Alberta Stroke Program Early CT Score Region Segmentation

Rafael de Freitas Brito*, Paulo Antonio Guimarães Bettero , Antônio Cláudio Paschoarelli Veiga and João Batista Destro Filho

Universidade Federal de Uberlândia, Brazil

Abstract.

Stroke is one of the leading causes of death and disability in the world, affecting low and middle-income countries in particular. The Alberta Stroke Program Early CT Score (ASPECTS) quantifies the extent of early ischemic changes in computed tomographies (CTs) of stroke patients and is widely used as a patient selection tool in stroke care. The score, however, isn't very reliable (has low interrater agreement), such that computational methods for assisted or even automated diagnostics could improve its reliability and, consequently, stroke care. Some automated ASPECTS scoring tools have already been developed to this day, both in an academic and a commercial context, and although almost all of them perform some sort of region segmentation, this step is often very briefly discussed in ASPECTS automation papers. Given this gap in literature, this paper proposes and evaluates an ASPECTS region segmentation algorithm, using a public CT template, CT database (CQ500) and a public library for image processing (SimpleITK). The proposed method is composed by four steps: pre-processing/slice selection, linear registration, non-linear registration and display/evaluation. Overall, it obtained a mean Dice coefficient of 0.6587 with a standard deviation of 0.0595 and a mean Hausdorff distance of 14.3903 with a standard deviation of 4.4366, for the 10 CTs evaluated.

Keywords: ASPECTS; Atlas Registration; Computer-aided Diagnosis; Computed Tomography; Digital Image Processing

1. Introduction

According to the World Health Organization, stroke is one of the leading causes of death and disability in the world, specially in low and middle income countries, where 70% of strokes and 87% of stroke related deaths occur (Johnson et al., 2016).

There are two main types of stroke: ischemic and hemorrhagic stroke. Ischemic stroke is characterized by the blockage of a blood vessel by a clot, while hemorrhagic stroke, as the name suggests, happens when a blood vessel ruptures causing hemorrhage. Overall, ischemic stroke is the most common (around 80% of cases), but also the least lethal, with a death rate of 13-23% against 25-35% presented by the hemorrhagic variety (Feigin et al., 2009).

Computed Tomography (CT) imaging is very relevant in stroke care, as it is used for distinguishing the stroke type as well as the extent of the lesion caused by it, allowing patient selection for stroke care (Puetz et al., 2009).

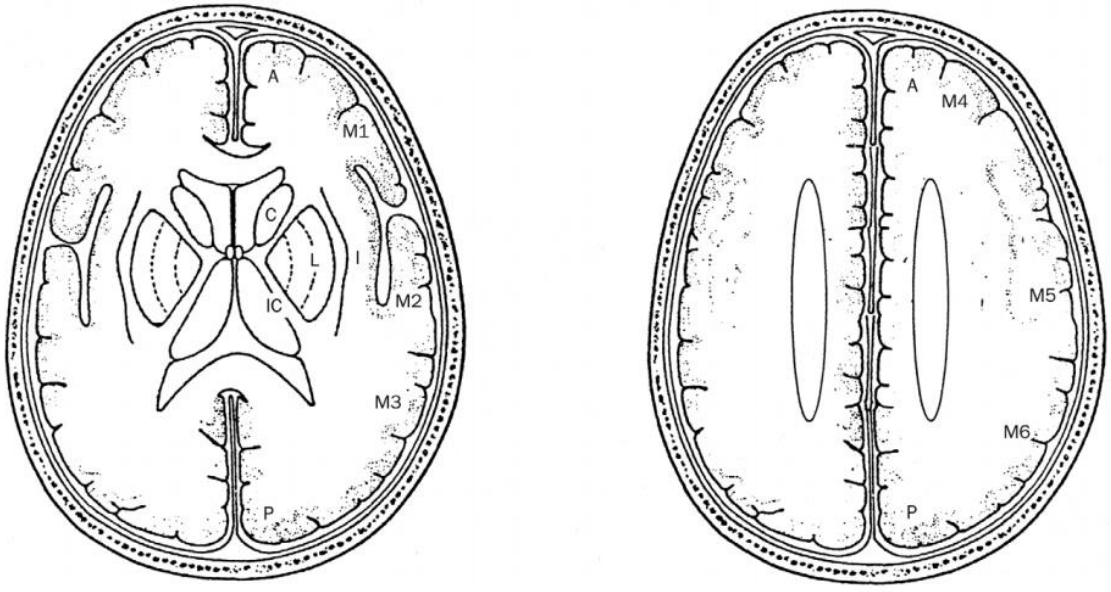
Thrombolytic therapy is the main treatment for acute ischemic stroke, in which a thrombolytic medicine (usually rt-PA alteplase) is administered as to break down clots blocking blood vessels and restore cerebral blood flow. This therapy, however, increases the risk of hemorrhages, and as such patient selection becomes very important to assess risk and deliver proper care (Wardlaw et al., 2012).

