

SCIENTIFIC REPORTS



OPEN

MRI Compatibility: Automatic Brain Shunt Valve Recognition using Feature Engineering and Deep Convolutional Neural Networks

Luca Giancardo¹, Octavio Arevalo^{2,3}, Andrea Tenreiro^{2,3}, Roy Riascos^{2,3} & Eliana Bonfante^{2,3}

The aim of this study is to evaluate whether we could develop a machine learning method to distinguish models of cerebrospinal fluid shunt valves (CSF-SVs) from their appearance in clinical X-rays. This is an essential component of an automatic MRI safety system based on X-ray imaging. To this end, a retrospective observational study using 416 skull X-rays from unique subjects retrieved from a clinical PACS system was performed. Each image included a CSF-SV representing the most common brands of programmable shunt valves currently used in US which were split into five different classes. We compared four machine learning pipelines: two based on engineered image features (Local Binary Patterns and Histogram of Oriented Gradients) and two based on features learned by a deep convolutional neural network architecture. Performance is evaluated using accuracy, precision, recall and f1-score. Confidence intervals are computed with non-parametric bootstrap procedures. Our best performing method identified the valve type correctly 96% [CI 94–98%] of the time (CI: confidence intervals, precision 0.96, recall 0.96, f1-score 0.96), tested using a stratified cross-validation approach to avoid chances of overfitting. The machine learning pipelines based on deep convolutional neural networks showed significantly better performance than the ones based on engineered image features (mean accuracy 95–96% vs. 56–64%). This study shows the feasibility of automatically distinguishing CSF-SVs using clinical X-rays and deep convolutional neural networks. This finding is the first step towards an automatic MRI safety system for implantable devices which could decrease the number of patients that experience denials or delays of their MRI examinations.

Magnetic Resonance Imaging (MRI) is an essential tool for the evaluation and diagnosis of several medical conditions which uses a powerful magnetic field and radiofrequency stimulation. Implanted medical devices and metallic foreign bodies with ferromagnetic properties interact with the magnetic field and react to radiofrequency stimulation, resulting in undesirable changes in the implantable devices, such as malfunction, displacement, or thermal effects on the tissues surrounding the implant. Prior to an MRI examination, patients need to undergo a screening process to determine if they have devices or foreign bodies that might cause problems, or devices that need to be scanned under specific parameters to assure safety of the patient and integrity of the implant. If patients are able to provide the information, they will fill a screening sheet with the device information. Implant cards or surgical records complement the information provided by the patient. In routine practice, a significant percentage of patients do not have accurate information about medical devices implanted in their bodies. Based on 50 interviews with staff members involved in scheduling and preparing patients for MRI at Memorial Hermann-Texas Medical Center (one of the leading hospitals in the South and Southwest USA), we estimated that 5–10% of patients referred for MRI do not have conclusive information regarding implanted medical devices. As a consequence, radiologists need to infer the device implanted from visually inspecting the X-rays, which leads to delay in the MRI exam and inefficient use of their time. This problem is exacerbated in the emergency room

¹Center for Precision Health, School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, USA. ²McGovern Medical School, University of Texas Health Science Center, Houston, TX, USA. ³Memorial Hermann Hospital, Houston, TX, USA. Correspondence and requests for materials should be addressed to L.G. (email: luca.giancardo@uth.tmc.edu)

