

Fully Automated Volumetric Classification in CT Scans for Diagnosis and Analysis of Normal Pressure Hydrocephalus

Angela Zhang^{1*}, Po-Yu Kao¹, Ashutosh Shelat³, Ronald Sahyouni², Jefferson Chen², and B.S. Manjunath^{1*}
¹University of California, Santa Barbara, ²University of California, Irvine, ³Santa Barbara Cottage Hospital

Abstract – Normal Pressure Hydrocephalus (NPH) is one of the few reversible forms of dementia. Due to their low cost and versatility, Computed Tomography (CT) scans have long been used as an aid to help diagnose intracerebral anomalies such as NPH. However, because CT imaging presents 2-dimensional slices of a 3-dimensional volume, recapitulating the ventricular space in 3-dimensions to facilitate the diagnosis of NPH poses numerous challenges such as head rotation and human error. As such, no well-defined and effective protocol currently exists for the analysis of CT scan-based ventricular, white matter and subarachnoid space volumes in the setting of NPH. The Evan’s ratio, an approximation of the ratio of ventricle to brain volume using only one 2D slice of the scan, has been proposed but is not robust. Instead of manually measuring a 2-dimensional proxy for the ratio of ventricle volume to brain volume, this study proposes an automated method of calculating the brain volumes for better recognition of NPH from a radiological standpoint. The method first aligns the subject CT volume to a common space through an affine transformation, then uses a random forest classifier to mask relevant tissue types. A 3D morphological segmentation method is used to partition the brain volume, which in turn is used to train machine learning methods to classify the subjects into non-NPH vs. NPH based on volumetric information.

Keywords: Normal Pressure Hydrocephalus, segmentation, classification, machine learning, morphological contours

I. INTRODUCTION

NPH presents as ventriculomegaly accompanied with symptoms of dementia, specifically cognitive dysfunction, changes in gait, urinary incontinence (Shprecher, et al. 2008). It is estimated that more than 700,000 Americans have normal pressure hydrocephalus (NPH), however, due to the nonspecific and indolent nature of NPH, the majority of cases are under- or misdiagnosed (Jaraj et al. 2014). Without appropriate diagnostic testing, NPH is often misdiagnosed as Alzheimer’s disease or Parkinson’s disease, or the symptoms are attributed to the aging process. NPH is one of few reversible causes of dementia in the elderly, making correct diagnosis important, as shunt placement has been demonstrated to be a safe and effective treatment for NPH (Shprecher, et al. 2008).

Current diagnostic methods for NPH involve a mixture of clinical and imaging approaches (Shprecher, et al. 2008). Although MRI volumetric data may, in ideal cases, provide better detail of the borders of the ventricles than standard CT imaging it is important to remember that many cases are far from ideal. Patient motion, dental or other hardware artifact and different technique often results in poor definition of certain structures often including the ventricular system (Lin and Alessio 2009). Additionally, distinctions of the skull to define the peripheral border of the subarachnoid space is much easier achieved using CT images. Shorter acquisition time, use for follow-up, and the sheer number of data points also lead to the superiority of CT imaging in understanding and classifying NPH on a broad scale.

In this context, this paper presents a fully automated volumetric screening of NPH from CT scans obtained in clinical settings. The proposed algorithm overcomes these difficulties, using a combination of statistical and mathematical pre-processing methods, *a-priori* knowledge, and morphological operations, to accurately classify the ventricles, white matter, and subarachnoid space in CT scans of the head acquired in clinical settings. A random forest classifier is then trained to predict the diagnosis of each patient given the corresponding volumetric information of their CT scan. This method could be used to alert medical health professionals of possible cases of NPH, thereby facilitating diagnosis and expediting treatment.

The current methods available for analyzing brain scans for possible NPH, such as finding the Evan’s ratio, are time-intensive, manual, and prone to error (Toma et al. 2011). Evan’s index is the ratio of the transverse diameter of the anterior horns of the lateral ventricles to the greatest internal diameter of the skull in a single slice of a 3D volume. This is illustrated in Figure 1.

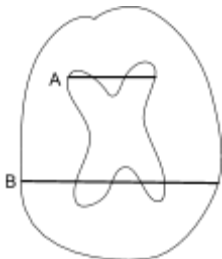


Fig. 1: Illustration of Evan’s ratio method. The ratio takes the length of the widest part of the frontal horns (line A) over the length of the widest part of the inner skull (line B). The location of the slice in the z dimension is the location where the parietal lobe appears to be widest.

* corresponding authors at the Vision Research Laboratory, emails angela00@ucsb.edu (Angela Zhang), manj@ucsb.edu (B.S. Manjunath)

Current guidelines state that an Evan's index of greater than 0.3 indicates NPH. However, recent findings have shown that the Evan's ratio in fact varies greatly depending on the level (slice location) of the brain CT scan image at which the frontal horns and maximal inner skull diameters are measured (Toma et al. 2011). The Evan's index was shown to have a correlation of only 0.619 with the ventricle volume and 0.498 with the ventricle volume divided by the total intracranial volume (Toma et al. 2011).

A 3-dimensional, volumetric method of measuring the relevant regions of the brain could help to mitigate these challenges. In (Moore et al. 2012) the authors suggest that volumetric measurements hold promise for improving NPH differential diagnosis. While there are some semi-automated methods of segmenting regions of the brain, these methods are still labor intensive and require human expertise and intervention at each step.

