Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

MRI-Based Radiomics and Machine Learning for Predicting Refractory Outcomes in Ménière's Disease: A Multicenter Preliminary Study

yufeng zhong
The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University
Xiaochang Xiang
Xiushui County People's Hospital of Jiangxi Province
Yulong Tao
Nanchang University
Yingchuan Zhu
Nanchang University
Wen Xie
The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University
Aimei Yuan
The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University
Zijian Gong
Nanchang University
Lianggeng Gong
The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University
Ting Shu
shuting1020@163.com
The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University

Research Article

Keywords: Ménière's disease, radiomics, machine learning, gadolinium-enhanced magnetic resonance imaging (MRI), Endolymphatic hydrops, Prognosis prediction

Posted Date: October 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10396270/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objective

To evaluate the value of a machine learning model based on radiomics features derived from multicenter gadolinium-enhanced inner ear magnetic resonance imaging (MRI) for predicting refractory clinical outcomes in patients with Ménière's disease (MD).

Methods

This retrospective study enrolled patients with unilateral Ménière's disease diagnosed at two medical centers between January 2019 and December 2023. A total of 154 patients from Center 1 constituted the internal dataset and were divided into a training set (n = 123) and an internal test set (n = 31) at an 8:2 ratio, while 32 patients from Center 2 were included as an independent external validation set. According to established clinical guidelines, patients were categorized into good-control and poor-control groups. Radiomics features were extracted from the cochlea and vestibule on gadolinium-enhanced 3D-FLAIR-CUBE images, and bilateral difference (Diff) features were generated by subtracting contralateral ear features from those of the affected ear. Feature selection was performed using correlation filtering and least absolute shrinkage and selection operator (LASSO) regression. Multiple machine learning algorithms were used to construct predictive models, which were evaluated in both internal and external datasets. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, and F1 score, and clinical utility was evaluated using decision curve analysis (DCA).

Results

In both the training and internal test sets, the Diff radiomics model achieved higher AUCs than the symptomatic-ear radiomics models (training set: 0.815 vs. 0.682; internal test set: 0.730 vs. 0.650), along with improved sensitivity and F1 scores. Feature importance analysis indicated that vestibular-related radiomics features contributed more prominently to the prediction of refractory outcomes. Although model performance declined in the external validation set, the Diff model maintained high sensitivity (0.957) and F1 score. DCA demonstrated greater clinical net benefit of the Diff model in the training and internal test sets.

Conclusion

A machine learning model based on bilateral difference radiomics features can effectively predict refractory outcomes in MD and demonstrates superior performance compared with single-ear radiomics models. This approach may provide objective imaging support for prognostic assessment and individualized management of patients with MD.

Introduction

Ménière's disease (MD) is a chronic vestibular disorder characterized by endolymphatic hydrops (EH) as its underlying pathological substrate. Clinically, MD is characterized by recurrent vertigo, fluctuating hearing loss, tinnitus, and aural fullness. The diagnostic criteria have been standardized by the Bárány Society consensus [1], and management recommendations were further summarized in the AAO-HNS clinical practice guideline [2].

With the widespread application of delayed gadolinium-enhanced inner ear MRI, particularly 3D-FLAIR/related protocols, *in vivo* visualization and semi-quantification of EH have become feasible. Connor et al. systematically evaluated delayed post-gadolinium MRI descriptors and summarized their diagnostic performance across clinical MD classifications [3], while Orimoto et al. compared the diagnostic value of hydrops MRI against audiovestibular testing in a systematic review [4]. These advances have enabled radiologic phenotyping, grading, and treatment monitoring in MD.

However, imaging-defined EH is not a disease-specific or outcome-specific marker. For instance, Liu et al. reported that EH could also be detected in the asymptomatic contralateral ears of patients with unilateral MD [5]. van Steekelenburg et al. further demonstrated that although EH and perilymphatic signal intensity may facilitate MD detection, imaging overlap remains with other vertigo-associated inner ear disorders [6]. In addition, studies by Cho et al. and Xiao et al. showed that the correlation between EH grading and audiovestibular

function is inconsistent, with limited agreement across different scoring methods [7, 8]. Therefore, relying solely on unilateral absolute EH grading may be insufficient to comprehensively reflect individual disease burden and long-term clinical risk.

Most patients with MD improve with conservative treatments (e.g., dietary sodium restriction, diuretics, corticosteroids), as recommended in the AAO-HNS guideline [2]. Nevertheless, a subset progresses to poor symptom control or refractory disease and may require invasive interventions. Refractory Ménière's Disease (RMD) is defined as a clinical condition unresponsive to standard conservative management (e.g., salt restriction, diuretics, betahistine) and/or intratympanic therapies. Its associated risk factors can be categorized as follows: (1) inherent disease severity[9–11]; (2) high risk of refractoriness, such as the presence of "Tumarkin" otolithic crises [11, 12]; (3) historical treatment failure, [9, 10]; and (4) underlying pathophysiology, including endolymphatic sac dysfunction (e.g., hypoplasia or malabsorption) acting as a mechanical driver of intractable hydrops [13], as well as immune and inflammatory mechanisms may also play a significant role in some refractory cases [14, 15]. Therefore, early identification of patients at high risk for refractory outcomes remains an important, unmet need in MD management.

Radiomics enables high-throughput extraction of quantitative features to characterize signal distribution and spatial heterogeneity, providing a data-driven link between imaging phenotypes and clinical outcomes. Lambin et al. outlined radiomics as a framework for building predictive imaging biomarkers [16], and the IBSI initiative led by Zwanenburg et al. standardized radiomics feature definitions to improve reproducibility [17]. In MD, radiomics has mainly focused on disease detection rather than outcome prediction: van der Lubbe et al. performed a multicenter feasibility study using radiomics and machine learning on conventional MRI for automated MD diagnosis [18], while earlier exploratory work also evaluated radiomics-based detection on conventional MRI [19]. However, models optimized for predicting refractory outcomes remain largely unexplored.