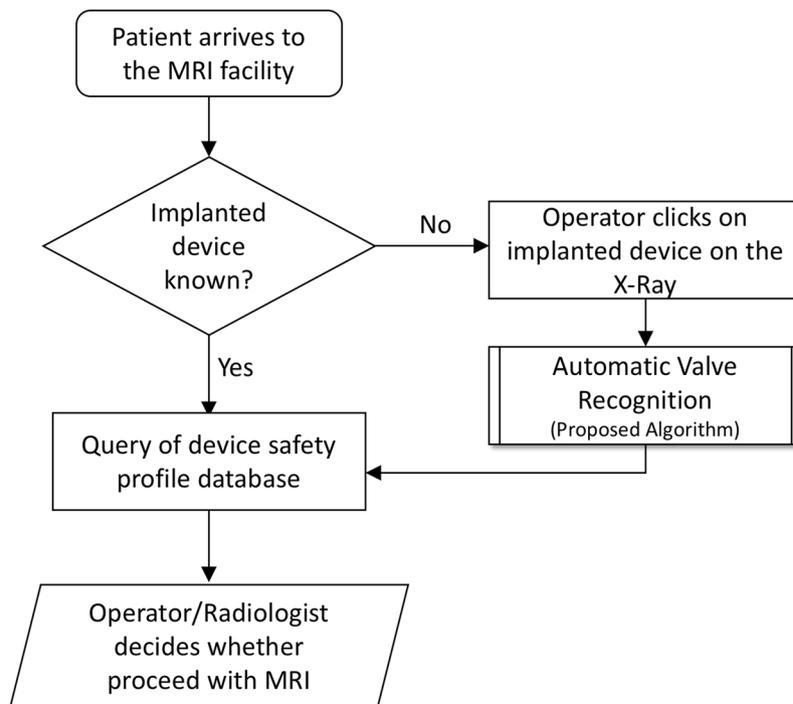


Figure 1. Pipeline of the envisioned system.

setting, where many patients are unable to provide information regarding implantable devices due to acute medical conditions and a timely MRI can be essential for life saving procedures. A streamlined method of identifying implantable medical devices is not currently available.

In this work, we focus on cerebrospinal fluid shunt valves (CSF-SVs) a type of implanted device important for hydrocephalus treatment. The current standard of care already requires radiologists to visually compare X-rays to manufacturer's product manuals or other medical publications^{1–3}. Correct identification of CSF-SVs is essential for verifying the compatibility with the MRI machine and to monitor for potential setting changes due to the MR magnetic field^{4–8}. In fact, exposing patients with some programmable CSF-SVs to MR imaging procedures poses a significant risk of unintentional changes in shunt settings, even if such devices are not MR incompatible^{9–12}. Lavinio *et al.*¹³ tested 5 different types of CSF-SV (Codman Hakim, Miethke ProGAV, Medtronic Strata, Sophysa Sophy and Polaris) and found that, with the exception of the Polaris and ProGAV models, all are prone to unintentional reprogramming when exposed to heterogeneous magnetic fields stronger than 40 mT. For this reason, these valves are considered “MR conditional” according to the American Society for Testing and Materials (ASTM)¹⁴ and require monitoring and readjusting valve setting after the MRI examination.

In the current standard of care, CSF-SV identification requires coordinated interaction between the MR technologist, the X-ray technologist, the radiologist interpreting the X-ray, and personnel from the referring service. However, this process is suboptimal due to multiple factors, including that non-urgent X-rays have a low level of priority in busy radiology departments, and that the identification process is time consuming and tedious¹⁵. There are currently 4 different manufacturers of shunt valves in the US market. An image-based system able to automatically identify the main valve model from the X-ray image would allow to immediately display all the safety information required to the radiologist. Based on personal experience from 5 radiologists in our institutions, we estimate that such system has the potential to decrease the X-ray interpretation time by 50–70%, expedite clearance for MRI imaging in emergency conditions, and provide an extra layer of safety for patients and health care providers. We envision an integrated system where a radiologist or X-ray operator could click on the image of the implanted device, a machine learning-based algorithm identifies the type of implanted device, and a database is automatically queried to show the appropriate safety profile (see Fig. 1).

In this study, we compared 4 machine learning pipelines to identify 5 different valves classes on a dataset containing 416 valve X-ray images collected from our clinical PACS system. Two pipelines were based on predefined texture features (Local Binary Patterns and Histogram of Oriented Gradients) and two based on features learned by a deep convolutional neural network architecture. We believe that we are the first group to propose such system and demonstrate its feasibility on clinical data.

Materials and Methods

Dataset. The dataset was acquired as a part of a quality improvement project approved by our institution (Quality Improvement Project Registry, no. 2017–017). The UTHealth IRB Office reviewed the study protocol and established that it does not meet the regulatory definition of human subjects research. The methods were carried out in accordance with the relevant guidelines and regulations.

