Machine Learning-Based Radiomic Analysis of Pre-Treatment MRI Predicts Symptomatic Improvement in Patients with Venous Malformations Following a Single Sclerotherapy

Alan J. Kim, BS[1]; Samuel Lefcourt, MS[2]; Adham Khalil, MD[3]; Suizhi Ma, BS[2]; Craig K. Jones, PhD[1,2]; Clifford R. Weiss, MD[1]

[1] Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD; [2] Department of Computer Science, Johns Hopkins University, Baltimore, MD; [3] Medstar Georgetown University/Washington Hospital Center, Washington, D.C.

Introduction

- Venous malformations (VMs) are congenital low-flow vascular malformations that arise from the abnormal development of veins[1]
- VMs cause significant morbidity, including pain, swelling, functional impairment, and cosmetic disfigurement[2]
- Percutaneous sclerotherapy (PS) is first-line treatment, and involves image-guided intra- or perilesional injection of a sclerosing agent[3]
- However, predicting at baseline which patients will achieve meaningful symptomatic improvement, rather than mere volumetric reduction, remains a significant challenge

Objective

To develop a radiomics-based machine learning (ML) model that uses pre-treatment multiparametric MRI features to predict patient-reported outcomes after a single session of sclerotherapy

Materials and Methods

- We performed an IRB-approved retrospective review of patients with VMs treated with percutaneous sclerotherapy at our institution between November 2005 and August 2024
- All patients included in our study received:
  - Multiparametric MRI—including T2-weighted imaging (T2WI) with fat saturation—before their first session of PS
  - ≥ 1 session of percutaneous sclerotherapy from our institution's division of interventional radiology
- Patients without pre-treatment T2WI, not receiving PS at our institution, or lacking 3- to 4-week follow-up after treatment were excluded
- From the eligible cohort, we randomly selected 100 head-and-neck, 100 extremity, and 37 trunk VM cases
- We collected data for assessment of the following parameters:
  1) Demographics and Laboratory Values
  2) Presenting Symptoms
  3) Sclerotherapy Procedural Details
  4) Clinical Response Following First Session of Sclerotherapy
  5) Pre-Treatment Multiparametric Images
- Clinical Response following first sclerotherapy was classified as optimal vs. suboptimal based on patient-reported % improvement in symptoms and/or changes in patient-reported pain scores recorded prior to and following treatment (**Table 1**)

Materials and Methods

Clinical Response	Criteria
Optimal	Patient states ≥66% improvement. If patient did not directly state % change, ≥6.6 reduction in visual analogue pain score
Suboptimal	Patient states <66% improvement. If patient did not directly state % change, <6.6 reduction in visual analogue pain score

Table 1. Clinical Response Grading Criteria

- Lesion segmentation was performed manually on T2WI using 3D Slicer[4] and validated by two board-certified radiologists **(Figure 1)**
- Radiomic features were extracted at 1, 3, and 5 mm scales to account for texture differences at both finer (i.e., 1mm) and broader (i.e., 5 mm) levels
  - At each spatial scale, 102 features were extracted
    - 18 first-order features (e.g., mean, median, mode voxel intensity)
    - 14 shape features (e.g., sphericity, major axis length, surface area)
    - 70 texture features (24 GLCM, 14 GLDM, 16 GLRLM, and 16 GLSZM features)
- Non-radiomic imaging features extracted included number of phleboliths and number of fluid-fluid levels

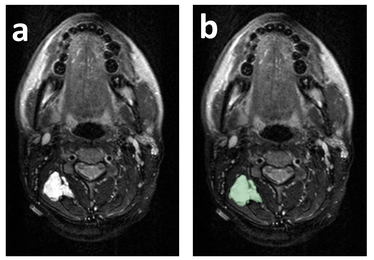


Figure 1. Example of a manually segmented venous malformation. (a) Fat-saturated T2-weighted MRI image of VM in head and neck. (b) Segmented mask (green) distinguishes VM from surrounding tissue

A machine learning model was built in Python using the mljar-supervised package[5]

- **Goal**: Predict optimal vs. suboptimal outcomes using clinical, radiomic, and non-radiomic imaging features
- **Framework**: Compete mode with unlimited time allocation
- **Validation Strategy**: Stratified 10-fold cross-validation
- **Model Search & Hyperparameter Tuning**:
  - **Random Start:** Train 10 models with randomly chosen algorithms, hyperparameters, and preprocessing steps
  - **Hill-Climbing:** Pick the best performers and, for two rounds, make incremental changes to their hyperparameters, keeping only changes that boost accuracy.
  - **Top-3 Refinement:** Applied a finer, more intensive hyperparameter search around the three strongest models from hill-climbing
- **Ensembling**: Generated stacked ensemble of top performing models

Results

Best Performing Model: Stacked Ensemble with LightGBM Meta-Learner

- **Architecture:** Base learners (e.g., XGBoost, CatBoost) each output class probabilities. These are combined into a new feature matrix, which is fed into a LightGBM classifier acting as the meta-learner. The meta-learner is trained to optimally weight and integrate the base predictions
- **Cross-Validation Performance Metrics:** All metrics were computed using pooled out-of-fold (OOF) predictions from 10 stratified folds, providing an unbiased estimate of model performance (**Table 2**)
- **Performance:** Achieved high specificity (0.92) and moderate sensitivity (0.78) (**Figure 2**)

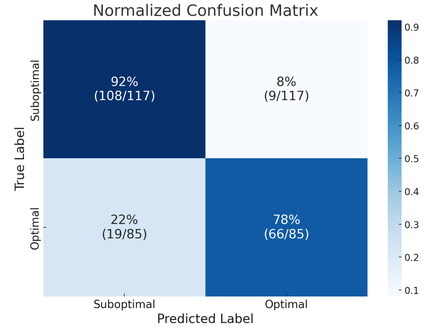


Figure 2. Normalized confusion matrix illustrating the performance of the model. The model correctly identified 92% of Suboptimal cases and 78% of Optimal cases, with respective misclassification rates of 8% and 22%.

Cross-Validated Performance of Best Performing Model					
Evaluation Metric	Accuracy	AUC	F_1-score	Sensitivity	Specificity
Value	0.86	0.85	0.83	0.78	0.92

Table 2. Cross-validated performance of the top-ranked model. Metrics—accuracy, area under the ROC curve (AUC), F_1-score, sensitivity and specificity—were calculated on the pooled out-of-fold (OOF) predictions obtained from each of the ten stratified folds.

Conclusion

A radiomics-based ML model using pre-treatment MRI may, at baseline, predict symptomatic improvement in VMs following single-session PS, potentially offering a promising tool for clinical decision-making and prognostication.

References

1. Eifert, S., et al. Prevalence of deep venous anomalies in congenital vascular malformations of venous predominance. *J. Vasc. Surg.* 31 (2000).
2. Nguyen, H. L., et al. Vascular malformations and health-related quality of life: A systematic review and meta-analysis. *JAMA Dermatol.* 154 (2018).
3. Kim, K. R. Percutaneous sclerotherapy of venous malformations. *Tech. Vasc. Interv. Radiol.* 27, 100960 (2024)
4. Fedorov A, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*. 2012;30(9).
5. Płoński P. *mljar-supervised: Automated Machine Learning for Humans*. GitHub. https://github.com/mljar/mljar-supervised. Published 2020.

Acknowledgements

- **Funding Sources**: DOD Grant Multi-Site Automated Segmentation and MultiParametric MRI Quantification Grant HT9425-23-1-0032; The Johns Hopkins Malone Center for Engineering in Healthcare