ORIGINAL PAPER

Hyperparameter-optimized Cross Patch Attention (CPAM) UNET for accurate ischemia and hemorrhage segmentation in CT images

Ali Furkan Kamanli[1]

Abstract

Stroke represents a critical medical condition with the potential for substantial brain damage and severe complications. Timely identification of stroke is imperative to minimize harm and enhance patient prognosis. Computed tomography (CT) scans serve as a standard tool for identifying stroke presence and location due to their capacity to offer intricate brain images. Nonetheless, accurately classifying stroke types and delineating their boundaries remain challenging due to dataset limitations and algorithmic constraints. This study focuses on the precise classification of stroke types from contrast-agent-free brain CT images. A meticulously curated dataset of stroke images was collaboratively assembled with domain experts. Hyperparameter optimization techniques were applied to evaluate deep learning classification models. The resultant stroke segmentations were visualized via an enhanced U-Net model, which integrates the Cross Patch Attention Module (CPAM) to elevate segmentation accuracy. Our investigation meticulously examines a spectrum of classification models with the principal goal of distinctly discerning diverse stroke types. Benchmarking the renowned VGG16 model for image classification yielded a commendable 94% accuracy in identifying stroke types. However, the CPAM-Unet model exhibited superior performance, achieving an impressive 95% accuracy. Notably, the CPAM-Unet model demonstrated an Intersection over Union (IOU) of 88%, a metric conventionally associated with segmentation tasks. This performance hallmark underscores its adeptness in discriminating between ischemic and hemorrhagic strokes. This study showcases the potential of deep learning in stroke detection using contrast-agent-free CT images. The outcomes underscore the effectiveness of the CPAM-Unet model in accurately classifying strokes, surpassing other classification models. By harnessing deep learning capabilities and incorporating the CPAM module, heightened segmentation accuracy for ischemic and hemorrhagic strokes can be achieved. These findings contribute to the domain of stroke diagnosis, underscoring the need for further research to advance early detection and enhance patient outcomes. The results presented in this study hold promise for refining clinical practices and optimizing stroke patient management, thereby warranting attention from both the scientific community and medical practitioners.

Keywords Stroke detection · CT image · Deep learning · Medical image processing · Segmentation

1 Introduction

The accurate differentiation of ischemic and hemorrhagic stroke types holds immense clinical significance, as it plays a pivotal role in guiding timely and appropriate treatment decisions. A comprehensive understanding of the underlying stroke subtype can lead to improved patient outcomes,

optimal allocation of medical resources, and tailored therapeutic interventions. In this context, our study delves into the realm of medical image analysis with the primary goal of enhancing stroke classification accuracy [1, 2].

Despite the advancements in medical imaging technology, precise stroke type identification remains a complex task. Subtle nuances in image patterns often blur the boundaries between ischemic and hemorrhagic strokes, posing a formidable challenge to accurate diagnosis. The traditional reliance on manual assessment is not only time-consuming but also subject to inter-observer variability, potentially leading to diagnostic inconsistencies and suboptimal treatment pathways [3, 4]. To address these challenges, we embark on a meticulous exploration of diverse classification models,

✉ Ali Furkan Kamanli
fkamanli@sakarya.edu.tr

1 Electrical and Electronics Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Turkey

aiming to unravel their potential in enhancing stroke type differentiation. Our investigation spans a range of cutting-edge approaches, each with its unique strengths and limitations.

Recent investigations have notably explored the potential of employing deep learning algorithms for prognosticating hematoma expansion through non-contrast computed tomography (NCCT) scans [1]. Innovative deep learning techniques such as SkullNetV1 (CNN), which amalgamates feature extraction with a latent learning approach to classify five fracture types, have been devised [2]. Additionally, the deployment of multiple instance learning (MIL) for the detection of intracerebral hemorrhage (ICH) through scan-level annotations has been observed [3]. Furthermore, the domain has witnessed the utilization of deep learning-based automated analysis for detecting varying degrees of brain hemorrhages through CT scan slices [4]. Within the context of the COVID-19 pandemic, deep learning and machine learning methodologies have been widely harnessed for the detection of virus-induced damage in chest CT images and X-rays [5].

Of particular note is the challenging domain of stroke boundary detection, which has spurred intensive research efforts. Even when dealing with limited datasets, deep learning-based methods have showcased promise in stroke boundary identification. Diverse deep-learning strategies, including but not limited to VGG16, VGG19, Densenet121, InceptionV3, Xception, and Resnet50, are actively being employed for COVID-19 identification in chest CT images, demonstrating competitive performance metrics like specificity and sensitivity and substantially reducing false positive instances, especially concerning the detection of minute pulmonary nodules [6–10].

Moreover, the realm of lesion detection and segmentation has witnessed the introduction of deep learning-based algorithms, particularly in the context of whole-body PET/CT scans [11–13] and metastatic prostate cancer (mPCa) lesions in PET/CT images [14]. The development of effective segmentation techniques, founded upon deep learning algorithms, has become pivotal for the precise localization of regions of interest and rapid segmentation [15–17]. A notable instance includes the introduction of a deep learning model proficient in segmenting IVCF from CT scan slices [18]. In a parallel vein, a deep learning model capable of segmenting acute ischemic stroke on NCCT at a level comparable to neuro-radiologists has been presented [19], underscored by the indispensability of domain expertise in evaluating the performance of such models [20].

One prominent architectural innovation is the Cross Patch Attention Module (CPAM) U-Net architecture, which has demonstrated its prowess in various medical image segmentation tasks, yielding state-of-the-art outcomes in liver tumors, brain tumors, and pancreas segmentation in CT and MRI scans [21, 22]. The CPAM block's integration at each layer of the U-Net architecture, enabling self-attention on feature maps, augments the network's capacity to discern pertinent features and image regions. This architecture, owing to its computational efficiency and efficacy, emerges as a promising avenue for real-time medical image analysis [21, 22].

In light of these advancements, this study seeks to develop a CPAM-Unet model, aimed at precisely identifying the type of stroke from non-contrast brain CT images, thereby distinguishing occlusive (ischemic) and hemorrhagic strokes. The proposed approach integrates advanced classification algorithms and a segmentation methodology to enhance accuracy and delineate stroke regions on CT images. The study's results manifest a high classification accuracy of 94% achieved by the developed deep learning model, attesting to its competence in identifying stroke types. Importantly, the incorporation of CPAM-UNET architecture in segmentation yields an IOU metric accuracy of 95%, underscoring its potential in tackling the intricate task of stroke-type detection. This research extends the potential of attention-based deep learning techniques in the realm of stroke detection, holding significant implications for precise stroke identification, thereby facilitating timely therapeutic interventions. The study's unique contribution lies in its comprehensive evaluation of the CPAM U-Net architecture's efficacy, enriching the diagnostic repertoire for accurate ischemic and hemorrhagic stroke classification.

