NeuroImage: Clinical 21 (2019) 101605

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Automated detection of imaging features of disproportionately enlarged subarachnoid space hydrocephalus using machine learning methods

Nathaniel B. Gunter[a,b], Christopher G. Schwarz[a], Jonathan Graff-Radford[c], Jeffrey L. Gunter[a,*], David T. Jones[c], Neill R. Graff-Radford[d], Ronald C. Petersen[c], David S. Knopman[c], Clifford R. Jack Jr.[a]

[a] Department of Radiology, Mayo Clinic and Foundation, Rochester, MN, USA
[b] University of Oklahoma, Norman, OK, USA
[c] Department of Neurology, Mayo Clinic and Foundation, Rochester, MN, USA
[d] Department of Neurology, Mayo Clinic and Foundation, Jacksonville, FL, USA

ARTICLE INFO

Keywords:
Normal pressure ydrocephalus
Disproportionately enlarged subarachnoid hydrocephalus
Support vector machines
Computer-aided diagnosis
Tight high-convexity

ABSTRACT

Objective: Create an automated classifier for imaging characteristics of disproportionately enlarged subarachnoid space hydrocephalus (DESH), a neuroimaging phenotype of idiopathic normal pressure hydrocephalus (iNPH).

Methods: 1597 patients from the Mayo Clinic Study of Aging (MCSA) were reviewed for imaging characteristics of DESH. One core feature of DESH, the presence of tightened sulci in the high-convexities (THC), was used as a surrogate for the presence of DESH as the expert clinician-defined criterion on which the classifier was trained. Anatomical MRI scans were automatically segmented for cerebrospinal fluid (CSF) and overlaid with an atlas of 123 named sulcal regions. The volume of CSF in each sulcal region was summed and normalized to total intracranial volume. Area under the receiver operating characteristic curve (AUROC) values were computed for each region individually, and these values determined feature selection for the machine learning model. Due to class imbalance in the data (72 selected scans out of 1597 total scans) adaptive synthetic sampling (a technique which generates synthetic examples based on the original data points) was used to balance the data. A support vector machine model was then trained on the regions selected.

Results: Using the automated classification model, we were able to classify scans for tightened sulci in the high convexities, as defined by the expert clinician, with an AUROC of about 0.99 (false negative ≈ 2%, false positive ≈ 5%). Ventricular volumes were among the classifier's most discriminative features but are not specific for DESH. The inclusion of regions outside the ventricles allowed specificity from atrophic neurodegenerative diseases that are also accompanied by ventricular enlargement.

Conclusion: Automated detection of tight high convexity, a key imaging feature of DESH, is possible by using support vector machine models with selected sulcal CSF volumes as features.

1. Introduction

Idiopathic normal pressure hydrocephalus (iNPH) is a treatable disorder, with clinical symptoms that present as the triad of gait disturbance, urinary incontinence, and dementia (Hakim and Adams, 1965). Standard treatment for iNPH is shunt placement to drain cerebrospinal fluid (CSF) (Mori et al., 2012). Prevalence of iNPH is estimated clinically at between 1.1% and 2.9% among the elderly (Hiraoka et al., 2008; Jaraj et al., 2014; Mori et al., 2012). Disproportionately enlarged subarachnoid space hydrocephalus (DESH) is a neuroimaging variant of iNPH defined by enlarged Sylvian fissures, ventriculomegaly, and, most specifically, constricted CSF spaces in the high convexities of the brain (Hashimoto et al., 2010). The Japanese Society of Normal Pressure Hydrocephalus guidelines for management of iNPH now include radiologic signs of DESH as a supporting factor, making automatic detection potentially impactful in treatment of patients (Mori et al., 2012). The frequency of DESH has not been systemically studied outside of iNPH. In the current study, the presence of a tight high-convexity (THC) was visually assessed on all participants by an experienced neurologist, blinded to any clinical information about the

* Corresponding author.
E-mail address: gunter.jeffrey@mayo.edu (J.L. Gunter).

https://doi.org/10.1016/j.nicl.2018.11.015
Received 4 May 2018; Received in revised form 11 September 2018; Accepted 18 November 2018
Available online 19 November 2018

Table 1
Population characteristics for the dataset used in this study. Summaries are given as mean (std deviation) [min, max] for continuous variables and count (percent) for categorical variables. Total subjects: 1597.

