Neurosense: deep sensing of full or near-full coverage head/brain scans in human magnetic resonance imaging

Baris Kanber[1,2,3], James Ruffle[4], Jorge Cardoso[5], Sebastien Ourselin[5], Olga Ciccarelli[3,6]
[1] Centre for Medical Image Computing, University College London, London, UK.
[2] Department of Medical Physics and Biomedical Engineering, University College London, London, UK
[3] National Institute for Health Research University College London Hospitals Biomedical Research Centre, London, UK
[4] Department of Radiology, University College London Hospitals NHS Foundation Trust, London UK
[5] School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK
[6] Department of Neuroinflammation, Queen Square Institute of Neurology, University College London, London, UK

Corresponding Author: Dr Baris Kanber, UCL Centre for Medical Image Computing, 90 High Holborn, Holborn, London WC1V 6LJ, UK. E-mail: b.kanber@ucl.ac.uk. Telephone: +44 (0)20 7679 2000.

Keywords: Neurosense; MRI; classification; sensing; detection; head; brain.

Introduction

The application of automated algorithms to imaging requires knowledge of its content, a curatorial task, for which we ordinarily rely on the *Digital Imaging and Communications in Medicine (DICOM)* header as the only source of image meta-data. However, identifying brain MRI scans that have full or near-full coverage among a large number (e.g. >5000) of scans comprising both head/brain and other body parts is a time-consuming task that cannot be automated with the use of the information stored in the DICOM header attributes alone. Depending on the clinical scenario, an entire set of scans acquired in a single visit may often be labelled "BRAIN" in the DICOM field *0018,0015 (Body Part Examined)*, while the individual scans will often not only include brain scans with full coverage, but also others with partial brain coverage, scans of the spinal cord, and in some cases other body parts. DICOM field *0018,1250 (Receive Coil Name)* can be used to determine the type of receiver coil used, however, the use of a *head coil* does not guarantee a full-brain MRI scan. At other times another type of receiver coil such as a *'dual coil'* may have been used, or the *0018,1250 (Receive Coil Name)* attribute may be blank. Similarly, determining the *field-of-view* from the DICOM attributes is not sufficiently informative to detect full or near-full coverage brain scans, even if a brain scan is guaranteed. Furthermore, some datasets may have been converted to other formats such as the *Neuroimaging Informatics Technology Initiative (NIFTI)* format, with no access to the source DICOM files, and the associated tags. In other cases, DICOM files may be available but with no useful information in their headers, for example, if the DICOM files have been produced by the conversion of old plain-films, or as in legacy datasets. The imaging catalogue of the average hospital is too large to be manually checked, let alone labelled: a comprehensive framework for rendering imaging robustly auditable requires a fully automated process.

In this article, we introduce Neurosense, a trained, convolutional, deep neural network that efficiently performs this task. The network was trained and validated on 31556 image volumes (primarily T1-weighted MRI) collated from: The Cancer Imaging Archive (*TCIA*)[1], a public imaging database of low grade gliomas[2],

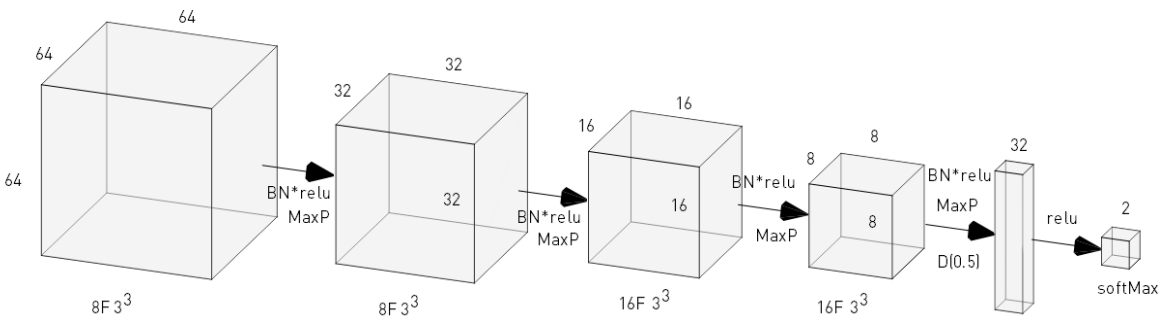
[1] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. Journal of Digital Imaging, 26(6), 1045-1057.
[2] Erickson, B., Akkus, Z., Sedlar, J., Korfiatis, P. (2017). Data From LGG-1p19qDeletion. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2017.dwehtz9v

