Using a Siamese Network to Accurately Detect Ischemic Stroke in Computed Tomography Scans

Ana Beatriz Gil Vieira
anabeatrizvieira@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

June 2022

Abstract

The diagnosis procedure of stroke, a leading cause of death in the world, involves the acquisition of images using computed tomography scans, making possible the assessment of the severity of the incident and the type and location of the lesion. The fact that the brain has two hemispheres with a high level of anatomical similarity, exhibiting significant symmetry, has led to extensive research based on the assumption that a decrease in symmetry is directly related to the presence of pathologies. This work is focused on the analysis of the symmetry (or lack of it) of the two brain hemispheres, and on the use of this information for the classification of computed tomography brain scans of stroke patients. The objective is to contribute to the process of automatic identification of brain lesions caused by stroke events. To perform this task, we used the Siamese Network architecture, which uses two parallel neural networks that share the same weights. The composed network receives a double image (the original image and the mirrored one) and a label that reflects the existence or not of stroke. The network then extracts the relevant features and classifies the images taking into account their similarity. The resulting network can be used to classify unseen scans, depending on the perceived level of symmetry into one of two existing classes: evidence of stroke or absence of stroke. The accuracy of the proposed method is approximately 72%, significantly outperforming a standard convolutional network architecture, which was used as a baseline.

Keywords: Ischemic Stroke, Computed Tomography, Symmetry Detection, Image Classification, Convolutional Neural Network, Siamese Network.

1. Introduction

According to the World Stroke Organization, stroke is the second leading cause of death and disability in the world, with about 13 million cases annually [1]. In this work, only the ischemic strokes were considered since they represent the majority of all strokes.

Ischemic stroke occurs when a vessel supplying blood to the brain is obstructed, leading to damage or death of brain cells. It is usually caused by blood clots in a brain vessel or by narrowing of the blood vessels that irrigate the brain due to a process of atherosclerosis [2].

The diagnosis is made based on the evaluation of the symptoms and images resulting from Magnetic Resonance Imaging (MRI) or Computed Tomography (CT - used in this paper) exams. CT is most often chosen due to its widespread availability and short imaging time, but the resulting images are harder to interpret by automated means.

The exams should be performed immediately upon admission, in order to identify and evaluate as rapidly as possible the existing lesions. The time elapsed since the onset of the stroke is crucial for the treatment and recovery of the patient. Recovery depends on the severity of the stroke, and the faster the diagnosis and treatment, the greater the chance of recovering the penumbra, which is the area around the ischemic region that can be recovered [2].

During the first few hours after the onset of stroke, CT scans should be used since they can detect the presence of ischemic or hemorrhagic lesions, pinpoint the location and extent of the lesion, and exclude situations that can be confused with strokes, known as stroke mimic [3].

Several types of CT scans can detect different changes in brain structures. The first exam most commonly performed is the Non-Contrast Computed Tomography (NCCT), since it enables the quick identification of the stroke type (ischemic or hemorrhagic) [4]. Depending on the type, other exams may follow. In our case, since only patients with ischemic stroke are considered, the exams that are usually performed later are: Computed Tomog-

raphy Perfusion (CTP) and Computed Tomography Angiography (CTA).

A human brain has two roughly symmetrical hemispheres, separated by the longitudinal fissure, which is a membrane filled with Cerebro-Spinal Fluid (CSF) [5]. This is usually known as the Mid-Sagittal Plane (MSP), because it is a virtual plane perpendicular to the brain, which divides it into left and right parts [5, 6].

It is widely acknowledged that hemispheres display both anatomical and functional asymmetries and this has been a topic of research over the years [7]. Several results have shown that the asymmetry of hemispheres is correlated with the presence of brain injuries [8, 9], tumors [10, 11] or mental illness [12].

The results presented in this paper are based on an approach that uses the level of symmetry between the two brain hemispheres to detect the probable presence or absence of stroke. The approach is based on the use of a Siamese Network, a particular neural network architecture developed to identify similarities between images, in this case, symmetric ones.

2. Related Work

Chin *et al.* [13] proposed a method to assist neurologists in the diagnosis, since the sooner the patient is treated, the better the chances of recovery. This work developed a Convolutional Neural Network (CNN) for the early detection of ischemic stroke, which was composed by five layers: two convolutional layers, one pooling layer, one fully connected layer and a softmax layer. The dataset is composed by CT images of the brain, with size 512x512, enriched with MRI data. However, since only the stroke area is considered, the authors captured a size 32x32 patch image, which is used as model input. Skull bone and other structures, such as, CSF that might mislead the model were removed, since these structures exhibited pixel intensities as the ischemic area. The authors report an accuracy of 93% in the classification task, where 128 patch images were used for training and 128 patch images were used for test, for a total of 256. Since this approach uses also MRI data, the results are not directly comparable with ours.