While there are methods to obtain ventricle and white matter volumes in MRI, these methods cannot find the subarachnoid space, as it does not show up in MRI. The method detailed in (Gunter et al. 2018) automatically detects features of disproportionately enlarged subarachnoid space hydrocephalus in MRI, without differentiating between ventricles and subarachnoid space. (Coupé et al. 2011) uses expert priors to aid in patch based segmentation of the lateral ventricles in MRI. In (Takahashi et al. 2017; Yepes-Calderon, Nelson, and McComb 2018), automated ventricular volume measurement in MRI is implemented through the PACS system. Another method of lateral ventricle segmentation in MRI is presented in (Kobashi et al., n.d.). The authors in (Shao et al. 2018) explore challenges in ventricle segmentation using neural networks in MRI. An automated method of computing the Evan's ratio from CT is presented in (Takahashi et al. 2017; Yepes-Calderon et al. 2018), but this method loses the volumetric advantage of directly computing the volumes from CT scans.

This paper proposes a novel method to automatically classify in 3 dimensions the lateral ventricles, white matter, and subarachnoid space from CT scans and use these volumes to predict possible NPH. The classification algorithm achieved a Dice score of 85.31 ± 6.16 % for the lateral ventricles and 91.03 ± 2.38 % for white matter. The NPH prediction method achieved a sensitivity of 84 ± 6 % and a specificity of 92 ± 8 %. The proposed method has improved performance compared to predicting NPH using Evan's Ratio. This can be used as an aid in screening for potential NPH in patients who might otherwise be misdiagnosed.

III. MATERIALS and METHODS

Data

The de-identified subject data comes from two sources: the University of California Irvine Medical Center and the Santa Barbara Cottage Hospital. This is a retroactive study, with all images de-identified as specified by the IRB agreement between each medical center and the University of California, Santa Barbara.

There was no protocol determining the number of slices, orientation, or other imaging parameters for the data used in this study. 177 CT scans of 79 patients from the University of California, Irvine and Santa Barbara Cottage Hospital were included in the study, with 49 scans (43 patients) having a diagnosis of normal, 80 scans (18 patients) having a diagnosis of NPH, 18 scans (11 patients) having a diagnosis of dementia (non-NPH), and 30 scans (7 patients) having a diagnosis of non-NPH hydrocephalus. Scans were acquired as part of the treatment process, and the number of slices varies from 25 to 207. The greater the number of slices, the higher the resolution in the z-dimension.

30 manual segmentations of 9 subjects were performed by members of the research team under direct supervision and validation by a neurological surgeon. The Evan's ratio, as measured by a neurological surgeon, was provided for 54 subjects, with 11 scans (5 patients) having a diagnosis of normal, 34 scans (5 patients) having a diagnosis of NPH, and 9 scans (2 patients) having a diagnosis of dementia.

Of the subjects with known age ($N = 23$), the average age of the normal subjects is 71.6 ± 12.4 years and the average age of the NPH subjects is 71.3 ± 18.8 years. The subjects without known age ($N = 56$) are all between the ages of 60-80 years. For patients with multiple scans, the scans were taken over a period of time as follow-ups to the initial scan. All NPH subjects underwent treatment with a ventricular shunt.

Algorithm Overview

There are several major steps involved in the proposed method for NPH prediction. The first step is preprocessing, followed by automatic classification of the lateral ventricles, white matter, and subarachnoid space. These volumes are then used to train machine learning models to predict possible NPH. Implementation details are provided in Supplemental Materials.

The code to the algorithm is available at https://github.com/UCSB-VRL/NPH_Prediction.

I. Voxel-wise Classification

The following diagram provides an overview of the automated ventricle, white matter, and subarachnoid space classification process.

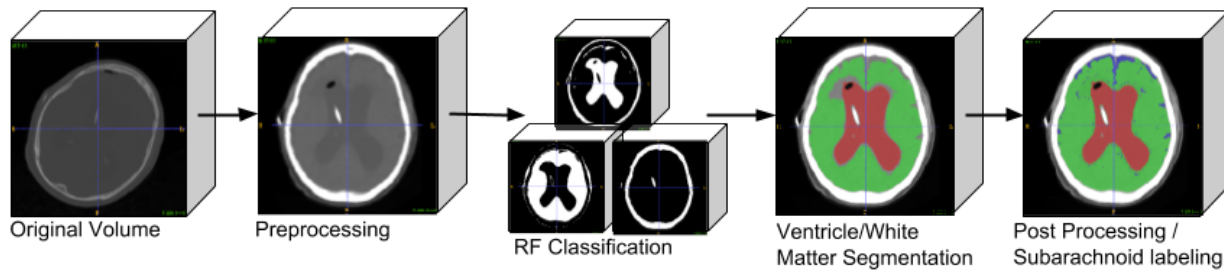


Fig. 2: Workflow of automatic classification algorithm. From left to right, 1) the original volumetric image. 2) An affine transformation is computed to transform the original volume to a template, correcting for skew and rotation. 3) Random forest classification is used to classify the types of tissue in the volume. 4) Segmentation of the ventricle and white matter. 5) Post processing and segmentation of the subarachnoid space.

a) Preprocessing

First, the skull of each scan is extracted using thresholding, since the intensity value of the skull (~ 1000 x-ray interactions) is much higher than any other tissue (>100 x-ray interactions) in the CT scan. The skull volume is used to compute an affine transformation with the standard MRI in MNI152 space (Jenkinson et al. 2012). The MNI152 space is a 3-dimensional coordinate system for stereotactic localization in neurosurgery made from the average MRI of 152 scans from the Montreal Neurological Institute of McGill University Health Centre (Grabner et al. 2006). The computed affine transform is then applied to its corresponding CT scan.