Characteristic	Summary
Male	870 (51%)
Age	78.5 (5.2) [70, 94]
Education (y)	13.9 (2.9) [6, 20]
APOE e4 carrier	438 (27%)
Cognitively Unimpaired	1353 (85%)

participants. Manual grading for THC was selected as a criterion standard indicator of DESH, because it is a radiologic marker that does not overlap with any common neurodegenerative disorder, unlike ventriculomegaly. We developed an automatic classifier aimed at reproducing the manual THC assessments based on the volume of CSF in individual sulci. Such a classifier may serve as a rapid, fully automated, high-throughput pre-screening test for DESH. Because CSF expansion in DESH can be confused with tissue atrophy on in-vivo imaging, DESH may be an unrecognized confounding factor in imaging studies of neurodegenerative disorders. Automated screening for DESH features may assist in correctly classifying individuals with iNPH.

2. Materials and methods

Participants were identified from the Mayo Clinic Study of Aging (MCSA) (Petersen et al., 2010; Roberts et al., 2008). Demographics for the sample population are shown in Table 1 (Graff-Radford et al., 2016). The Mayo Clinic Study of Aging is an epidemiologically recruited sample of participants from Olmsted County, Minnesota (Roberts et al., 2008).

All images were acquired on GE 3 T scanners using an 8-channel receiver array. T1-weighted MRI scans were performed using a sagittal 3D magnetization prepared rapid acquisition gradient recalled echo (MP-RAGE) sequence with no acceleration. Repetition time (TR) was ≈ 2300 ms, echo time (TE) ≈ 3 ms, and inversion time (TI) = 900 ms. The MP-RAGE was used for atlas-based anatomic parcellation. Voxel dimensions were ≈1.20 × 1.015 × 1.015 mm. FLAIR imaging was performed using 3 mm thick axial slices, TR/TI/TE of 11,000/2250/150 ms with 1Nex, a 256 × 192 acquisition matrix, and three-fold interleave.

FLAIR images from the dataset were visually assessed for imaging factors of iNPH by an experienced neurologist (NGR). All participants were graded by the same neurologist, eliminating inter-rater variability. Expert evaluation of high tight sulci was made by reviewing the FLAIR images of each patient. The images were reviewed in three orientations (coronal, axial and sagittal). For subjects with tight high sulci MRI scans showed the posterior hemispheres were tightly opposed. Sample images are shown in Fig. 1. Tight high convexity was typically found to co-exist with ventriculomegaly and enlarged Sylvian fissures.

A four-dimensional probabilistic atlas of named sulcal regions was obtained from the BrainVISA software (Mazziotta et al., 2001), then warped to the Mayo Clinic Adult Lifespan Template space (Schwarz et al., 2018) using the ANTS package (Avants et al., 2008). At this point the probabilistic maps of each sulcus were smoothed using an empirically chosen 2 mm isotropic kernel to ensure complete coverage of the intracranial vault.

The atlas was warped to each subject's individual anatomic T1-weighted MRI, again using the ANTS package. We then collapsed the atlas into three dimensions by assigning each voxel to the region with