Recently, Herzog and Magoulas [14] proposed a method based on brain asymmetry to identify early dementia and its diverse stages, such as amnestic early mild cognitive impairment and Alzheimer's Disease. They analyzed the structural and functional cerebral changes in both hemispheres using supervised machine learning algorithms (Naive Bayes, Linear Discriminant, Support Vector Machine, K-Nearest Neighbor) and a convolutional neural network, AlexNet. The dataset was composed of brain asymmetries features, extracted from 750 MRI scans from the Alzheimer's Disease Neuroimaging Initiative database, which was normalized and resized to 256x256x3. For running the supervised machine learning algorithms, the 10-fold cross-validation was used, while for running the CNN, the images were resized to 227x227x3 (a requirement of the AlexNet architecture) and divided into training, testing and validation sets. The proposed pipeline achieved an accuracy that ranged from 75% to 93% for the mentioned diseases. Furthermore, this method offers a promising low-cost alternative for the classification of dementia and could potentially be useful in other brain degenerative disorders that are accompanied by visible changes in the brain symmetry. Again, since it uses MRI data it cannot be directly compared with ours.

To perform brain symmetry studies on large neuroimaging archives, reliable and automatic detection of the MSP is required to extract the brain hemispheres first. However, traditional planar estimation techniques fail when the MSP presents a curvature caused by existing pathology or a natural phenomenon known as brain torque. As a result, midline estimates can be inaccurate. To address this issue, Gibicar *et al.* [15] suggested an unsupervised midline estimation method that consisted of three main stages: head angle correction, control point estimation and midline generation. The technique was applied on a slice-to-slice basis for more accurate results and also provides accurate delineation of the midline in the septum pellucidum (exactly in the middle of the brain), which is a source of failure for traditional approaches. A midline validation dataset was created by sampling 75 Fluid-Attenuated Inversion Recovery (FLAIR) MRI volumes (around 3000 images) from three different databases. To validate the performance of the proposed midline estimation algorithm, the work was evaluated over a set of ground truth images and compared with to two other methods for midline estimation [16, 17]. Performance is quantified with three different validation metrics: mean Hausdorff distance, mean absolute distance, and mean volume-difference, which all compare the automated midline to the ground truth delineations.

Inspired by the characteristics of the Siamese Network, Barman *et al.* [18] proposed a Siamese Neural Network (DeepSymNet) for the detection of ischemic stroke from CTA volumes. This method enabled them to detect the changes in symmetry of vascular and brain tissue texture of the two brain hemispheres in parallel. The model was tested on a clinical dataset of 217 patients with two different approaches: original CTA and brain tissue only. With the first alternative, the authors tested the

network's ability to recognize strokes with the original 3D images, which contain asymmetries in both vascular structures and brain tissue. Then, removing the vasculature, they evaluated the network's ability to recognize strokes by analyzing only the brain tissue, i.e., detecting the darker textures in the brain tissue. In these two approaches, the data was resized to 29x73x20 using bilinear interpolation. The proposed model achieved a Area Under the ROC Curve (AUC ROC) of 0.89 for CTA volumes and 0.91 using only the brain tissue.

Tummala [19] proposed a deep learning framework that uses Siamese neural networks for computer aided diagnosis of autism spectrum disorder (ASD) using T1-weighted MRI. The dataset is composed of 102 control and 112 ASD patients. Preprocessing of the images involved reorienting them to a standard space, followed by applying the cropping technique to remove the neck regions, where FSL interfaces were used. After that, affine registration was applied, and finally, the images were reduced and resized to 224x224x3 to match the ResNet-50 input size. It was implemented a Siamese network with pre-trained ResNet-50 model to train and test 1070 positive and negative images pairs. Afterwards, the L1-norm was computed between the embeddings of the two inputs which is further used to backpropagate the error computed using the contrastive loss function. The model was trained using 5-fold stratified cross-validation, achieving an accuracy, recall, precision and f1-score of 0.99 within 50 epochs in the validation set.

Our approach differs from these and other related works in that we use only CT data obtained at the time of hospital admission. This makes the problem harder, since the changes in the brain images are more subtle, but also potentially more relevant in a clinical setting.

3. Implementation

3.1. Dataset

The data used in this work was collected from ischemic stroke patients at *Hospital de Santa Maria*. The dataset is composed of non-contrast computed tomography images in DICOM format, which is the most commonly used format for storing medical images. The data was properly anonymized to preserve participant privacy and fulfill all ethical requirements.

Since we decided to use only the most relevant slices of each NCCT, i.e. those in which the ischemic stroke can be detected, we converted the 3D data to 2D slices, each with the size of 512x512 pixels. The final dataset used in the tests consists of slices chosen from different brain regions of different patients, in order to make the model more robust.

The dataset is balanced and includes 340 images of each class, where one class corresponds to the absence of stroke (and therefore symmetry between both hemispheres) and the other class corresponds to the visible effects of a stroke event (and lack of symmetry).

Since all CT scan images correspond to patients with ischemic stroke, the selection of slices for each class was made based on the ASPECTS (Alberta Stroke Program Early Computer Tomography Score) assigned to each of the patients. This is a quantitative score used by neurologists to assess the severity of an ischemic stroke, where the Middle Cerebral Artery (MCA) is divided into 10 defined regions, and for each area showing early ischemic changes on at least two consecutive slices, one point is subtracted from the overall score [20]. The scale ranges from 0 to 10, where 10 means that the CT scan is normal.