Besides other criteria, as glucose level, initially the "⅓ rule" was used for patient selection, determining as eligible anyone with signs of ischemia (hypoattenuation) in up to one third of the Middle Cerebral Artery (MCA) territory, as these patients benefited less from the therapy. The rule, however, wasn't very reliable (low interrater agreement), and as an alternative, a new CT scoring method was developed: the Alberta Stroke Program Early CT Score (ASPECTS) (Barber et al., 2000).

1.1 ASPECTS

ASPECTS evaluation is done by first dividing the MCA territory in 10 regions, selected from 2 standard axial CT slices. The first slice is the at the level of the thalamus and basal ganglia, and is divided in 7 regions: caudate nuclei (C), lentiform nuclei or putamen (L), internal capsule (IC), insular ribbon or insula cortex (I), anterior MCA cortex (M1), MCA cortex lateral to the insular ribbon (M2) and posterior MCA cortex(M3). The second slice is rostral to the ganglionic structures, and is divided in 3 regions: anterior MCA cortex (M4), lateral MCA cortex (M5) and posterior MCA cortex (M6). Both slices and its regions can be seen in Fig. 1.

Figure 1: First and second slices used for ASPECTS evaluation



Source: (Barber et al., 2000).

Each of these individual regions is later assessed for early ischemic changes (EICs), such as focal swelling, and parenchymal hypoattenuation, defined as a region with abnormally low attenuation ("darker") when compared to the same region of the contralateral hemisphere. Finally, for each region with EICs a point is subtracted from 10. A healthy CT scan has an ASPECTS of 10, while an CT with EICs throughout the entire MCA territory has a score of 0 (Barber et al., 2000).

The score was also found to be a very good prognostics tool for stroke patients: as it decreases, the odds of a good clinical outcome fall, especially when crossing the threshold of 7. It is also used as a patient selection tool, as thrombolysis is most effective for patients with ASPECTS over 7, while with anything from 5 to 7 its benefits are uncertain (Hill, 2005).

Although an improvement over the "⅓ rule", ASPECTS still had a limited reliability (interrater agreement). In (Barber et al., 2000) it had a kappa coefficient of 0.69 for neurologists, 0.39 for trainees and 0.47 for neuroradiologists, while in (Gupta et al., 2012) a dichotomized ASPECTS (>7) had a kappa of 0.53 when evaluated by two neuroradiologists.

In order to overcome this limitation and improve stroke care, many automated ASPECTS scoring methods have been developed, both in an academic and commercial context.

1.2 Automated ASPECTS

In a systematic review, (Mikhail et al., 2020) found a total of 20 studies centered around automated ASPECTS. Most of those studies (15) focused on testing previously developed commercial ASPECTS scoring software, such as e-ASPECTS (from Brainomix), Frontier ASPECTS and RAPID ASPECTS. According to the author, these studies presented "major inconsistency in the presentation of data" and some even came from a group with potential conflicts of interest.

In a similar, but smaller review, (de Freitas Brito et al., 2020) found 9 studies related to ASPECTS automation, 3 of those evaluating commercial software, and 6 novel academic ASPECTS scoring methods. Besides the same data presentation inconsistencies, this review found that region segmentation was very briefly discussed in these papers, and although it was usually done by some sort of template registration, there wasn't an evaluation of how well this technique performed or how ASPECTS region segmentation was done in general.

Given the lack of academic papers focusing on the development of ASPECTS automation software, and the lack of papers centring around the segmentation stage of automation, this paper aims to discuss, implement and evaluate an ASPECTS region segmentation method, using a public CT template, a public registration tool (SimpleITK with python) and a public CT database (CQ500).

2. Methods

2.1 Image registration

According to (Oliveira & Tavares, 2012), image registration can be defined as the process of aligning two or more images, that requires the selection of a feature space, a similarity measure, a transform type and a search strategy.

The main idea around the process is to search for a transformation that when applied to an image, called "moving image", maximizes the given similarity measure to another image, called "fixed image".

Image registration is a very important step in many medical imaging analysis algorithms, and one of its main applications is atlas-based segmentation, where a reference (atlas) is aligned to a target image, that is segmented by transferring it the expert-annotated labels. (Viergever et al., 2016)

In this study, ASPECTS regions were segmented by registering a manually labeled CT template (atlas) to different CTs from a public database.