b) Tissue Classification

The affine transformed scans are used to train a random forest classifier (Pedregosa et al. 2011; Breiman 2001) on a small subset of labeled data ($\sim 10,000$ voxels) to recognize the intensity values of 3 different tissue types - cerebrospinal fluid (CSF), white matter, and skull. CSF serves as a proxy of the ventricular and subarachnoid spaces within the brain. The classifier then aggregates the votes from the different decision trees to decide the final class of the voxel. There are 4 classes total, the last of which is the background class. The trained classifier is then used to classify each voxel of every scan. Since the intensity values of CT scans are generally constant given tissue density and type, no further intensity normalization of the scans is done. The trained random forest classifier is used to select relevant regions of the volume and mask the other regions based on tissue type.

c) Ventricle / White Matter Segmentation

Each masked volume of the CSF is then seeded at the center of the anatomical ventricular region based on anatomical prior knowledge of the average ventricle location in MNI152 space. The seeds are then grown using the 3-dimensional Morphological Chan-Vese (MCV) algorithm (Chan and Vese 2001; van der Walt et al. 2014). This algorithm is a level set evolution algorithm with the goal of minimizing the energy function defined by:

$$F = \mu_1 * \text{Surface Area}(C) + \nu * \text{Volume}(\text{inside}(C)) + \lambda_1 \int_{\text{inside}(C)} |u_0(x,y,z) - c_1|^2 dx dy dz + \lambda_2 \int_{\text{outside}(C)} |u_0(x,y,z) - c_2|^2 dx dy dz,$$

where $u_0(x,y,z)$ is the volumetric image, $\lambda_1, \lambda_2 > 0$ are parameters that can adjust the comparisons between average intensity inside and outside of 3-dimensional surface contour C (c_1 and c_2 , respectively), and $\mu_1 > 0$ is a regularizing parameter to promote evolution. A 2D illustration of the algorithm is shown in Fig. 3.

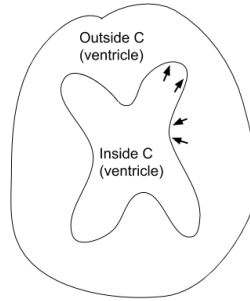


Fig. 3: 2-dimensional illustration of the inside and outside of the contour for lateral ventricles, with arrows showing example evolutions in one iteration.

Likewise, each masked volume of the white matter is seeded at three points: on the top, back, and front of the head next to the skull. The boundaries are then found using the MCV algorithm. An example is shown in Fig. 4.

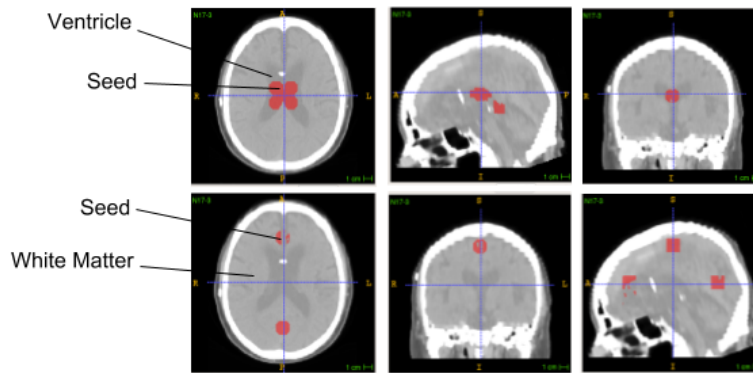


Fig. 4: Example seed points for the lateral ventricle (top row) and white matter (bottom row).

Because the ventricular space is separate from the subarachnoid space in terms of fluid flow, this method allows for the separation of the two, even though their tissue classes are the same from the random forest classifier.

d) Post Processing / Subarachnoid Space Labeling

Following segmentation of the lateral ventricles and white matter, the remaining voxels (classified as ‘fluid’) are labeled as subarachnoid space. The volumes of each class are computed by multiplying the total number of labeled voxels with the 3D voxel spacing information in the metadata of the scan.

In cases of a previous or current stroke, there is a prevalence of darker tissue surrounding the ventricle. This tissue would be classified as ventricle in the above algorithm, but should actually be classified as white matter. Hence, the segmentation is modified in cases where the ventricles are larger than 90 mL. First, the original volume is denoised with the non-local means method (Buades et al. 2011) to better separate the two intensity distributions, eliminating most outliers within a region. Regions marked 'ventricle' are re-marked as ‘white matter’ if their intensity exceeds a threshold T as determined by the lower peak in histogram of intensities for white matter.

The histograms in Fig. 5 show the difference in intensity frequencies in the white matter region in a subject with previous stroke and a normal subject. From the histogram, a threshold intensity of 10 was chosen to relabel stroke affected regions of the brain from ventricle to white matter after denoising.

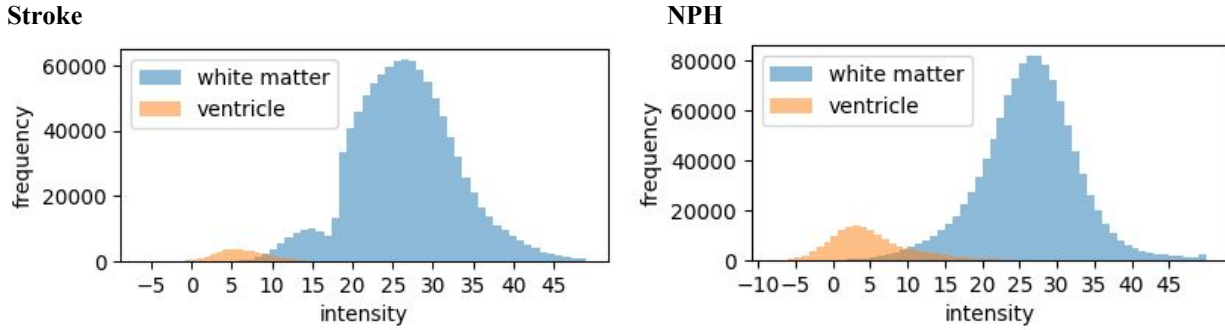


Fig. 5: Histogram of intensities in white matter and lateral ventricle regions for patients with stroke and NPH. There is a bump in intensities from the range 8-20 in the white matter of stroke subjects, which overlaps with their ventricle intensities. This translates to an optimal threshold of 10 to differentiate between ventricular and white matter space.

