

Radiomics Based on mDixon for Predicting the Severity of Primary Lower Limb Lymphedema

Xingpeng Li

Beijing Tiantan Hospital, Capital Medical University

Yimeng Zhang

Sichuan Provincial People's Hospital

Jie Ren

Beijing Shijitan Hospital

Kun Hao

Beijing Shijitan Hospital

Mengke Liu

Shandong Provincial Hospital, Shandong First Medical University

Jia Guo

Beijing Shijitan Hospital

Yuming Yang

Beijing Shijitan Hospital

Rengui Wang

Beijing Shijitan Hospital

Yunlong Yue

yueyunlong@bjshjth.cn

Beijing Shijitan Hospital

Research Article

Keywords: Lymphedema, Prediction, Magnetic resonance imaging, Dixon, Radiomics

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10905618/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Title Page:

Title: Radiomics Based on mDixon for Predicting the Severity of Primary Lower Limb

Lymphedema

Article type: original article

All authors: Xingpeng Li, Yimeng Zhang, Jie Ren, Kun Hao, Mengke Liu, JiaGuo, Yuming Yang,

Rengui Wang, Yunlong Yue

Affiliation:

Corresponding author: Yunlong Yue

Academic degree of Yunlong Yue: Doctor of Medicine

Correspondence to Yunlong Yue: Department of Medical Imaging, Beijing Shijitan Hospital,

Capital Medical University

Address: Yangfangdian Tieyiyuan Road No.10, Haidian District, Beijing City, China,100038

Tel: 86-010-63926133

Fax: 86-010-66953359

E-mail: yueyunlong@bjsjth.cn

First author: Xingpeng Li

Academic degree of Xingpeng Li: Doctor of Medicine

Institution of Xingpeng Li: Department of Radiology, Beijing Tiantan Hospital, Capital Medical

University, China

Address: No. 119 South Fourth Ring West Road, Fengtai District, Beijing City, China, 100070

E-mail: lxp17634986228@163.com

Other authors:

Yimeng Zhang

Academic degree of Yimeng Zhang: Doctor of Medicine

Institution of Yimeng Zhang: Department of Radiology, Sichuan Provincial People's Hospital,

China

Address: No. 32, West Section 2, First Ring Road, Qingyang District, Chengdu City, China,

610072

E-mail: zym12151115@163.com

Jie Ren

Institution of Jie Ren: Department of Medical Imaging, Beijing Shijitan Hospital, Capital Medical

University, China

Academic degree of Jie Ren: Bachelor's degree in Medicine

Address: Yangfangdian Tieyiyuan Road No.10, Haidian District, Beijing City, China,100038

E-mail: renjie3324@163.com

Kun Hao

Academic degree of Kun Hao: Doctor of Medicine

Institution of Kun Hao: Department of Lymph Surgery, Beijing Shijitan Hospital,

Capital Medical University, China

Address: Yangfangdian Tieyiyuan Road No.10, Haidian District, Beijing City, China,100038

E-mail: haokun3056@bjsjth.cn

Mengke Liu

Academic degree of Mengke Liu: Doctor of Medicine

Institution of Mengke Liu: Department of Radiology, Shandong Provincial Hospital, Shandong

First Medical University, China

Address: No. 324, Jingwu Road and Weiqi Road, Huaiyin District, Jinan City, Shandong, China,

250021

E-mail: 805587047@qq.com

Jia Guo

Academic degree of Jia Guo: Master of Medicine

Institution of Jia Guo: Department of Medical Imaging, Beijing Shijitan Hospital, Capital

Medical University, China

Address: Yangfangdian Tieyiyuan Road No.10, Haidian District, Beijing City, China,100038

E-mail: 649489127@qq.com

Yuming Yang

Institution of Yuming Yang: Department of Medical Imaging, Beijing Shijitan Hospital, Capital

Medical University, China

Academic degree of Yuming Yang: Doctor degree of Medicine

Address: Yangfangdian Tieyiyuan Road No.10, Haidian District, Beijing City, China,100038

E-mail: 1079634403@qq.com

Rengui Wang

Academic degree of Rengui Wang: Doctor of Medicine

Institution of Rengui Wang: Department of Medical Imaging, Beijing Shijitan Hospital,

Capital Medical University, China

Address: Yangfangdian Tieyiyuan Road No.10, Haidian District, Beijing City, China,100038

E-mail: wangrg@bjsjth.cn

Radiomics Based on mDixon for Predicting the Severity of Primary Lower Limb Lymphedema

Purpose: This study aimed to develop a model based on mDixon combined with radiomics to predict the severity of primary lower limb lymphedema (PLE).

Methods: A retrospective analysis was conducted on 172 patients with PLE from November 2019 to July 2023. According to the classification criteria outlined in the International Society of Lymphology (ISL) guidelines, the patients were divided into a non-severe group (mild, moderate) and a severe group. The semiautomatic thresholding method in ITK snap software was used to delineate four images of mDixon (including in phase, out of phase, water-only phase and fat-only phase), filling the area between the skin and muscle. The PyRadiomics software package was used to extract the radiomics features of four images separately, and integrate the radiomics features of the in phase and water-only phase images, defined as fusion phase (FP). Five machine learning models: Logistic Regression, Support Vector Machine, Random Forest, ExtraTrees and Light Gradient Boosting Machine were used. The predictive performance was evaluated via receiver operating characteristic curve curves, and the sensitivity and specificity of these measures were calculated. Decision curve analysis was used to evaluate the performance of the predictive model in guiding decisions for nonsevere and severe lymphedema patients.

Results: The ExtraTrees model of fusion phase has the best results among all models. The AUC of the training set was 0.899 (95% CI: 0.846-0.953), the accuracy was 0.817, the sensitivity was 0.808, and the specificity was 0.824. The corresponding values for the test set were 0.907 (95% CI: 0.824-0.990), 0.846, 0.760 and 0.926, respectively. The decision curve revealed that when the predicted probability was between 18% and 58%, the net benefit of the ExtraTrees model was greater than that

of the two extreme curves.

Conclusion: This study successfully developed a model based on mDixon combined with radiomics to predict the severity of PLE. The ExtraTrees algorithm model exhibited good performance in predicting the severity of PLE in the fusion phase, and the features had good interpretability.

Keywords: Lymphedema; Prediction; Magnetic resonance imaging; Dixon; Radiomics

Introduction

Primary lymphedema refers to the abnormal development or dysfunction of the lymphatic system, which results in the inability of lymphatic fluid in the interstitial spaces to flow back into the bloodstream, leading to edema. Usually, this swelling can occur in any part of the body, but it is most common in the limbs. Lymph retention can lead to inflammation, fibrosis, and fat deposition in local tissues and significantly increase the risk of infection[1]. Primary lower limb lymphedema (PLE) is unpredictable, but how to predict its severity is a key focus of clinical attention. Otherwise, patients' symptoms may gradually worsen into chronic incurable diseases. The grading criteria for lymphedema in the International Society of Lymphology (ISL) guidelines are based on the difference between the affected limb and the healthy side and are used to evaluate the degree of swelling in the affected limb. The grading is closely related to the severity of the disease and is divided into mild, moderate and severe[2].