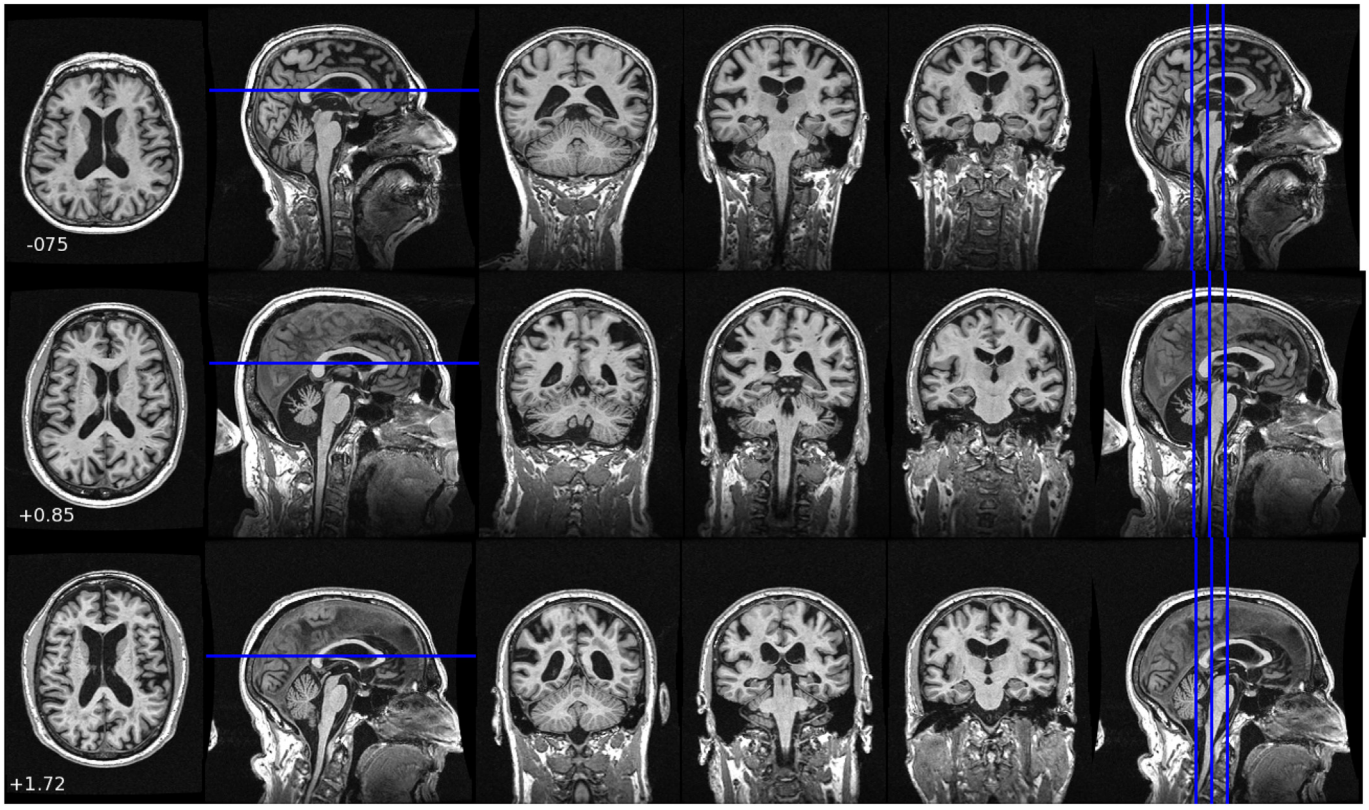

Fig. 1. Sample MP-RAGE images from subjects visually identified as having tight high convexity. CSF spaces near the mid-sagittal line at the high convexity of the brain appear compressed. Each row presents images from a different subject. An axial slice chosen to be near the maximum A/P extent of the corpus callosum is shown. Three coronal images spaced 5 mm apart in regions where tight high convexity is observed are also shown. The numeric values are the machine learning system output score as described in the text.