breast cancer[3], prostate cancer[4], renal clear cell carcinomas[5], and melanomas[6]; the *IXI* Brain MRI dataset[7], a public imaging database of healthy controls; the Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer's Disease dataset[8] (*OASIS-3*); NCIGT-Harvard-Brigham Neurosurgical Intraoperative Image Database (*NeuroMRT*); and a heterogeneous set of MRI volumes from University College London Hospitals (UCLH), obtained at different field strengths, using different scanners, and a variety of protocols. These volumes have been manually verified as being full or near-full coverage brain scans (46%) or brain scans with partial coverage or scans of other body parts (54%).

Methods

Neurosense comprises a convolutional neural network of the following architecture: 4x(3D Convolution -> Batch Normalisation -> Activation -> Max Pooling) -> Flatten -> Dropout -> 2x(Dense -> Activation) (Figure 1). The layers within brackets are repeated 4, and 2 times, respectively, as indicated. The first two 3D convolutions have 8 filters of size 3^3 voxels, while the last two 3D convolutions have 16 filters of the same size. All activations of the convolutional layers are *rectified linear units* (relu), and all max-pooling layers are of size 2^3. The dropout layer has a drop-out probability of 0.5. Finally, the dense layers have a simple structure where the number of neurons are 32, and 2. The first dense layer is followed by relu activation, while the output layer has a *normalized exponential function* (softMax) activation.

Figure 1 – Network architecture. BN=batch normalization, relu=rectified linear unit, MaxP=max pooling, D=dropout, softMax= *normalized exponential function*.



MRI volumes are put into a standard orientation (e.g. x-axis: right to left, y-axis: posterior to anterior, z-axis: inferior to superior) using FSL's *fslreorient2std*[9], scaled down to a size of 64^3 voxels, and normalized to the range -127 to +127 before being input into the network. The latter normalization technique is adopted instead of the more conventional method of normalization to a zero mean and unit standard deviation, so that data can be held in memory using 8-bit integer, instead of 32-bit floating-point matrices. At 8-bit resolution, 31556 volumes of 64^3 voxels occupy 7.7 GB of computer memory, compared with 30.8 GB at 32-bit resolution. This is a 75% reduction in the amount of computer memory used for each image volume, and is useful for avoiding streaming from disk, and speeding up network training. However, GPU processing was at 32-bit floating point resolution.

[3] David, N., Nola, H. (2016). Single site breast DCE-MRI data and segmentations from patients undergoing neoadjuvant chemotherapy. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.QHsyhJKy

[4] Choyke, P., Turkbey, B., Pinto, P., Merino, M., Wood, B. (2016). Data From PROSTATE-MRI. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.6046GUDv

[5] Akin, O., Elnajjar, P., Heller, M., Jarosz, R., Erickson, B. J., Kirk, S., et al. (2016). Radiology Data from The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma [TCGA-KIRC] collection. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.V6PBVTDR

[6] Anti-PD-1 Melanoma Dataset. The Cancer Imaging Archive.

[7] IXI dataset, Imperial College London. https://brain-development.org/ixi-dataset/

[8] The Open Access Series of Imaging Studies (OASIS). http://www.oasis-brains.org/

[9] Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M. (2012). FSL. NeuroImage, 62, 782-90, 2012.

The data used for training and validation was augmented by randomly cropping volumes down to half size, either from the top, the bottom, or the middle of the volume, along either one of the three primary axes randomly, and by randomly cropping down to individual slices in the primary acquisition direction. Cropped volumes were subsequently scaled back up to 64^3 voxels without the use of padding.