The classification system requires two core types of data: lower limb circumference, and the volume calculated based on the measured circumference. Lower limb circumference is measured manually by clinicians using a flexible tape measure, while volume is computed via a formula that sums the volumes of multi-segmented truncated frustums. Although this traditional measurement approach has advantages such as simple operation and low cost, it presents non-negligible limitations in terms of measurement accuracy, operational efficiency and assessment dimension: (1) Low precision of measurement and calculation, with high susceptibility to multiple interfering factors: The accuracy of manual measurement and subsequent volume calculation is directly affected by three key factors, namely measurement technique, segmental division, and limb morphology; any deviation

occurring at any stage will lead to inaccurate classification results[3-5]; (2) Time-consuming and labor-intensive procedures cannot meet the efficiency requirements of clinical practice: According to clinical practice, the complete process of circumference measurement and volume calculation for a lymphedema patient typically consists of six steps—"patient positioning, measurement point marking, segmental circumference measurement, data recording, formula calculation, and result verification"—which takes approximately 30 minutes and substantially occupies clinicians' diagnosis and treatment time; (3) Single assessment index fails to reflect in-depth disease characteristics: Lower limb circumference and volume difference only reflect the severity of limb swelling at the macro level, and are essentially "surface-level indicators" that cannot capture the intrinsic pathological features of lymphedema lesions[6,7].

The Dixon technique for water-fat separation was first proposed by Thomas Dixon in 1984. As a type of chemical shift imaging, it processes proton information contained in different tissues such as water and fat through computation and image information processing. A single scan yields four sets of images (in phase, out of phase, water-only phase, fat-only phase) [8]. Subcutaneous adipose tissue in lymphedema contains multiple fatty acid components, so the unimodal fat model of traditional Dixon technique can hardly fully capture its resonance characteristics. This issue is particularly prominent in high-field magnetic resonance equipment, where fat signal fitting deviation is prone to occur. The fifth-generation water-fat separation technology, mDixon XD, innovatively adopts a 7-peak fat model, which significantly improves water-fat separation performance and improves the reliability of quantitative analysis, providing an accurate quantitative basis for PLE disease grading. In addition, mDixon XD is equipped with an improved B0 field correction algorithm, which

effectively eliminates artifacts or water-fat signal confusion that may occur during large field scanning, and can maintain a uniform lipid compression effect in full lower limb scanning for lymphedema [9].

Radiomics employs feature screening to eliminate redundant information, when combined with machine learning techniques, uncovers crucial information related to lymphedema severity from high-dimensional data. This approach ultimately enables the construction of robust, interpretable, and reproducible prediction models, thereby providing an objective quantitative basis for precise diagnosis [10].

Therefore, this study sought to investigate the potential of machine learning combined with radiomics based on mDixon-sequence MR images for predicting the severity of PLE.

Materials and Methods

Sample Size Estimation Method

This study adopted a machine learning approach based on radiomics extracted from MRI mDixon sequences to predict PLE grading. To guarantee the accuracy and reliability of the constructed prediction model, the sample size calculation formula applied in this study is presented as follows [11]:

$$n = \left(\frac{1.96}{\delta} \right)^2 \hat{\phi}(1 - \hat{\phi})$$

where n denotes the required sample size, $\hat{\phi}$ represents the expected incidence of the outcome event, and δ indicates the allowable error, which is defined as the maximum permissible deviation between the sample statistic and the population parameter. The constant 1.96 corresponds to the two-sided quantile of the standard normal distribution at a 95% confidence level.

Based on historical data previously collected by our research team, the expected incidence of the outcome event (severe PLE) is approximately 44.77%. Considering that PLE is a rare disease with limited overall available sample size, which leads to constrained statistical precision, this study sets the allowable margin of error as $\pm 8\%$. Under this assumption, the preliminary calculated sample size is 148 cases. After adjusting for an expected 10% missing data rate, the final estimated sample size required for this study is 165 cases.

Patients

A retrospective review was undertaken of 653 patients with PLE managed at Beijing Shijitan Hospital between November 2019 and July 2023. All patients received a confirmed diagnosis of lymphedema based on clinical assessment and lymphoscintigraphy (LSG) findings. Following exclusion of 295 patients with bilateral edema, 181 with incomplete clinical or MRI data, and 5 with concomitant swelling of vascular origin, a final set of 172 patients with unilateral PLE was enrolled in the study. The final sample size aligned closely with the a priori sample size estimate.

Inclusion criteria: (1) Presence of unexplained lower limb swelling, with lower limb lymphoedema confirmed by LSG; (2) No identifiable aetiology for the oedema (e.g., post-malignancy surgery, trauma, filariasis, or infection), with PLE verified by clinical evaluation and LSG; (3) Completion of bilateral lower limb MRI acquired using the mDixon sequence.

Exclusion criteria: (1) Incomplete clinical data precluding accurate clinical grading; (2) MRI images of insufficient quality to permit radiomics-based region-of-interest delineation; (3) Presence of bilateral lower limb lymphoedema;

(4) Concomitant vascular disorders, such as Klippel-Trenaunay syndrome.

Limb volume is calculated using the formula V [12]:

$$V = \frac{1}{3}\pi h (R_1^2 + r_1^2 + R_1 \times r_1) + \frac{1}{3}\pi h (R_2^2 + r_2^2 + R_2 \times r_2) + \dots + \frac{1}{3}\pi h (R_n^2 + r_n^2 + R_n \times r_n)$$

where R denotes the radius at the inferior landmark of the measured segment, calculated as the corresponding limb circumference divided by 2π ; and r denotes the radius at the superior landmark of the measured segment, derived in the same manner as the corresponding limb circumference divided by 2π , and h is the axial distance between the two reference points along the measured segment.

Six anatomical planes were selected for assessment: the ankle region, the distal third of the lower leg, the proximal third of the lower leg, the knee joint, the distal third of the thigh, and the proximal third of the thigh.

Percentage volume difference between the affected and unaffected limbs was calculated using Formula X, in accordance with the grading criteria outlined in the 2023 ISL guidelines [2] (Table 1):

$$X = [(Affected\ limb\ volume - Unaffected\ limb\ volume) / Unaffected\ limb\ volume] \times 100\%$$

On the basis of this calculation, enrolled patients were categorised into two sets: a non-severe group (mild or moderate disease) and a severe group. The patient enrolment workflow is illustrated in Figure 1. Clinical data collected for each patient included sex, congenital onset, age at diagnosis, disease duration, age at hospital admission, and non-pitting edema, elevated skin temperature, cutaneous erythema, reduced skin elasticity, and pain.

This retrospective study was approved by the institutional review board, which waived the need for informed consent and clinical trial review.