Taking into account that lower scores correspond to more severe strokes and, consequently, with more visible lesions, the respective slices were assigned to the "presence of stroke" class. On the other hand, slices corresponding to patients with an ASPECTS score close to 10, and who did not present readily identifiable lesions were assigned to the "absence of stroke" class. The images from the class where stroke effects were visible have been annotated by an expert neurologist at *Hospital de Santa Maria*.

Given that the dataset is limited in size and in order to avoid model overfit, we applied data augmentation techniques, namely Gaussian blur, to the training set. The idea of using this type of data augmentation was due to the fact that we want to "force" the model to learn to identify the images that contain asymmetries.

In addition, we decided to use stratified 5-fold cross-validation, where the whole dataset is equally divided into 5 folds, one-fold is used for testing and the remaining are used for training. This process is repeated 5 times until all folds are used for testing.

In this paper, we have two distinct approaches when dividing the data. The first approach is to split the data randomly, while the second approach is to split the samples by group, in this case by subject, i.e. the data is split in a way that ensures that all samples from the same patient are in the same fold, avoiding overfitting and data leakage.

The performance of the cross-validation is obtained by calculating the mean and standard deviation of the model performance on the test set in each iteration.

3.2. Image Preprocessing

We used image preprocessing techniques to transform raw image data into a clean image data. In this work, this process was one of the most time-

consuming since CT scan images are very complex and present undesired characteristics. The main preprocessing steps will be described next. All the applied techniques were developed with Python 3.7 and using the PyTorch framework, which is one of the preferred platforms for deep learning research.

Windowing. A CT image consists of a set of points called voxels. The voxels represent the three-dimensional tissue volumes. The 'color' of each voxel is conventionally displayed as a shade of grey, corresponding to the density of brain tissue at that location, and is called the attenuation value, which is given in Hounsfield Units (HU) [21, 22]. It is a relative quantitative measurement of radio density used by radiologists in the interpretation of CT images, and is calculated as follows:

$$HU = \frac{(\mu_v - \mu_w)}{\mu_w} \times K \tag{1}$$

where μ_v is the calculated voxel attenuation coefficient, μ_w is the water attenuation coefficient and K is an integer constant, standardized as taking the value 1000 [21].

To set the appropriate window for the color range, we define two parameters: Windows Level (WL), also known as Windows Center, and Windows Width (WW). The windows level is the midpoint of the range of the CT numbers and is responsible for the image brightness, while windows width is the range of the CT numbers that an image contains and is responsible for the image contrast.

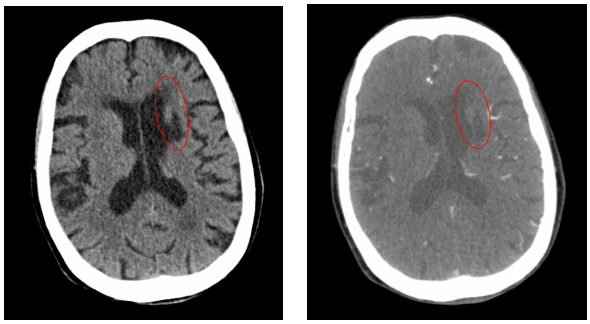
The upper and lower grey levels in the window are given by:

$$UpperLevel = WL + (WW/2) \tag{2}$$

$$LowerLevel = WL - (WW/2) \tag{3}$$

All values above the upper level will be viewed as white and all values below the lower level will be viewed as black [23]. In our case we want to see the details of the brain, so we need to choose a window that maximizes contrast while not losing information. If the range is chosen poorly some gradations may be not visible [23]. We found out that picking a range between 0 and 80 for CTs, i.e., WL=40 and WW=80 provided best results, which coincides with the window most commonly selected for stroke analysis.

Figure 1 shows two sample images from an NCCT (WL:40, WW:80) and CTA (WL:60, WW:360), respectively, where a lesion caused by an ischemic stroke can be seen near the left ventricle. The lesion is marked by a red circle. As is clear from the figure, the choice of the correct window enables us to see more details and, consequently, makes it easier to detect the asymmetry between the two hemispheres.



(a) NCCT image (b) CTA image

Figure 1: Ischemic stroke lesions after the implementation of two different windowing techniques.

Head Tilt Correction. A common phenomenon in medical imaging devices is that they produce distorted brain images, which can mislead visual inspection and lead to false clinical interpretation. The main reasons for this are the mobility of the patients, the inexperience of the technicians, and the imprecision of calibration systems [8].

In order to correct the head tilt it is necessary to find the rotation angle, in order to align the MSP with the y-axis [14, 5]. For that, the external contours of the brain were determined and an ellipse that best fits these contours was computed. Finally, the angle between the ellipse and the vertical axis was computed. Figure 2 illustrates the same brain slice with and without head tilt correction.

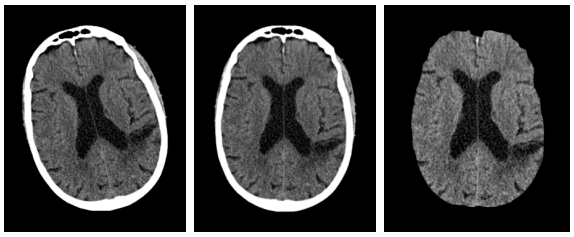
Since the design of the external contours does not always fit the head optimally, there are images where the orientation of the head is not correctly aligned with the y-axis. This happens because of the slight (white) lines around the brain that causes interference in determining external contours.