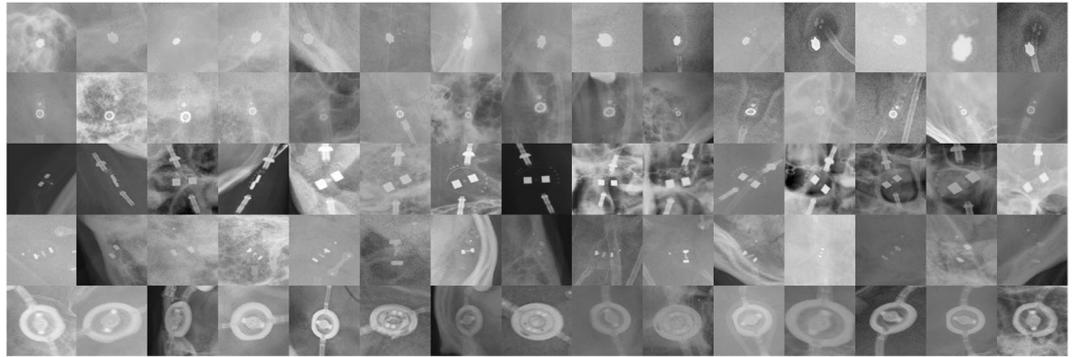


Figure 2. Examples of CSF-SV X-ray valve images in our dataset. Each row shows 15 random samples for each of the 5 classes used. From the top, row 1: *Medtronic Strata II - NSC* (n = 164); row 2: *Codman-Hakim* (n = 106); row 3: *Sophysa Polaris SPV* (n = 82); row 4: *Codman-Certas* (n = 42); row 5: *Miethke ProGAV* (n = 22). In parentheses, the number of samples contained in the dataset. The dataset contains all of the shunt valves brands currently used in US.

A total of 416 skull X-rays that included a CSF-SV image were collected from the institutional PACS. We grouped together different versions of the same CSF-SV models. The specific 5 class grouping is as follows: *Codman-Certas*[®] (42), *Codman-Hakim*[®] (106), *Miethke ProGAV*[®] (22), *Sophysa Polaris SPV*[®] (standard/140/400) (82) and *Medtronic Strata*[®] (II/NSC) (164). All images were acquired from different subjects and were selected regardless of acquisition perspective, scale or any demographic information. All images were anonymized and an expert radiologist was asked to select an image window of 300 × 300 pixels containing the valve. Such image windows included: valves from different perspectives, scales, brightness, as well as confounding background objects, such as: bone structures, craniotomy hardware or catheters, as shown in Fig. 2. This dataset simulated a system where the X-ray operator or radiologist clicks on an implanted device identified in the X-ray images to retrieve the relevant implanted device safety profile.

Machine learning pipelines. We developed and tested 4 different machine learning-based pipelines. Each pipeline can be broadly split into an image feature extraction and classification phase. During the image feature extraction phase, a compact numerical representation of the image is generated by encoding the visual information into a fixed size feature vector whose size is significantly less than the number of pixels in the image. These vectors can be created using feature engineering, where measurements are based on a predetermined set of imaging operators or learned, by using representation learning approaches such as the ones implemented by Deep Convolutional Neural Networks (DCNNs)^{16,17}. In the classification phase, feature vectors are used as input for a machine learning classifier that will predict the most likely valve type.

Image feature extraction with feature engineering. The images were pre-processed with histogram normalization which ensures that the dynamic range of the image histogram is between 0 and 1. Then, we implemented two feature extraction pipelines using validated feature engineering approaches, one based on Local Binary Patterns (LBP)¹⁸ and one based on Histogram of Oriented Gradients (HOG)¹⁹.

LBP features are computed by splitting the image in local windows. Each pixel in the window is compared to its neighbors in order to generate a unique code that represents a characteristic of the texture, such as edges, corners or flat areas. The histogram of these codes is the actual feature vector that will be used as input for the classification phase. More information is available in Ojala *et al.*¹⁸. In our experiments, we used a multiscale LBP approach using neighbor radius of 6 and 12 pixels, histograms of 30 bins and the LBP codes invariant to image rotations.

HOG features are based the image gradients, i.e. the direction/intensity of the image edges. The distribution of these image gradients is computed in local windows that are then concatenated to create the final feature vector. More information is available in Dalal *et al.*¹⁹. In our experiments we used local windows (or cells) of 20 × 20 pixels.