MRI

All lower-limb MRI examinations in this study were performed using a 1.5 T MR scanner (Ingenia; Philips, the Netherlands) in conjunction with a 16-channel body coil. Scans were acquired with patients in the supine, feet-first position. The imaging volume extended from the level of the femoral head distally to cover the entire foot. The modified Dixon (mDixon, 5th generation) sequence was acquired with the following parameters: repetition time (TR) 5.8 ms, first echo time (TE1) 1.8 ms, second echo time (TE2) 4.0 ms, field of view (FOV) 420 mm × 260 mm, slice thickness 8 mm, inter-slice gap 0.2 mm, matrix size 280×175, voxel size 1.5 mm³, and number of excitations (NEX) 1-2.

VOI

DICOM-formatted mDixon sequence images were exported from the Picture Archiving and Communication System (PACS) by a specialist MR radiographer (J.R.). Four image phases were derived: in phase (IP), out of phase (OP), water-only (W) and fat-only (F). Volumetric segmentation was performed using ITK-SNAP software (<https://www.itksnap.org>). For each of the four phase datasets, ten consecutive axial slices at the matching anatomical level, corresponding to the region of maximum swelling in the affected limb, were selected, with articular surfaces avoided wherever possible. The selected image data were subsequently converted to NIfTI format. A threshold-based segmentation tool built into ITK-SNAP was applied to delineate and fill the tissue compartment between the skin surface and muscle fascia. All software-generated segmentation results were independently reviewed by two physicians with over 8 years of experience in musculoskeletal MRI diagnosis (XPL, MKL). In the event of discrepancy between the two reviewers, a final decision was

reached via adjudication by a senior professor (RGW or YLY). Finally, the definitive volume of interest (VOI) was generated and the corresponding segmentation mask file was saved.

Radiomics and Machine Learning

Radiomic feature extraction was performed on the IP, OP, W and F phase images using PyRadiomics, an open-source package implemented in Python. Radiomic features derived from the combined W and IP phases were also computed to form a fusion phase (FP) dataset. All feature modelling was conducted using scikit-learn version 1.0.2. The detailed methodology is described below.

Three categories of radiomic features were extracted: (1) Shape features: These capture the morphological properties of the VOI, including volume, surface area and diameter. (2) First-order statistics: These describe the distribution of voxel intensity values within the VOI, including metrics such as median, mean, kurtosis and root mean square. (3) Texture features: These characterise the spatial distribution of voxel intensities, and comprise five feature families: grey-level co-occurrence matrix (GLCM), grey-level run length matrix (GLRLM), grey-level size zone matrix (GLSZM), grey-level dependence matrix (GLDM), and neighbourhood grey-tone difference matrix (NGTDM). All extracted features were normalised using Z-score standardisation. Feature selection was performed in three sequential stages. First, *Mann-Whitney U* testing was applied to all features, and only those with a P value < 0.05 were retained. Next, *Pearson* correlation coefficient was used to filter the remaining features, with a single feature retained from each pair of highly correlated features. Finally, the least absolute shrinkage and selection operator (LASSO) algorithm was employed to identify an optimal feature subset for final model

construction.

For the LASSO procedure, the tuning parameter λ was optimised via 10-fold cross-validation. Features with non-zero coefficients at the optimal λ value were retained as the final feature set. These selected features were then fed into five machine learning classifiers: logistic regression (LR), support vector machine (SVM), random forest (RF), extratrees (ET), and light gradient boosting machine (LightGBM). The full radiomics modelling workflow is summarised in Figure 2.

Statistical analysis

Data analysis software: Python 3.7.12 and SPSS 26.0. The *Shapiro–Wilk* test was used to test for a normal distribution of the data. The data with a normal distribution were subjected to *t* tests. The *Mann–Whitney U* test was used for nonnormally distributed data. *Pearson* correlation analysis is used to determine the correlation between variables. The categorical variables were tested via the chi-square test. $P < 0.05$ was considered statistically significant. ROC curves were used to evaluate the diagnostic performance of different machine learning models in distinguishing between the severe and nonsevere groups, and the AUC, accuracy, sensitivity, and specificity were calculated. And validate the applicability of the model in treatment decision-making through decision curve analysis (DCA).

Results

Clinical characteristics

A total of 172 patients were randomly allocated to a training set and a test set in a 7:3 ratio. The training set comprised 120 patients (68 in the non-severe group and 52

in the severe group), whilst the test set included 52 patients (27 in the non-severe group and 25 in the severe group).

A statistically significant difference in disease duration between the severe and non-severe groups was observed in both sets ($P=0.025$, $P<0.001$, $P=0.002$). In the training set, cutaneous erythema, as well as reduced skin elasticity, showed significant between-group differences (all $P<0.05$). In the test set, statistically significant differences were identified for disease duration ($P=0.006$) and elevated skin temperature ($P=0.041$). These findings are summarised in Table 2.

Radiomic features

A total of 1197 radiomic features were extracted per patient, comprising 14 shape features, 234 first-order statistics, and 949 texture features. The texture features fell into five families: GLCM ($n=286$), GLRLM ($n=208$), GLSZM ($n=208$), GLDM ($n=182$) and NGTDM ($n=65$). Following initial screening with the *Mann-Whitney U* test and *Pearson* correlation analysis, 127, 95, 100, 95 and 226 features were retained for the IP, OP, W, F and FP datasets respectively. The distribution of retained features across the five phases is illustrated in the violin plot in Figure 3. Subsequent feature selection was conducted using LASSO regression, with the tuning parameter λ optimised via 10-fold cross-validation. For the fusion phase model, the optimal λ was 0.0596, and 14 features were ultimately retained for final model construction (Figure 4). Figure 5 displays a bar chart of the coefficients for the 14 retained imaging biomarkers, demonstrating the magnitude and direction of association between each biomarker and severe PLE.

Modelling

Models were constructed using radiomic features from the IP, OP, W, F and FP datasets following feature selection and dimensionality reduction. All five machine learning classifiers exhibited robust discriminative performance in differentiating between the non-severe and severe PLE groups across all five phase datasets. Performance metric comparisons across the five classifiers for each phase are detailed in Tables 3-7, with comparative radar charts presented in Figure 6. ROC curves demonstrating the discriminative capacity of the five models are shown in Figure 7. The ET classifier trained on FP features achieved the highest overall performance. In the training set, it yielded an AUC of 0.899 (95% CI 0.846-0.953), a sensitivity of 0.808 and a specificity of 0.824. In the test set, it achieved an AUC of 0.907 (95% CI 0.824-0.990), a sensitivity of 0.760 and a specificity of 0.926.

Feature weight map of ExtraTrees model

The ET model was selected for subsequent variable importance analysis of the FP feature set (Figure 8). Of the 14 features incorporated in the final model, the three most influential predictors, ranked in descending order of importance, were: wavelet LLL glszm SmallAreaLowGrayLevelEmphasis from the W, log Sigma 5 0 mm 3D glszm SmallAreaEmphasis from the IP, and wavelet LLL glgm LargeDependencyHighGrayLevelEmphasis from the IP. The highest-weighted feature, wavelet LLL glszm SmallAreaLowGrayLevelEmphasis, showed a negative correlation with lymphedema grading.

Decision curve analysis

Decision curve analysis was conducted to assess the clinical utility of the FP-based ET model in guiding treatment decisions for patients with severe PLE

(Figure 9). The model exhibited a broad clinically applicable range across a threshold probability interval of 0%-98%. The optimal balance between clinical benefit and potential harm was identified within the 18%-58% threshold probability range, where a relatively high net benefit (>20%) was maintained.