Skull Stripping. Skull stripping is a fairly common process when using CT imaging because we can focus only on the brain tissue, which is where the lesions are, and we can perform a better segmentation of the different brain areas [13].

In fact, before we did the skull stripping we noticed that the network attention was mainly focused on the eye region or the bones around the brain. To get a better system performance and following expert advice, we decided to remove the bone leaving only the brain tissue.

The process of skull stripping consisted of doing a subtraction of the original image with an image where only the bone was visible. Using a windowing process, a mask was created with the bone and this

image was subtracted from the original one, creating an image with only the brain tissue. Figure 2 illustrates a slice before and after skull stripping.



(a) Original slice (b) Head tilt correction (c) Skull stripping

Figure 2: Main steps in CT images processing.

3.3. Symmetry detection

Originally proposed for signature verification, similarity detection has been used widely for various computer vision applications [18, 24, 25, 26, 27].

This network is composed of two twin neural networks, which share the same weights and parameters, but are joined by a loss function on top [24]. Unlike traditional machine learning classifiers that receive one input and predict its class, Siamese networks receive two distinct inputs and predict whether or not they belong to the same class. Therefore, instead of learning a classifier, this network learns a similarity function between inputs.

Similarity detection can be used to perform brain symmetry detection by using as inputs the original image and a mirrored image. Figure 3 shows the schematic of our Siamese Network.

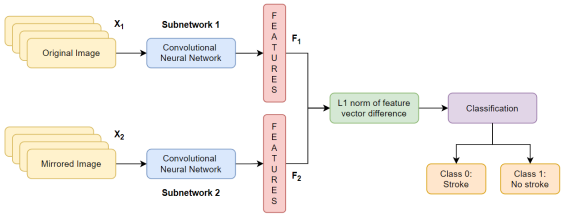


Figure 3: Siamese Network schematic.

The first step is the feature extraction, where we used two CNNs (subnetwork 1 and subnetwork 2) to extract the features of each image. They are trained with a label that reflects whether the two input images correspond to a visible stroke event (and are therefore approximately symmetric) or not. The output of each subnetwork was subsequently flattened (F_1 and F_2).

The second step is the comparison between the two feature vectors. The difference of each subnetwork output is computed, and the L1 norm is used to generate a new feature vector given by:

$$D = \|F_1 - F_2\|_1 \tag{4}$$

Finally, the network output is the probability of the two input images being symmetric or not, reflecting the probability of stroke. The smaller the value of D, the more likely the images are to be symmetric and, therefore, to not exhibit evidence of stroke.

3.4. Architecture

Since the Siamese network (SiameseNet) is made up of two equal networks, each of them has a set of images as input. One of the networks, in figure 3, subnetwork 1, takes as input an (original) set of slices with the size 512x512, while subnetwork 2 takes as input the same set of mirrored slices. The purpose of this is to make it possible to compare the two hemispheres (right and left) simultaneously.

Figure 4 shows in detail the characteristics of the different layers The feature extraction process is composed of 2D convolutional layers, 2D batch normalization layers and max pooling layers and a flatten layer, generating a feature vector with a size of 8192. The entire network uses ReLU activations, except for the prediction layer that uses softmax activation.

After feature extraction and subtraction, two fully connected layers were used, the first generating a vector of size 1024 and the second generating a vector of size 2, one output unit for each class.

The model was trained and evaluated using 5-fold cross-validation, for 200 epochs with a batch size of 32, using the ADAM optimizer with an initial learning rate $\gamma = 0.00001$. Since high weight values increase the chances of overfitting, L2 regularization with a weight factor $\lambda = 0.0005$ was used, applying a penalty for high weight values. The model required 10M trainable parameters.

3.5. Loss Function

Usually when working with SiameseNet the preferred loss function is the contrastive loss since it gives a similarity score by calculating the Euclidean distance between the feature vectors [24]. However, once this is a classification task, we choose to use the cross-entropy loss and, as we have only two classes, is given by:

$$L(y_i, \hat{y_i}) = -(y_i \log \hat{y_i} + (1 - y_i) \log(1 - \hat{y_i})) \tag{5}$$

where y_i and $\hat{y_i}$ are the actual class and the predicted class, respectively.

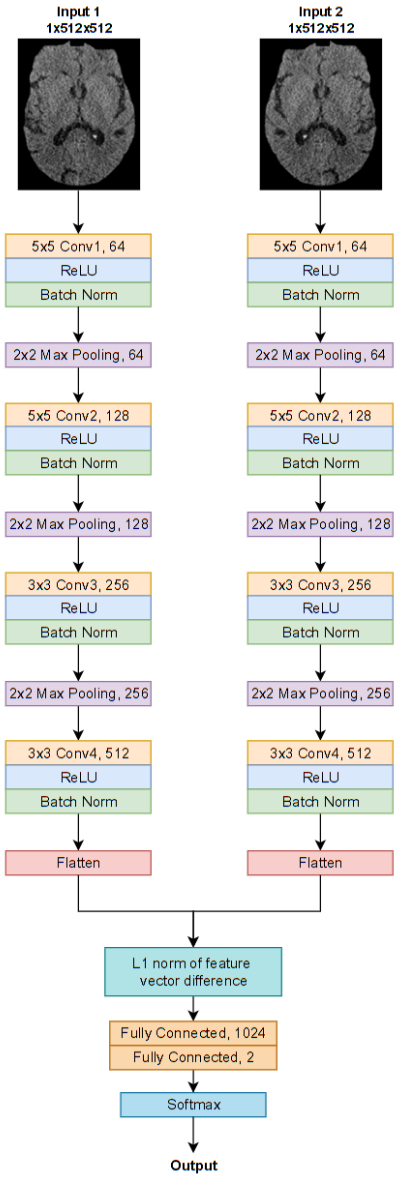


Figure 4: Siamese Network architecture.