2.1 CT Database

The public CT database CQ500 was used for testing and evaluating the proposed segmentation method. It has a total of 491 CT scans along with the manual evaluation of three specialists of findings such as hemorrhage, fracture, midline shift and mass effect (Chilamkurthy et al., 2018).

A total of 10 CT scans were randomly selected from the available healthy (no findings) CTs in the database. The selected CT numbers are as follows: 34, 54, 85, 122, 232, 321, 323, 350, 441 and 467.

After selecting the thinnest available scan for each of these CTs, it went through image contrast enhancement, or windowing, with a standard 0-100 Hounsfield Units (HUs) window. An experienced neurologist then chose two slices from each scan and manually labelled each of its 10 ASPECTS regions .

2.2 CT Template

Unlike MRI templates, it is not possible to take CTs of volunteers, due to the risk of radiation exposure and its ethical implications. As such, CT templates can only be done after retrospective searches of medical records, trying to identify scans without pathology or damage. According to (Muschelli, 2019) the first publicly available CT template was released by (Rorden et al., 2012), almost 20 years after the first MNI atlases (MRI) were released.

In this study, the template presented by (Rorden et al., 2012) was manually segmented by an experienced neurologist after being downsampled to keep half of its initial axial resolution. A total of 4 ganglionic and 6 supraganglionic slices were obtained from this step to be then used for registration.

2.3 Software

The Insight Segmentation and Registration Toolkit (ITK) is an open-source biomedical library, developed by the National Library of Medicine (NLM). Written in C++ it provides important tools and frameworks for biomedical image analysis, such as segmentation, registration, transforms, image filters, etc.

Although very powerful, ITK was sometimes too complex, lacking in usability. To solve these issues, SimpleITK was developed, providing wrapping to other languages such as R and Python and many other practical conveniences (Lowekamp et al., 2013).

SimpleITK version 1.2.4 with python 3.7.6 was used for almost everything done related to software development in this study: reading and writing DICOM files, windowing, linear and non-linear registration and segmentation evaluation. Minor code parts were also written with standard python libraries, such as numpy and matplotlib.

2.4 Proposed Method

The proposed segmentation method can be divided into four steps: pre-processing/slice selection, linear registration, non-linear registration and display/evaluation.

Initially, the chosen CT goes through contrast enhancement, or windowing, with a standard 0-100 HU window. After contrast enhancement, the CT slices are manually chosen

by a neurologist for registration, according to the ASPECTS slice standards: one ganglionic and one supraganglionic slice.

The chosen slices are then processed in an initial linear registration phase, where the available template slices are registered with a rigid transform into the selected slices. After being registered, the best template slices are chosen, based on its registration evaluation metric, and then go through two other linear registration stages: one with a scaling transform and one with an affine transform.

The linearly registered template cuts, as well as the chosen CT slices, are then segmented with a simple bone removal mask, and then are non-linearly registered with a demons based method.

With non-linear registration completed, the transforms and displacement fields obtained are applied to each of the template's regions, generating a label map, used to display each of the ASPECTS regions, as well as evaluate segmentation performance.

2.5 Segmentation performance metrics

Two segmentation metrics were used to evaluate the proposed methods' performance: Dice coefficient and Hausdorff distance.

Dice coefficient is an overlap based metric that is the most commonly used to evaluate segmentation performance. Given a number of true positives (TP), false positives (FP) and false negatives (FN), it is defined as in Eq. 1: (Taha & Hanbury, 2015)

$$DICE = \frac{2TP}{2TP+FP+FN} \quad (1)$$

Hausdorff distance, on the other hand, is a distance based evaluation metric. In an easy and intuitive way, it can be understood that if the Hausdorff distance between two sets of points, A and B, is d, then every point of A is within a maximum distance d of B and vice versa. (Min et al., 2007)

3. Results

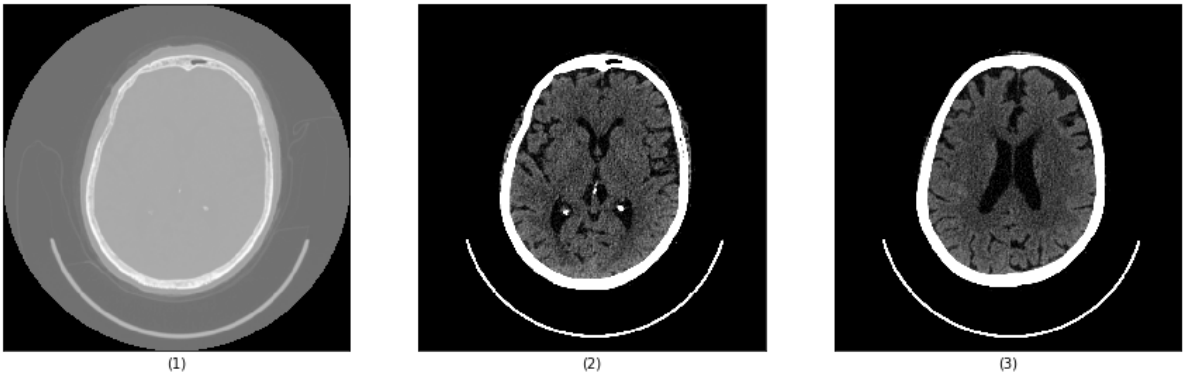
In this section results of each of the proposed methods' four steps, as well as overall results of its evaluation (Dice coefficient and Hausdorff distance) are shown.