Waterfall plot

Waterfall plot presents the classification performance of the ET model in differentiating severe from non-severe PLE. The results confirm that the ET model achieves favourable discriminative capacity between the two groups (Figure 10).

Discussion

This study developed a risk prediction modelling framework for PLE grading using radiomic features derived from MRI mDixon sequences. Following comparative analysis across multiple phase datasets and machine learning classifiers, the FP-based ET model achieved the best overall discriminative performance and showed substantial clinical utility for PLE severity grading.

Lymphedema typically presents on MRI as swelling of the subcutaneous adipose layer with abnormal signal on T1 and T2 weighted sequences, while muscle and bone remain unaffected [13]. This pathological feature forms the key rationale for ROI placement in this study. The complementary signal characteristics of the water-only and in-phase mDixon sequences provide a robust basis for radiomic feature fusion, which can compensate for the limitations of each single phase in lymphedema assessment [13,14].

In this study, the ET model constructed with fused FP features exhibited

particularly strong performance. The model achieved the highest AUC among all compared classifiers (including SVM, LightGBM and LR), with the notable observation that AUC in the test set exceeded that in the training set. This finding mitigates the commonly recognised limitation of traditional machine learning models, whereby strong training set performance is often accompanied by degraded performance on independent test data, and indicates robust feature learning capacity and generalisability for PLE assessment. Based on the study design and inherent properties of the model, we propose the following explanations for this finding: (1) The FP feature set aligns closely with the core pathophysiological features of PLE. Rather than simply combining features from the water-only and in-phase datasets, FP achieves deep integration of radiomic information from both modalities through systematic feature selection. This feature set captures the three core tissue components of PLE: retained lymphatic fluid, hypertrophied adipose tissue and hyperplastic fibrous tissue. Water-only images depict the distribution and signal intensity of interstitial lymphatic fluid, while in-phase images differentiate adipose and fibrotic signals. The complementary FP features thus reflect PLE severity grading from a pathophysiological perspective, providing robust input for model training. (2) The ET model has inherent strong generalisability [15-19]. As a tree-based ensemble learning method, the ET model offers advantages in overfitting resistance and noise robustness. It balances bias and variance through full-sample training and increased randomness. Each decision tree is built using the complete training dataset, with split thresholds selected randomly at each node. This high degree of randomness reduces the model's dependence on training data, lowering overfitting risk. For minor noise in lymphedema radiomic data (such as scanner variability or movement artefacts), the model aggregates predictions from multiple ExtraTrees via a voting mechanism.

Compared with noise-sensitive classifiers such as SVM, this approach produces more robust results and supports cross-sample generalisation. (3) Close alignment exists between feature dimensionality and model architecture. The high-dimensional FP feature set (226 retained features) is highly compatible with the ET model's feature processing mechanism. The model's random feature selection strategy automatically filters redundant information in the high-dimensional feature space (such as overlapping tissue signal features between the two phases) and identifies features with strong discriminative power for PLE grading. This model-driven selection of effective features avoids learning bias from single-modality features and maximises the multi-dimensional advantages of the FP feature set, allowing the model to capture underlying pathological characteristics of PLE beyond superficial data patterns. In addition, randomness from sample splitting and stochastic model training may contribute to the higher AUC observed in the test set.

Feature importance analysis further revealed that three of the 14 retained FP features made the greatest contribution to PLE grading, all showing a negative correlation with disease severity. This pattern supports the validity of the model's predictions and confirms the rationale of the FP feature set at the micro-level. The highest-ranked feature, wavelet LLL glszm SmallAreaLowGrayLevelEmphasis from the water-only phase, quantifies the relative proportion of small low-gray-level regions, corresponding predominantly to adipose tissue components on water-only imaging. Its negative correlation with PLE grading may be explained as follows: a lower value indicates a smaller proportion of small low-gray-level regions, reflecting reduced adipose tissue volume, while the proportion of high-gray-level lymph-rich interstitial spaces increases with reduced signal homogeneity. These changes are consistent with pathological features such as lymphatic fluid retention and adipose

tissue fibrosis, which are characteristic of progressive PLE. The second and third highest-weighted features (log Sigma 5 0 mm 3D glszm SmallAreaEmphasis and wavelet LLL glgm LargeDependencyHighGrayLevelEmphasis, both from the in-phase phase) reflect lesion size distribution and lymphatic fluid dispersion patterns respectively. Their negative correlation with severity aligns with the pathophysiological process of progressive lymphatic accumulation and extensive fibrosis in advanced PLE. These three features reflect PLE progression across three dimensions: adipose tissue proportion, lesion size, and lymphatic distribution pattern. All are key texture features derived from water-only and in-phase imaging within the FP set, and their high weighting further validates the effectiveness of the dual-modality feature fusion strategy, supporting the model's strong generalisation performance.

This study has several limitations. First, it used a retrospective, single-centre dataset, which limits the external generalisability of the model. Second, the sample size was relatively modest. Finally, only radiomics features were included, without clinical features and laboratory data. Therefore, further multi-centre prospective studies are required to verify the stability and generalisability of the model.

Conclusion

This study innovatively developed a radiomic model for PLE severity grading using mDixon MRI sequences, with the FP-based ET model achieving optimal discriminative performance. The imaging biomarkers identified by the model demonstrate clinical interpretability and reflect underlying pathophysiological alterations associated with disease severity at the micro-level, providing a novel technical pathway for the precise grading of PLE and holding promise as an adjunct

tool to support clinical treatment decision-making.

Declarations:

1.Funding: This study was supported by Beijing Natural Science Foundation (No. L2601003). The sponsor is Professor Yunlong Yue, the corresponding author of this article.

2.No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication.

3.Ethics approval: This study was approved by the Medical Ethics Committee of Beijing Shijitan Hospital, Capital Medical University (approval number: IIT-2024-059-001). This study was performed in accordance with the principles of the Declaration of Helsinki.

4.Consent to Participate, and Consent to Publish declarations: not applicable. Written informed consent was waived by the above-mentioned ethics committee due to the retrospective nature of the study.