Notably, MD is characterized by bilateral susceptibility and asymmetric, dynamic progression. Frejo et al. reviewed clinical subgroups in bilateral MD and summarized reported rates and timelines of contralateral involvement [20], and Hudson et al. quantified the long-term risk of conversion from unilateral to bilateral MD, emphasizing implications for treatment decisions [21]. Given these disease-specific characteristics, single-ear imaging may fail to capture the balance between lesion severity and contralateral compensation within an individual. Intra-individual interaural comparison therefore provides an intrinsic reference that may better reflect individualized disease dynamics. Therefore, the present multicenter study aimed to extract radiomics features from the cochlea and vestibule on gadolinium-enhanced 3D-FLAIR-CUBE MRI and to develop a machine learning model centered on bilateral difference features for predicting refractory outcomes in MD, with both internal and external validation.

Materials and methods

Patients

This multicenter retrospective study was approved by the Institutional Ethics Committee of the Second Affiliated Hospital of Nanchang University, and the requirement for informed consent was waived because of the retrospective study design. Patients who underwent gadolinium-enhanced inner ear magnetic resonance imaging (MRI) and were diagnosed with unilateral Ménière's disease (MD) between January 2019 and December 2023 were consecutively enrolled from the Second Affiliated Hospital of Nanchang University (Center 1) and Xiushui County People's Hospital of Jiangxi Province (Center 2).

Inclusion criteria

Patients were included if they met all of the following criteria:

1. A definite diagnosis of unilateral MD according to the 2020 American Academy of Otolaryngology–Head and Neck Surgery (AAO-HNS) clinical practice guideline;

2. Availability of complete 3D-FLAIR-CUBE MRI images;

3. Complete clinical follow-up data.

Exclusion criteria

Patients were excluded if they met any of the following conditions:

1. Bilateral MD;

2. Severe motion artifacts or inadequate image quality;

3. Coexisting middle ear disease, vestibular schwannoma, temporal bone trauma, enlarged vestibular aqueduct syndrome, or other congenital cochlear abnormalities;

4. Presence of severe systemic diseases.

A total of **186 patients (186 ears)** were ultimately included in the study. Among them, **154 patients from Center 1** constituted the internal dataset and were divided into a training set (n = 123) and an internal test set (n = 31) at an **8:2 ratio**, whereas **32 patients from Center 2** were included as an independent external validation set. The study workflow is illustrated in Figure 1

MRI Acquisition

All MR examinations at Center 1 were performed on a 3.0T Signa scanner (GE Healthcare, USA) using a 16-channel head coil. Gadolinium contrast (Gadopentetate Dimeglumine, Magnevist) was diluted 1:8 with saline and injected transtympanically. Patients maintained a lateral decubitus position with the injected ear facing upward for 30 minutes, and MRI was performed 24 hours later. The scan sequences included 3D-Fiesta, 3D-FLAIR-CUBE and 3D-MPRAGE, and the imaging parameters are listed in Table 1.

Table 1.Parameters of magnetic resonance imaging scan

	3D-FIESTA	3D-FLAIR-CUBE	3D-MPRAGE
Repetition time (TR), ms	4.6	6000	11.4
Echo time (TE), ms	Min	124.9	Min
Inversion time (TI), ms	–	1865	–
Flip angle (FA), °	60	Variable, min: 20°, max: 120°	15
Echo train length (ETL)	–	140	–
Field of view (FOV), cm	16	24	20
Slice thickness, mm	1.4	1.2	1.2
Gap, mm	0	0	0
Scan time	2'49"	5'39"	4'46"
Bandwidth	62.5	31.25	15.63

The same protocol was used at Center 2.

Clinical Outcome Definition

Clinical outcomes were defined according to the 2020 clinical practice guideline for the diagnosis and management of Ménière's disease. Vertigo control was classified into five categories based on the vertigo control score:
Class A, 0 points (complete control);
Class B, 1–40 points (substantial control);
Class C, 41–80 points (limited control);
Class D, 81–120 points (insufficient control); and
Class E, >120 points (worsening).

The vertigo control score was calculated as follows:

$$\text{Score} = \left(\frac{\text{Number of vertigo episodes during 1-year follow-up}}{\text{Number of episodes in the year prior to treatment}} \right) \times 100$$

In the present study, patients classified as Class A or B were assigned to the good-control group, whereas those classified as Class C–E were defined as having poor control (refractory group). The following clinical information was also collected, including patient age, disease duration, and degree of hearing impairment.

Radiomics Feature Extraction , Feature Selection, and Model Construction

All MRI images were exported in Digital Imaging and Communications in Medicine (DICOM) format and resampled to isotropic voxels of 1 × 1 × 1 mm^3. Manual segmentation of the cochlea and vestibule was performed on gadolinium-enhanced 3D-FLAIR-CUBE images using ITK-SNAP software (version 4.0.1) by a radiologist with 8 years of experience in head and neck imaging and was completed under the supervision of another radiologist with 12 years of experience in head and neck imaging.

Radiomics features were extracted using the PyRadiomics library. Extracted features included first-order statistics, shape features, texture features, and wavelet-transformed features. Feature normalization and discretization were performed using a fixed bin width of 25. Texture features were derived from gray-level co-occurrence matrix (GLCM), gray-level dependence matrix (GLDM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighboring gray tone difference matrix (NGTDM).

To capture intra-individual asymmetry, bilateral difference (Diff) features were calculated by subtracting radiomics features of the contralateral ear from those of the symptomatic ear.

Feature selection was conducted in a stepwise manner. First, all features were standardized using Z-score normalization. Second, Pearson correlation analysis was applied to remove highly correlated features (correlation coefficient $|r| > 0.95$). Finally, least absolute shrinkage and selection operator (LASSO) regression was used to identify the most predictive features.

Based on the selected features, multiple machine learning algorithms were employed to construct predictive models. Among them, logistic regression (LR) demonstrated the most stable performance and was therefore selected as the primary model for further analysis .