3.1 Pre-processing

As previously discussed, a total of 10 CTs from the CQ500 database were randomly selected for our study. After windowing with a standard 0-100 HU window, a neurologist chose two slices from each one to be processed and segmented.

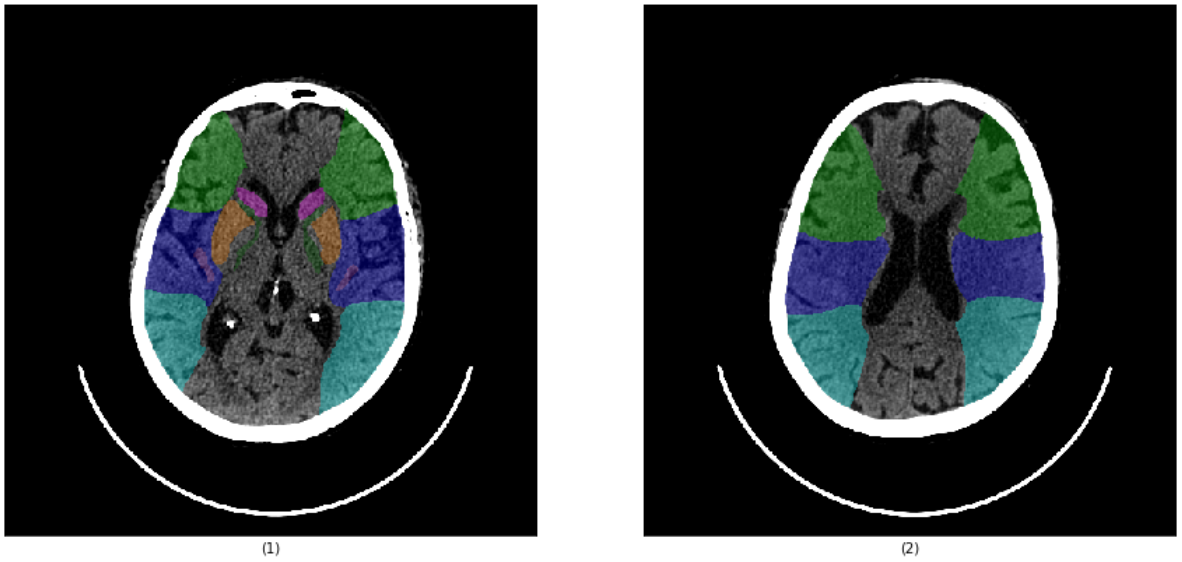
In Fig. 2 it is possible to see some of the preprocessing results for CT number 54: (1) shows the raw CT data, while (2) and (3) show the chosen CT slices after contrast enhancement.

Figure 2: Preprocessing and slice selection for CT number 54



Each of the chosen slices were also manually segmented by a neurologist, as to provide a reference, or ground truth, for the method's evaluation. Fig 3 shows the result of manual segmentation of the ganglionic (1) and supraganglionic (2) slices for CT number 54.

Figure 3: Manual segmentation of CT number 54



3.2 Linear Registration

Before the linear registration itself, a downsampled version of the (Rorden et al., 2012) CT template was manually segmented to be used in the registration process.

A total of 10 template slices were manually labeled: 4 ganglionic slices, as seen in Fig. 4 and 6 supra ganglionic slices, as seen in Fig 5. These manually segmented slices were used as atlases for the segmentation process.