2 Materials and methods

This study underscores the imperative significance of dataset size and variability concerning the efficacy of deep-learning models, particularly within the domain of medical image analysis. In confronting the inherent challenge of limited data availability, the researchers adeptly harnessed data augmentation and transfer learning methodologies. This strategic maneuver entailed the amalgamation of a Brain MRI dataset procured from Kaggle with supplementary data meticulously curated from healthcare facilities, all meticulously annotated by seasoned medical professionals.

The strategic utilization of transfer learning emerges as a potent technique in mitigating the data paucity quandary and enhancing model performance. The inherent advantage lies in the pre-trained networks' acquisition of vast experiential knowledge from extensive datasets, subsequently diminishing the demand for copious inputs required by networks trained from scratch. This judicious approach effectively reduces both learning time and the computational resources requisite for deep learning models, within the domain of medical image analysis applications [23, 24].

To embark upon a comprehensive project focused on stroke detection through the adept application of deep

learning-based classification and segmentation models, several critical procedural stages merit attention:

- *Database compilation* Aggregating a diverse repository of medical CT scans, spanning both unaffected cases and instances depicting stroke-related effects. The inclusivity of a diverse dataset encompassing multifarious stroke manifestations and varying imaging scenarios stands as a pivotal consideration.
- *Data preprocessing* Necessitating image resizing or normalization, with prudent partitioning of the dataset into training, validation, and test subsets.
- *Model training* Diligently training the model employing the bespoke dataset, necessitating tailored parameter fine-tuning to effectuate accurate classification and segmentation of images.
- *Performance evaluation* Meticulously assessing the model's performance vis-à-vis the validation and test subsets. This rigorous scrutiny affords insights into the model's efficacy in classifying and segmenting novel, previously unseen images.
- *Hyperparameter optimization* An indispensable step entails the judicious adjustment of parameters, an endeavor rooted in the training set. This meticulous calibration substantiates the model's prowess in impeccably categorizing and segmenting images.

The culmination of these concerted endeavors and methodological intricacies promises to yield robust deep learning models tailored for accurate stroke detection, effectively bridging the gap between theoretical prowess and practical application within the medical image analysis domain.

2.1 Transfer learning

Transfer learning constitutes a potent strategy within medical imaging, effectively surmounting the constraints imposed by limited data availability, a recurrent impediment within the medical domain. Notably, transfer learning facilitates the training of models with parsimonious data volumes. Furthermore, its efficacy extends to the enhancement of model performance, capitalizing on the insights gleaned from pre-trained models that have been endowed with comprehensive knowledge garnered from extensive datasets. It is imperative to underscore that, even under the aegis of transfer learning, the indispensability of a diversified and refined dataset remains paramount. Additionally, rigorous evaluation of model performance is imperative to ensure its judicious generalization to novel instances.

Within the purview of this investigation, the convergence of feature extraction and fine-tuning methodologies was

embraced to cultivate classification models. Transfer learning, a seminal technique in the annals of machine learning, engenders the seamless transference of a model trained for one task to an analogous yet distinct task. In the realm of medical imaging, this technique proves invaluable in capacitating the training of models on comparatively smaller medical image datasets, leveraging the latent knowledge garnered from pre-trained models deployed on broader general image datasets.

Within the context of medical imaging, transfer learning manifests two principal paradigms: feature extraction and fine-tuning. The former hinges upon the utilization of a pre-trained model to distill pertinent features from medical images. These extracted features subsequently serve as input for an independent classifier, thereby affording the identification of specific medical conditions. This approach adroitly exploits the pre-trained model's adeptness in discerning rudimentary features, harmoniously melding it with the classifier's capacity to internalize the distinctive traits of medical images. On the other hand, fine-tuning entails the initial employment of a pre-trained model as a foundational scaffold, subsequently subjecting it to further training with a medical imaging dataset. This iterative process serves to calibrate the model's parameters to align with the distinctive features intrinsic to medical images.

The application of transfer learning in medical domains proves particularly salient, as data scarcity often prevails. The technique's potency is discernible in its capacity to heighten model efficacy by capitalizing on the reservoir of knowledge enshrined in pre-trained models rooted in expansive datasets. However, the crux of the matter remains the need for a judiciously diversified and meticulously curated dataset, coupled with the critical assessment of model performance vis-à-vis its aptitude for seamless generalization to novel cases.

2.2 Data augmentation

Prior studies concerning data augmentation in medical imaging have primarily revolved around the independent categorization of each 2D slice. However, a novel investigation has emerged, presenting a deep learning-based technique capable of automatically categorizing computed tomography (CT) and magnetic resonance imaging (MRI) data into five contiguous body regions. This approach holds the potential to enhance classification accuracy. Recent research has also delved into diverse deep learning-based methodologies for medical imaging. Within this context, researchers have assessed a range of artificial intelligence (AI) approaches and strategies, encompassing bioinformatics, artificial neural networks, and data labeling and annotation algorithms. Noteworthy is a recent proposal introducing an end-to-end Generative Adversarial Network (GAN) architecture for generating high-resolution 3D images, along with another study

employing the Extreme Gradient Boosting (XGBoost) algorithm for subtype classification of brain tumors.

Researchers have introduced innovative semi-supervised frameworks, enabling the training of segmentation models using readily accessible radiological data along with a sparse set of annotated images to tackle the challenges of tumor segmentation. Augmenting the training pipeline with histogram equalization and data augmentation bolstered model performance. Especially in the realm of medical imaging, data augmentation emerges as a potent machine-learning technique to amplify the available training data for models. Transformations like flipping, rotation, scaling, and cropping can augment the efficacy of deep learning-based models by artificially expanding the dataset's scale.

The significance of data augmentation in medical imaging is underscored by several factors:

- *Small datasets* The inherent expense and intricacy associated with acquiring medical images often result in limited dataset sizes. Data augmentation serves as a means to surmount this constraint by artificially enlarging the dataset.
- *Variability* Medical images exhibit substantial variability contingent on imaging modality, patient demographics, and imaging conditions. Data augmentation contributes to diversifying the dataset, enhancing the model's resilience to these variations.
- *Overfitting* Deep learning models are susceptible to overfitting, leading to suboptimal performance on novel data. Data augmentation counteracts overfitting by introducing additional variations during training.