II. NPH Prediction

Support Vector Machines (SVM) (Cortes and Vapnik 1995) with linear and radial basis function (RBF) kernels are trained and tested in this paper to determine the extent to which the brain volumes are linearly separable features for NPH prediction and obtain an optimal prediction scheme. A Random Forest classifier is also trained and tested on the same features.

In order to determine whether using all of the volumetric information results in better performance compared to using volumetric information analogous to obtaining the Evan’s Ratio, only the ventricle and whole brain volumes are first used and compared to models using the ventricle, white matter, subarachnoid space, and total brain volumes for prediction. The prediction procedure is illustrated in Fig. 6.



Fig. 6: Procedure for NPH prediction. The first row depicts the training stage and the second row depicts the prediction stage.

Brain size differs significantly based on gender (Lüders et al. 2002). In order to account for this variability, the total brain volume is calculated by adding the ventricle, white matter, and subarachnoid space volumes. This calculation underestimates the total brain volume due to the cutoff occurring at the brain stem. The ventricle, white matter, subarachnoid space, and total brain volumes are used as input features to train a linear SVM and rbf SVM, as well as a random forest classifier. Stratified k-fold cross-validation was used to create training and validation datasets. For each method, 100 classifiers are trained to obtain an average and standard deviation for classifier performance.

IV. RESULTS

Classification Validation and Comparison

To return the segmentation to the original patient space for purposes of visualisation and segmentation verification, the inverse affine transform from the MNI152 space to the patient space was computed and applied to the automatic segmentation using nearest neighbor interpolation. Each automatic segmentation was then compared to its corresponding manual segmentation in the original

patient space. The segmentations performed by the algorithm are compared with 30 manual segmentations of 9 subjects performed by members of the research team under the direct supervision and validation by a neurological surgeon.

Some basic machine learning methods are implemented to compare with the proposed method. Some of these methods are used as one step in our method. They include random forest classification, 3D morphological geodesic active contours, and 3D morphological Chan-Vese.

The implementation of the alternative methods of ventricle segmentation use thresholding to find the skull region and remove any labels outside of this region. All implementations first compute and apply the affine transformations into MNI152 space, then computes and applies the inverse transformation after completing segmentation. For the morphological chan-veve and morphological geodesic active contour methods, the volumes are seeded in the same manner as the proposed algorithm. The regions are then grown according to their perspective algorithms. Finally, the regions inside the skull not labeled as white matter are then labeled as ventricular space.

For the scores in Table 1, the *Dice Score*, $2|X \cap Y| / (|X| + |Y|) = 2TP / (2TP + FP + FN)$, where X and Y are two classes (positive and negative for each class), is used.

| Method | Ventricle (Dice) | White Matter (Dice) |
|---|-----------------------|-----------------------|
| Proposed Method | 85.31 ± 6.16 % | 91.03 ± 2.38 % |
| Random Forest | 57.34 ± 16.52 % | 88.16 ± 3.19 % |
| 3D Morphological Geodesic Active Contours | 20.51 ± 19.87 % | 84.68 ± 9.13 % |
| 3D Morphological Chan Vese | 15.05 ± 18.36 % | 85.71 ± 3.65 % |

Table 1: Comparison of Dice Scores for various ventricle and white matter segmentation algorithms for CT scans. The scores are reported as mean ± standard deviation.

Example segmentation results for normal and NPH subjects are included in Fig. 7.

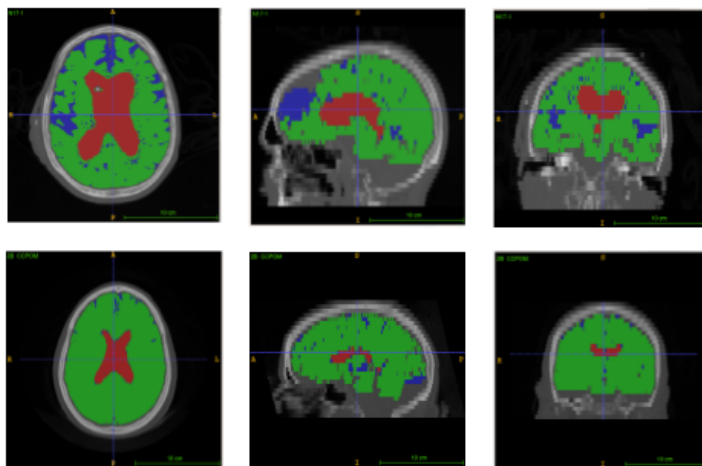


Fig. 7: Examples of segmentations generated by our algorithm. The first (top) set of images consists of cross sections of a subject diagnosed with NPH and the second (bottom) set of images are cross sections of a normal subject.

It is important to note that the proposed method is unique in that it allows for separation of ventricle space and subarachnoid space, due to the combination of prior anatomical knowledge, intensity-based tissue classifier, and level set evolution.

Volume Statistics

The lateral ventricle, subarachnoid, white matter, and total volumes are shown in Fig. 8 and Table 2.