We leveraged the implementation provided in the scikit-image Python library²⁰ for both LBP and HOG.

Image feature extraction with deep convolutional neural networks (DCNN). In contrast to feature engineering, representation learning using approaches like DCNN allow for the creation of feature vectors that are not based on a predefined set of rules, but rather learned from the data at hand¹⁶. However, modern DCNNs often require hundreds of thousands of images for a complete training, also called end-to-end training. Transfer learning is a strategy where a DCNN is first trained on a large dataset containing images unrelated to the problem of interest and then adapted to a smaller dataset. In our experiments, we adopted transfer learning strategies with a modern DCNN architecture as a feature vector generator. Transfer learning had already been used successfully for multiple computational medical imaging problems²¹. We used the Xception network architecture²², a DCNN inspired by Inception V3 where convolutional filters are depthwise separable. This network was composed of 126 layers for a total of 22,910,480 trainable parameters (or network weights). A full description of the network is beyond the scope of this article, and we refer the reader to the original publication. In our experiments, we used an Xception network pre-trained on the Imagenet dataset (available at www.image-net.org) which contains over

		Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score
Feature Engineering Methods	Local Binary Patterns (LBP)	64% [60–69]	0.64 [0.60–0.69]	0.64 [0.60–0.69]	0.64 [0.60–0.68]
	Histogram of Oriented Gradients (HOG)	56% [53–63]	0.57 [0.53–0.63]	0.56 [0.52–0.61]	0.51 [0.47–0.56]
Deep Convolutional Neural Networks (Transfer Learning)	Xception Network	95% [94–97]	0.95 [0.94–0.97]	0.95 [0.94–0.97]	0.95 [0.93–0.97]
	Enhanced-Xception Network	96% [94–98]	0.96 [0.95–0.98]	0.96 [0.94–0.98]	0.96 [0.94–0.98]

Table 1. Classification performance of the 4 machine learning pipelines. In brackets the confidence intervals are shown. The accuracy column shows the absolute percentage of images correctly classified. The precision, recall and f1-score columns represent the performance metrics averaged over the 5 classes. The performance metric for each class is available in Table 2. Deep convolutional networks trained with a transfer learning strategy clearly outperform the two feature engineering methods tested. Specifically, the “Enhanced-Xception Network” is the one showing the best performance overall. (*Precision* is also known as *positive predictive value*; *recall* is also known as *sensitivity*; *F1-score* is the harmonic mean of precision and recall).

	N	Precision	Recall	F1-score	
Strata II - NSC	164	0.98	0.96	0.97	***p < 0.001
Codman-Hakim	106	0.92	0.98	0.95	***p < 0.001
Sophysa Polaris SPV	82	0.98	1.00	0.99	***p < 0.001
Codman-Certas	42	0.95	0.86	0.90	***p < 0.001
Miethke proGAV	22	1.00	0.95	0.98	***p < 0.001

Table 2. Class-level performance metrics using the Enhanced-Xception Network. N = number of samples per class. (*Precision* is also known as *positive predictive value*; *recall* is also known as *sensitivity*; *F1-score* is the harmonic mean of precision and recall). The p-value is computed using the two-tailed Mann–Whitney U test to reject the null hypothesis that each valve is indistinguishable from the others in a 1 vs all framework.

14 million of hand-annotated natural color images for 1000 classes. The last fully-connected layer of the network was removed and a max pooling layer was added to generate the feature vector. The images were pre-processed using the same histogram normalization as discussed in the previous section, with an additional step to encode monochrome X-ray images into a 3-channel image.

Since the network pre-training was performed on color images, we would normally have to artificially replicate the monochrome intensity value of the X-ray images into 3 channels which the network interprets as RGB. This is the typical approach used by multiple research groups²¹. However, tissues, bones and implanted devices in X-ray images have well defined absorption rates as the X-ray beams traverse matter. Therefore, global thresholds can be effective in differentiating prominent structures from the background. Using this insight, we devised a novel pre-processing strategy allowing us to input additional domain relevant information into the network. In this strategy, we created an image containing a rough estimation of the foreground objects, which was placed into the red and blue channels, while the original X-ray image was left untouched in the green channel. The threshold for foreground objects was estimated using the non-parametric Otsu thresholding approach²³. In the rest of the manuscript, this step is referred as Enhanced-Xception Network. All DCNNs were implemented using the Keras (www.keras.io) and Tensorflow (www.tensorflow.org) Python libraries.