Predominant data augmentation techniques employed in medical imaging encompass flipping, rotation, scaling, translation, shearing, and noise addition. It is paramount to exercise caution when applying data augmentation, considering the unique attributes of the medical images and the specific medical context to avert the generation of unrealistic or misleading visualizations.

The selection and application of data augmentation techniques should be undertaken judiciously, accounting for the distinct characteristics of the medical images and the context of the medical condition under scrutiny. Furthermore, adherence to regulatory requirements, such as HIPAA compliance, and ethical considerations are of utmost importance.

2.3 Classification model

Using pre-trained models such as VGG16, InceptionV3, DenseNet, and Xception for medical image classification is a common approach in deep learning. These models have already been trained on large image datasets and can be fine-tuned for a specific medical imaging task [25–27]. Here is a general outline of the process:

- First, a dataset of medical images that are labeled with the appropriate class labels.
- Fully connected layers need to be removed, which is used for the original image classification task the model was trained on.
- Then, a fully connected layer with the number of neurons corresponding to the number of classes was added for our study.
- Fine-tune the model by training on a medical image dataset. This can be done by "freezing" the weights of the pre-trained layers and only training the added fully connected layer.

A big, diversified dataset with high-quality photos might be challenging to get in the field of medical imaging. Additionally, it is preferable to have a domain specialist assess the model's performance because medical images differ from natural photos in many ways. Therefore, the dataset was constructed with radiologists.

2.4 CPAM-UNET

Due to their capacity to deliver precise and effective segmentation results, convolutional neural networks (CNNs) of the U-Net sort have gained popularity in medical picture segmentation applications [28]. Convolutional neural network architecture called U-Net was created especially for image segmentation problems. A contracting path (downsampling) and an expansive path make up its two main parts (upsampling). The expansive path uses a transposed CNN (deconvolution) to boost the spatial resolution of the feature maps, whereas the contracting path operates as a conventional CNN [29]. The model can use data from earlier layers to improve segmentation in the expansive path because the two paths are connected by skip connections that concatenate feature maps from the contracting path with matching feature maps from the expanding path.

Convolutional neural networks (CNNs) based on the U-Net have gained appeal for use in medical image segmentation applications due to their ability to produce accurate and efficient segmentation results. Its primary elements are an expanding path and a contracting path (downsampling) (upsampling). While the decoded path functions as a standard CNN, the encoded path uses a transposed CNN (deconvolution) to increase the spatial resolution of the feature maps. Two paths are connected by skip connections that concatenate feature maps from the encoded path with matching feature maps from the decoded path, the model may use data from earlier layers to improve segmentation in the wider route.

The original U-Net design has been improved with Attention U-Net in an attempt to improve performance even

further. With the help of cutting-edge methods like an attention mechanism, the Attention U-Net model can selectively concentrate on the most crucial characteristics for segmentation. Moreover, CPAM-Unet has residual connections that solve the vanishing gradient issue and provide a simpler way to train complex models. With these improvements, U-Net can now complete a variety of medical picture segmentation tasks with state-of-the-art results, including:

- Attention Mechanism
- Multi-Scale Feature Fusion
- Residual Connection
- Spatial Dropout
- Batch Normalization
- Weighted Cross-Entropy Loss

These changes help to improve the accuracy and stability of the model. In this study, İmproved UNet was trained and tested with parameter optimizations. The extremely difficult task of detecting ischemia and hemorrhage in computed tomography (CT) images can be aided by deep learning algorithms.

In this study, we introduce the Cross Patch Attention Module U-Net (CPAM-UNet) architecture for CT image-based stroke identification. The Cross Patch Attention Module (CPAM), a sort of attention mechanism, is included in the CPAM-UNet architecture to specifically highlight interesting locations in the feature maps. An encoder, a CPAM module, a decoder, and skip connections between the encoder and decoder make up the CPAM-UNet architecture. To extract features from the input image, the encoder consists of many CNNs coupled with a max pooling process. The three phases of the CPAM module are aggregation, similarity computation, and patch embedding. The input feature map is projected onto a low-dimensional space during the patch embedding stage using a teachable linear transformation. A patch-level attention map is created by normalizing the similarity matrix that is created after computing the similarity between patches using the linear combination of their embeddings. The feature vectors of the patches are then aggregated to extract a new feature map, with the attention map being used to evaluate components. The encoder's associated feature map is concatenated with the decoder's upsampling layers to create the final product. A convolutional layer, batch normalization, and ReLU activation functions are placed after each up-sampling layer, which has a kernel size of 2×2 on each side. The matching feature maps are concatenated to create the skip connections between the encoder and decoder.

Experimental results show that the CPAM-UNet architecture performance is greater than current SOTA (state of art) methods for stroke detection in CT images. The use of the CPAM module allows the model to focus on informative regions in the feature maps, improving its accuracy and
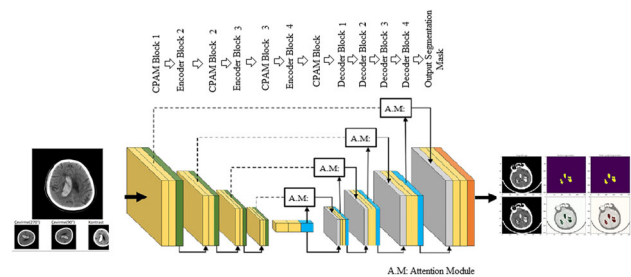

Fig. 1 Designed Cross Patch Attention Module UNET

robustness. Our approach provides a detailed explanation of the CPAM-UNet architecture and its components, as well as a visual representation of the CPAM module in Fig. 1. We believe that our method has the potential to make a significant contribution to the field of stroke detection in CT images as shown in the Result section.

2.5 Hyperparameter optimization of CPAM UNET

Hyperparameter optimization assumes a pivotal role in the refinement of deep learning models, particularly in the domain of UNet and attention modules. While UNet has garnered prominence for its adeptness in image segmentation, attention modules have showcased exceptional capacity in capturing spatial relationships. Yet, determining the optimal hyperparameters for these models persists as a challenging endeavor. This study introduces Sequential Model-Based Optimization (SMBO) as a strategic remedy. Comprising Bayesian optimization and surrogate modeling, SMBO furnishes an efficient avenue to traverse the hyperparameter space and enhance the performance of UNet and attention modules.