Predictive models were constructed using multiple machine learning algorithms, including logistic regression (LR), support vector machine (SVM), AdaBoost, multilayer perceptron (MLP), decision tree, extra trees, gradient boosting, k-nearest neighbors (KNN), Light Gradient Boosting Machine (LightGBM), random forest, and extreme gradient boosting (XGBoost). All models were trained using the training dataset and subsequently evaluated in the internal test set and the independent external validation set. In addition, a multivariable logistic regression model was developed by combining radiomics features derived from different feature sets.

Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, and F1 score. Comparisons between AUCs were performed using DeLong's test. Calibration curves were generated to assess the agreement between predicted probabilities and observed outcomes, and decision curve analysis (DCA) was conducted to evaluate the clinical utility of the models across a range of threshold probabilities.

Statistical Analysis

Statistical analyses were performed using Python (version 3.7) and R software (version 4.3.2). A two-sided P value < 0.05 was considered statistically significant. Normality of continuous variables was assessed using the Kolmogorov–Smirnov test. Normally distributed variables were reported as mean ± standard deviation (SD), whereas non-normally distributed variables were expressed as median with interquartile range (IQR). Between-group comparisons were conducted using the independent-samples t test or the Mann–Whitney U test, as appropriate.

Results

A total of 186 patients with unilateral MD were included. Of these, 154 patients constituted the internal dataset and 32 patients formed the external validation set. Patients were classified into good-control and poor-control (refractory) groups based on vertigo control during follow-up. Baseline clinical characteristics did not differ significantly between groups in either dataset Table 2.

Table 2
Baseline characteristics of patients in the internal dataset and external validation set

Characteristics	Internal dataset (n = 154)	Refractory (n = 129)	Non-refractory (n = 25)	χ^2 / Z value	P value	External validation set (n = 32)	Refractory (n = 27)	Non-refractory (n = 5)	χ^2 / Z value	P value
Gender, n (%)				0.172	0.679				0.000	1.000
Male	59 (38.3)	48 (81.4)	11 (18.6)			16 (50.0)	14 (87.5)	2 (12.5)		
Female	95 (61.7)	81 (85.3)	14 (14.7)			16 (50.0)	13 (81.3)	3 (18.8)		
Age (years)		56.00 (48.00–67.00)	61.00 (51.00–65.00)	−0.703	0.483		63.63 ± 11.21	63.17 ± 9.04	0.094	0.926
Affected side, n (%)				0.033	0.855				0.674	0.412
Left		85 (65.9)	16 (64.0)				13 (48.1)	4 (66.7)		
Right		44 (34.1)	9 (36.0)				14 (51.9)	2 (33.3)		

Feature Selection

From the gadolinium-enhanced 3D-FLAIR-CUBE images, radiomics features were extracted from the cochlear and vestibular regions for both the symptomatic ear and the bilateral difference (Diff) feature sets, with 1,702 features initially generated for each feature set. After removal of highly correlated features using Pearson correlation analysis, 731 symptomatic-ear features and 785 Diff features were retained for subsequent analysis.

Further feature reduction was performed using least absolute shrinkage and selection operator (LASSO) regression, resulting in 13 optimal features for each model.

Analysis of feature weights demonstrated that vestibular-derived radiomics features accounted for a larger proportion of the selected features in the Diff model, suggesting that vestibular imaging phenotypes may provide greater discriminative value for predicting refractory Ménière's disease. The selected features and their corresponding weights are presented in Fig. 2.

Model Construction and Comparison

Based on the selected optimal feature sets, predictive models were constructed using multiple machine learning algorithms. Among the symptomatic-ear radiomics models, logistic regression (LR) demonstrated the best overall performance, with AUCs of 0.682 in the training set, 0.650 in the internal test set, and 0.696 in the external validation set. Similarly, for the Diff radiomics models, LR achieved the highest performance, yielding AUCs of 0.815, 0.730, and 0.600 in the training, internal test, and external validation sets, respectively. Therefore, LR was selected as the primary model for subsequent analyses and combined model construction.

In the training set, the Diff radiomics model outperformed the symptomatic-ear model, as evidenced by a higher AUC (0.815 vs. 0.682), along with improved sensitivity and F1 score. Consistent results were observed in the internal test set, where the Diff model again achieved a higher AUC (0.730 vs. 0.650). Detailed performance metrics of all models are summarized in Table 3.

Internal and External Validation Performance

In the internal test set, the Diff radiomics model achieved higher accuracy, sensitivity, and F1 score than the symptomatic-ear radiomics model, while specificity was comparable between the two models. These results indicate that the Diff model demonstrated superior overall discriminative ability for identifying patients with refractory Ménière's disease.

In the external validation cohort, the AUCs of all models decreased compared with those observed in the internal validation. The AUC of the symptomatic-ear radiomics model was 0.696, whereas that of the Diff model was 0.600. Although the Diff model showed a slightly

lower AUC in the external validation set, it maintained high sensitivity (0.957) and F1 score, suggesting potential utility for screening patients at high risk of refractory disease.

Overall, the Diff radiomics model achieved higher AUCs than the symptomatic-ear model in both the training set (AUC: 0.815 vs. 0.682, $P = 0.103$) and the internal test set (AUC: 0.730 vs. 0.650, $P = 0.660$), whereas its performance was slightly inferior in the external validation set (AUC: 0.600 vs. 0.696, $P = 0.233$). Across all three datasets, the Diff model consistently demonstrated higher accuracy, sensitivity, and F1 score, with comparable specificity relative to the symptomatic-ear radiomics model.

