
Bayesian Trees Match Radiologist on MRI Lipoma Classification



Distinguishing atypical lipomatous tumours (ALT) from simple lipomas (SL) on magnetic resonance imaging (MRI) influences biopsy decisions, surgical planning and follow-up. While MRI is the preferred modality for lipomatous masses, day-to-day accuracy can vary, particularly for specificity. A machine learning approach using Bayesian additive regression trees (BART) applied to MRI radiomics was evaluated on a large, multi-centre cohort and compared head-to-head with an experienced musculoskeletal radiologist. Pathology served as the reference. The work also explored how results transfer across sites, how many image features are truly needed and where errors tend to arise. The findings indicate that BART can sit close to expert interpretation and offer calibrated uncertainty, suggesting a potential role as a second reader to support decisions on when to biopsy or how aggressively to manage a lesion.

Multicentre Dataset and Practical Pipeline

The analysis drew on a retrospective cohort from five centres, including patients who had pre-operative MRI and subsequent complete resection of a lipomatous mass with pathology confirming ALT or SL. Axial T1-weighted sequences without contrast on 1.5 T or 3.0 T scanners provided the common imaging ground across varied protocols. Tumour regions were manually segmented on three-dimensional T1-weighted images, then processed to reduce scanner-related intensity variation and resampled to a uniform voxel size to create consistent inputs for feature extraction.

Must Read: [AI-driven MRI Classification of Ovarian Lesions](#)

Radiomics features characterising intensity, texture and shape were computed from the original images and from standard filtered variants. This generated a rich but high-dimensional description of each lesion that captured patterns known to differ between ALT and SL, such as heterogeneity, septation and edge conspicuity. Because the data originated from multiple centres with different scanners and protocols, a harmonisation step was used to dampen site effects while preserving signal relevant to diagnosis.

BART, implemented as a probabilistic ensemble of decision trees, was trained and tuned with cross-validation. A random forest (RF) model served as a comparator. External validation used a leave-one-cohort-out approach, rotating each centre as an unseen test set to test transferability. Beyond full-feature models, the team assessed parsimony by retraining with only the top features ranked by importance, checking whether a concise subset could preserve performance. Variable importance was quantified and used to probe which imaging attributes provided the most discriminative value.

Comparable Performance to Expert Reading

Across cohorts, BART performed at a level that closely matched the experienced reader when pathology was used as the reference. Discrimination measured by receiver operating characteristics was nearly identical, and accuracy, sensitivity and specificity fell within a similar range. Harmonisation improved calibration and helped BART maintain specificity in several centres, although sensitivity could shift modestly depending on the cohort. The RF baseline trailed BART slightly, underscoring that the probabilistic, uncertainty-aware trees were a good fit for this task.

Agreement analyses reinforced this picture. BART's classifications and the expert's scores showed moderate concordance, with no significant difference in the proportion of correct calls. When validation rotated through the five centres, performance varied by site for both human and model, reflecting residual differences in imaging characteristics and case mix that persisted despite harmonisation. Even so, the separation between BART and the expert remained small, suggesting that the model can approximate expert judgement under changing conditions without requiring site-specific retraining.

A practical advantage of BART is its well-calibrated probability output. Rather than delivering a hard yes or no, it offers a posterior estimate of class membership that can be interpreted as case-level risk. In workflows where uncertainty matters—such as setting biopsy thresholds or discussing management options in multidisciplinary meetings—this probabilistic view can increase diagnostic confidence, particularly when the imaging appearance is borderline. The alignment with the expert reader supports consideration of BART as a second-reader tool to reduce equivocal findings and standardise reporting across teams.

What Drives Errors and How to Use the Tool

Reducing the feature set preserved performance, with small, carefully chosen subsets performing on par with much larger sets. The most informative features came from families that describe texture and local intensity relationships, which capture the heterogeneous, nodular or septated patterns more typical of ALT compared with the more uniform signal often seen in SL. Shape measures contributed supportively, but the discriminative core lay in how the lesion's internal signal varies across space. This parsimony helps streamline pipelines, lowers computational needs and eases future deployment.

Inspection of mismatches between predictions and pathology highlighted how imaging appearance can mislead both humans and algorithms. False positives often involved deep or large lesions where ALT is common or masses with thick septa and complex internal architecture that can also occur in SL. Conversely, ALT without pronounced heterogeneity risked being labelled as SL. Recognising these recurring patterns can help clinicians anticipate where human and model assessments might diverge and where an uncertainty-aware output should prompt closer scrutiny, additional sequences or a lower threshold for tissue sampling.

Because performance shifted across centres, harmonisation emerges as a crucial step for real-world roll-out. Site-to-site variability influenced accuracy and the balance between sensitivity and specificity, and while harmonisation mitigated this, it did not erase it. Continuous monitoring after deployment, coupled with retraining or recalibration when protocols change, will be important to maintain consistency. The path to routine use also includes automating segmentation to reduce manual effort and expanding validation beyond a single expert reader to gauge inter-reader variability and generalisability across different experience levels.

Applying BART to MRI radiomics for ALT versus SL classification delivered accuracy and discrimination comparable to an experienced musculoskeletal radiologist and ahead of a strong RF baseline. Performance held up across centres and with reduced feature sets, indicating feasibility for a second-reader role that quantifies uncertainty and supports consistent decision-making. Variability between sites and the need for automation and broader reader assessment remain as practical considerations. With careful harmonisation, parsimony in feature use and attention to deployment monitoring, this approach can help standardise interpretation, inform biopsy triage and refine surgical planning for patients with lipomatous masses.

Source: [Radiology Advances](#)

Image Credit: [iStock](#)

Published on : Sat, 25 Oct 2025