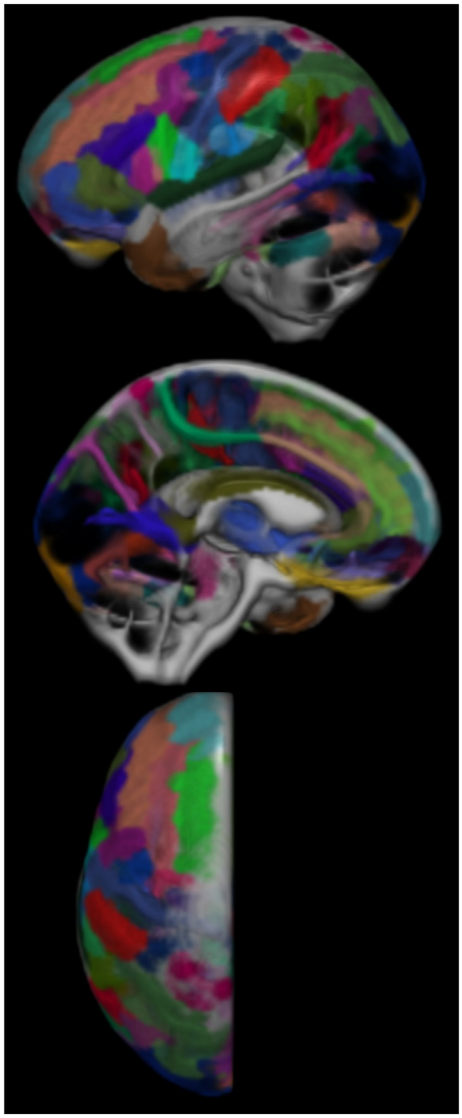

Fig. 2. The BrainVISA atlas laid on the left side of the Mayo Clinic Adult Lifespan Template brain's CSF segmentation. Regions in the atlas are paired left-right, so only one half of the brain is shown. Not pictured is the segmentation of anterior ventricle horn volumes.

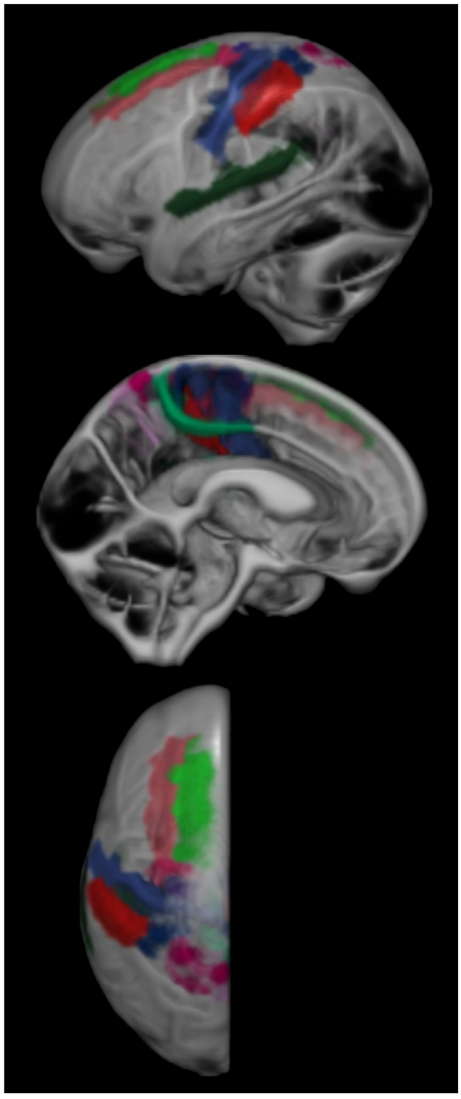

Fig. 3. Regions with individual AUROC values > 0.7 shown on the CSF segmentation of the Mayo Clinic Adult Lifespan Template brain. These regions are used to construct the composite model. Not pictured is our segmentation of anterior ventricle horn volumes.

the highest probability. This ensured that each voxel in the atlas only belonged to one region and allowed us to straightforwardly extract sulcal regions of interest (ROIs) in the subject-space T1-weighted image coordinate system. For reference, a collapsed version the atlas of 123 named sulcal regions is shown in Fig. 2. The T1-weighted images were segmented using in-house population-specific customizations to SPM12 Unified Segmentation (Ashburner and Friston, 2005; Schwarz et al., 2016). The overlap between the probabilistic CSF segmentation and each atlas region was summed to give a measure of CSF volume per sulcal region. Each sulcal volume was then normalized by the subject's total intracranial volume (TIV) to correct for head size (Barnes et al., 2010).