Reference

- [1] Senger JB, Kadle RL, Skoracki RJ. Current Concepts in the Management of Primary Lymphedema. *Medicina (Kaunas)*. 2023;59(5):894.
- [2] Executive Committee of the International Society of Lymphology. The Diagnosis and Treatment of Peripheral Lymphedema: 2023 Consensus Document of The International Society of Lymphology. *Lymphology*. 2023;56(4):133-151.
- [3] Dylke E. Measurement of breast cancer-related lymphoedema. *J Physiother*. 2022;68(4):238-243.
- [4] Jönsson C, Johansson K, Bjurberg M, Brogårdh C. Circumferential Measurements to Calculate Lower Limb Volume in Persons with Lymphedema: What Segment Length Is to Be Recommended? *Lymphat Res Biol*. 2023;21(3):275-282.
- [5] Eng CM, Abrams GD, Smallwood LR, Lieber RL, Ward SR. Muscle geometry affects accuracy of forearm volume determination by magnetic resonance imaging (MRI). *J Biomech*. 2007;40(14):3261-3266.
- [6] International Society of Lymphology. The diagnosis and treatment of peripheral lymphedema. 2009 Concensus Document of the International Society of Lymphology. *Lymphology*. 2009;42(2):51-60.
- [7] Lee S, Lee DG, Kim KT. Temporal Changes in Subcutaneous Fibrosis in Patients with Lower Extremity Lymphedema Following Surgery for Gynecologic Cancer: A Computed Tomography-Based Quantitative Analysis. *Diagnostics (Basel)*. 2022;12(8):1949.
- [8] Dixon WT. Simple proton spectroscopic imaging. *Radiology*. 1984;153(1):189-194.
- [9] REN Jie, LI Xingpeng, LIU Mengke, et al. The Value of mDIXON Histogram Parameters in Grading Primary Lymphedema of Lower Extremity. *Journal of Clinical Radiology*, 2024, 43(9): 1587-1590
- [10] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446.
- [11] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, Moons KGM, Collins G, van Smeden M. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
- [12] Brorson H, Höijer P. Standardised measurements used to order compression garments can be used to calculate arm volumes to evaluate lymphoedema treatment. *J Plast Surg Hand Surg*. 2012;46(6):410-415.
- [13] Aström KG, Abdsaleh S, Brenning GC, Ahlström KH. MR imaging of primary, secondary, and mixed forms of lymphedema. *Acta Radiol*. 2001;42(4):409-416.
- [14] Nowak S, Henkel A, Theis M, Luetkens J, Geiger S, Sprinkart AM, Pieper CC, Attenberger UI. Deep learning for standardized, MRI-based quantification of subcutaneous and subfascial tissue volume for patients with lipedema and lymphedema. *Eur Radiol*. 2023;33(2):884-892.

- [15] Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective Heart Disease Prediction Using Machine Learning Techniques. *Algorithms*. 2023;16(2):88.
- [16] Song, P. Random Forest vs Extremely Randomized Trees (Extra Trees): When to Choose Each. *ML Journey*. 2026;11(1).
- [17] Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18-22.
- [18] Petmezas G, Papageorgiou VE, Passman RS, Rogers JA, Stefanopoulos L, Katsaggelos AK, Maglaveras N. Optimizing atrial fibrillation detection through ECG feature selection using Extra-Trees and statistical association measures. *Journal of Electrocardiology*. 2026;95:154199.
- [19] Zhang J, Hu G, Zhang J, et al. Prediction of the Unconfined Compressive Strength of One-Part Geopolymer-Stabilized Soil Under Acidic Erosion: Comparison of Multiple Machine Learning Models. *Materials*. 2026;19(1):209.

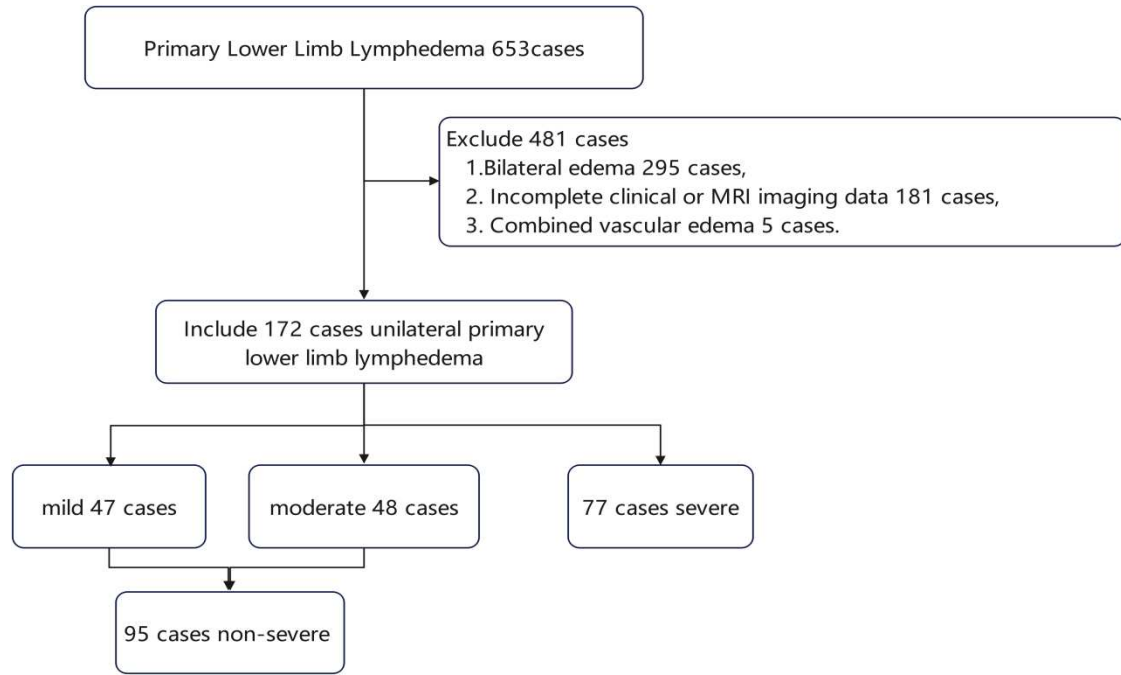


Figure 1: Flowchart for the inclusion and exclusion criteria for primary lower limb lymphedema.

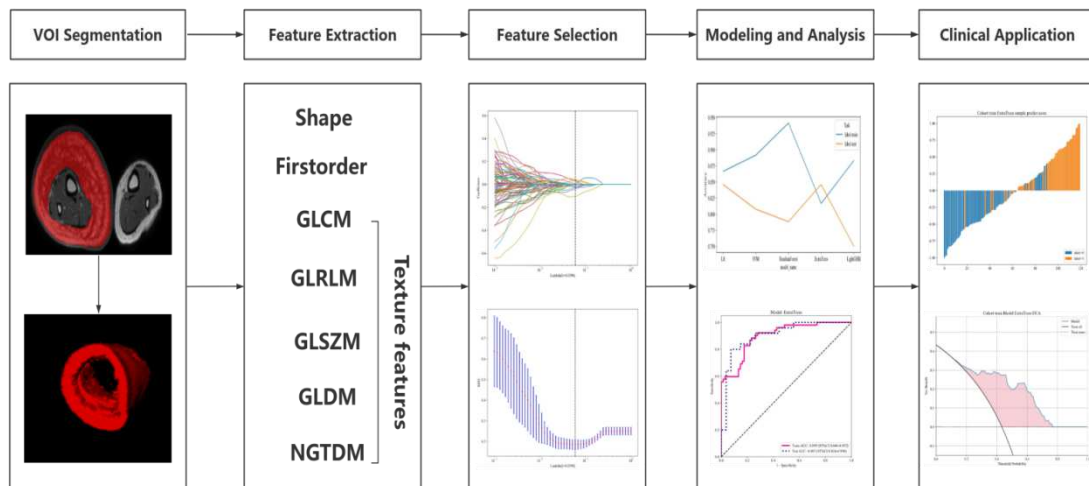


Figure 2. Flowchart illustrating the workflow of radiomic model construction and validation in this study.