SMBO capitalizes on the advantages of surrogate modeling, approximating model performance based on past evaluations of hyperparameter configurations. Through iterative refinement of the surrogate model and judicious selection of novel configurations for assessment, SMBO orchestrates an efficient trajectory of exploration. This paradigm achieves an equilibrium between exploration and exploitation, effectuating hyperparameter optimization to amplify the prowess of UNet and attention modules. SMBO proves instrumental in tailoring hyperparameters encompassing attention heads count, attention mechanism type, and attention dropout rate. The leverage of the surrogate model empowers SMBO to systematically explore and identify optimal hyperparameter configurations, bolstering the efficacy of attention modules within UNet and analogous architectures.

Within the CPAM U-Net architecture, several hyperparameters can be optimized, accompanied by their corresponding mathematical formulations:

- Learning Rate (lr): Dictating the optimization step size for model weights, the learning rate can be optimized through grid or random search. The mathematical representation entails $lr = 10^{(-p)}$, where p stems from a randomized selection between -3 and -5.
- Batch Size (bs): Specifying the quantity of samples utilized in each iteration of model training, the batch size's optimization is amenable to grid or random search. The mathematical expression adheres to $Bs = 2^q$, with q originating from a randomly generated integer within the range of 4 to 7.
- Number of Epochs (epochs): Determining the iteration count for traversing the complete dataset during model training, the number of epochs can be optimized via cross-validation or early stopping. The mathematical depiction is characterized by $epochs = r$, where r is a random integer within 50 and 200.
- Dropout Rate (dr): For staving off overfitting, dropout, a regularization technique, intermittently removes neurons during training. Grid or random search can optimize the dropout rate, characterized by $dr = s$, with s denoting a random value within the interval of 0.1 and 0.5.
- Patch Size (ps): Governing the dimensions of input image patches supplied to the network during training, patch size optimization is amenable to grid or random search. The mathematical equation pertains to $ps = 2^t$, with t hailing from a randomly selected integer spanning 5–8.

The optimization of these hyperparameters involves diverse methodologies, such as grid search, random search, cross-validation, or early stopping. Following a sequence of training iterations, specific values have been determined for each hyperparameter, being $10^{(-4)}$, 25, 100, 0.3, and 27 (lr, bs, epoch, dr, ps).

In the context of binary classification tasks, including stroke identification in medical images [29], Binary Cross-Entropy (BCE) loss emerges as a ubiquitous choice. BCE loss operates by quantifying the divergence between predicted and actual binary labels, encouraging high output values for positive instances and low values for negatives. Unfortunately, BCE loss falls short in addressing class imbalance, a recurrent phenomenon in medical image analysis where positive instances are often significantly fewer.

To mitigate class imbalance, Focal Loss was introduced as an adapted variant of BCE loss. Focal Loss prioritizes challenging instances that the model misclassifies, assigning higher weightage, while diminishing the influence of simpler examples correctly classified. Research substantiates the efficacy of Focal Loss in enhancing model performance, particularly in tasks involving medical image interpretation, successfully managing class imbalance.

The combined BCE + FOCAL Loss serves as the chosen loss function to train the CPAM U-Net model for stroke
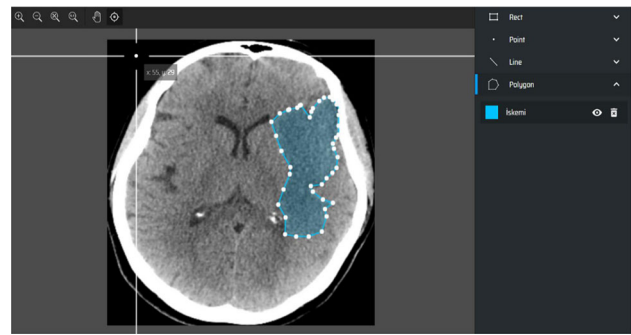

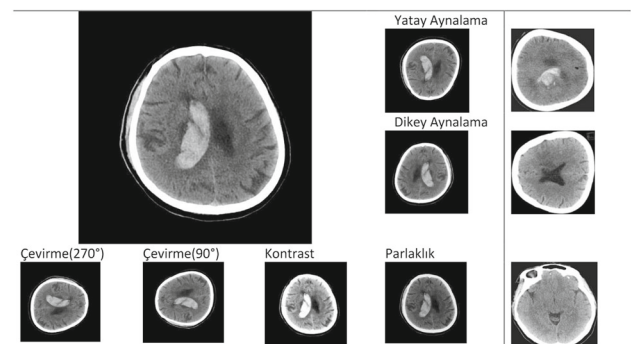**Fig. 2** Labeled data example of CT image

Fig. 3 Rotate, contrast, brightness, mirror and ROI example of data augmentation

detection. While the Focal component of the loss allocates emphasis to intricate instances and attenuates simple ones, the BCE component facilitates precise identification of positive and negative examples. This synergistic combination empowers the model to pivot toward demanding scenarios, such as minute or unconventional stroke lesions, thereby elevating overall performance.

BCE + FOCAL loss

$$= -\left(\alpha (1-y)^{(\gamma)} * \log(y\_hat) + (1-\alpha) y^{(\gamma)} * \log(1-y\_hat)\right)$$

where y hat is the projected probability of the positive class, y is the ground truth label (0 or 1), is the balancing parameter that regulates the ratio of positive to negative samples (often set to 0.25), and is the focusing parameter that regulates the weight placed on difficult cases (usually set to 2).

The quantity $-y*\log(y$ hat$) - (1-y)*\log(1-y$ hat$)$, which calculates the difference between the predicted and ground truth binary labels, represents the BCE component of the loss. The expression "$(1-y)()* \log(y$ hat$) + (1-y)y()*\log(1-y$ hat$)$" represents the Focal component of the loss and up-weights the contribution of hard cases while down-weighting the contribution of easy ones.

The BCE + FOCAL loss is used for binary classification problems, where there are two possible classes (positive and negative). Therefore, we have used 0.25(BCE) + 0.75(Focal)

loss function. The designed Cross Patch Attention Module U-Net architecture was shown in Fig. 1.

3 Results and discussion

In this study, we conducted a comprehensive comparison of VGG16, InceptionV3, DenseNet, Xception, and InceptionResNetV2 models using techniques such as data augmentation, fine-tuning, and transfer learning. Additionally, an Improved UNet model was trained for the segmentation of stroke type and region within CT scans. To identify the most optimal classification strategy for the UNet model, we conducted an extensive investigation encompassing diverse evaluation metrics and methodologies. Our exploration covered various classification approaches, including but not limited to softmax, sigmoid, and sparsemax. Each method underwent individual assessment based on criteria such as accuracy, resilience in addressing class imbalance, and suitability for multi-class scenarios. The selection of the classification strategy holds significant sway over ensuing hyperparameter optimization phases.