A receiver operating characteristic curve (ROC) analysis was performed using the pROC package (Robin et al., 2011) in R (R Core Team, 2013) to determine the ability of each individual sulcal volume to predict the expert neurologist's classification of tightened high convexity in the training data set. Regions with an individual area under receiver operating characteristic curve (AUROC) of > 0.7 were selected for inclusion in the composite model. We also included the right posterior subcentral ramus of the lateral fissure (AUROC = 0.604) to preserve left-right symmetry in the selected regions.

Due to the small number of participants identified as having THC by the expert reviewer, there was a problem of class imbalance between selected (n = 73) and unselected (n = 1522) classes. This was an issue because machine learning models with small minority classes will occasionally learn to simply classify everything as the majority class and accept the small error. Adaptive synthetic sampling (ADASYN) (He et al., 2008) was applied to balance the classes by generating synthetic positive samples of the minority class. ADASYN is an extension of synthetic majority oversampling technique (SMOTE) (Chawla et al., 2002) that weights instances of minority classes by the number of similar majority-class instances. It then generates new data points using the original instances as a base, generating more points the more majority classes are nearby. This shifts the decision boundary toward examples that are more difficult to classify, improving the model by ensuring that it can classify the difficult examples correctly.

Several different models were trained on the ADASYN balanced data using the MATLAB classification learner toolbox (MathWorks, 2015). We investigated discriminant analyses, logistic regression, support vector machines, and random forest models. Many models performed well, with model output AUROC's in excess of 0.95. Of the models