Valve classification and statistical analysis. We used the same classification phase for all four feature extraction strategies. The generated feature vector was classified using a linear logistic regression classifier with L2 regularization (the default 1.0 was used as regularization strength). The model was extended to multiclass using a 1 vs all strategy and the coordinate descent optimizer implemented by LIBLINEAR/scikit-learn²⁴ was used for training.

We used stratified 10-fold cross-validation to evaluate the performance of the machine learning pipelines. In summary, the dataset was split into 10 chunks (or folds), each fold maintained the same class distribution of the complete dataset. One fold was left out as testing set and the classifier was trained on the remaining 9 folds. This operation was iteratively performed for all folds making sure that the classifier is reset at each iteration. This strategy avoided overfitting and allowed us to robustly estimate the classification performance.

We evaluated the classification performance using *precision* (also known as *positive predictive value*), *recall* (also known as *sensitivity*) and *F1-score*, i.e. the harmonic mean of precision and recall. All confidence intervals were computed using the non-parametric bootstrap procedure, using 1000 repetitions and reporting the 5th and 95th percentiles. P-values are computed using the two-tailed Mann–Whitney U test to reject the null hypothesis that each valve is indistinguishable from the others in a 1 vs all strategy.

Results

Table 1 shows the classification performance of the four machine learning pipelines. Deep convolutional networks trained with a transfer learning strategy clearly outperform the two feature engineering methods tested by achieving an accuracy of 95–96% (confidence intervals (CI) [94–97]/[94–98]) vs 56–64% (CI [53–63]/[60–69]). Specifically, the Enhanced-Xception Network performed best for all metrics evaluated (precision: 0.96 CI [0.95–0.98], recall: 0.96 [0.94–0.98]).

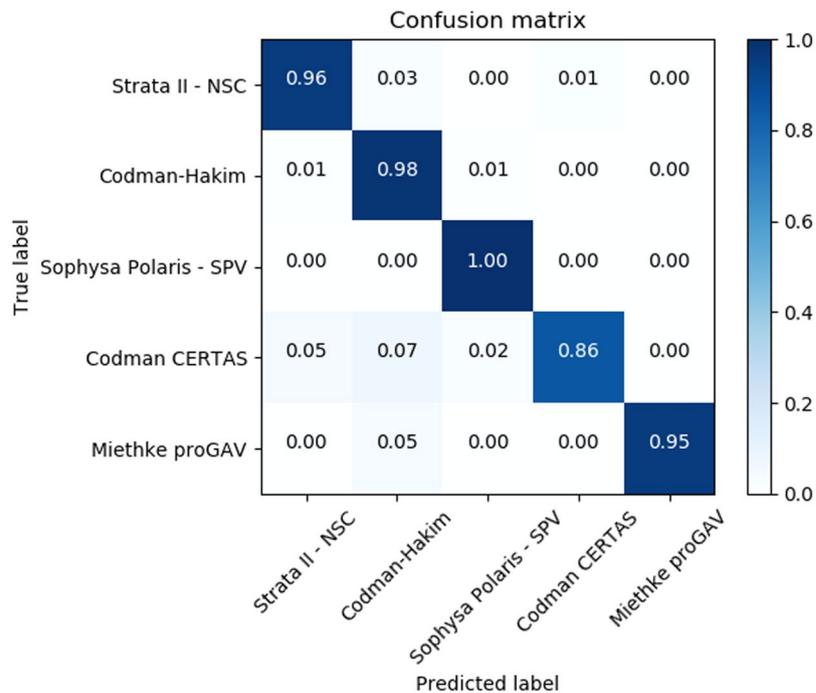


Figure 3. Confusion matrix for the Enhanced-Xception Network. Each cell shows the ratio of images of a given class (true label) classified into the class indicated by the column (predicted label). A perfect classifier will only have 1.00 in the matrix diagonal.