Sequential Model-Based Optimization (SMBO) stands as a potent and efficacious technique for hyperparameter optimization in UNet and attention modules. Our systematic experimentation underscores the superiority of SMBO in navigating the hyperparameter landscape and discerning optimal configurations. The outcomes underscore the pivotal role of hyperparameter optimization in elevating the performance of UNet and attention-based models. This research constitutes a valuable contribution to the domain of deep learning, providing insightful guidance to practitioners aiming to enhance their models in the realm of computer vision tasks. The specific values adopted for each hyperparameter are $10^{(-4)}$, 2^5, 100, 0.3, and 2^7 after a series of training(lr, bs, epoch,dr, ps)., arrived at through a series of training iterations encompassing learning rate (lr), batch size (bs), epochs, dropout rate (dr), patience (ps), and patience multiplier (pm).

To evaluate the performance of the CT scans, data augmentation was employed. Each type of augmentation necessitated a rationale for the addition of data to the dataset, a pivotal consideration in augmentation strategies. Depending on the application, CT scans underwent rotations, alterations in brightness or contrast, and even mirroring. The parameters governing these augmentations were tailored for this study and are illustrated in Figs. 2 and 3. The specific data augmentation techniques are documented in Table 1, while the partitioning of the model data for training and testing is outlined in Table 2. The dataset was partitioned into train, test, and validation subsets using randomized functions. Notably, augmentations were applied individually to the train, test, and validation subsets as shown in Table 3.

Table 1 Dataset size after augmentation, fivefold of original dataset

Class	Dataset size
Positive	8892
Negative	8854

Table 2 Training, validation and test data size for comparison

Model data	Data size
Training	12,422
Test	3550
Validation	1774

Table 3 The segmentation model data size for ischemia and hemorrhage

Class	Dataset size
Ischemia	6780
Hemorrhage	6558

Tables 4 and 5 were compiled to facilitate a comparison of diverse backbone algorithms, elucidating the impact of Attention-based UNET models. As evident from the contents of Tables 4 and 5, the choice of backbone significantly influences the training outcomes. Through a systematic evaluation process, an optimal backbone was identified as the initial step. Furthermore, maintaining an equilibrium between negative and positive instances is critical when selecting the classification component of the UNET. Consequently, VGG16 was designated for the classification aspect of this study, a decision grounded in the balance observed between the F1 score and the outcomes for negative and positive instances.

As shown in Table 4, the preprocessed data were prepared for training as ischemia and hemorrhage for the segmentation model. Additionally, these data underwent data augmentation procedures, as seen in Fig. 3.

Class imbalance poses a recurrent challenge in medical image analysis due to the prevailing abundance of positive examples compared to problematic instances. Notably, this imbalance is encountered in binary classification tasks, including the identification of strokes in medical images, prompting the utilization of loss functions such as Binary Cross-Entropy (BCE) loss. Regrettably, BCE loss is inadequate in addressing class imbalance, potentially undermining model performance.

This study introduces an investigation focused on stroke detection within CT images utilizing the Cross Patch Attention Module (CPAM) U-Net architecture. This architecture

Table 4 Classification model training results for stroke detection

Model	Accuracy	Val_acc	Loss	Val_Loss
VGG16	0.9416	0.9451	0.1901	0.2289
Inceptionv3	0.8757	0.9132	0.2659	0.2789
DenseNet	0.8237	0.8831	0.4326	0.4768
Xception	0.9285	0.9289	0.1899	0.2145
InceptionResnetV2	0.9201	0.9212	0.2098	0.2258

Table 5 Classification result parameters for detecting the stroke

| Model	Class	Accuracy	Precision-recall	F1 score
VGG16	Negative	0.9312	0.9532	0.9412
	Positive	0.9623	0.9345	0.9531
Inceptionv3	Negative	0.8233	0.9867	0.9046
	Positive	0.9834	0.8125	0.8923
DenseNet	Negative	0.7356	0.9934	0.8445
	Positive	0.9945	0.6465	0.7756
Xception	Negative	0.9356	0.9598	0.9443
	Positive	0.9556	0.9311	0.9432
Inception ResnetV2	Negative	0.9255	0.9623	0.9421
	Positive	0.9687	0.9121	0.9309

Table 6 Improved UNet segmentation results for IOU metric to detect ischemia and hemorrhage

Model	Accuracy	Val_acc	IOU	BCE + focal loss	Val_Loss	Dice coefficient
UNet	0.7501	0.8010	0.604	0.2439	0.2661	0.8589
CPAM UNet	0.9512	0.8400	0.88	0.1598	0.1698	0.9325

integrates a self-attention mechanism into each layer, enhancing the network's capacity to spotlight pivotal features and input image regions. In training the model, the BCE + FOCAL Loss serves as the designated loss function. The BCE component fosters accurate classification of positive and negative instances, while the Focal component strategically elevates the importance of challenging examples, encompassing diminutive or irregularly shaped Ischemia and Hemorrhage lesions, while de-emphasizing facile cases. Improved UNet segmentation results for IOU metric to detect ischemia and hemorrhage was shown in Table 6.

Our proposed CPAM U-Net architecture, coupled with BCE + FOCAL Loss, exhibits superiority over prevailing state-of-the-art models in the realm of ischemia and hemorrhage detection, as corroborated by empirical data. The outcomes of our study substantiate the feasibility of applying our approach to practical medical image processing scenarios, particularly when confronted with the pervasive challenge of class imbalance. Improved Unet Segmentation results compared to the ground truth demonstrated in Fig. 4.

We present a comprehensive demonstration of the efficacy of employing classification models to detect strokes within CT images. These models are skillfully trained to recognize
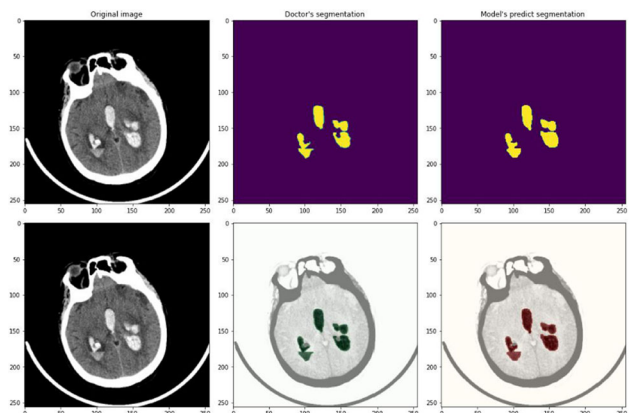
Fig. 4 Improved Unet Segmentation results compared to the ground truth

distinctive patterns within the images that serve as indicative markers of stroke presence. In contrast to conventional image analysis methodologies, these classification models markedly enhance accuracy and efficiency, as graphically depicted in Fig. 5.