4. Results

To assess the performance of the Siamese network architecture on this dataset, we performed an empirical comparison with two alternative baseline architectures, the SimResNet-18 and the ResNet-50. All models were trained using a PowerEdge C41402 server with an NVIDIA 32GB Tesla V100S.

4.1. ResNet-50

Convolutional neural networks exhibit a high performance on image classification tasks but can be difficult to train due to the complexity and the vanishing gradient problem - caused by the derivative/gradient of the activation function that decreases exponentially as it propagates down to the initial layers, which means that the weights and bias of the initial layers will not be updated properly - and, also to the difficulty in optimizing the training process as a result of the large number of parameters [28, 29].

Thereby, to fight these problems the Residual Network (ResNet) was proposed, which presents (residual) connections in parallel with the convolutional networks [29].

The name residual is based on the fact that the layers learn not the true result, $H(x)$, as in traditional networks, but the residual, $R(x)$. Consider a neural network block, where the input is denoted by x and is intended to learn the true output $H(x)$. Then, the difference (residual) between the two is given by:

$$\begin{aligned} R(x) = Output - Input = H(x) - x \\ \Leftrightarrow H(x) = R(x) + x \end{aligned} \tag{6}$$

The residual block is trying to learn the true output H(x), but since the connection is an identity matrix, then the layers are actually learning the residual R(x).

In contrast to convolutional layers, these connections are always active and therefore allow gradients to propagate through them, leading to a faster training process [28].

Inspired by the VGG network philosophy, the researchers created the ResNet, which consists of convolutional layers, usually with 3x3 filters. The downsampling process is performed directly through the convolutional layers with stride 2. The network ends with an average pooling layer and a fully connected layer with softmax [29].

Since ResNet exhibits high performance in image classification tasks, we decide to use it as baseline architecture. In this way, we tested ResNet with 18 layers, 34 layers, and 50 layers (three common variations of ResNet) and found that the best performance was obtained using ResNet-50. In addition, the choice of ResNet-50 over the other two models tested is due to the fact that it has more layers and therefore can learn more features, which can make a difference considering the complexity of the images. With this model, we required 174 minutes of training time using a network with 23M parameters.

4.2. SimResNet-18

Initially motivated by the idea of combining ResNet and the Siamese Network architecture [27], we also tested an alternative approach where the ResNet is used only for the feature extraction step. As we have chosen the ResNet, the main difference between the previous architecture and this one is the residual connections presented in the ResNet model.

The Siamese Network consists of two ResNet, where one of them accepts the original brain im-

age and the other accepts the mirrored one and extracts its features. Since ResNet is composed of a Global Average Pooling (GAP) layer after the last convolution layer, the resulting output is a feature vector with a size of 512 which is used to calculate the absolute value.

The GAP layer is designed to replace fully connected layers in classical CNNs. The idea is to generate one feature map for each corresponding category of the classification task in the last convolutional layer [30]. The GAP computes the average of each feature map until each spatial dimension is one. This technique avoids the model overfit.

The remaining architecture is similar to that described for the SiameseNet. The feature vectors are subtracted, the absolute value of the difference is computed, and the images are classified into one of the two existing classes using a fully connected layer.

In this architecture, we tested different ResNet models, and after analyzing the number of parameters, the training time, and the results obtained, we opted for the 18-layer model. Figure 5 illustrates the SimResNet-18 architecture. The hyperparameters used were the same as for the Siamese network. This approach required 154 minutes to train and 11M parameters.

4.3. Experimental Results

Table 1 summarizes the classification performance of the two developed architectures, SiameseNet and SimResNet-18, and the baseline architecture, ResNet-50.

Table 1: Model classification performance.

Model	Accuracy	Precision	Recall	F1-Score
ResNet-50	0.62 ± 0.086	0.71 ± 0.140	0.47 ± 0.268	0.57 ± 0.166
SimResNet-18	0.64 ± 0.040	0.73 ± 0.124	0.48 ± 0.123	0.58 ± 0.070
SiameseNet	0.72 ± 0.127	0.75 ± 0.165	0.69 ± 0.252	0.72 ± 0.144

The results presented in the table show that, although the accuracy achieved is not very high (72%), the Siamese network outperforms the other two models. Moreover, it is the best result known to us for this specific problem, ischemic stroke classification from CT scans. The runtime for training the SiameseNet in the proposed dataset was 144 minutes.