Figure 4: Manually segmented ganglionic template slices

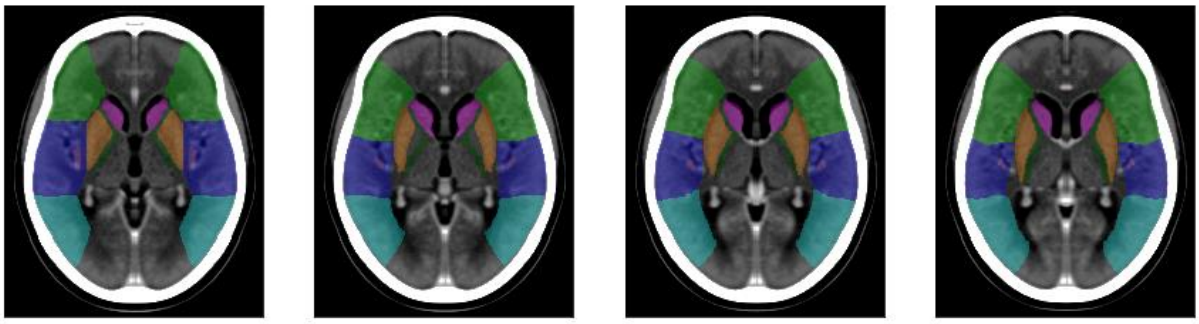
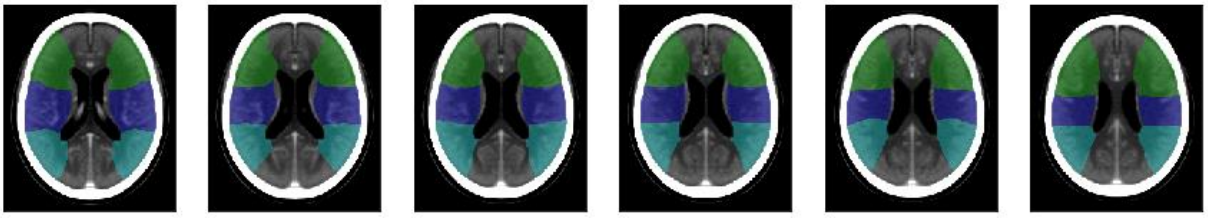


Figure 5: Manually segmented supraganglionic template slices



Each of these individual template ganglionic slices were linearly registered to the chosen ganglionic slice from the CT database. This initial registration was done with a simple translation/rotation transform (sitk.Euler2DTransform()), using Mattes Mutual Information as an evaluation metric, gradient descent as an optimization strategy and performed with the multi-resolution framework provided by SimpleITK.

After this initial registration, the best performing template slice (evaluated by Mattes Mutual Information) is chosen for two more linear registration steps: one performed with a anisotropic scale transform (sitk.ScaleTransform(2)) and finally a full translation/rotation/scaling transform (sitk.Similarity2DTransform()), done in an almost redundant way as to refine the two previous registrations. The entire process is then repeated for the supraganglionic slices.

Results for the ganglionic and supraganglionic slices of CT number 54 can be seen in Fig. 6 and Fig 7, respectively: (1) displays the chosen CQ500 slice, (2) shows the best template slice after translation/rotation, (3) shows that same slice after the anisotropic scaling registration, and (4) is the final output after the last registration.

Figure 6: Linear registration of ganglionic template slice to CT number 54

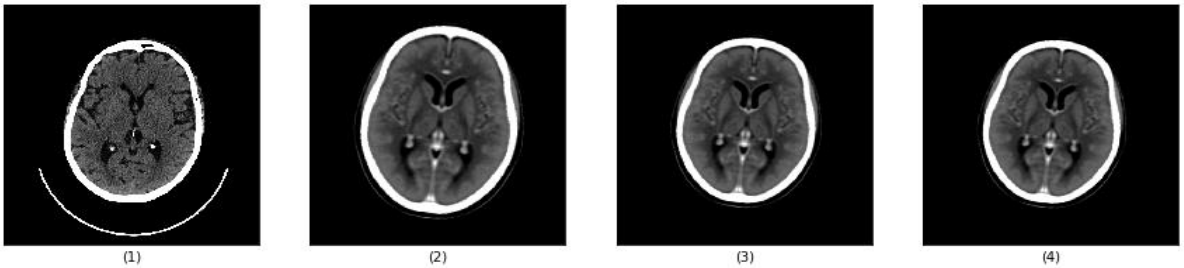
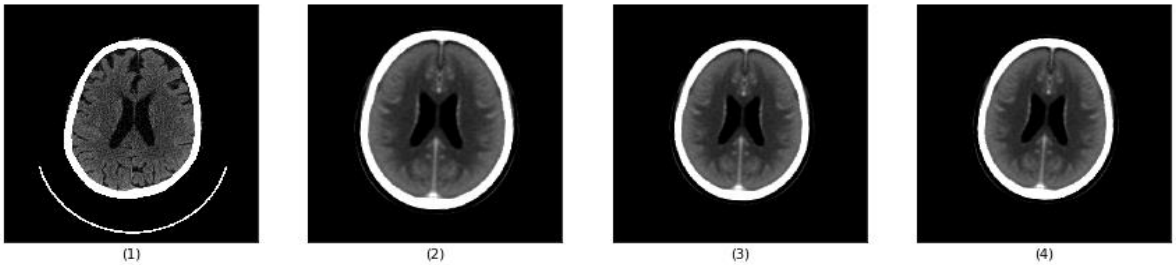