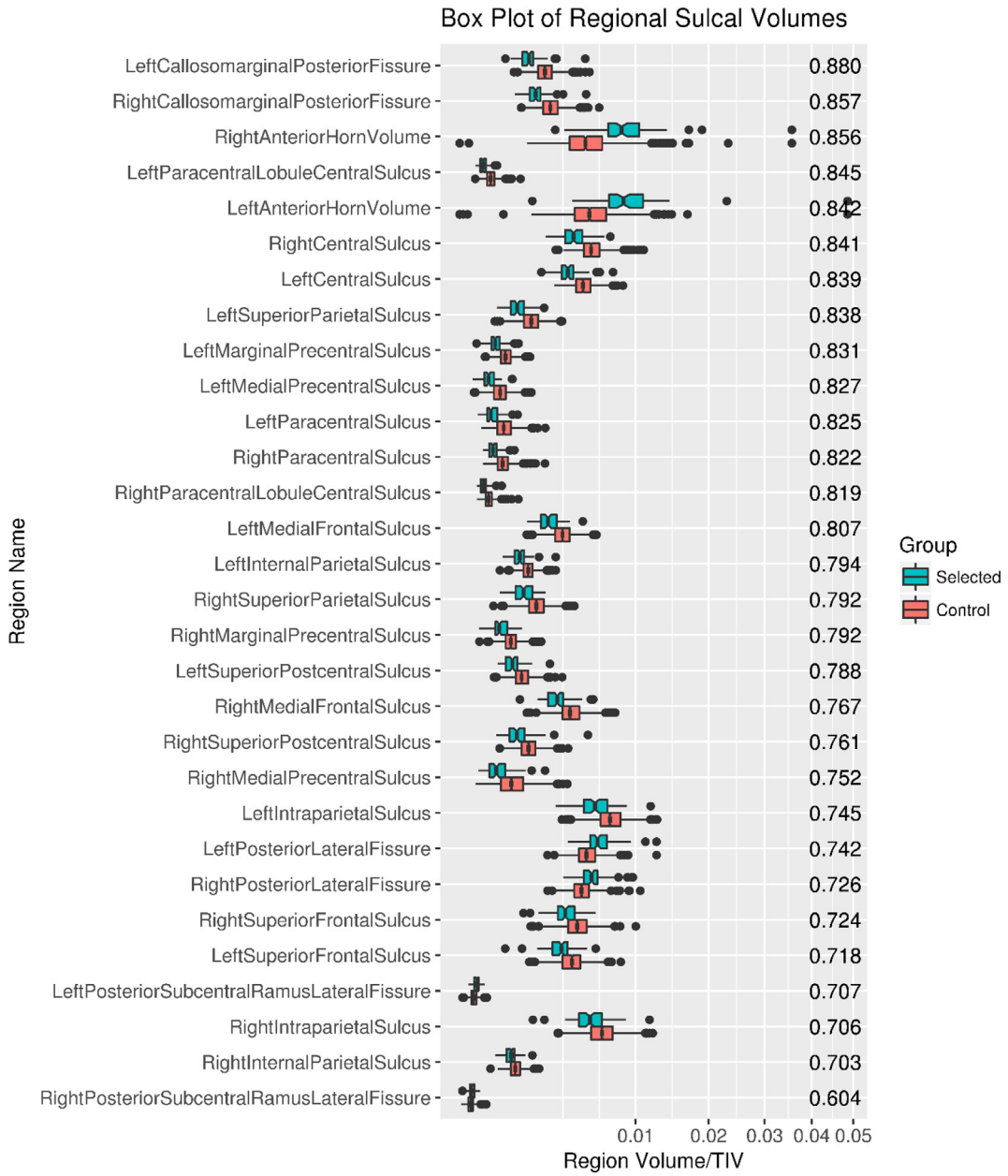

Fig. 4. Box plot of individual sulcal region volumes. AUROC values by region are displayed along the right side.

tested, a support vector machine (SVM) (Noble, 2006) with a quadratic kernel was empirically chosen for a balance of accuracy and consistency, as well as ease of conceptual understanding. After the model was chosen, its validity on the balanced data was tested by training it on a dataset consisting of half of the original negative class and an equal number of ADASYN generated instances of the positive class. The original positive class data were not used in training. The model was subsequently tested on the other half of the negative class and the original positive class. By doing so, the model was tested on data not present in the training data.

3. Results

The number of participants selected for THC was low compared to the size of the study ($n_{selected} = 72$, $n_{total} = 1595$, frequency $\approx 4.5\%$) (Table 1). Regions with individual AUROCS > 0.7 (selected for inclusion in the model), are shown in Fig. 3. Upon visual inspection, we considered the selected regions consistent with typical radiologic regions of interest for DESH. In Fig. 4, we show boxplots of CSF volume by sulcal region, ordered from top to bottom by individual region AUROC. The regions selected (those with AUROC > 0.7) included both expanded and constricted regions. The most discriminant single regions were the left and right posterior callosomarginal fissures where relative constriction was found in participants identified as having THC. The posterior portions of the callosomarginal fissures are superior to the callosal gyri and inferior to the superior portions of the sensory-motor cortex. The quadratic SVM model composed of the selected regions had an AUROC of 0.99, much higher than the best individual region's AUROC of 0.88 (Fig. 5).

4. Discussion

Several biologically plausible regions showed reasonable capabilities to differentiate participants with and without tight high convexity as determined by a blinded expert rater. For example, the posterior parts of the callosomarginal sulci sit between the callosal gyrus

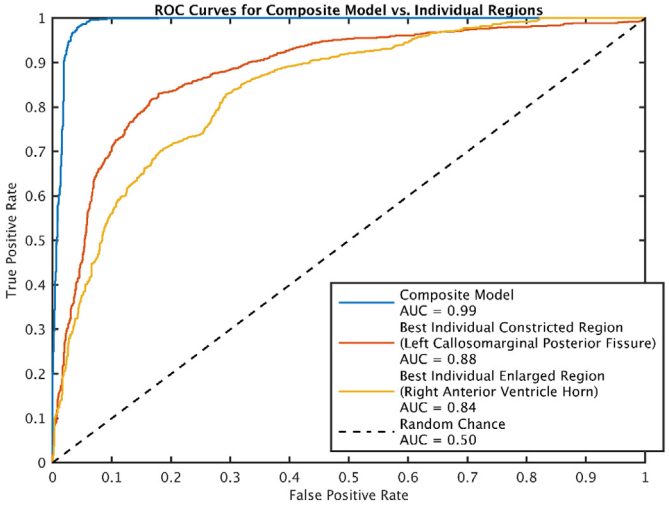

Fig. 5. ROC curves for the model and best individual regions that were relatively expanded or contracted. The composite model perfoms much better than either of the individual regions.