Table 3
Predictive performance of different models in the training set ,Internal test set and External validation set

Model	LR injured	LR diff	Combined
Training set			
AUC	0.682	0.815	0.816
(95% CI)	(0.544–0.820)	(0.711–0.910)	(0.715–0.910)
Cut-off	0.422	0.348	0.691
Accuracy	0.784	0.845	0.794
Sensitivity	0.844	0.896	0.818
Specificity	0.550	0.650	0.700
Precision	0.878	0.908	0.913
Recall	0.844	0.896	0.818
F1 score	0.861	0.902	0.863
Internal test set			
AUC	0.650	0.730	0.730
(95% CI)	(0.322–0.970)	(0.431–1.000)	(0.431–1.000)
Cut-off	0.453	0.499	0.798
Accuracy	0.760	0.800	0.800
Sensitivity	0.800	0.850	0.850
Specificity	0.600	0.600	0.600
Precision	0.889	0.895	0.895
Recall	0.800	0.850	0.850
F1 score	0.842	0.872	0.872
External validation set			
AUC	0.696	0.600	0.626
(95% CI)	(0.464–0.920)	(0.350–0.850)	(0.370–0.882)
Cut-off	0.454	0.277	0.545
Accuracy	0.788	0.788	0.818
Sensitivity	0.870	0.957	1.000
Specificity	0.600	0.400	0.400
Precision	0.833	0.786	0.793
Recall	0.870	0.957	1.000
F1 score	0.851	0.863	0.885

Combined Model Performance Evaluation

We further evaluated a combined model integrating the symptomatic-ear radiomics model and the Diff radiomics model. As shown in Figs. 3, the combined model achieved AUCs of 0.816, 0.733, and 0.626 in the training set, internal test set, and external validation set, respectively. Although these values were slightly higher than those of the Diff model alone, the differences did not reach statistical significance ($P > 0.05$).

Decision curve analysis demonstrated that, in both the training set and the internal test set, the Diff radiomics model and the combined model yielded a greater clinical net benefit across a wide range of threshold probabilities compared with the symptomatic-ear radiomics model. In the external validation cohort, the net benefit curves of all models tended to converge, indicating comparable clinical utility among the models (Fig. 4).

Discussion

In this multicenter study, we developed and validated a radiomics-based machine learning model using gadolinium-enhanced 3D-FLAIR-CUBE MRI to predict refractory outcomes in patients with MD. The main findings were: (1) radiomics models based on bilateral difference (Diff) features showed overall superior predictive performance compared with single-ear models in internal datasets; (2) vestibular-derived radiomics features contributed more prominently; and (3) although discrimination declined in the external cohort, the Diff model retained high sensitivity and F1 score, supporting potential value for high-risk screening.

Our results demonstrate that the Diff radiomics model consistently outperforms both the training set and internal test set in terms of AUC difference, accuracy, sensitivity, and F1-score, indicating that Diff features alone may provide superior prediction of treatment resistance risk compared to symptomatic-ear model. This finding is consistent with a growing body of hydrops-MRI evidence indicating that EH is neither strictly disease-specific nor linearly prognostic. Liu et al. reported EH in asymptomatic contralateral ears in unilateral MD[5]. van Steekelenburg et al. evaluated EH and perilymph signal intensity in a diagnostically heterogeneous cohort and demonstrated added value for MD diagnosis while illustrating overlaps with other vertigo-associated pathology[6]. In VM/MD differentiation, Oh et al. compared EH characteristics and neurotologic testing among vestibular migraine(VM), MD, and VM-MD overlap groups, reinforcing that EH patterns alone do not map one-to-one to symptom phenotypes[22]. More recently, Bernaerts et al. further explored hydrops MRI for differentiating MD from vestibular migraine in clinically ambiguous cases[23]. Importantly, studies assessing the relationship between EH grade and audiovestibular measures have yielded mixed results: Cho et al. examined correlations between IV-Gd EH grades and audiovestibular tests[7], and Xiao et al. compared common EH scoring systems and their association with auditory functions, highlighting variability across methods and cohorts[8]. These observations collectively weaken the prognostic specificity of absolute unilateral EH grading and support the rationale for an asymmetry-centered strategy.

Diff features use the contralateral ear as an internal reference, potentially reducing inter-individual variability (e.g., anatomy, contrast kinetics) and emphasizing disease-related asymmetry rather than absolute intensity. This concept aligns with quantitative/descriptor-based hydrops MRI work that has focused on robust descriptors and their reliability across classifications[3]and supports our internal performance gains.

Vestibular-derived radiomics features accounted for a greater proportion of selected features in the Diff model. This is clinically plausible because vertigo control is central to defining refractory MD. Multiple hydrops-MRI studies emphasize vestibular compartment behavior (e.g., saccule/utricle changes) as clinically informative. Attyé et al. introduced and evaluated the saccule-to-utricle ratio inversion (SURI) concept, highlighting saccular analysis as crucial in MD imaging assessment[24]. Deng et al. compared saccule-to-utricle ratio patterns between early- and late-stage MD, further supporting stage-dependent vestibular compartment alterations[25]. In addition, non-contrast approaches targeting vestibular anatomy have emerged: Venkatasamy et al. reported that saccular abnormalities on fast 3D T2-weighted sequences could assist diagnosis without contrast administration[26], and Fukutomi et al. demonstrated that optimized non-injected 3D-FLAIR parameters can visualize the saccule and utricle, paving the way for contrast-free inner-ear phenotyping[27]. Together, these studies provide biological plausibility for why vestibular-region heterogeneity captured by radiomics may be informative for refractory-risk modeling.

In the external validation cohort, the model demonstrated a decrease in AUC compared with the internal validation cohort, a phenomenon that is not uncommon in multicenter MRI-based radiomics studies. Radiomic features are susceptible to variations in scanner vendors, acquisition parameters, spatial resolution, image reconstruction, preprocessing procedures, and feature extraction implementations, all of which may impair the cross-center generalizability of radiomics models[17, 28, 29]. Traverso et al. emphasized that the repeatability and reproducibility of radiomic features remain major challenges limiting their clinical translation[28]. Furthermore, an MRI phantom study by Lee et al. demonstrated that variations in scanning parameters and scanner platforms could substantially alter certain MRI-based radiomic features[29]. Zwanenburg et al., through the Image Biomarker Standardisation Initiative (IBSI), provided standardized definitions and reference values for commonly used radiomic features, further highlighting the importance of standardized feature extraction workflows[17]. Therefore, the reduced external performance observed in our study more likely reflects real-world inter-center heterogeneity in imaging acquisition, feature stability, and patient composition, rather than simple model failure. Future studies should further validate the proposed model in larger prospective multicenter cohorts and improve model robustness and clinical

applicability through standardized MRI acquisition protocols, consistent image preprocessing procedures, rigorous feature stability selection, and standardized reporting practices.Notably, despite reduced AUC externally, our Diff model maintained high sensitivity and F1 score, indicating suitability for screening/risk-stratification rather than standalone diagnosis. This positioning is consistent with how hydrops MRI is frequently framed as adjunctive and descriptor-dependent in clinical practice[3, 32].