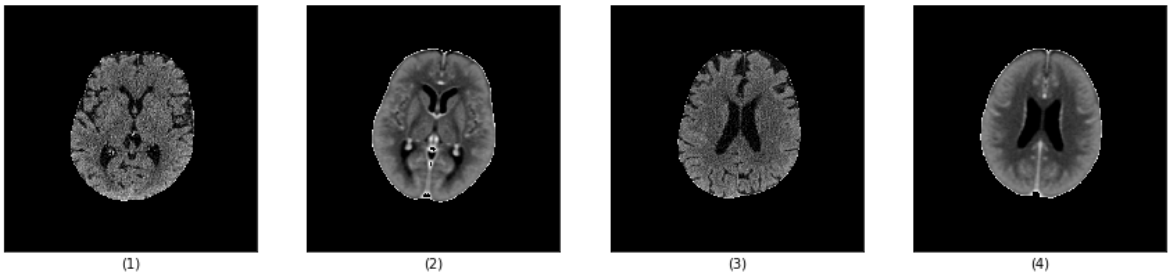
Figure 7: Linear registration of supraganglionic template slice to CT number 54



3.3 Nonlinear Registration

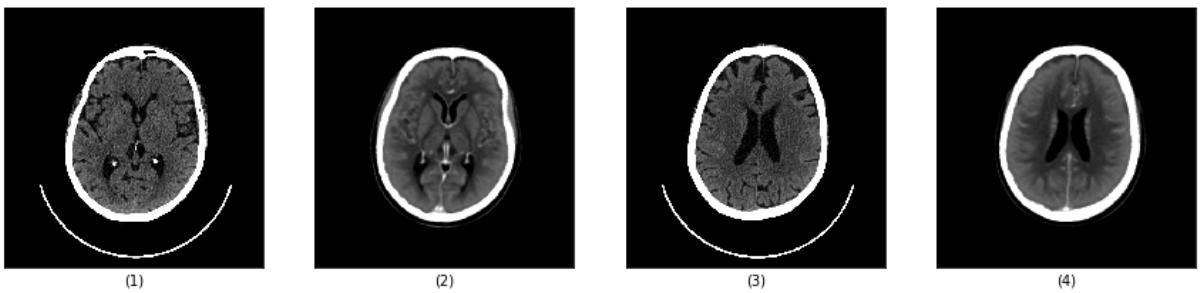
Before performing nonlinear registration, a simple bone removal mask is obtained for each of the four slices with thresholding and region growth. In Fig. 8 it is possible to see the effect of the bone removal mask in the ganglionic slice of CT number 54 (1), ganglionic slice of the template (2), supraganglionic slice of CT number 54 (3) and supraganglionic slice of the template (4).

Figure 8: Effect of the bone removal mask on CT slices



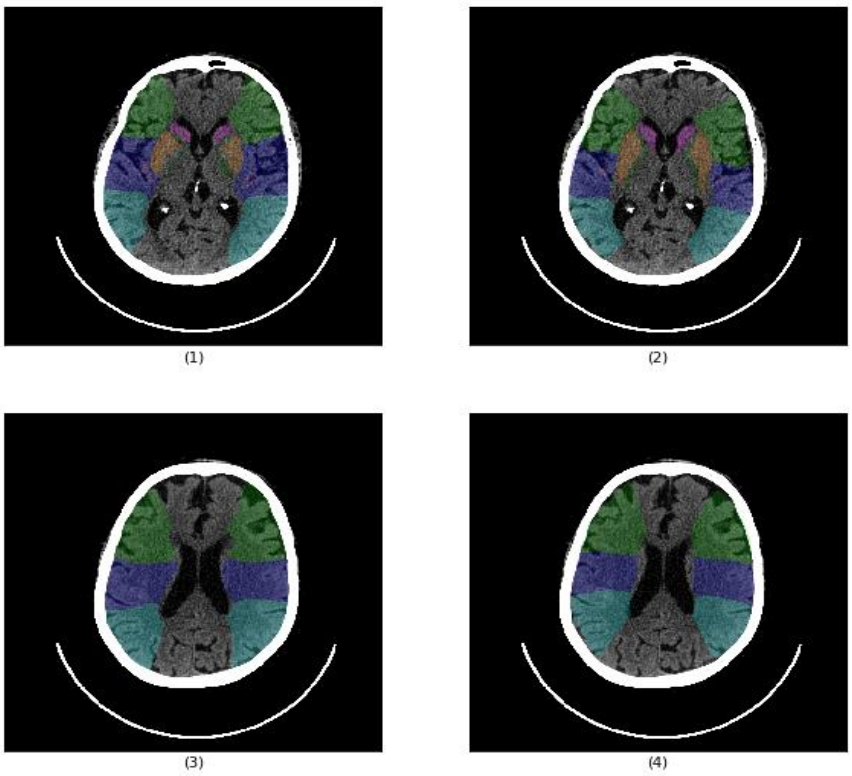
These masked slices then go through a nonlinear registration step, performed with the demons registration filter (sitk.FastSymmetricForcesDemonsRegistrationFilter()), with a gaussian smoothing standard deviation of 1.5, and a total of 1000 iterations. The results of the nonlinear registration and the registration step as a whole can be seen in Fig. 9 for CT number 54: in (1) and (3) it is shown the original ganglionic and supraganglionic slices and in (2) and (4) the ganglionic and supraganglionic slices of the registered template.