and sensory-motor areas. While several individual regions showed some effectiveness as predictors of tightened high convexity, a composite model using both constricted and enlarged regions shows a much better AUROC (best individual = 0.88, model composite = 0.99). When the model was run, it produced several false positives (predicted positives = 101, false positives = 29, frequency ≈ 29%). The false positive participants were visually reviewed. One particpant image had a poor-quality segmentation result. The remaining 28 subject images were found to have tight sulci along the mid-sagittal line and moderately enlarged Sylvian fissures but only modest ventricular enlargement.

Identifying DESH has treatment implications (Narita et al., 2016). The development of automated, non-invasive screening criteria may be helpful. A less obvious impact is in the study of neurodegenerative disorders. Ventriculomegaly is commonly encountered in dementia studies as a sensitive but non-specific measurement. We show that it is possible to create an automated classification for THC. By including regions outside of enlarged areas such as the ventricles, we increase the selectivity of our model without decreasing the sensitivity. Such an automated classification could potentially assist in the exclusion of participants with possible DESH from studies of AD dementia and other neurodegenerative disorders.

This study's primary strengths lie primarily in its large dataset and completely automated approach. Additionally, consistency between participants is a strength, as all participants were graded by the same reviewer and scanned using consistent imaging protocols. While others (Narita et al., 2016) have shown THC as indicative of shunt efficacy, a limitation of this study is its purely retrospective nature. We have restricted the scope to focus on image-based measurements. Epidemiological interpretation and age-dependent prevalence is a work in progress along with studies of THC and clinically observed factors such as gait. Our use of a single rater and imaging protocol also presents some limitations. We have no estimate of interrater reliability, as there was only one rater, and the robustness of this method against variations in image acquisition protocols is unknown.

Acknowledgements

This work was supported by the grants from the U.S. National Institutes of Health: AG006786, AG016574, AG034676, AG011378, AG04185, NS097495 as well the Gerald and Henrietta Rauenhorst Foundation, the Elsie and Marvin Dekelboum Family Foundation, the Alexander Family Alzheimer's Disease Research Professorship of the Mayo Clinic, the Liston Award, the Schuler Foundation, and the Mayo Foundation for Medical Education and Research.

References

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41. https://doi.org/10.1016/j.media.2007.06.004.

Barnes, J., Ridgway, G.R., Bartlett, J., Henley, S.M.D., Lehmann, M., Hobbs, N., Clarkson, M.J., MacManus, D.G., Ourselin, S., Fox, N.C., 2010. Head size, age and gender adjustment in MRI studies: a necessary nuisance? NeuroImage 53, 1244–1255. https://doi.org/10.1016/j.neuroimage.2010.06.025.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Graff-Radford, N., Jones, D., Gunter, J., Thomas, C., Crook, J., Boeve, B., Knopman, D., Jack, C., Petersen, R., 2016. The prevalence and prognosis of patients with enlarged ventricles (P4.034). Neurology 86.

Hakim, S., Adams, R.D., 1965. The special clinical problem of symptomatic hydrocephalus with normal cerebrospinal fluid pressure. Observations on cerebrospinal fluid hydrodynamics. J. Neurol. Sci. 2, 307–327. https://doi.org/10.1016/0022-510X(65)90016-X.

Hashimoto, M., Ishikawa, M., Mori, E., Kuwana, N., 2010. Diagnosis of idiopathic normal pressure hydrocephalus is supported by MRI-based scheme: a prospective cohort study. Cerebrospinal Fluid Res. 7, 18. https://doi.org/10.1186/1743-8454-7-18.