We used a *Stochastic Gradient Descent* optimiser with a time-based learning schedule, and a *Nesterov momentum* of 0.9. The learning rate decayed from an initial value of 0.001. After 8 epochs, a step reduction in learning rate of 1/10 was introduced. *Cross entropy* was adopted as the cost function, and the batch size was 8.

The network was trained on a randomly selected 70% subset of the dataset (22089 scans) and validated on the remaining 30% (9467 scans). The process was bootstrapped five times, with different values for the random number generator.

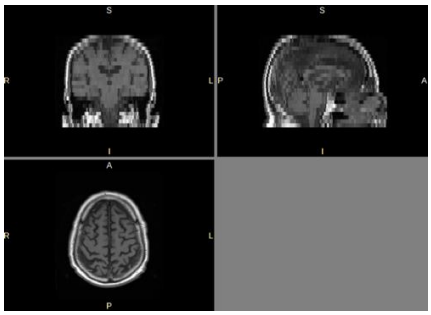
Testing was performed, on three separate test sets, collated from the same sources, having no overlap amongst, or with respect to the main training and validation dataset. These three test sets, test sets A, B, and C, comprised 3592, 1000, and 494 scans, respectively. In test set A, the proportion of full and near-full coverage brain MRI scans was 53%, while that of brain scans with partial coverage and scans of other body parts was 47%. Test set B comprised 50% full and near-full coverage brain MRI scans, and 50% brain scans with partial coverage and scans of other body parts. Test set C comprised 64% full and near-full coverage brain MRI scans, and 36% brain scans with partial coverage and scans of other body parts. Additionally, test set C comprised not only T1-weighted MR images, but also images of other MRI contrasts (11.1% FLAIR, 27.7% T2), as well as images that were not MRI (5.3% CT).

Results

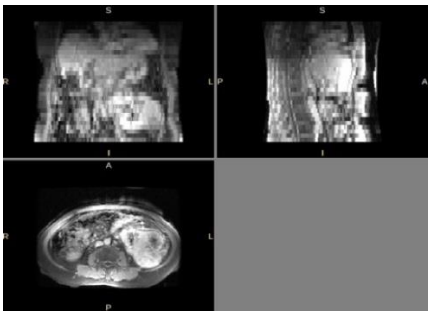
Network training took approximately 7.8 hours per bootstrap on an NVIDIA GeForce GTX 960 GPU with 4GB of memory (NVIDIA, Santa Clara, CA, USA). After 9 epochs, the network achieved an accuracy of 99.991% (95% confidence interval: 99.986-99.995%) on the training dataset, with a mean residual cross-entropy loss of 10^{-4}. On the validation dataset, the accuracy was 99.973% (95% confidence interval: 99.951-99.994%), with a similarly small mean cross-entropy loss of 10^{-3}.

Accuracy on test set A was 99.733% (95% confidence interval: 99.694-99.772). There were an average of ten (out of 3592) misclassifications per model (resulting from each of the five bootstrap runs), all of which were related to limited coverage brain scans, which were labelled as not having sufficient brain coverage but were classified as having enough coverage by the network, and vice versa. Examples of MRI volumes classified as full or near-full coverage brain scans and limited coverage brain scans or scans of other body parts are shown in Figure 2.

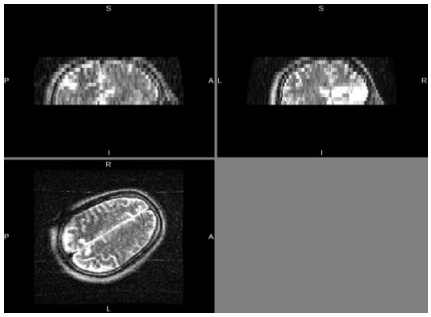
Figure 2 – Examples of MRI volumes classified as brain scans with full or near-full coverage (accepted), and those not so classified (rejected). Volumes are shown in their original orientation.