Figure 9: Final results of the registration stage



Finally, the transforms obtained by the linear registration and the displacement field obtained by the nonlinear registration step are applied to template slices' manually segmented regions, that are subsequently used as masks over the original (CQ500) CTs for ASPECTS region segmentation. Fig. 10 shows the final segmentation results: (1) and (3) are the manually segmented ASPECTS regions of the ganglionic and supraganglionic slices, respectively, while (2) and (4) are the obtained regions after the registration process, both for CT number 54.

Figure 10: Manually and automatically segmented regions



3.4 Segmentation evaluation

When the registration stage is completed, the obtained segmented regions are compared to their respective manually segmented regions for each of the 10 chosen CTs from CQ500. Segmentation was evaluated with two metrics: Dice coefficient and Hausdorff distance.

Overall, the segmentation process yielded a mean Dice coefficient of 0.6587±0.0595 and a mean Haussdorff distance of 14.3903±4.4366.

Tab. 1 presents the results obtained by ASPECTS region. As expected, when looking at mean Dice coefficient, large areas like the M1-M6 cortex regions perform better than smaller regions. The Internal Capsule and Insular Cortex in particular had very low overlap (Dice) scores, which was also somewhat expected, since these are very small and thin regions, therefore small displacements result in large overlap mismatches.

For these smaller regions, the Hausdorff distance is a better representing metric, and when evaluating it, there wasn't a big difference in performance between cortical (M1-M6) and deep (I, IC, L, C) regions. The first had a mean Hausdorff distance of 15.3706±4.5889, while the latter had a mean of 12.92±4.3760.

Table 1: Results obtained by region

Region	Dice	Hausdorff
Internal Capsule	0.3339	12.0721
Insular Cortex	0.2087	19.1166
Caudate	0.6899	8.8267
Putamen	0.5414	11.6646
M1	0.8055	14.2306
M2	0.7535	14.2562
M3	0.8178	24.2441
M4	0.8536	11.3128
M5	0.801	12.63
M6	0.7818	15.5497

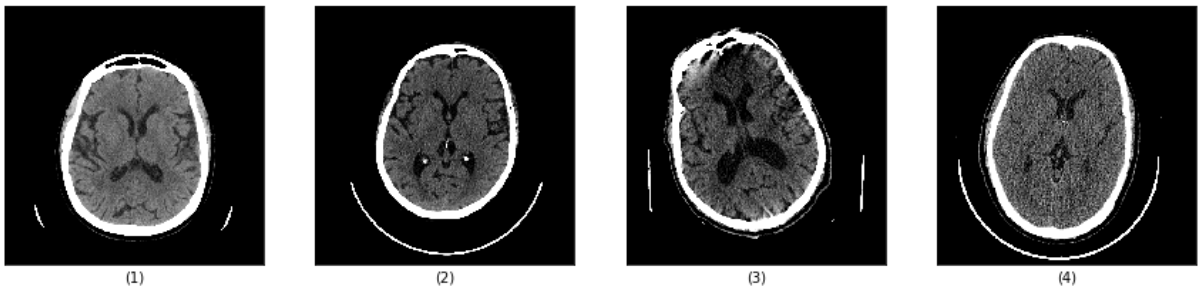
In Tab. 2 it is possible to see the results obtained by CT number. When inspecting these results, it is evident that CT quality and contrast affect how well segmentation is performed.