The results obtained with ResNet-50 are due to the fact that CT images, although labeled, have several features that make model learning difficult. In fact, the model looks for changes in the brain hemispheres, but most of the time it cannot distinguish them from anatomical asymmetries, which means that the extracted features are not the most relevant for this specific problem.

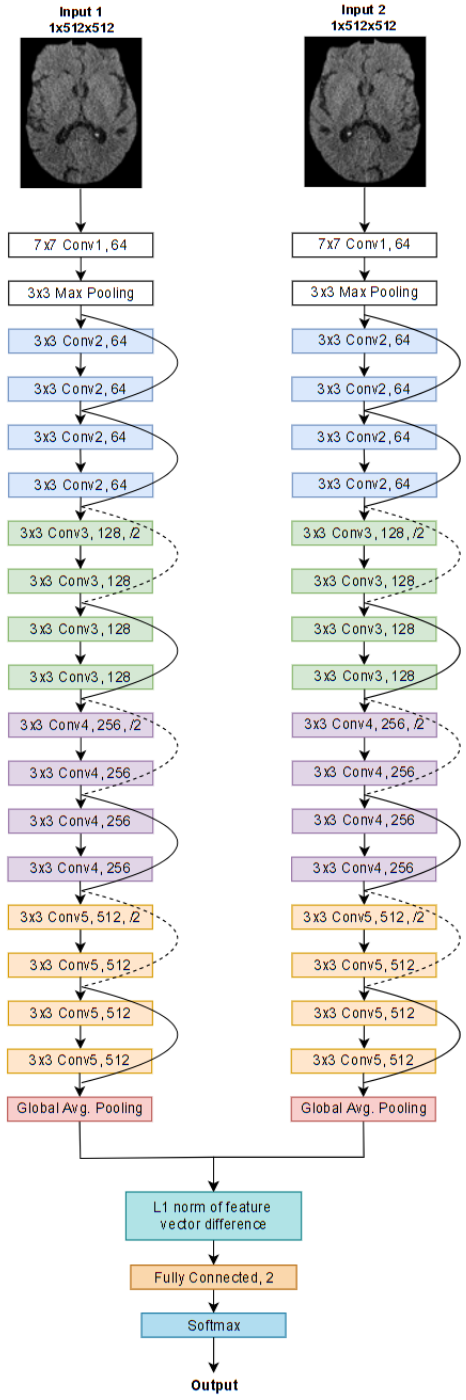


Figure 5: SimResNet-18 architecture.

On the other hand, when adding ResNet to the Siamese network (SimResNet-18), we see an improvement in performance, although not significant. Finally, when using the Siamese network, we see that there was a considerable improvement in model performance. However, these low accuracy results are mainly due to the fact that the dataset is small, considering the complexity of the problem.

The high value of the standard deviation of the

performance measures is essentially due to two points. First, since the dataset is small, with few labeled images, it cannot be guaranteed that all patients have the same number of samples. In fact there are patients with about 45 slices and others with about 3 slices. Also, the guarantee that all slices from the same patient were in only one of the two sets (training or testing), led the model to learn better the cases where there is a high number of sequential slices, such as the case of 45 slices from the same patient. In this way, we can say that the model contains a sampling bias that affects the performance of the model. This is one of the most common types of dataset bias and occurs when randomization is not properly achieved during data collection, resulting in poor generalization of the learning algorithm.

On the other hand, if it is not guaranteed that all samples from the same patient are in only one of the two sets, then there is data leakage, leading to the model not learning the features relevant to the problem but learning features to relate/sequence the slices (of the same patient).

In this case, splitting the dataset randomly, and training the same three models, we obtained the results shown in the table 2.

Table 2: Model classification performance with data leakage.

Model	Accuracy	Precision	Recall	F1-Score
ResNet-50	0.89 ± 0.037	0.87 ± 0.082	0.92 ± 0.069	0.89 ± 0.045
SimResNet-18	0.91 ± 0.033	0.89 ± 0.039	0.93 ± 0.040	0.91 ± 0.034
SiameseNet	0.97 ± 0.017	0.96 ± 0.015	0.97 ± 0.028	0.97 ± 0.017

The results show that not only SiameseNet achieved very high accuracy results (97%) but also SimResNet-18 (91%) and ResNet-50 (89%).

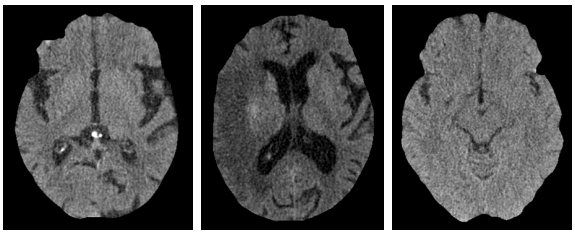
Thus, we can state that in this case, the results presented are not reliable, and that the network is using features that are not relevant to the problem, such as the dimension of the slices, or some other features that are not correlated with the stroke effects.

Chronologically, the first results obtained were those in which the dataset was randomly split, i.e., the results represented in the table 2. After observing the results, there were some doubts that they were correct and it was finally concluded that there was data leakage. Only then the 5-fold cross-validation by subject was implemented leading to the results in table 1.

4.4. Examples of predictions

Figure 6 illustrates some examples of good predictions by the SiameseNet architecture, using the stratified 5-fold cross-validation by subject.