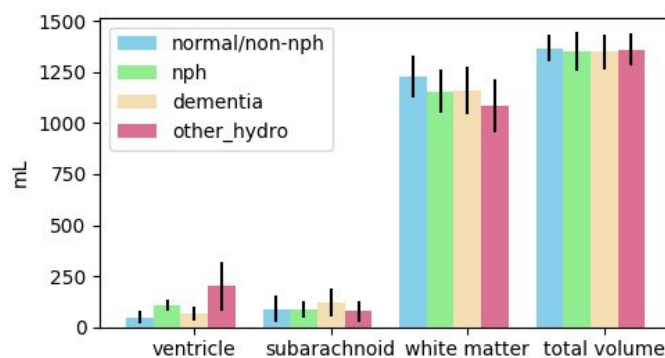


Fig. 8: Mean and standard deviation of ventricle, subarachnoid, white matter, and total volumes.

| | Ventricle | Subarachnoid | White Matter |
|----------------|------------------|-----------------|-------------------|
| Normal | 47.7 ± 30.2 mL | 89.1 ± 64.8 mL | 1229.8 ± 102.6 mL |
| NPH | 108.7 ± 28.5 mL | 85.6 ± 39.3 mL | 1154.8 ± 105.6 mL |
| Dementia | 68.3 ± 35.6 mL | 120.6 ± 67.4 mL | 1158.7 ± 115.1 mL |
| Hydro - no NPH | 200.0 ± 121.2 mL | 76.9 ± 49.3 mL | 1081.7 ± 129.6 mL |

Table 2: Mean and standard deviation of ventricular and white matter volumes.

Fig. 8 and Table 2 show that the ventricles volumes are greater for NPH and Hydrocephalus, as expected. The subarachnoid space is mostly consistent across subjects but with slightly higher mean in subjects with dementia. The white matter volumes roughly inversely correspond to ventricle volume. It is notable that the standard deviation for the ventricular volumes overlap for NPH and dementia. The ventricular volumes of normal patients are consistent with the average MRI-derived ventricular volumes in the age range of 69.5 ± 4.8 years (Jovicich et al. 2009).

Diagnosis Scores and Comparison with Evan's Ratio

A subset of the data are labeled with the Evan's ratio as measured by a neurological surgeon. The subset contains 54 subjects, with 11 scans (5 patients) having a diagnosis of normal, 34 scans (5 patients) having a diagnosis of NPH, and 9 scans (2 patients) having a diagnosis of dementia. NPH prediction on the labeled subset using only the Evan's ratio are first computed by the current guidelines, with subjects having an Evan's ratio greater than or equal to 0.3 classified as NPH, and the remaining subjects classified as non-NPH. The sensitivity and specificity of these methods are compared to a linear SVM, RBF SVM, and RF classifier trained and tested on the volumetric information of the subset of data. Because the RBF SVM and RF classifier performed better than the linear SVM, we relegate the linear SVM scores to Supplemental Materials.

| | Sensitivity (Train/Test) | Specificity (Train/Test) |
|------------------------------|--------------------------|--------------------------|
| Evan's Index, Thresholding | 89% (all data) | 85% (all data) |
| Volumetric Features, RBF SVM | 91 ± 3 / 89 ± 9% | 90 ± 4 / 89 ± 15% |
| Volumetric Features, RF | 97 ± 2 / 89 ± 10% | 98 ± 3 / 86 ± 17% |

Table 3: Thresholding and 5-fold cross-validation scores for NPH and Dementia prediction using an average of 100 linear SVMs, RBF SVMs, and random forest classifiers using Evan's Index or ventricle, white matter, subarachnoid space, and total brain volumes.¹ The thresholding of Evan's Index performs the worst in terms of specificity, but performs comparably with the fully automated methods in terms of sensitivity.

NPH diagnosis prediction scores using automatically segmented volumetric information on the full dataset is as follows:

| | Sensitivity (Train/Test) | Specificity (Train/Test) |
|----------------------|--------------------------|--------------------------|
| NPH (rbf SVM) | 85 ± 2 / 84 ± 6% | 94 ± 2 / 92 ± 8% |
| Dementia (rbf SVM) | 25 ± 40 / 3 ± 1% | 89 ± 2 / 88 ± 2% |
| NPH (RF) | 89 ± 2 / 83 ± 7% | 98 ± 1 / 91 ± 9% |
| Dementia (RF) | 94 ± 7 / 45 ± 40% | 95 ± 1 / 90 ± 3% |

Table 4: 5-fold cross-validation scores for NPH and Dementia prediction using an average of 100 linear SVMs, RBF SVMs, and random forest classifiers using ventricle, white matter, subarachnoid space, and total brain volumes.¹

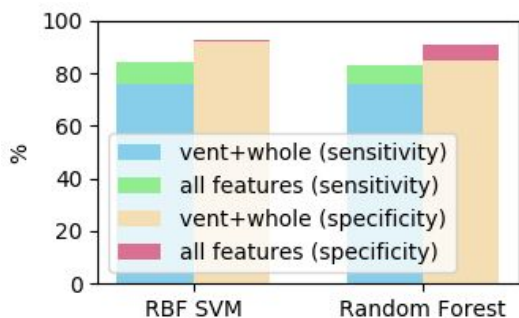


Fig. 9: Increase in performance when using all features (ventricle, white matter, subarachnoid space, and whole brain volumes) compared to using only ventricle and white matter volumes. The RBF SVM saw the most increase in sensitivity while the Random Forest Classifier saw the most increase in specificity.

Fig. 10 shows the average, standard deviation, min and max of the significance of each feature for each predictive model.