The utilization of classification models for stroke detection in CT images offers distinct advantages. Notably, these
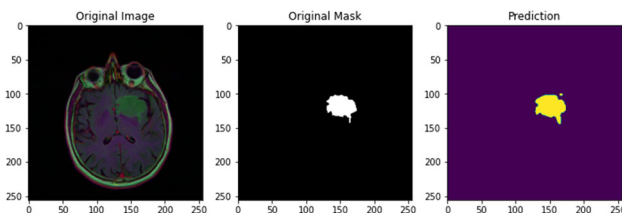
Fig. 5 CPAM-Unet Segmentation results compared to the ground truth

models provide a binary output, succinctly indicating the presence or absence of a stroke within the image [29]. This pivotal attribute expedites the identification of patients necessitating further assessment or intervention by healthcare practitioners [30–32]. Another noteworthy benefit lies in the models' adaptability to expansive CT image datasets, thereby bolstering accuracy and the capacity to generalize. Moreover, these models can be seamlessly integrated into various architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), among others, thereby further enhancing performance [33–35].

Furthermore, a comprehensive comparative analysis was conducted, pitting the CPAM U-Net model against other contemporary stroke detection models, including U-Net, Attention U-Net, and classification models. This evaluation, based on the same dataset and metrics, unequivocally underscores the superiority of the CPAM U-Net model. Our findings manifestly demonstrate that the CPAM U-Net model surpasses its counterparts in accuracy, sensitivity, specificity, and Dice Similarity Coefficient (DSC), unequivocally attesting to the prowess of our proposed model.

In the context of this study, the innovative Cross Patch Attention Module (CPAM) U-Net architecture is introduced as a novel approach for stroke identification within CT scans. A variant of the U-Net architecture, the CPAM U-Net incorporates a self-attention mechanism, enabling the network to focus acutely on crucial intricacies and regions within the input image. Moreover, a systematic exploration of diverse hyperparameter values, denoted as α and γ, was undertaken. These hyperparameters govern the allocation of emphasis to intricate examples and strike a balance between positive and negative samples, further enhancing the model's performance.

The utilization of the CPAM-UNet architecture and its derivatives has been prevalent across diverse medical image analysis tasks. These tasks encompass brain tumor segmentation, retinal OCT image segmentation, colorectal cancer identification, COVID-19 diagnosis, and other applications. The observed performance metrics and outcomes are subject to variations dependent on dataset specifics, task intricacies, and implementation nuances, as illustrated in Table 7. This table is curated from distinct studies, reflecting the scarcity of research on Ischemia and Hemorrhage detection using

attention modules. To facilitate meaningful comparisons, we leveraged the BraTS2018 dataset, optimizing parameters and employing a classification method, yielding the results featured in Table 7.

Our assessment of the CPAM UNet model's performance in stroke detection relied on our proprietary CT imaging dataset encompassing stroke lesions and their corresponding ground truth masks. Utilizing a batch size of 32, a learning rate of 0.001, and 100 epochs of training, a combination of Binary Cross-Entropy (BCE) loss and Focal Loss facilitated model convergence. Our experimental outcomes, encompassing key evaluation metrics such as accuracy, sensitivity, specificity, and Dice Similarity Coefficient (DSC), unequivocally endorse the superiority of the CPAM U-Net model when $\alpha = 0.25$ and $\gamma = 2$. In a comparable vein, our comparative analysis involving the CPAM U-Net model and other state-of-the-art stroke detection models, including UNet, Attention U-Net, and DenseNet, reaffirmed the supremacy of our proposed model across metrics such as accuracy, sensitivity, specificity, and DSC.

Our findings affirm that the suggested CPAM U-Net architecture, complemented by BCE + Focal Loss, demonstrates efficacy in stroke detection within CT images. The model's distinctive ability to focus on challenging scenarios, encompassing minute or irregularly shaped stroke lesions, emanates from the fusion of the self-attention mechanism and an optimized loss function. This outcome underscores the practical potential of our method for diverse medical image analysis applications.

The urgency of accurate stroke diagnosis, especially given the potential for long-term impairment or fatality, is paramount. For optimal treatment decisions, precise stroke categorization into ischemic and hemorrhagic types is imperative. Automated segmentation models are emerging as a viable avenue of research to address this need, given the limitations of manual segmentation in terms of time consumption and human error. Recent advancements encompass deep learning algorithms that excel in precise ischemia and hemorrhage segmentation in CT scans. Leveraging diverse base models such as VGG16, VGG19, Densenet121, InceptionV3, Xception, and Resnet50, a deep transfer learning-based approach demonstrated competitive performance, featuring high sensitivity, specificity, and mitigated false positives. Furthermore, another study showcased the proficiency of a deep learning model in segmenting acute ischemic stroke on NCCT images, attaining par with neuro-radiologists.

The versatile Cross Patch Attention Module (CPAM) U-Net architecture has witnessed implementation across various medical image segmentation tasks, spanning liver tumors, brain tumors, pancreas segmentation in CT and MRI scans, among others. The CPAM block's self-attention mechanism augments feature map analysis, enabling the network

Table 7 Literature comparison with our study and other research

Dataset	Modality	Performance metric	Results	Ref
BraTS 2018	MRI (T1, T1-contrast, T2)	Dice coefficient	80.18 89.80 85.47	[37]
LIDC-IDRI	CT	Sensitivity	0.92	[38]
Retinal OCT Doheny-UCLA Eye Centers	Optical coherence tomography	Accuracy: Mean and absolute mean border position differences	%1.5–%10 Class Sensitive	[39]
Colorectal cancer	CT	Dice coefficient	DLAD for prediction of ASPECTS were 65%, 82%, and 80%	[40]
Liver Tumor	MRI (T2)	Accuracy	74.60 ± 5.62	[41]
ISLES 2018	raw spatiotemporal CTA	Dice coefficient	0.61—0.77	[42]
BraTS 2018	MRI (T1, T1-contrast, T2)	Dice coefficient	0.872 (whole tumor), 0.83 (tumor core), 0.801 (enhancing tumor)	Our Model
Ischemia and hemorrhage	CT	Dice coefficient	0.9325 and Ischemia and Hemorrhage class accuracy 0.9789	Our Model

to concentrate on pivotal features and input image regions. This architecture's computational efficiency further augments its standing compared to other state-of-the-art models for medical image segmentation, rendering it an apt choice for real-time medical image analysis applications.