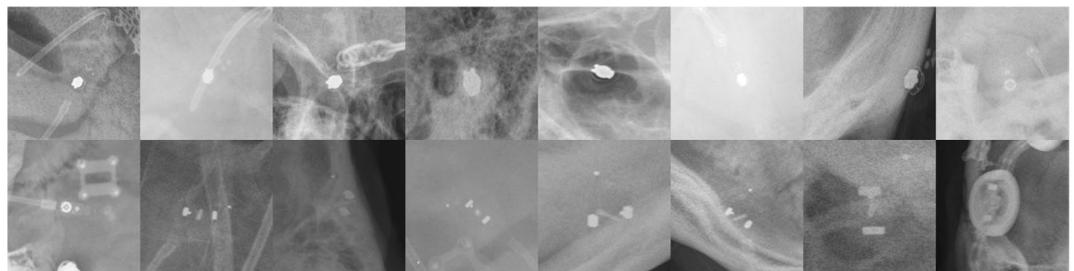


Figure 4. All 16 misclassified samples by the Enhanced-Xception Network. Large foreign object, low contrast or acquisition angles not well represented in the dataset are visible.

In Table 2, we investigate the performance metric for each valve. All valves were classified with a F1-score equal to or above 0.90, which ranged from 0.99 on the Sophysa Polaris SPV class to 0.90 with the Codman-Certas. The F1-score is computed as the harmonic mean of precision and recall. Figure 3 shows the confusion matrix indicating the correct and wrong classification by true valve class and predicted valve class. No obvious misclassification bias was apparent in the matrix.

After training, all four machine pipelines could be run in under 1 second per image on a 3.0 Ghz Xeon desktop with a Titan X GPU. Specifically, LBP: ~0.13 seconds/image; HOG: ~0.08 seconds/image; Xception Network ~0.53 seconds/image; Enhanced-Xception Network: ~0.54 seconds/image. The code to replicate our comparative feature analysis is available at <https://github.com/lgiancaUTH/shunt-valve-mri-compatibility>.

Discussion

In this work, we described a X-rays-based automatic implanted device identification system for MRI safety and devised a pilot study to test the feasibility of the automatic CSF-SV recognition component. We designed four different machine learning pipelines and our results indicate that a deep learning based algorithm (Enhanced-Xception Network) can achieve a very high accuracy (96%) in identifying the valves correctly. We visually inspected the 16 (out of 416 images) that were misclassified (see Fig. 4). In all cases, we noticed large foreign objects, low contrast or acquisition angles not well represented in the dataset. These issues are likely to be solved by increasing the dataset size and including an automatic quality assurance algorithm that would warn the user if the quality of the image is too low for a reliable valve identification. In general, the algorithm was able to classify very challenging samples of valves imaged at skewed angles, scales and location in the skull as shown in Fig. 2.

The best performing algorithm can be run in real time on commodity hardware, thereby making it possible to integrate it on X-ray machines, hospital picture archiving and communication systems (PACS) or software-as-a-service (SaaS) cloud services. Additionally, the proposed approach does not require any type of protected health information (PHI), thereby drastically reducing security concerns in the translation of the project to clinical practice.

This study presents some limitations. First, being a retrospective study based on clinical data, we have a different number of samples for each valve type. In order to address the issue, we combined different versions of the same valve type if discrimination between versions was not clinically useful. In our analysis, we could not identify clear classification biases due to class imbalances. Secondly, the automatic image recognition component, while achieving excellent classification performance, did not achieved the performance required for a truly automatic MRI compatibility system. However, in the envisioned workflow shown in Fig. 1, we will always have a human visually comparing the valve models identified by the automated system with the actual X-ray image. Therefore, perfect accuracy would not be strictly required. Additional studies applying our valve identification software in clinical practice are underway.

This technology has the potential to be applied to other medical devices thereby enabling a fast and precise MRI clearance process for busy radiology departments and decrease the number of patients that experience denials or delays of their MRI exams. In fact, the algorithm proposed is not specific to a particular type of implanted device and could be readily adapted by retraining the algorithm on other datasets.