The combined model integrating symptomatic-ear and Diff features yielded only marginal internal AUC gains without clear external advantage, suggesting the Diff model may already capture principal discriminative information. This underscores a key principle in radiomics: biologically aligned feature design often matters more than indiscriminately increasing complexity[16, 17]. For transparent reporting and reproducibility, TRIPOD + AI provides updated guidance for ML-based prediction model studies[31].

Limitations include retrospective design, limited sample size, manual segmentation variability, relatively small external cohort, and lack of systematic multimodal clinical variables. Future work should emphasize prospective multicenter validation, automated segmentation/quantification (e.g., deep learning-based hydrops measurement methods[32, 33]), protocol standardization, harmonization, and multimodal integration.

Conclusion

Machine learning models based on bilateral difference radiomics features provide a noninvasive and reproducible approach for objective prognostic assessment of refractory MD. Further validation in large prospective multicenter cohorts is required to clarify clinical applicability.

Abbreviations

3D-FLAIR-CUBE: Three-dimensional fluid-attenuated inversion recovery CUBE

3D-FIESTA: Three-dimensional fast imaging employing steady-state acquisition

3D-MPRAGE: Three-dimensional magnetization-prepared rapid acquisition gradient echo

AAO-HNS: American Academy of Otolaryngology–Head and Neck Surgery

AUC: Area under the receiver operating characteristic curve

CI: Confidence interval

DCA: Decision curve analysis

Diff: Difference between the symptomatic and contralateral ears

DICOM: Digital Imaging and Communications in Medicine

EH: Endolymphatic hydrops

ETL: Echo train length

FOV: Field of view

FA: Flip angle

GLCM: Gray-level co-occurrence matrix

GLDM: Gray-level dependence matrix

GLRLM: Gray-level run-length matrix

GLSZM: Gray-level size zone matrix

IBSI: Image Biomarker Standardisation Initiative

IQR: Interquartile range

ITK-SNAP: Insight Segmentation and Registration Toolkit

KNN: K-nearest neighbors

LASSO: Least absolute shrinkage and selection operator

LR: Logistic regression

LightGBM: Light Gradient Boosting Machine

MD: Ménière's disease

ML: Machine learning

MLP: Multilayer perceptron

MRI: Magnetic resonance imaging

NGTDM: Neighboring gray-tone difference matrix

ROC: Receiver operating characteristic

SVM: Support vector machine

SD: Standard deviation

TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis plus Artificial Intelligence

TR: Repetition time

TE: Echo time

TI: Inversion time

VM: Vestibular migraine

XGBoost: Extreme Gradient Boosting

Declarations

Ethics approval and consent to participate

This multicenter retrospective study was approved by the Institutional Ethics Committee of the Second Affiliated Hospital of Nanchang University. The study included retrospective data from the Second Affiliated Hospital of Nanchang University and Xiushui County People's Hospital of Jiangxi Province under the same approved multicenter study protocol. All data included in this study were collected before 2023. At the time of ethics approval, the ethics approval documents issued by the Institutional Ethics Committee confirmed ethical approval but did not include a formal ethics approval reference number. The study was conducted in accordance with the principles of the Declaration of Helsinki. Because of the retrospective nature of the study, the requirement for informed consent to participate was waived by the Institutional Ethics Committee of the Second Affiliated Hospital of Nanchang University .

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because they contain patient information and are subject to institutional privacy regulations. De-identified data may be made available from the corresponding authors upon

reasonable request and with permission from the participating institutions.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceptualization: Ting Shu, Lianggeng Gong. Methodology: Ting Shu, Lianggeng Gong. Data curation: Yufeng Zhong, Xiaochang Xiang, Yulong Tao, Yingchuan Zhu, Wen Xie, Aimei Yuan. Formal analysis: Yufeng Zhong, Ting Shu, Aimei Yuan. Writing – original draft: Yufeng Zhong, Ting Shu. Writing – review & editing: Ting Shu. All authors read and approved the final manuscript.

Acknowledgements

The authors sincerely thank all clinicians, radiologists, and technical staff from the Second Affiliated Hospital of Nanchang University and Xiushui County People's Hospital of Jiangxi Province for their support in patient recruitment, image acquisition, and data management.

Clinical trial number

Not applicable.

Funding

This work was supported by the Jiangxi Provincial Health Commission (Grant No. 202610325), 'Exploration of a Precision Diagnosis and Prognostic Model for Ménière's Disease Based on Multimodal Gadolinium-Enhanced Endolymphatic Imaging.'

Nanchang University Scientific Research Project (2025) "Exploration of constructing a precise diagnostic and prognostic model for Ménière's disease based on multimodal gadolinium contrast membrane labyrinth imaging"

Conflict of interest: The authors declare that they have no competing interests.

Author contributions (CRediT)

Conceptualization: Ting Shu, Lianggeng Gong. Methodology: Ting Shu, Lianggeng Gong. Data curation: Yufeng Zhong, Xiaochang Xiang, Yulong Tao, Yingchuan Zhu, Wen Xie, Aimei Yuan. Formal analysis: Yufeng Zhong, Ting Shu, Aimei Yuan. Writing – original draft: Yufeng Zhong, Ting Shu. Writing – review & editing: Ting Shu. All authors read and approved the final manuscript.