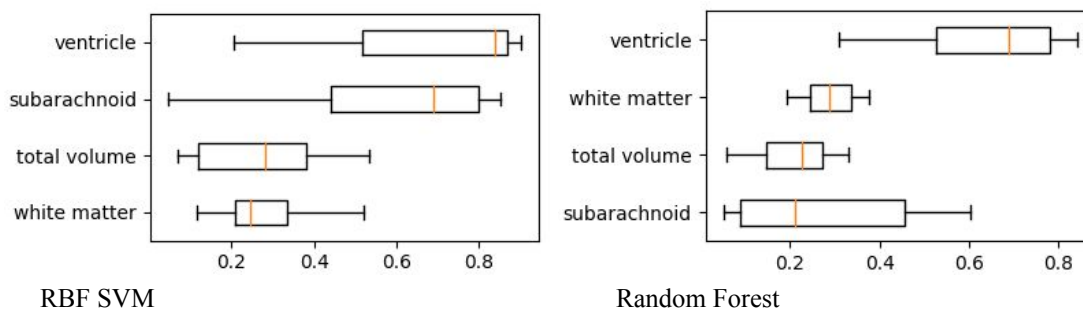


Fig. 10: Box-and-whisker plots showing the significance of each feature for each model. It is evident that the ventricular volume is the most important feature, which follows expectation. The significance of the other volumes change depending on the model.

As seen in Table 4, the SVM with RBF kernel had the best performance in NPH sensitivity and specificity. However, in terms of Dementia sensitivity and specificity, RF outperformed the SVMs. The misclassification of subjects with dementia may be due to the overlap between the ventricular volumes of NPH and dementia patients. In a clinical setting, the NPH detector may serve to alert clinicians to the possibility of NPH, with further clinical testing needed to validate the potential diagnosis.

From Fig. 9, the sensitivity of NPH diagnosis was higher when all of the volumetric data are included, while the specificity of NPH diagnosis was higher with just the ventricle and white matter volumes. It seems, then, that including the additional volume information is helpful in screening for NPH, as the potential application of our algorithm would serve to alert medical health professionals about the possibility of NPH so that they can conduct further investigation. Non-NPH hydrocephalus was not included in the diagnosis algorithm, due to the overlap in volumes with NPH. However, NPH is often misdiagnosed as other forms of dementia or missed entirely, and is not often confused with hydrocephalus from other causes.

V. DISCUSSION AND FUTURE WORK

The paper presents a fully automated, volumetric method of lateral ventricles, subarachnoid space and white matter segmentation in CT scans which outperforms the thresholding method using Evan’s ratio. Additionally, this paper proposes a fully automated, volumetric method to predict NPH diagnosis, which in conjunction with the clinical symptomatology, can facilitate the diagnosis of

NPH and rule-out patients who do not meet the radiographic criteria of an NPH diagnosis. This technological system can be used as a screening tool to identify or stratify possible NPH cases in a clinical setting.

In this paper gender is not explicitly used as a feature for NPH prediction, although the ratio of each tissue type to the overall brain volume was included in addition to raw volumetric information. Nonetheless, the influence of age-related atrophy may contribute to the observed absence of gender-based differences in brain volume presented in this study.

Additionally, it may be beneficial to explore the combination of MRI and CT scans to create a multimodal method for more fine-tuned diagnosis as well as for symptom and treatment outcome prediction. This algorithm could be further refined by incorporating relevant demographic and medical variables.

VI. ACKNOWLEDGEMENTS

This paper was partially supported by the National Institutes of Health [grant number T32-GM08620].

REFERENCES

- Andersson, J., M. Rosell, K. Kockum, L. Söderström, and K. Laurell. 2017. "Challenges in Diagnosing Normal Pressure Hydrocephalus: Evaluation of the Diagnostic Guidelines." *eNeurologicalSci* 7 (June): 27–31.
- Bijsterbosch, Janine, Stephen M. Smith, and Christian F. Beckmann. 2017. *Introduction to Resting State FMRI Functional Connectivity*. Oxford University Press.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.
- Buades, Antoni, Bartomeu Coll, and Jean-Michel Morel. 2011. "Non-Local Means Denoising." *Image Processing On Line* 1: 208–12.
- Chan, T. F., and L. A. Vese. 2001. "Active Contours without Edges." *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 10 (2): 266–77.
- Cortes, Corinna, and Vladimir Vapnik. 1995. "Support Vector Machine." *Machine Learning* 20 (3): 273–97.
- Coupé, Pierrick, José V. Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles, and D. Louis Collins. 2011. "Patch-Based Segmentation Using Expert Priors: Application to Hippocampus and Ventricle Segmentation." *NeuroImage* 54 (2): 940–54.
- Daniel Jaraj, Katrin Rabiei, Thomas Marlow, Christer Jensen, Ingmar Skoog, Carsten Wikkelsø. 2014. "Prevalence of Idiopathic Normal-Pressure Hydrocephalus." *American Academy of Neurology* 82: 1449–54.
- Espay, Alberto J., Gustavo A. Da Prat, Alok K. Dwivedi, Federico Rodriguez-Porcel, Jennifer E. Vaughan, Michela Rosso, Johnna L. Devoto, et al. 2017. "Deconstructing Normal Pressure Hydrocephalus: Ventriculomegaly as Early Sign of Neurodegeneration." *Annals of Neurology* 82 (4): 503–13.
- Grabner, Günther, Andrew L. Janke, Marc M. Budge, David Smith, Jens Pruessner, and D. Louis Collins. 2006a. "Symmetric Atlas and Model Based Segmentation: An Application to the Hippocampus in Older Adults." *Medical Image Computing and Computer-Assisted Intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention* 9 (Pt 2): 58–66.
- Gunter, Nathaniel B., Christopher G. Schwarz, Jonathan Graff-Radford, Jeffrey L. Gunter, David T. Jones, Neill R. Graff-Radford, Ronald C. Petersen, David S. Knopman, and Clifford R. Jack Jr. 2018. "Automated Detection of Imaging Features of Disproportionately Enlarged Subarachnoid Space Hydrocephalus Using Machine Learning Methods." *NeuroImage. Clinical*, November. <https://doi.org/10.1016/j.nicl.2018.11.015>.
- Jaraj, Daniel, Katrin Rabiei, Thomas Marlow, Christer Jensen, Ingmar Skoog, and Carsten Wikkelsø. 2014. "Prevalence of Idiopathic Normal-Pressure Hydrocephalus." *Neurology* 82 (16): 1449–54.
- Jenkinson, Mark, Christian F. Beckmann, Timothy E. J. Behrens, Mark W. Woolrich, and Stephen M. Smith. 2012. "FSL." *NeuroImage* 62 (2): 782–90.
- Jovicich, Jorge, Silvester Czanner, Xiao Han, David Salat, Andre van der Kouwe, Brian Quinn, Jenni Pacheco, et al. 2009. "MRI-Derived Measurements of Human Subcortical, Ventricular and Intracranial Brain Volumes: Reliability Effects of Scan Sessions, Acquisition Sequences, Data Analyses, Scanner Upgrade, Scanner Vendors and Field Strengths." *NeuroImage* 46 (1): 177–92.
- Kobashi, S., T. Takae, Y. T. Kitamura, Y. Hata, and T. Yanagida. 2001. "Fuzzy Medical Image Processing for Segmenting the Lateral Ventricles from MR Images." In *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, 3:1095–98 vol.3.
- Lin, Eugene, and Adam Alessio. 2009. "What Are the Basic Concepts of Temporal, Contrast, and Spatial Resolution in Cardiac CT?" *Journal of Cardiovascular Computed Tomography* 3 (6): 403–8.