GLCM: Gray-level co-occurrence matrix; GLRLM: Gray-level run-length matrix; GLSZM: Gray-level size region matrix; GLDM: Gray-level dependency matrix; NGTDM: Neighborhood gray-level difference matrix; ROC: Receiver Operating Characteristic curve; DCA: Decision Curve Analysis

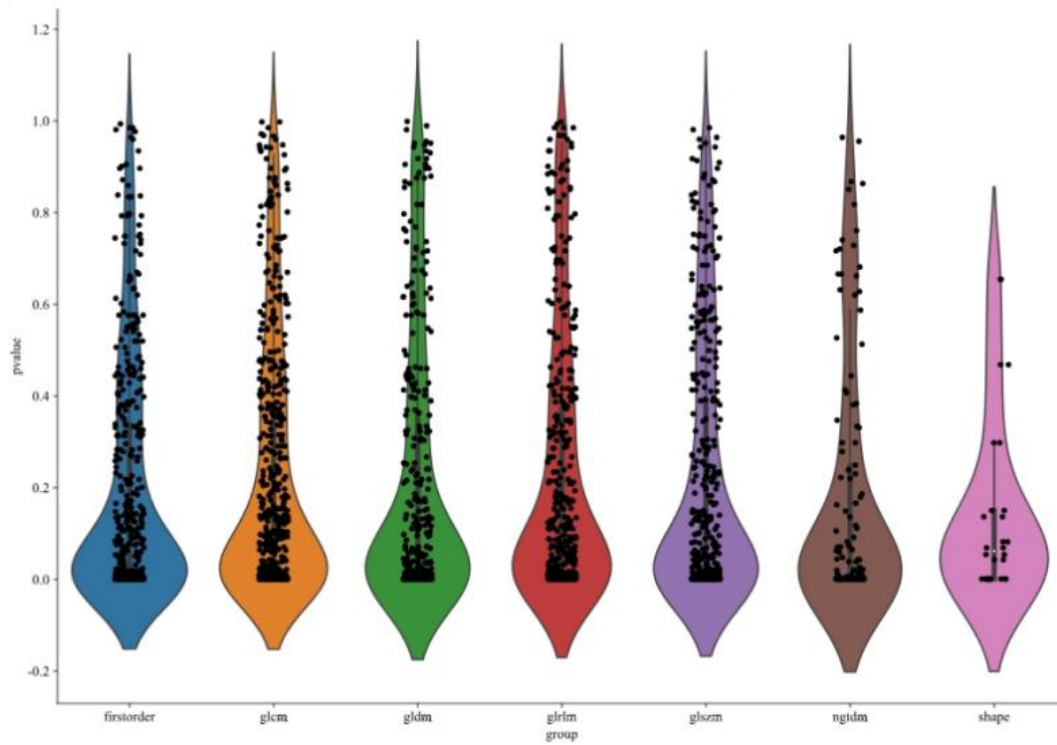


Figure 3: Violin plot of radiomics features.

GLCM: Gray-level co-occurrence matrix; GLRLM: Gray-level run-length matrix; GLSZM: Gray-level size region matrix; GLDM: Gray-level dependency matrix; NGTDM: Neighborhood gray-level difference matrix.

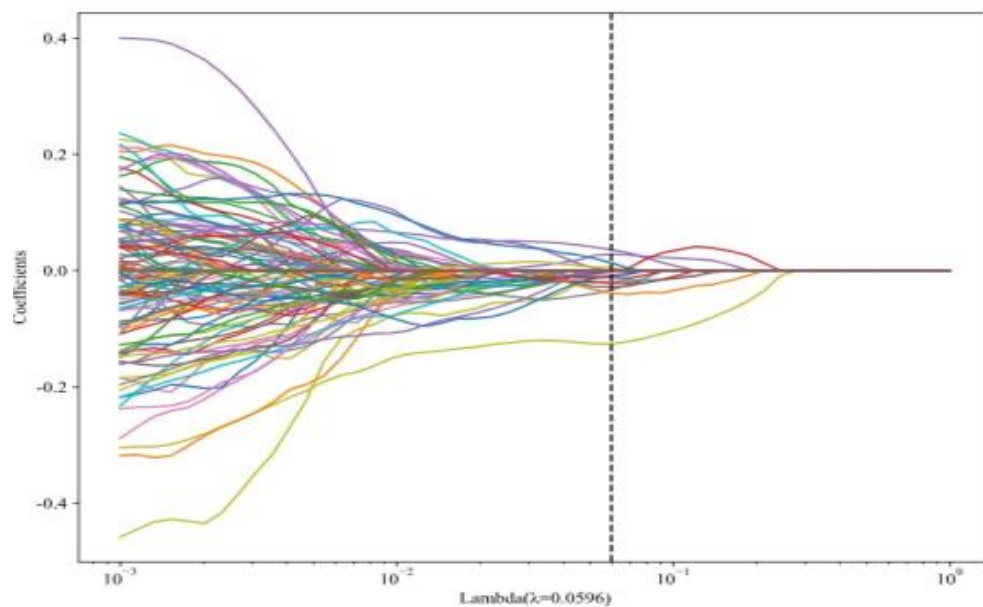


Figure 4: Distribution plot of LASSO coefficients for radiomics features.