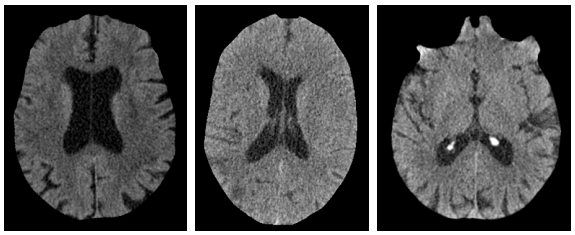
In the case of class 0 images, 6a and 6b, these ones contain strong visible changes in one of the hemispheres that help the model learn and correctly classify the samples. In the case of the image corresponding to class 1, i.e. the image present in 6c, the absence of variations in the hemispheres and the high anatomical similarity enabled the model to learn and correctly classify the image.



(a) (b) (c)

Figure 6: Examples of good predictions by the SiameseNet model. The images shown in a) and b) are examples of slices belonging to class 0 - presence of stroke and, the image shown in c) is an example of slices belonging to class 1 - absence of stroke.

On the other hand, figure 7 illustrates some examples of wrong predictions by the same model and in the same conditions.



(a) (b) (c)

Figure 7: Examples of wrong predictions by the SiameseNet model. The images shown in a) and c) are examples of slices belonging to class 1 - absence of stroke and, the image shown in b) is an example of slices belonging to class 0 - presence of stroke.

Considering images 7a and 7c belonging to class 1, i.e, absence of stroke, the model misclassified these images due to the high asymmetry present in the image, as a result of the brain anatomy itself. In this case, it can be stated that the model failed to learn to distinguish between stroke injuries and brain anatomy.

In the case of image 7b, which is an example of class 0, the presence of stroke (in the right hemisphere) is so subtle that the model has not learned to classify these less visible lesions.

5. Conclusions and Future Work

We proposed to use a Siamese Network architecture for the detection of ischemic stroke, based on the

fact that detected asymmetry of brain hemispheres is strongly correlated with evidence of stroke. The architecture obtained 72% accuracy, on average, on an independent test set. Although the model exhibits an accuracy insufficient for medical applications, this architecture may become the starting point for several other studies focusing on the symmetry of the brain hemispheres and how it is related to the existence of pathologies in the brain.

Although the results were not as positive as expected, a large part of this work was on the preprocessing of the images. Therefore, a labeled and preprocessed dataset is now available and can be used for future research.

During the development of the work, one of the concerns was the choice of images, since the model could learn unwanted features. By achieving very high performance, we realized that the model was learning something else since there are sequences of similar images in both sets (training and test).

Another reason that explains the results obtained is related to the complexity of the images. The hemispheres present a certain anatomical similarity that can mislead the learning of the features associated with stroke.

There are some aspects of the presented work that can be improved. One of the aspects that can be improved is image preprocessing. Although the skull stripping process is mostly successful there are slices where small portions of non-brain tissue are mistakenly removed. This happens because all the white pixels in the image are removed when only the white pixels around the brain (bone) should be removed.

Another possible improvement is the removal of the CSF area since the intensity of the pixels is the same as the ischemic zone and can mislead the model.

Finally, since the dataset is not large and extensive enough, the performance of the model falls short. In fact, only slices from 60 patients were used, with 35 patients associated with class 1 and 25 patients assigned to class 0. So, one possibility of improving the performance of the model would be to label more data and to select the same number of slices from each patient, avoiding overfitting, data leakage, and selection bias.

References

[1] M Patrice Lindsay, Bo Norrving, Ralph L Sacco, Michael Brainin, Werner Hacke, Sheila Martins, Jeyaraj Pandian, and Valery Feigin. World Stroke Organization (WSO): Global Stroke Fact Sheet 2019. *International Journal of Stroke*, 14(8):806–817, oct 2019.

[2] Michael S Phipps and Carolyn A Cronin. Management of acute ischemic stroke. *BMJ*, 368:l6983, feb 2020.

[3] Marietta Pohl, David Hesszenberger, Krisztian Kapus, Janos Meszaros, Andrea Feher, Imre Varadi, Gabriella Pusch, Eva Fejes, Antal Tibold, and Gergely Feher. Ischemic stroke mimics: A comprehensive review. *Journal of Clinical Neuroscience*, 93:174–182, nov 2021.

[4] Olli Öman, Teemu Mäkelä, Eero Salli, Sauli Savolainen, and Marko Kangasniemi. 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. *European Radiology Experimental*, 3(1):8, dec 2019.

[5] Guilherme C. S. Ruppert, Leonid Teverovskiy, Chen-Ping Yu, Alexandre X. Falcao, and Yanxi Liu. A new symmetry-based method for mid-sagittal plane extraction in neuroimages. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 285–288. IEEE, mar 2011.

[6] Huisi Wu, Xiujuan Chen, Ping Li, and Zhenkun Wen. Automatic Symmetry Detection From Brain MRI Based on a 2-Channel Convolutional Neural Network. *IEEE Transactions on Cybernetics*, 51(9):4464–4475, sep 2021.

[7] Maria Elena Miletto Petrazzini, Valeria Anna Sovrano, Giorgio Vallortigara, and Andrea Messina. Brain and Behavioral Asymmetry: A Lesson From Fish. *Frontiers in Neuroanatomy*, 14:11, mar 2020.