- Lüders, Eileen, Helmuth Steinmetz, and Lutz Jäncke. 2002. "Brain Size and Grey Matter Volume in the Healthy Human Brain." *Neuroreport* 13 (17): 2371–74.
- Martín-Láez, Rubén, Hugo Caballero-Arzapalo, Luis Ángel López-Menéndez, Juan Carlos Arango-Lasprilla, and Alfonso Vázquez-Barquero. 2015. "Epidemiology of Idiopathic Normal Pressure Hydrocephalus: A Systematic Review of the Literature." *World Neurosurgery* 84 (6): 2002–9.
- Moore, Dana W., Ilhami Kovanlikaya, Linda A. Heier, Ashish Raj, Chaorui Huang, King-Wai Chu, and Norman R. Relkin. 2012. "A Pilot Study of Quantitative MRI Measurements of Ventricular Volume and Cortical Atrophy for the Differential Diagnosis of Normal Pressure Hydrocephalus." *Neurology Research International* 2012: 718150.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research: JMLR* 12 (Oct): 2825–30.
- Shao, Muhan, Shuo Han, Aaron Carass, Xiang Li, Ari M. Bliz, Jerry L. Prince, and Lotta M. Ellingsen. 2018. "Shortcomings of Ventricle Segmentation Using Deep Convolutional Networks: First International Workshops, MLCN 2018, DLF 2018, and iMIMIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16-20, 2018, Proceedings." In *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*, edited by Danail Stoyanov, Zeike Taylor, Seyed Mostafa Kia, Ipek Oguz, Mauricio Reyes, Anne Martel, Lena Maier-Hein, et al., 11038:79–86. Lecture Notes in Computer Science. Cham: Springer International Publishing.
- Shinoda, Narihide, Osamu Hirai, Shinya Hori, Kazuyuki Mikami, Toshiaki Bando, Daisuke Shimo, Takahiro Kuroyama, Yoji Kuramoto, Masato Matsumoto, and Yasushi Ueno. 2017. "Utility of MRI-Based Disproportionately Enlarged Subarachnoid Space Hydrocephalus Scoring for Predicting Prognosis after Surgery for Idiopathic Normal Pressure Hydrocephalus: Clinical Research." *Journal of Neurosurgery* 127 (6): 1436–42.
- Shprecher, David, Jason Schwalb, and Roger Kurlan. 2008. "Normal Pressure Hydrocephalus: Diagnosis and Treatment." *Current Neurology and Neuroscience Reports* 8 (5): 371–76.
- Takahashi, Noriyuki, Toshibumi Kinoshita, Tomomi Ohmura, Eri Matsuyama, and Hideto Toyoshima. 2017. "Automated Method to Compute Evans Index for Diagnosis of Idiopathic Normal Pressure Hydrocephalus on Brain CT Images." In *Medical Imaging 2017: Computer-Aided Diagnosis*, 10134:101342C. International Society for Optics and Photonics.
- Toma, Ahmed K., Etienne Holl, Neil D. Kitchen, and Laurence D. Watkins. 2011. "Evans' Index Revisited: The Need for an Alternative in Normal Pressure Hydrocephalus." *Neurosurgery* 68 (4): 939–44.
- Walt, Stéfan van der, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and scikit-image contributors. 2014. "Scikit-Image: Image Processing in Python." *PeerJ* 2 (June): e453.
- Yepes-Calderon, Fernando, Marvin D. Nelson, and J. Gordon McComb. 2018. "Automatically Measuring Brain Ventricular Volume within PACS Using Artificial Intelligence." *PloS One* 13 (3): e0193152.

Supplementary Materials

Segmentation and Prediction Implementation Details

The optimal hyperparameters for the SVM are $C=1.0$ and $\gamma=0.25$. For the random forest classifier, the optimal hyperparameters are 20 trees, max depth=5, bootstrapping=True, minimum sample split=3, max features=2, and gini criterion (impurity measure).