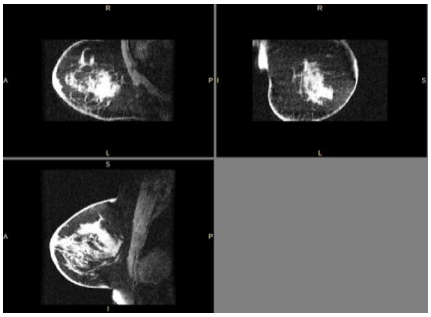
Accepted (good coverage, *TCIA*)



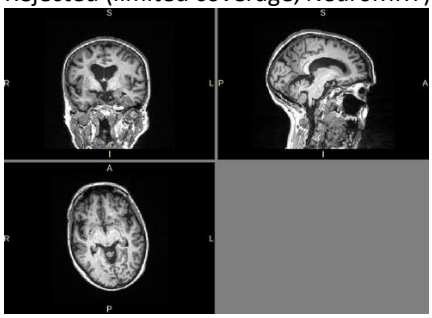
Rejected (non-brain, *TCIA*)



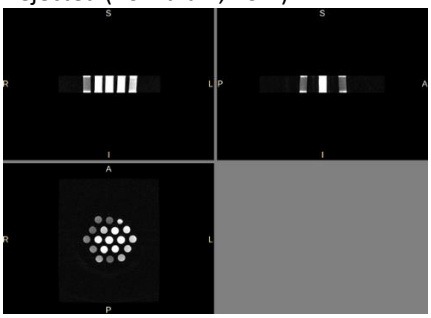
Rejected (limited coverage, *NeuroMRT*)



Rejected (non-brain, *TCIA*)



Accepted (good coverage, *OASIS-3*)



Rejected (non-brain, *TCIA*)

Accuracy, sensitivity, and specificity on test set B, using the model from the first bootstrap run, were 99.5%, 99.8%, and 99.4%, respectively. The corresponding precision, and recall values were 99.4%, and 99.8%. The optimal threshold value was 0.01.

Accuracy, sensitivity, and specificity on test set C, also using the model from the first bootstrap run, were 95.3%, 95.5%, and 96.1%, respectively. The corresponding precision, and recall values were 97.7%, and 95.5%. The optimal threshold value was 0.01, as for test set B. We note that this dataset comprised not only T1-weighted MR image volumes (as the network was primarily trained on), but also scans of other MRI contrasts (e.g. T2 and FLAIR), as well as image volumes that were not MRI (e.g. CT).

The average speed of testing was 963 seconds per 1000 volumes in a virtual machine which had access to 4 Intel Core i7-4870HQ CPU cores running at 2.50GHz and 4096 MB of RAM. Only 88 seconds were spent on the actual network inference with the remaining 875 seconds spent on data loading and preparation (e.g. reorientation to a standard orientation).

Summary

We present Neurosense, a trained convolutional neural network that efficiently identifies full and near-full coverage brain MRI scans in large MRI datasets. The network displays value in classifying MRI volumes acquired using a diverse range of acquisition parameters. We believe Neurosense could be an important tool for the curation of large-scale neuroimaging data. It is made available for use as a virtual machine from *CMIClab*.

Compliance with Ethical Standards

This study was approved by the University College London Hospitals NHS Trust. The study was classified as a service evaluation and optimization project using irrevocably anonymized data, which does not require ethical approval or consent.

Acknowledgements

We are grateful to Professor Frederik Barkhof, Professor Daniel Alexander, Dr Parashkev Nachev, Dr Robert Gray, Dr Ferran Prados, and Dr Eugenio Iglesias for their help and support of this study, and for providing comments on previous versions of the article. The data used were provided in part by OASIS and were part based upon data generated by the TCGA Research Network. This work was funded by the National Institute for Health Research University College London Hospitals Biomedical Research Centre, and the Wellcome Trust.

Information Sharing Statement

Neurosense is available for use as a virtual machine from *CMIClab*. The datasets used during development are publicly available except for the UCLH dataset which is not a public database.

Conflicts of Interest

The authors declare that they have no conflicts of interest.