[8] Sheena Xin Liu. Symmetry and asymmetry analysis and its implications to computer-aided diagnosis: A review of the literature. *Journal of Biomedical Informatics*, 42(6):1056–1064, dec 2009.

[9] Anusha Vupputuri, Susheelkumar Dighade, P S Prasanth, and Nirmalya Ghosh. Symmetry determined superpixels for efficient lesion segmentation of ischemic stroke from MRI. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 742–745. IEEE, jul 2018.

[10] Zeynab Barzegar and Mansour Jamzad. Fully automated glioma tumour segmentation using anatomical symmetry plane detection in multimodal brain MRI. *IET Computer Vision*, 15(7):463–473, oct 2021.

[11] Chen-Ping Yu, Guilherme Ruppert, Dan Nguyen, Alexandre Falcão, and Yanxi Liu. Statistical Asymmetry-based Brain Tumor Segmentation from 3D MR Images. In *Biosignals*, 2012.

[12] Michele Ribolsi, Zafiris J Daskalakis, Alberto Siracusano, and Giacomo Koch. Abnormal Asymmetry of Brain Connectivity in Schizophrenia. *Frontiers in Human Neuroscience*, 8:1010, dec 2014.

[13] Chiun-Li Chin, Bing-Jhang Lin, Guei-Ru Wu, Tzu-Chieh Weng, Cheng-Shiun Yang, Rui-Cih Su, and Yu-Jen Pan. An automated early ischemic stroke detection system using CNN deep learning algorithm. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, pages 368–372. IEEE, nov 2017.

[14] Nitsa J Herzog and George D Magoulas. Brain Asymmetry Detection and Machine Learning Classification for Diagnosis of Early Dementia. *Sensors*, 21(3):778, jan 2021.

[15] Adam Gibicar, Alan R Moody, and April Khademi. Automated Midline Estimation for Symmetry Analysis of Cerebral Hemispheres in FLAIR MRI. *Frontiers in Aging Neuroscience*, 13:644137, apr 2021.

[16] Felipe P G Bergo, Alexandre X Falcão, Clarissa L Yasuda, and Guilherme Ruppert. Fast, accurate and precise mid-sagittal plane location in 3D MR images of the brain. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 278–290. Springer, 2008.

[17] Hugo J Kuijf, Susanne J van Veluw, Mirjam I Geerlings, Max A Viergever, Geert Jan Biessels, and Koen L Vincken. Automatic extraction of the midsagittal surface from brain MR images using the Kullback–Leibler measure. *Neuroinformatics*, 12(3):395–403, 2014.

[18] Arko Barman, Mehmet E Inam, Songmi Lee, Sean Savitz, Sunil Sheth, and Luca Giancardo. Determining Ischemic Stroke From CT-Angiography Imaging Using Symmetry-Sensitive Convolutional Networks. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1873–1877. IEEE, apr 2019.

[19] Sudhakar Tummala. Deep Learning Framework using Siamese Neural Network for Diagnosis of Autism from Brain Magnetic Resonance Imaging. In *2021 6th International Conference for Convergence in Technology (I2CT)*, pages 1–5, 2021.

[20] Julian Schröder and Götz Thomalla. A Critical Review of Alberta Stroke Program Early CT Score for Evaluation of Acute Stroke Imaging. *Frontiers in Neurology*, 7:245, jan 2017.

[21] Lee W Goldman. Principles of CT and CT Technology. *Journal of Nuclear Medicine Technology*, 35(3):115–128, sep 2007.

[22] Tami D DenOtter and Johanna Schubert. *Hounsfield Unit*. jan 2022.

[23] Zhiyun Xue, Sameer Antani, L Rodney Long, Dina Demner-Fushman, and George R Thoma. Window classification of brain CT images in biomedical articles. *AMIA Annual Symposium proceedings. AMIA Symposium*, 2012:1023–1029, 2012.

[24] Sounak Dey, Anjan Dutta, Juan Ignacio Toledo, Suman K Ghosh, Josep Lladós, and Umapada Pal. SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification. *CoRR*, abs/1707.0, 2017.

[25] Xuning Liu, Yong Zhou, Jiaqi Zhao, Rui Yao, Bing Liu, and Yi Zheng. Siamese Convolutional Neural Networks for Remote Sensing Scene Classification. *IEEE Geoscience and Remote Sensing Letters*, 16(8):1200–1204, aug 2019.

[26] Iaroslav Melekhov, Juho Kannala, and Esa Rahtu. Siamese network features for image matching. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 378–383. IEEE, dec 2016.

[27] Kai Qiu, Yunfeng Ai, Bin Tian, Bin Wang, and Dongpu Cao. Siamese-ResNet: Implementing Loop Closure Detection based on Siamese Network. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 716–721. IEEE, jun 2018.

[28] Mohammad Sadegh Ebrahimi and Hossein Karkeh Abadi. Study of Residual Networks for Image Recognition. *CoRR*, abs/1805.0, 2018.

[29] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[30] Min Lin, Qiang Chen, and Shuicheng Yan. Network In Network. *arXiv preprint arXiv:1312.4400*, dec 2013.