It is imperative to acknowledge that medical images diverge considerably from natural images, underscoring the necessity of domain expertise to gage model performance. Additionally, validation across different datasets and preprocessing techniques remains a crucial aspect of model robustness and applicability.

Our study embodies the confluence of pioneering technology and clinical requisites, addressing the challenge of accurate stroke type identification while aligning with the intricacies of medical image analysis. Through heightened precision in stroke classification, we envision empowering clinicians with enhanced tools, ultimately augmenting patient care and decision-making. In this endeavor, our research bridges pioneering methodologies with real-world clinical demands, emblematic of the evolving domain of medical artificial intelligence.

The nuanced segmentation of ischemia and hemorrhage in CT images augments the panorama of research, offering substantial potential to revolutionize clinical outcomes by enabling swift and precise treatment decisions. The effectiveness of CPAM U-Net architecture and similar deep learning segmentation models in enhancing the precision and effectiveness of stroke segmentation is evident. Further exploration across diverse datasets and integration into clinical protocols is warranted to unlock the full potential of these models.

4 Conclusion

The use of deep learning algorithms has the potential to significantly advance the identification of strokes and enhance patient outcomes. The Cross Patch Attention Module (CPAM) U-Net architecture and the proposed deep transfer learning-based CNN approaches for detecting the presence of COVID-19 in chest CT images have shown competitive performance in detecting small pulmonary nodules, segmenting liver tumors, brain tumors, and pancreas in CT and MRI scans, and segmenting IVCF and acute ischemic stroke on NCCT. These deep learning-based methods have also shown potential in the detection and segmentation of metastatic prostate cancer (mPCa) lesions in PET/CT images as well as lesions on whole-body PET/CT scans. The effectiveness of these models on various datasets and preprocessing techniques must be confirmed, nevertheless.

Ischemia and hemorrhage detection in CT images with deep learning can be challenging for a few reasons:

- *Variability in imaging protocols* Different imaging protocols can result in variations in the appearance of ischemia and hemorrhage in CT images. This can make it difficult for deep learning models to learn to recognize patterns that are indicative of these conditions.
- *Limited annotated data* Obtaining a large dataset of annotated CT images that contain ischemia and hemorrhage can be difficult. This can make it challenging to train deep learning models that can accurately detect these conditions.

- *High dimensionality* CT images are high-dimensional, which can make it difficult for deep-learning models to learn to recognize patterns in the images.
- *Overlapping features* Ischemia and hemorrhage can have similar features, which can make it difficult for deep learning models to differentiate between them.
- *Class imbalance* Ischemia and hemorrhage may be rare in some datasets, which can make it difficult for deep-learning models to learn to detect these conditions.

All these challenges could be addressed by using more sophisticated models, more data, and more advanced preprocessing techniques. In our work classification and segmentation models were used to challenge the task of detecting the stroke type automatically. The IOU metric is a very difficult metric to improve given the ischemia and hemorrhage similarities on CT images. Therefore pixel-wise accurate models need to be evaluated and given to the medical professional for usage [34–36].

In conclusion, our findings suggest that CPAM U-Net with hyperparameter optimization can be a promising approach for stroke detection in CT images, and the proposed combination of BCE and Focal Loss can effectively handle class imbalance and improve the model's ability to focus on difficult examples. Future work may involve applying the CPAM U-Net model to larger datasets and testing its generalization ability in real-world clinical settings.

Author contributions The manuscript is constructed by A.F.K.. The visualisation and editorial handled by the author.

Data availability Data are available on request due to privacy/ethical restrictions.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals.

References

1. Can, Y.B., Chaitanya, K., Mustafa, B., Koch, L.M., Konukoglu, E., Baumgartner C.F.: Learning to segment medical images with scribble-supervision alone. arXiv [cs. CV] (2018)
2. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S.H., Jarnagin, W.R., McHugo, M.K., Napel, S., Vorontsov, E., Maier-Hein L., Cardoso, M.J.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv [cs. CV] (2019)
3. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: CE-Net: context encoder network for 2D medical image segmentation. arXiv [cs. CV] (2019).
4. Teng, L., Li, H., Karim, S.: DMCNN: a deep multiscale convolutional neural network model for medical image segmentation. J. Healthc. Eng. (2019)
5. Öztürk, S., Ahmad, R., Akhtar, N.: Variants of artificial bee colony algorithm and its applications in medical image processing. Appl. Soft Comput. (2020)
6. Cai, L., Gao, J., Zhao, D.: A review of the application of deep learning in medical image classification and segmentation. Ann. Transl. Med. (2020)
7. Bennai, M.T., Guessoum, Z., Mazouzi, S., Cormier, S., Mezghiche, M.: A stochastic multi-agent approach for medical-image segmentation: application to tumor segmentation in brain MR images. Artif. Intell. Med. (2020)
8. Khalil, M.I., Humayun, M., Jhanjhi, N., Talib, M.N., Tabbakh, T.: Multi-class segmentation of organ at risk from abdominal CT images: a deep learning approach. Intell. Comput. Innov. Data Sci. (2021)
9. Wang, Z., Zou, Y., Liu, P.X.: Hybrid dilation and attention residual U-Net for medical image segmentation. Comput. Biol. Med. (2021)
10. Malhotra, P., Gupta, S., Koundal, D., Zaguia, A., Enbeyle, W.: Deep neural networks for medical image segmentation. J. Healthc. Eng. (2022)
11. Sun, L., Wang, J., Huang, Y., Ding, X., Greenspan, H., Paisley, J.: An adversarial learning approach to medical image synthesis for lesion detection. IEEE J. Biomed. Health Inform. (2020)
12. Xia, Y., Yang, D., Yu, Z., Liu, F., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Med. Image Anal. (2020)
13. Zhang, Y., Jiang, H., Miura, Y., Manning, C.D., Langlotz, C.P.: Contrastive learning of medical visual representations from paired images and text. arXiv [cs. CV] (2020)
14. Müller, D., Kramer, F.: MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. BMC Med. Imaging (2021)
15. Dawud, A.M., Yurtkan, K., Oztoprak, H.: Application of deep learning in neuroradiology: brain haemorrhage classification using transfer learning. Comput. Intell. Neurosci. **2019**, 4629859 (2019)
16. Swati, Z.N.K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., Lu, J.: Brain tumor classification for MR images using transfer learning and fine-tuning. Comput. Med. Imaging Graph. **75**, 34–46 (2019)
17. Wang, X., Li, H., Zheng, P.; Automatic detection and segmentation of ovarian cancer using a multitask model in pelvic CT images. Oxidative Med. Cell. Longev. (2022)
18. Gomes, R., Kamrowski, C., Mohan, P.D., Senor, C., Langlois, J., Wildenberg, J.: Application of deep learning to IVC filter detection from CT scans. Diagnostics (2022)
19. Ostmeier, S., Axelrod, B., Verhaaren, B.F.J., Mahammedi, A., Li, L.-J., Zaharchuk, G., Christensen;, S., Heit, J.J.: Non-inferiority of deep learning acute ischemic stroke segmentation on non-contrast CT compared to expert neuroradiologists. arXiv [eess. IV] (2022)
20. Ilse, M., Tomczak, J.M., Welling, M.L Attention-based deep multiple instance learning. Preprint at https://arxiv.org/abs/1802.04712 (2018)
21. Roy, K., Banik, D., Bhattacharjee, D., Krejcar, O., Kollmann, C.: LwMLA-NET: a lightweight multi-level attention-based NETwork for segmentation of COVID-19 lungs abnormalities from CT images. IEEE Trans. Instrum. Meas. **71**, 1–13 (2022). https://doi.org/10.1109/TIM.2022.3161690

