] Koska İÖ, Koska Ç, Fernandes A. Automatic stroke classification: Domain knowledge injection augmented transfer learning approach. Anatol Clin J Med Sci 2024; 29(3): 260-267.

[26] Çınar N, Kaya B, Kaya M. Classification of brain ischemia and hemorrhagic stroke using a hybrid method. In: 2023 4th International Conference on Data Analytics for Business and Industry (ICDABI); IEEE; 2023. pp. 279-284.

[27] Katar O, Yıldırım O, Eroğlu Y. Vision transformer model for efficient stroke detection in neuroimaging. In: 2023 4th International Informatics and Software Engineering Conference (IISEC); IEEE; 2023. pp. 1-6.
[28] Çınar N, Kaya B, Kaya M. Brain stroke detection from CT images using transfer learning method. In: 2023 13th International Conference on Advanced Computer Information Technologies (ACIT); IEEE; 2023. pp. 595-599.
[29] Azad R, Kazerouni A, Heidari M, Aghdam EK, Molaei A, Jia Y et al. Advances in medical image analysis with vision transformers: a comprehensive review. Med Image Anal 2024; 91: 103000.
[30] Henry EU, Emebob O, Omonhinmin CA. Vision transformers in medical imaging: A review. arXiv preprint arXiv:2211.10043, 2022.
[31] Li Z, Wang Y, Yu J, Guo Y, Cao L, Gao J et al. Brain MRI image classification for Alzheimer's disease diagnosis based on DenseNet and Vision Transformer. arXiv preprint arXiv:2103.03732, 2021.
[32] Liang S, Zhang W, Gu Y. A hybrid and fast deep learning framework for COVID-19 detection via 3D chest CT images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. pp. 508-512.
[33] Xia Y, Yao J, Lu L, Huang L, Xie G, Xiao J et al. Effective pancreatic cancer screening on non-contrast CT scans via anatomy-aware transformers. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021, 24th International Conference; 27 September–1 October 2021; Strasbourg, France. Springer International Publishing; 2021. Vol. 24. pp. 259-269.
[34] Ayoub M, Liao Z, Hussain S, Li L, Zhang CW, Wong KK. End-to-end vision transformer architecture for brain stroke assessment based on multi-slice classification and localization using computed tomography. Comput Med Imaging Graph 2023; 109: 102294.
[35] Abbaoui W, Retal S, Ziti S, El Bhiri B. Automated ischemic stroke classification from MRI scans: Using a vision transformer approach. J Clin Med 2024; 13(8): 2323.
[36] Koç U, et al. Artificial intelligence in healthcare competition (Teknofest-2021): Stroke data set. Eurasian J Med 2022; 54(3): 248-253.
[37] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019; 6(1): 1-48.
[38] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621, 2017.
[39] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S et al. Generative adversarial nets. In: Advances in Neural Information Processing Systems (NIPS); 2014. Vol. 27. pp. 2672-2680.
[40] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. pp. 4681-4690.
[41] Zhang K, Liang X, Gao H, Van Gool L, Timofte R. Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021. pp. 4791-4800.
[42] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004; 13(4): 600-612.
[43] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
[44] Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ et al. Attention mechanisms in computer vision: A survey. Comput Vis Media 2022; 8(3): 331-368.
[45] Ridnik T, Ben-Baruch E, Noy A, Zelnik-Manor L. ImageNet-21K pretraining for the masses. arXiv preprint arXiv:2104.10972, 2021.
[46] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems (NIPS); 2014. Vol. 27. pp. 3320-3328.
[47] Yalçın S, Vural H. Brain stroke classification and segmentation using encoder-decoder based deep convolutional neural networks. Comput Biol Med 2022; 149: 105941.
[48] Karataş AF, Doğan V, Kılıç V. Artificial intelligence-based cerebrovascular disease detection on brain computed tomography images. Avrupa Bilim ve Teknoloji Dergisi 2022; (41): 175-182.