References

- Shellock, F. G. Biomedical implants and devices: Assessment of magnetic field interactions with a 3.0-Tesla MR system. *J. Magn. Reson. Imaging* **16**, 721–732 (2002).
- Shellock, F. G., Habibi, R. & Knebel, J. Programmable CSF Shunt Valve: *In Vitro* Assessment of MR Imaging Safety at 3T. *Am. J. Neuroradiol.* **27**, 661–665 (2006).
- Shellock, F. G. & Cruess, J. V. MR Procedures: Biologic Effects, Safety, and Patient Care. *Radiology* **232**, 635–652 (2004).
- Hatlen, T. J. *et al.* Nonprogrammable and programmable cerebrospinal fluid shunt valves: a 5-year study. *J. Neurosurg. Pediatr.* **9**, 462–467 (2012).
- Li, M., Wang, H., Ouyang, Y., Yin, M. & Yin, X. Efficacy and safety of programmable shunt valves for hydrocephalus: A meta-analysis. *Int. J. Surg.* **44**, 139–146 (2017).
- Ortler, M., Kostron, H. & Felber, S. Transcutaneous pressure-adjustable valves and magnetic resonance imaging: an *ex vivo* examination of the Codman-Medos programmable valve and the Sophy adjustable pressure valve. *Neurosurgery* **40**, 1050–1057; discussion 1057–1058 (1997).
- Zuzak, T. J., Balmer, B., Schmidig, D., Boltshauser, E. & Grotzer, M. A. Magnetic toys: forbidden for pediatric patients with certain programmable shunt valves? *Childs Nerv. Syst. ChNS Off. J. Int. Soc. Pediatr. Neurosurg.* **25**, 161–164 (2009).
- Nakashima, K. *et al.* Programmable shunt valves: *in vitro* assessment of safety of the magnetic field generated by a portable game machine. *Neurol. Med. Chir. (Tokyo)* **51**, 635–638 (2011).
- Zabramski, J. M., Preul, M. C., Debbins, J. & McCusker, D. J. 3T magnetic resonance imaging testing of externally programmable shunt valves. *Surg. Neurol. Int.* **3** (2012).
- Inoue, T., Kuzu, Y., Ogasawara, K. & Ogawa, A. Effect of 3-tesla magnetic resonance imaging on various pressure programmable shunt valves. *J. Neurosurg.* **103**, 163–165 (2005).
- Lindner, D., Preul, C., Trantakis, C., Moeller, H. & Meixensberger, J. Effect of 3T MRI on the function of shunt valves—evaluation of Paedi GAV, Dual Switch and proGAV. *Eur. J. Radiol.* **56**, 56–59 (2005).
- Allin, D. M., Czosnyka, M., Richards, H. K., Pickard, J. D. & Czosnyka, Z. H. Investigation of the hydrodynamic properties of a new MRI-resistant programmable hydrocephalus shunt. *Cerebrospinal Fluid Res.* **5**, 8 (2008).
- Lavinio, A. *et al.* Magnetic field interactions in adjustable hydrocephalus shunts. *J. Neurosurg. Pediatr.* **2**, 222–228 (2008).
- Shellock, F. G. TheList. *MRI Safety.com* Available at: http://www.mrisafety.com/TheList_search.asp. (Accessed: 17th January 2018)
- Lollis, S. S., Mamourian, A. C., Vaccaro, T. J. & Duhaime, A.-C. Programmable CSF shunt valves: radiographic identification and interpretation. *AJNR Am. J. Neuroradiol.* **31**, 1343–1346 (2010).
- Bengio, Y., Courville, A. & Vincent, P. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–828 (2013).
- LeCun, Y. *et al.* Deep learning. *Nature* **521**, 436–444 (2015).
- Ojala, T., Pietikainen, M. & Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans PAMI* **24**, 971–987 (2002).
- Dalal, N. & Triggs, B. Histograms of oriented gradients for human detection. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* **1**, 886–893, (IEEE, 2005).
- van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
- Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).
- Chollet, F. Xception: Deep learning with depthwise separable convolutions. *ArXiv Prepr.* (2016).
- Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
- Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

We would like to thank Dr. Elmer Bernstam, Dr. Jeanie Choi for their insightful suggestions, and all of the collaborators at UTHealth and Memorial Hermann for making this project possible.

Author Contributions

L.G., R.R., E.B. planned and supervised the experiments; O.A., A.T. collected, cleaned and vetted the data; L.G. developed the algorithms and ran the computational experiments. All authors discussed the results and contributed to the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018