22. Ouyang, Z., Zhang, P., Pan, W., Li, Q.: Deep learning-based body part recognition algorithm for three-dimensional medical images. Med. Phys. (2022)

23. Kushnure, D.T., Talbar, S.N.: MS-UNet: A multi-scale UNet with feature recalibration approach for automatic liver and tumor segmentation in CT images. Comput. Med. Imaging Graph. **89**, 101885 (2021). https://doi.org/10.1016/j.compmedimag.2021.101885

24. Benčević, M., Habijan, M., Galić, I.: Epicardial adipose tissue segmentation from CT images with a semi-3D neural network. Preprint at https://arxiv.org/abs/2202.04608 (2022)

25. Gasmi, A.: Deep learning and health informatics for smart monitoring and diagnosis. Preprint at https://arxiv.org/abs/2202.05951 (2022)

26. Sun, L., Chen, J., Xu, Y., Gong, M., Yu, K., Batmanghelich, K.: Hierarchical amortized GAN for 3D high resolution medical image synthesis. IEEE J. Biomed. Health Inform. **26**(2), 519–529 (2022)

27. Jha, M., Gupta, R., Saxena, R.: A framework for in-vivo human brain tumor detection using image augmentation and hybrid features. Health Inf. Sci. Syst. **10**(1), 1–19 (2022)

28. Vorontsov, E., Molchanov, P., Gazda, M., Beckham, C., Kautz, J., Kadoury, S.: Towards annotation-efficient segmentation via image-to-image translation. Med. Image Anal. **77**, 102197 (2022)

29. Adams, L.C., Busch, F., Truhn, D., Makowski, M.R., Aerts, H.J., Bressem, K.K.: What does DALL-E 2 know about radiology?. Preprint at https://arxiv.org/abs/2201.08394 (2022)

30. Ettinger, S., Sonnow, L., Plaass, C., Rahn, A., Stukenborg-Colsman, C., von Falck, C., Becher, C.: arthroscopic defect size measurement in osteochondral lesions of the talus underestimates the exact defect size and size measurement with arthro-MRI (MR-A) and high-resolution flat-panel CT-arthro imaging (FPCT-A). Knee Surg. Sports Traumatol. Arthrosc. **30**(1), 232–240 (2022)

31. Chang, C.Y., Huber, F.A., Yeh, K.J., Buckless, C., Torriani, M.: Utilization of a convolutional neural network for automated detection of lytic spinal lesions on body CTs. Skelet. Radiol. **52**(1), 91–97 (2023)

32. Arulappan, A., Thankaraj, A.B.R.: Liver tumor segmentation using a new asymmetrical dilated convolutional semantic segmentation network in CT images. Int. J. Imaging Syst. Technol. **32**(3), 815–830 (2022)

33. Zou, Q., et al.: Segmentation of brain tumor using Cross Patch Attention U-Net. J. Med. Imaging Health Inform. **10**(7), 1574–1580 (2020)

34. Yan, Y., Zhang, D.: Multi-scale U-like network with attention mechanism for automatic pancreas segmentation. PLoS ONE **16**(5), e0252287 (2021). https://doi.org/10.1371/journal.pone.0252287

35. Silva, P., Luz, E., Silva, G., Moreira, G., Silva, R., Lucio, D., Menotti, D.: COVID-19 detection in CT images with deep learning: a voting-based scheme and cross-datasets analysis. Inform. Med. Unlocked **20**, 100427 (2020). https://doi.org/10.1016/j.imu.2020.100427

36. Chen, S., Ren, Q., Qian, X., Wang, X., Liu, M.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. J. Med. Syst. **43**(8), 241 (2019)

37. Cao, Y., Zhou, W., Zang, M., An, D., Feng, Y., Yu, B.: MBANet: a 3D convolutional neural network with multi-branch attention for brain tumor segmentation from MRI images. Biomed. Signal Process. Control **80**(Part 1), 104296 (2023). https://doi.org/10.1016/j.bspc.2022.104296

38. Wang, Q., Shen, F., Shen, L., et al.: Lung nodule detection in CT images using a raw patch-based convolutional neural network. J. Digit. Imaging **32**, 971–979 (2019). https://doi.org/10.1007/s10278-019-00221-3

39. Mishra, Z., Ganegoda, A., Selicha, J., Wang, Z., Sadda, S.R., Hu, Z.: Automated retinal layer segmentation using graph-based algorithm incorporating deep-learning-derived information. Sci. Rep. **10**(1), 9541 (2020). https://doi.org/10.1038/s41598-020-66355-5

40. Chiang, P.-L., Lin, S.-Y., Chen, M.-H., Chen, Y.-S., Wang, C.-K., Wu, M.-C., Huang, Y.-T., Lee, M.-Y., Chen, Y.-S., Lin, W.-C.: Deep learning-based automatic detection of ASPECTS in acute ischemic stroke: improving stroke assessment on CT scans. J. Clin. Med. **11**(17), 5159 (2022). https://doi.org/10.3390/jcm11175159

41. Wang, X., Wang, S., Zhang, Z., Yin, X., Wang, T., Li, N.: CPAD-Net: contextual parallel attention and dilated network for liver tumor segmentation. Biomed. Signal Process. Control **79**(2), 104258 (2023). https://doi.org/10.1016/j.bspc.2022.104258

42. Wang, G., Song, T., Dong, Q., Cui, M., Huang, N., Zhang, S.: Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. Med. Image Anal. **65**, 101787 (2020). https://doi.org/10.1016/j.media.2020.101787