The random forest tissue classifier used 20 trees and a max depth of 3, created with random subsets of the data and a gini impurity measure, $Gini(E) = 1 - \sum_{j=1}^c p_j^2$, where c is the number of classes and p_j is the fraction of items labeled with the class j in the set. For morphological chan vese, $\lambda_1 = \lambda_2 = 1$.

For diagnosis prediction, the linear and RBF SVMs had a c value of 1 and the RBF SVM had a γ value of 0.25. The random forest classifier had 20 trees, max depth of 5, bootstrapping, a minimum sample split or 10, a maximum feature of 2, and gini impurity measure. All other parameters are left at default values given by the sklearn (Pedregosa et al. 2011) and skimage (van der Walt et al. 2014) software.

Model optimization

The volumes are scaled to a range of $[-1, 1]$. To optimize the SVM parameters, a grid search was done with C and gamma (gamma for rbf kernel only) ranging logarithmically from 10^{-3} to 10^3 . The grid search shown in Table 3 was used to optimize the random forest parameters. Since there is some variance in the performance of each classifier trained, scores are averaged over 100 different trained classifiers for each method. Stratified k-fold cross validation was used to separate the training and testing datasets.

| Parameter | Values |
|------------------------|------------------|
| Number of estimators | 10, 20, 100, 200 |
| Max. depth | 1, 3, 5 |
| Max. features | [1:3] |
| Min. samples for split | [2:10] |
| Bootstrap | True, False |
| Criterion | Gini, Entropy |

Table 5: Grid search parameters for the random forest classifier

| | Sensitivity (Train/Test) | Specificity (Train/Test) |
|-----------------------|---------------------------------------|---------------------------------------|
| NPH (linear SVM) | 76 ± 1 / 76 ± 6% 79 ± 3 / 79 ± 6% | 94 ± 4 / 92 ± 10% 89 ± 3 / 88 ± 9% |
| Dementia (linear SVM) | 0 ± 0 / 0 ± 0% 13 ± 10 / 1 ± 7% | 88 ± 0 / 88 ± 1% 88 ± 0 / 88 ± 1% |
| NPH (rbf SVM) | 76 ± 2 / 76 ± 6% 85 ± 2 / 84 ± 6% | 94 ± 3 / 93 ± 9% 94 ± 2 / 92 ± 8% |
| Dementia (rbf SVM) | 21 ± 41 / 0 ± 0% 25 ± 40 / 3 ± 1% | 88 ± 0 / 88 ± 1% 89 ± 2 / 88 ± 2% |
| NPH (RF) | 84 ± 3 / 76 ± 7% 89 ± 2 / 83 ± 7% | 97 ± 2 / 85 ± 11% 98 ± 1 / 91 ± 9% |
| Dementia (RF) | 97 ± 2 / 5 ± 20% 94 ± 7 / 45 ± 40% | 92 ± 1 / 88 ± 2% 95 ± 1 / 90 ± 3% |

Table 6: 5-fold cross-validation scores for NPH prediction using an average of 100 linear SVMs, RBF SVMs, and random forest classifiers. The first row (in grey) are the scores using only ventricle and whole brain volumes, while the second row are scores using ventricle, white matter, subarachnoid space, and total brain volumes.

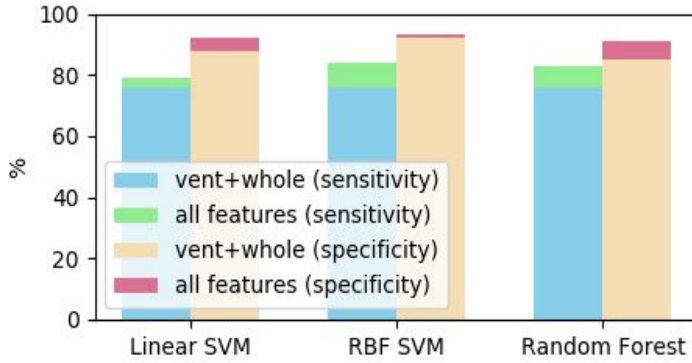


Fig. 11: Increase in performance when using all features (ventricle, white matter, subarachnoid space, and whole brain volumes) compared to using only ventricle and white matter volumes. Linear SVM, RBF SVM, and Random Forest are used.

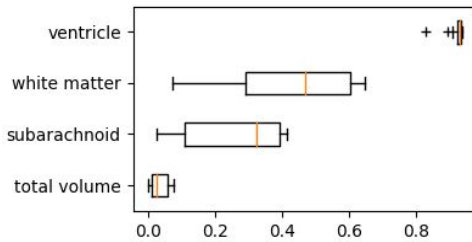


Fig. 12: Feature Significance plot for linear SVM.

Linear SVM

| | | | |
|----------|-----|-----|------|
| Normal | 37 | 11 | 0.34 |
| NPH | 5.9 | 74 | 0.13 |
| Dementia | 8.6 | 9.4 | 0.02 |

Normal NPH Dementia

RBF SVM

| | | | |
|----------|----|-----|------|
| Normal | 40 | 7.4 | 1.6 |
| NPH | 4 | 76 | 0.06 |
| Dementia | 10 | 7.4 | 0.28 |

Normal NPH Dementia

Random Forest Classifier

| | | | |
|----------|-----|-----|------|
| Normal | 36 | 9.6 | 3.3 |
| NPH | 3.9 | 75 | 0.68 |
| Dementia | 7.2 | 6.7 | 4.2 |

Normal NPH Dementia

Table 7: confusion matrices for the average of 100 linear SVMs, RBF SVMs, and Random Forest classifiers trained using ventricle, subarachnoid space, white matter, and overall brain volume as features. Values represent number of scans.