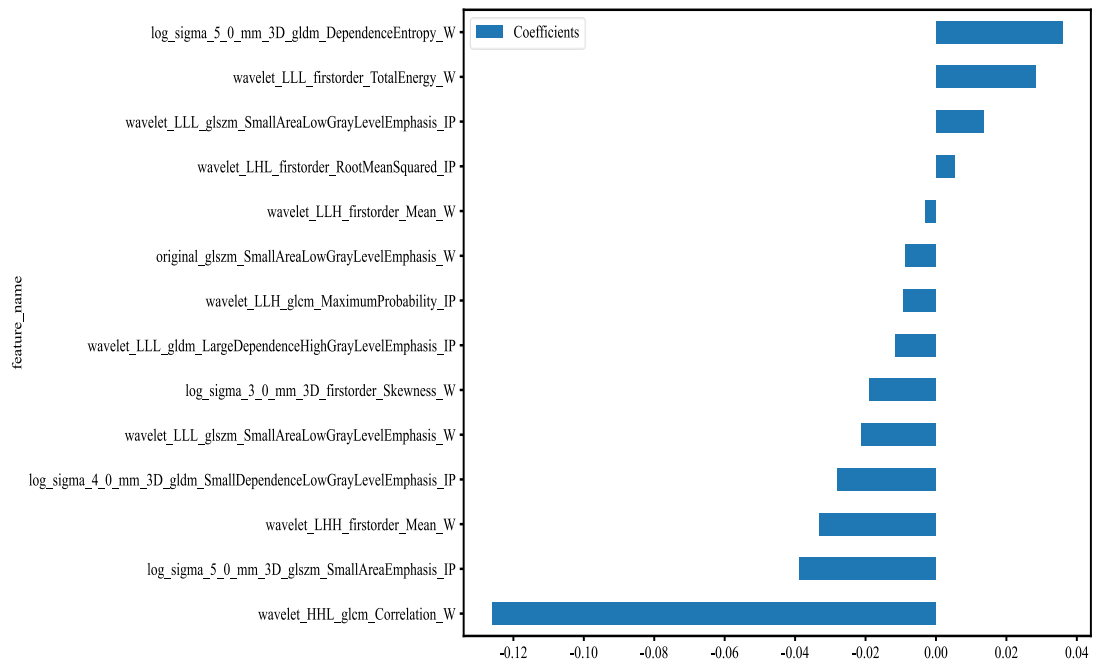
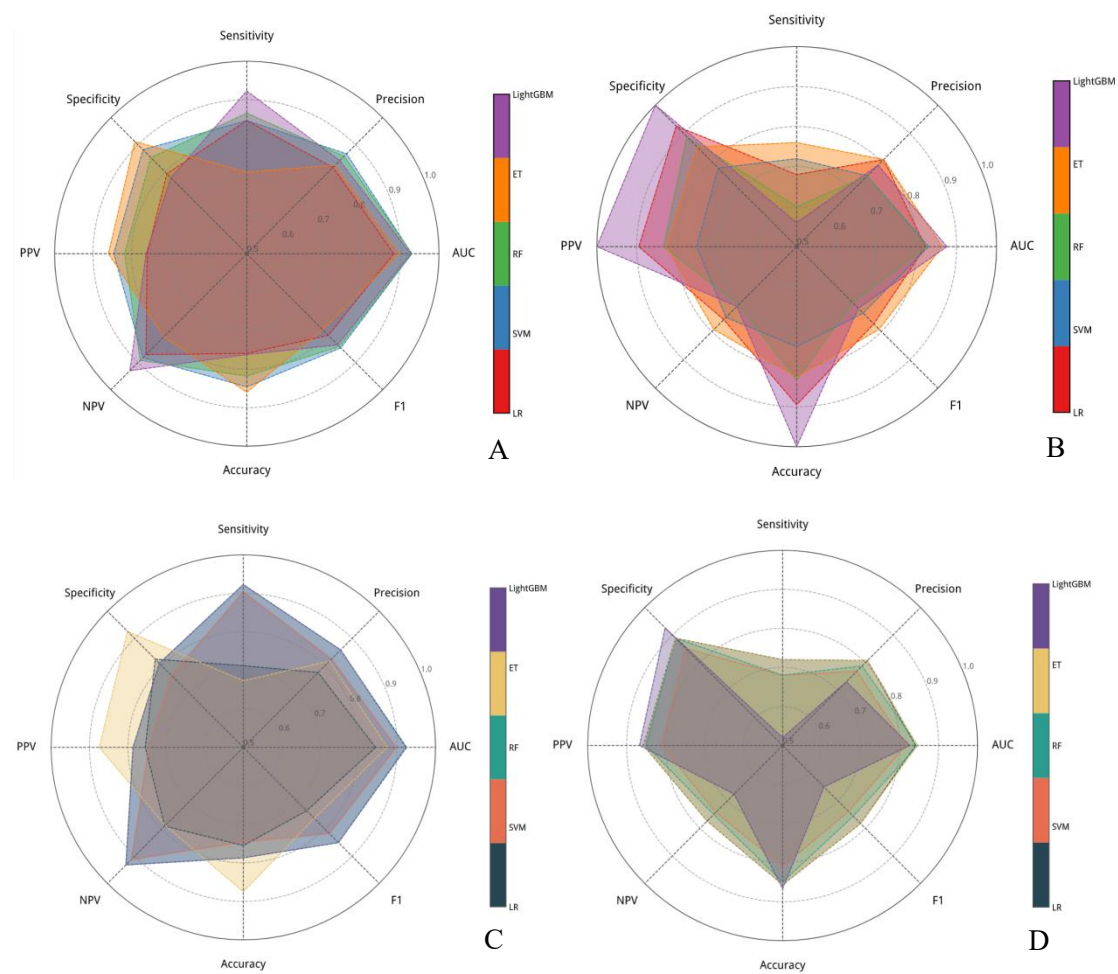


Figure 5: Bar chart illustrating the levels of radiomics feature coefficients. W: Water-only phase; IP: In phase.