References

1. Lopez-Escamez JA, Carey J, Chung WH. Diagnostic criteria for Ménière's disease. J Vestib Res. 2015; https://doi.org/10.3233/VES-150549
2. Basura GJ, Adams ME, Monfared A. Clinical practice guideline: Ménière's disease. Otolaryngol Head Neck Surg. 2020; https://doi.org/10.1177/0194599820909438
3. Connor SEJ, Grzeda MT, Jamshidi B. Delayed post-gadolinium MRI descriptors for Ménière's disease: a systematic review and meta-analysis. Eur Radiol. 2023; https://doi.org/10.1007/s00330-023-09651-8
4. Orimoto KY, Kariya S, Cureoglu S. Diagnostic value of endolymphatic hydrops MRI in Ménière's disease: a systematic review. Clin Radiol. 2022; https://doi.org/10.1016/j.crad.2022.05.012
5. Liu Y, Dai C, Liu B. Endolymphatic hydrops detected by 3D-FLAIR MRI in asymptomatic contralateral ears of patients with unilateral Ménière's disease. AJNR Am J Neuroradiol. 2015; https://doi.org/10.3174/ajnr.A4278
6. van Steekelenburg JM, van Weijnen A, de Pont LMH. Value of endolymphatic hydrops and perilymph signal intensity in suspected Ménière disease and other vertigo-associated pathology. AJNR Am J Neuroradiol. 2020; https://doi.org/10.3174/ajnr.A6413
7. Cho YS, Ahn JM, Lee HS. Usefulness of intravenous gadolinium-enhanced inner ear MRI in Ménière's disease: correlation of endolymphatic hydrops with audiovestibular tests. Sci Rep. 2018; https://doi.org/10.1038/s41598-018-34557-7
8. Xiao H, Xu X, Zhou Y. Comparison of common endolymphatic hydrops scoring methods and their associations with auditory function in Ménière's disease. Front Neurol. 2022; https://doi.org/10.3389/fneur.2022.907628

9. Lee SA, Lee YJ, Lee JD. Endoscopic Transcanal Labyrinthectomy for Intractable Meniere's Disease: An Alternative to Transmastoid Labyrinthectomy? Otol Neurotol. 2022;43:113–9. https://doi.org/10.1097/MAO.0000000000003355
10. Scarpa A, Ralli M, De Bonis E, Troisi D, Montanino A, Viola P, et al. Pharmacological, Surgical and Diagnostic Innovations in Meniere's Disease: A Review. Transl Med UniSa. 2020;23:48–52. https://doi.org/10.37825/2239-9747.1009
11. Kutlubaev MA, Xu Y, Manchaiah V, Zou J, Pyykkö I. Vestibular drop attacks in Ménière's disease: A systematic review and meta-analysis of frequency, correlates and consequences. J Vestib Res. 2022;32:171–82. https://doi.org/10.3233/VES-201514
12. Zhang D, Lyu Y, Fan Z, Wang H. Semicircular canal and vestibular plugging in patients with Meniere's disease: a preliminary study. Acta Otolaryngol. 2024;144:350–4. https://doi.org/10.1080/00016489.2024.2405004
13. Paparella MM. Endolymphatic sac revision for recurrent intractable Meniere's disease. Otolaryngol Clin North Am. 2006;39:713–21, vi. https://doi.org/10.1016/j.otc.2006.03.001
14. Ito T, Inui H, Miyasaka T. Three-Dimensional Magnetic Resonance Imaging Reveals the Relationship Between the Control of Vertigo and Decreases in Endolymphatic Hydrops After Endolymphatic Sac Drainage With Steroids for Meniere's Disease - PubMed [Internet]. [cited 2026 May 16]. https://doi.org/10.3389/fneur.2019.00046
15. Peng A, Hu J, Wang Q, Pan X, Zhang Z, Jiang W, et al. A comparison of endolymphatic duct blockage, endolymphatic sac drainage and endolymphatic sac decompression surgery in reversing endolymphatic hydrops in Meniere's disease. J Otolaryngol Head Neck Surg. 2021;50:70. https://doi.org/10.1186/s40463-021-00545-7
16. Lambin P, Rios-Velazquez E, Leijenaar R. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017; https://doi.org/10.1038/nrclinonc.2017.141
17. Zwanenburg A, Vallières M, Abdalah MA. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020; https://doi.org/10.1148/radiol.2020191145
18. van der Lubbe MFJA, Steens SCA, van den Burg EL. Non-invasive automated diagnosis of Ménière's disease using radiomics and machine learning on conventional MRI. Radiol Med. 2022; https://doi.org/10.1007/s11547-021-01447-6
19. van den Burg EL, van Hoof M, Postma AA. Exploratory radiomics and machine learning for detection of Ménière's disease on conventional MRI. Front Neurol. 2016; https://doi.org/10.3389/fneur.2016.00190
20. Frejo L, Soto-Varela A, Santos-Perez S. Clinical subgroups in bilateral Ménière's disease. Front Neurol. 2016; https://doi.org/10.3389/fneur.2016.00182
21. Hudson TJ, Tan LSH, Phillips V. Long-term risk of progression from unilateral to bilateral Ménière's disease: a systematic review and meta-analysis. Otol Neurotol. 2025; https://doi.org/10.1097/MAO.0000000000004491
22. Oh SY, Kim JS, Dieterich M. Endolymphatic hydrops in vestibular migraine and Ménière's disease. Front Neurol. 2021; https://doi.org/10.3389/fneur.2021.699712
23. Bernaerts A, Sluydts M, Liégeois V. Role of hydrops MRI in differentiating between Ménière's disease and vestibular migraine: a prospective study. Front Neurol. 2025; https://doi.org/10.3389/fneur.2025.1582754
24. Attyé A, Eliezer M, Boudiaf N. MRI of endolymphatic hydrops in Ménière's disease: saccule-to-utricle area ratio inversion. Radiology. 2017; https://doi.org/10.1148/radiol.2017162000
25. Deng W, Wu H, Chen Y. Comparing the saccule-to-utricle ratio in early- versus late-stage Ménière's disease. Laryngoscope. 2024; https://doi.org/10.1002/lary.31655
26. Venkatasamy A, Eliezer M, Boudiaf N. Imaging of the saccule for the diagnosis of endolymphatic hydrops using non-contrast 3D T2-weighted MRI. Eur Radiol. 2017; https://doi.org/10.1007/s00330-016-4444-8
27. Fukutomi H, Naganawa S, Kawai H. Visualization of the saccule and utricle using non-contrast-enhanced inner ear MRI. Eur Radiol. 2022; https://doi.org/10.1007/s00330-021-08249-7
28. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys. 2018; https://doi.org/10.1016/j.ijrobp.2018.05.053
29. Lee J, Steinmann A, Ding Y. Radiomics feature robustness as measured using an MRI phantom. Sci Rep. 2021; https://doi.org/10.1038/s41598-021-83593-3
30. Connor SEJ. Endolymphatic hydrops MRI in Ménière's disease: practical considerations. Clin Radiol. 2021; https://doi.org/10.1016/j.crad.2021.04.010
31. Collins GS, Moons KGM, Dhiman P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;