Table 2: Results obtained by CT number

CT number	Dice	Hausdorff
34	0.7206	18.2899
54	0.7453	12.6596
85	0.5921	18.4153
122	0.6785	13.1602
232	0.646	22.3117
321	0.6475	9.0859
323	0.7121	9.1719
350	0.5988	15.9847
441	0.5665	15.3953
467	0.6796	9.4287

Fig. 11 shows the difference between two CTs with good segmentation metrics, 34 in (1) and 54 in (2), and two CTs with worse metrics, 232 in (3) and 350 in (4). The very apparent difference in CT quality and contrast also suggest that more complex pre-processing techniques might improve segmentation performance.

Figure 11: CTs with different imaging quality



4. Conclusion

This paper proposed to evaluate an ASPECTS region segmentation through atlas registration, using a public CT database (CQ500), CT template and software (SimpleITK). The proposed method was implemented in python, and could be divided into four steps: pre-processing/slice selection, linear registration, non-linear registration and display/evaluation.

Overall, the segmentation method's evaluation resulted in a mean Dice coefficient of 0.6587±0.0595 and a mean Haussdorff distance of 14.3903±4.4366. There was a significant difference between overlap measures of cortical (M1-M6) regions and deep (I, IC, L, C) regions, but not between its Hausdorff distances, which was expected given the different nature of these regions, the first being larger, and the second being smaller and thinner, so that small displacements greatly affect overlap measures.

To the best of our knowledge this was the first paper evaluating ASPECTS region segmentation and although the proposed method performed reasonably well, it can certainly be improved, especially with more complex pre-processing techniques.

References

Barber, P. A., Demchuk, A. M., Zhang, J., and Buchan, A. M. (2000). Validity and reliability of a quantitative computed tomography score in predicting out- come of hyperacute stroke before thrombolytic therapy. *The Lancet*, vol. 355(9216), pp. 1670–1674.

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., Mahajan, V., Rao, P., and Warier, P. (2018). Development and validation of deep learning algorithms for detection of critical findings in head ct scans. ArXiv [Online]. Available: https://arxiv.org/abs/1803.05854

de Freitas Brito, R., Veiga, A. C. P., and Filho, J. B. D. (2020). Alberta stroke program early ct score automático: uma breve revisão. *Proceedings of XVIII Conferência de Estudos em Engenharia Elétrica*. Uberlândia, Brazil.

Feigin, V. L., Lawes, C. M., Bennett, D. A., Barker-Collo, S. L., and Parag, V. (2009). Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *The Lancet Neurology*, vol. 8(4), pp. 355–369.

Gupta, A., Schaefer, P., Chaudhry, Z., Leslie-Mazwi, T., Chandra, R., González, R., Hirsch, J., and Yoo, A. (2012). Interobserver reliability of baseline noncontrast CT alberta stroke program early CT score for intra-arterial stroke treatment selection. *American Journal of Neuroradiology*, vol. 33(6), pp.1046–1049.

Hill, M. D. (2005). Thrombolysis for acute ischemic stroke: results of the canadian alteplase for stroke effectiveness study. *Canadian Medical Association Journal*, vol. 172(10), pp. 1307–1312.

Johnson, W., Onuma, O., Owolabi, M. and Sachdev, S. (2016). Stroke: a global response is needed. Bulletin of the World Health Organization, vol. 94(9), pp. 634–634A.

Mikhail, P., Le, M. G. D., and Mair, G. (2020). Computational image analysis of nonenhanced computed tomography for acute ischaemic stroke: A systematic review. *Journal of Stroke and Cerebrovascular Diseases*, vol. 29(5), pp. 104715.

Min, D., Zhilin, L., and Xiaoyong, C. (2007). Extended hausdorff distance for spatial objects in GIS. *International Journal of Geographical Information Science*, vol. 21(4), pp. 459–475.

Muschelli, J. (2019). Recommendations for processing head CT data. *Frontiers in Neuroinformatics*, vol. 13.

Oliveira, F. P. and Tavares, J. M. R. (2012). Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17(2), pp. 73–93.

Puetz, V., Dzialowski, I., Hill, M. D., and Demchuk, A. M. (2009). The alberta stroke program early CT score in clinical practice: What have we learned? *International Journal of Stroke*, vol. 4(5), pp. 354–364.

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., and Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization. *NeuroImage*, vol. 61(4) pp. 957–965

Taha, A. A. and Hanbury, A. (2015). Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, vol. 15(1).

Viergever, M. A., Maintz, J. A., Klein, S., Murphy, K., Staring, M., and Pluim, J. P. (2016). A survey of medical image registration – under review. *Medical Image Analysis*, vol. 33, pp. 140–144.

Wardlaw, J. M., Murray, V., Berge, E., del Zoppo, G., Sandercock, P., Lindley, R. L., and Cohen, G. (2012). Recombinant tissue plasminogen activator for acute ischaemic stroke: an updated systematic review and meta-analysis. *The Lancet*, vol. 379(9834), pp. 2364–2372.