He, H., Bai, Y., Garcia, E.A., Li, S., 2008. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: Proceedings of the International Joint Conference on Neural Networks, pp. 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

Hiraoka, K., Meguro, K., Mori, E., 2008. Prevalence of idiopathic normal-pressure hydrocephalus in the elderly population of a Japanese rural community. Neurol. Med. Chir. (Tokyo) 48, 197–200. https://doi.org/10.2176/nmc.48.197.

Jaraj, D., Rabiei, K., Marlow, T., Jensen, C., Skoog, I., Wikkelsø, C., 2014. Prevalence of idiopathic normal-pressure hydrocephalus. Neurology 82, 1449–1454. https://doi.org/10.1212/WNL.0000000000000342.

MathWorks, 2015. Statistics and Machine Learning Toolbox Release Notes. MatLab.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Feidler, J., Smith, K., Boomsma, D., Pol, H.H., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A four-dimensional probabilistic atlas of the human brain. J. Am. Med. Inform. Assoc. 8, 401–430. https://doi.org/10.1136/jamia.2001.0080401.

Mori, E., Ishikawa, M., Kato, T., Kazui, H., Miyake, H., Miyajima, M., Nakajima, M., Hashimoto, M., Kuriyama, N., Tokuda, T., Ishii, K., Kaijima, M., Hirata, Y., Saito, M., Arai, H., Japanese Society of Normal Pressure, H, 2012. Guidelines for management of idiopathic normal pressure hydrocephalus: second edition. Neurol. Med. Chir. (Tokyo) 52, 775–809. https://doi.org/10.2176/nmc.52.775.

Narita, W., Nishio, Y., Baba, T., Iizuka, O., Ishihara, T., Matsuda, M., Iwasaki, M., Tominaga, T., Mori, E., 2016. High-convexity tightness predicts the shunt response in idiopathic normal pressure hydrocephalus. Am. J. Neuroradiol. 37, 1831–1837. https://doi.org/10.3174/ajnr.A4838.

Noble, W.S., 2006. What is a support vector machine? Nat. Biotechnol. 24, 1565–1567. https://doi.org/10.1038/nbt1206-1565.

Petersen, R.C., Roberts, R.O., Knopman, D.S., Geda, Y.E., Cha, R.H., Pankratz, V.S., Boeve, B.F., Tangalos, E.G., Ivnik, R.J., Rocca, W.A., 2010. Prevalence of mild cognitive impairment is higher in men: the Mayo clinic study of aging. Neurology 75, 889–897. https://doi.org/10.1212/WNL.0b013e3181f11d85.

Core Team, R., 2013. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing. Austria, Vienna. http://www.R-project.org.

Roberts, R.O., Geda, Y.E., Knopman, D.S., Cha, R.H., Pankratz, V.S., Boeve, B.F., Ivnik, R.J., Tangalos, E.G., Petersen, R.C., Rocca, W.A., 2008. The Mayo clinic study of aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology 30, 58–69. https://doi.org/10.1159/000115751.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12. https://doi.org/10.1186/1471-2105-12-77.

Schwarz, C.G., Gunter, J.L., Wiste, H.J., Przybelski, S.A., Weigand, S.D., Ward, C.P., Senjem, M.L., Vemuri, P., Murray, M.E., Dickson, D.W., Parisi, J.E., Kantarci, K., Weiner, M.W., Petersen, R.C., Jack, C.R., 2016. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. NeuroImage Clin. 11, 802–812. https://doi.org/10.1016/j.nicl.2016.05.017.

Schwarz, C.G., Gunter, J.L., Ward, C.P., Vemuri, P., Senjem, M.L., Wiste, H.J., Petersen, R.C., Knopman, D.S., Jack Jr., C.R., 2018. The Mayo clinic adult lifespan template: better quantification across the lifespan. Alzheimer's Dement. J. Alzheimer Assoc. 13, P792. https://doi.org/10.1016/j.jalz.2017.06.1071.