32. Cho YS, Ahn JM, Lee HS. Automated measurement of endolymphatic hydrops on MRI using deep learning. Sci Rep. 2020; https://doi.org/10.1038/s41598-020-75418-1
33. Ahmadi SA, Glaister J, Abderezaei J. Automated inner ear segmentation in MRI using deep learning. Med Image Anal. 2022; https://doi.org/10.1016/j.media.2021.102280

Figures

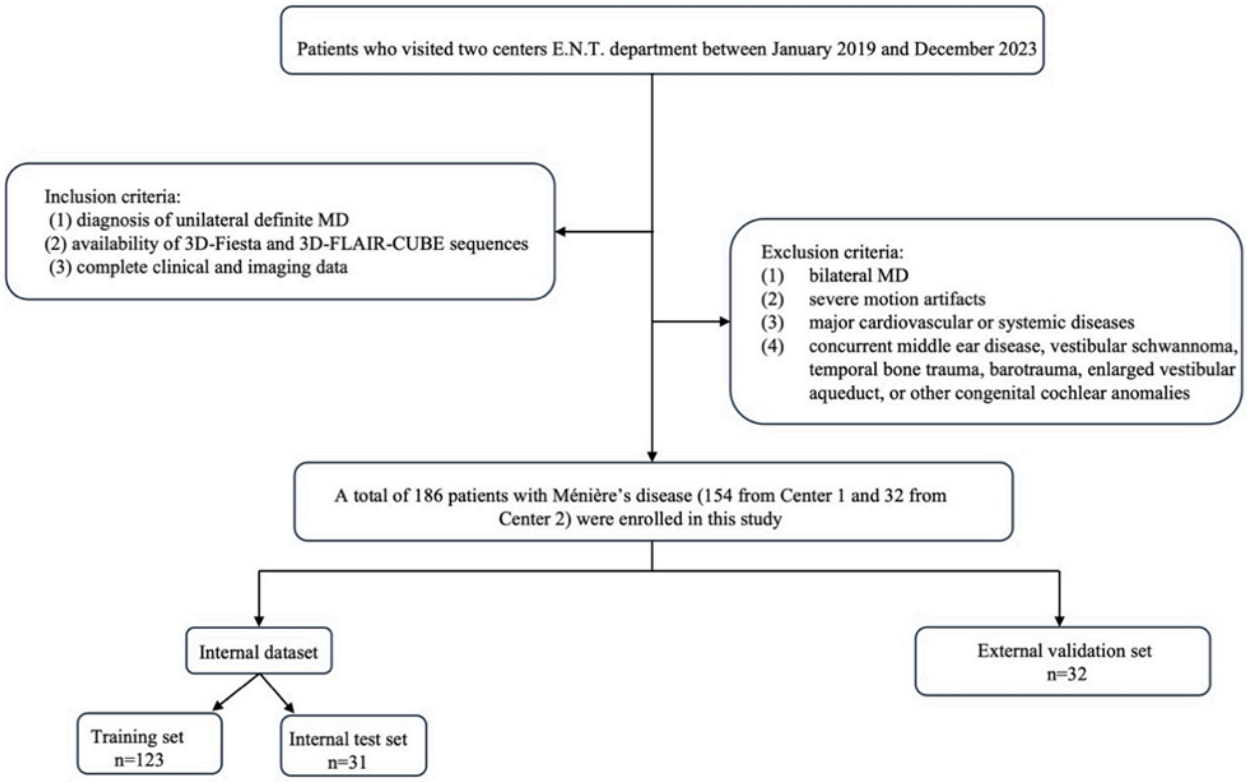

Figure 1

Study flow diagram of patient inclusion and dataset allocation

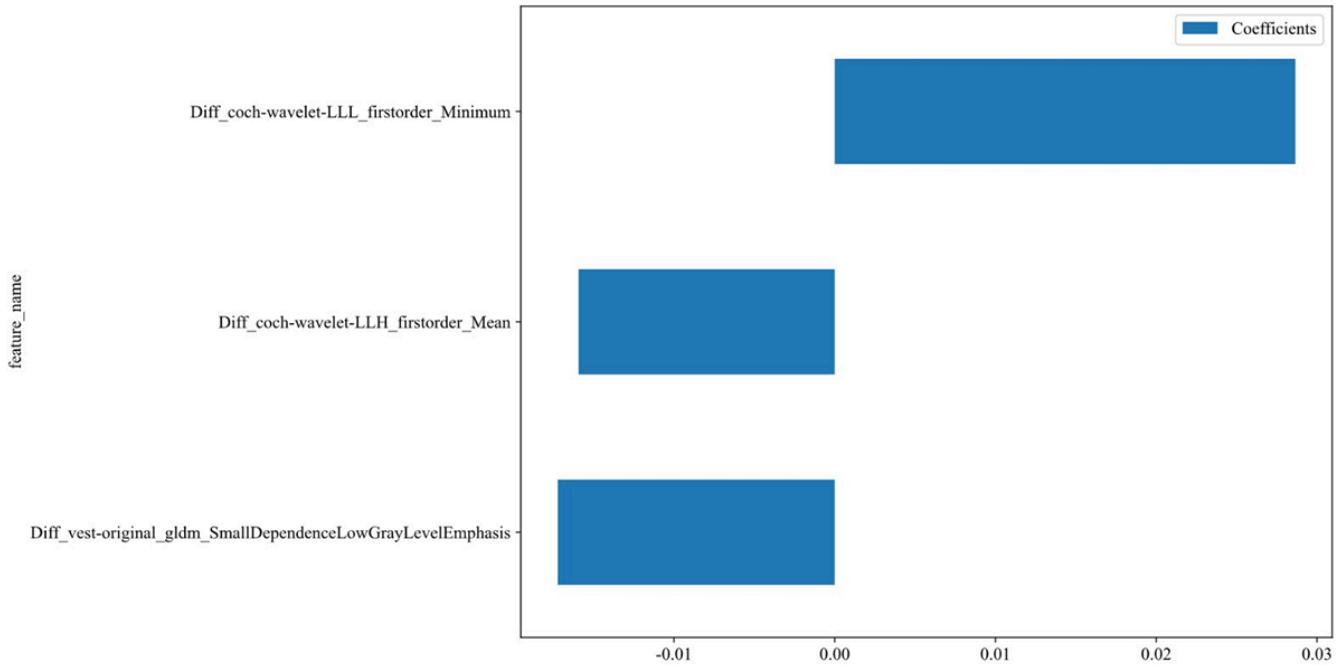

Figure 2

Remaining features and their corresponding weights

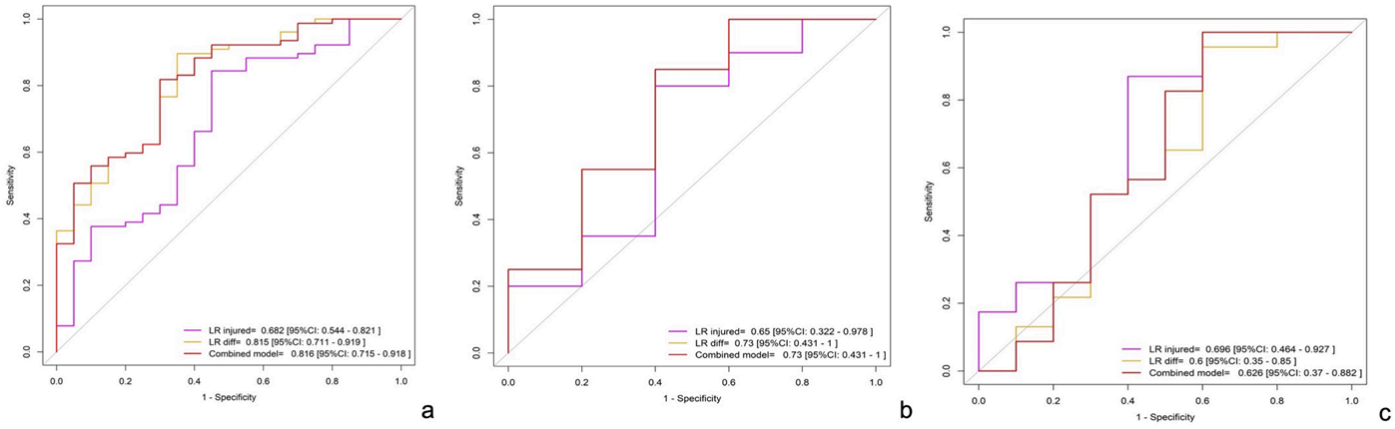

Figure 3

Receiver operating characteristic (ROC) curves of the radiomics nomograms in (a) the training set; (b) the internal test set; (c) the external test set

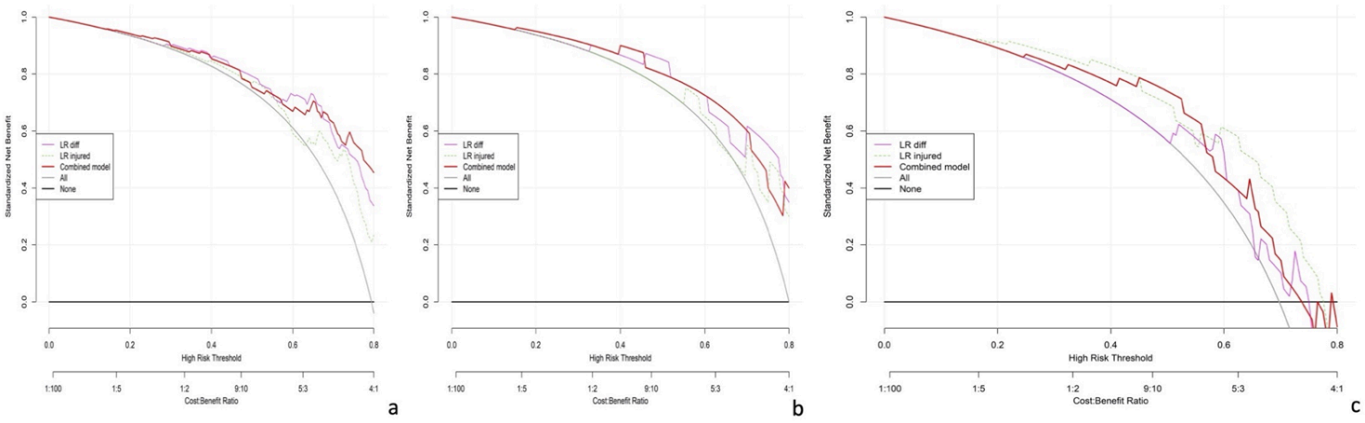

Figure 4

Decision curve analysis(DCA) with (a) the training set; (b) the internal test set; (c) the external test set