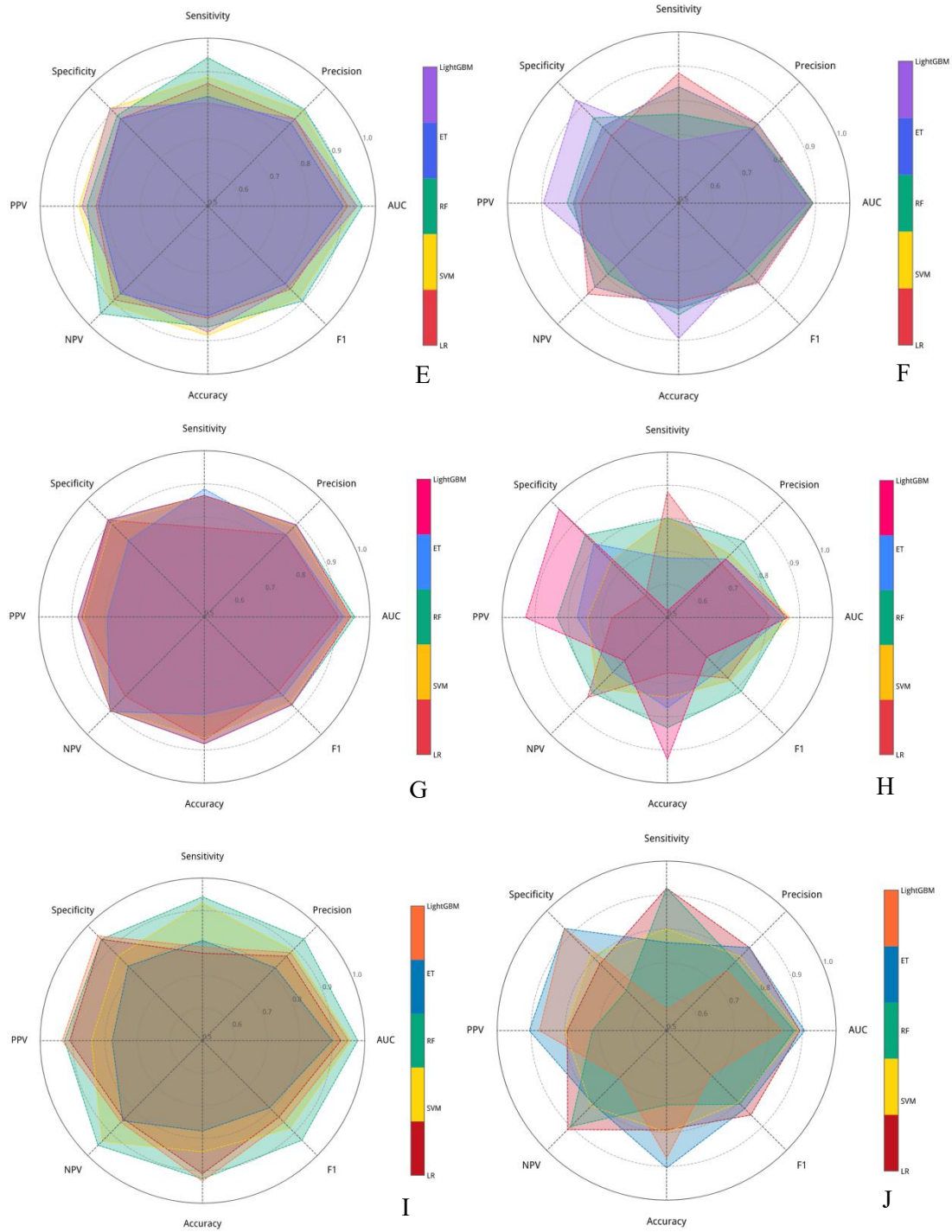


Figure 6: Radar chart illustrating the performance metrics (AUC, F1, recall, accuracy, negative predictive value, positive predictive value, specificity, sensitivity, and precision) of five machine learning models across different phases.

Figures A and B show the results for the in phase training and test sets, respectively;

Figures C and D show the results for the out of phase training and test sets, respectively;

Figures E and F show the results for the water-only phase training and test sets, respectively;

Figures G and H show the results for the fat-only phase training and test sets, respectively;

Figures I and J show the results for the fusion phase training and test sets, respectively.

LR: Logistic Regression; SVM: Support Vector Machine; RF: Random Forest; ET: ExtraTrees;

LightGBM: Lightweight Gradient Boosting Machine

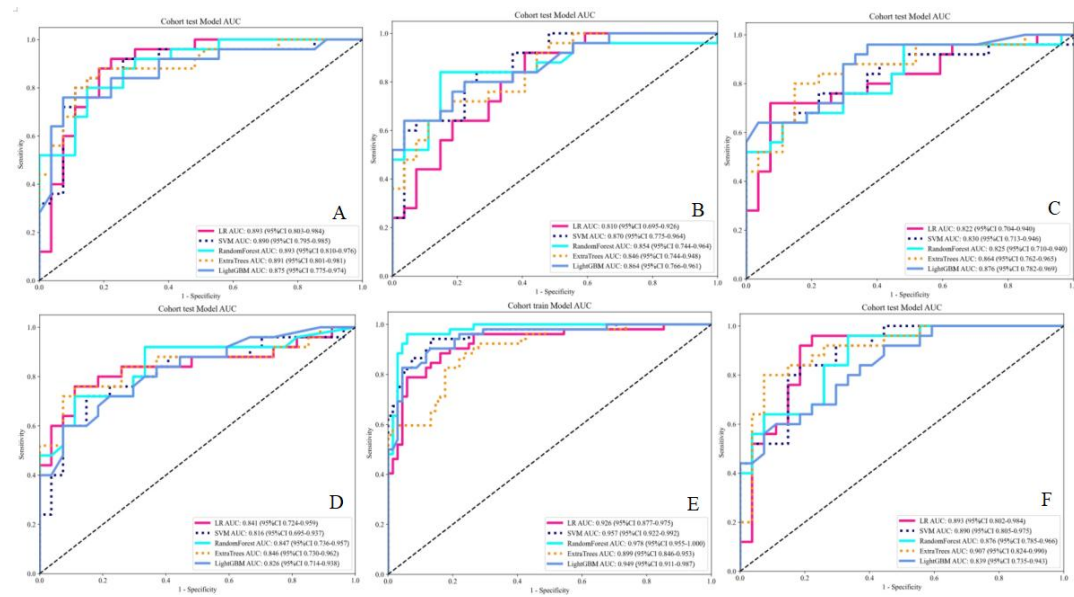


Figure 7: ROC curves of five machine learning models for distinguishing between the non-severe and severe primary lower limb lymphedema groups.

A-D corresponds to the results of the water-only phase, fat-only phase, in phase, and out of phase test sets, respectively.

E and F represent the results of the fusion phase training set and test set, respectively.

LR: Logistic Regression; SVM: Support Vector Machine; Random Forest: Random Forest; ET: ExtraTrees; LightGBM: Lightweight Gradient Boosting Machine

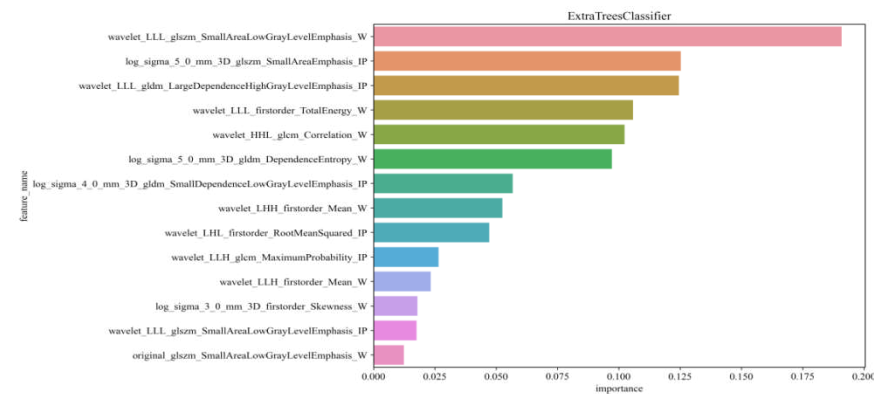


Figure 8. Feature weight map of the ExtraTrees model based on fusion-phase features. Bar length represents the relative importance of each radiomic feature in the classification model. Features are ranked in descending order of importance, and the source phase of each feature is indicated (W: water-only phase; IP: in phase).

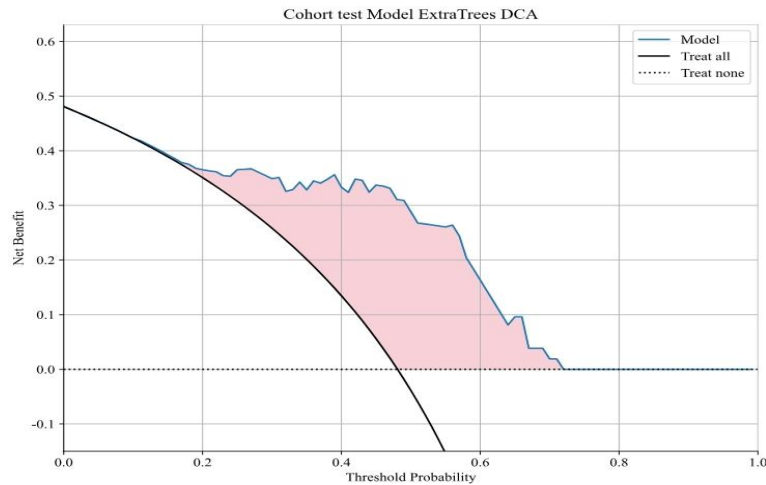


Figure 9. Decision curve analysis of the ExtraTrees model using FP features for severe primary lower limb lymphedema. The blue line represents the net benefit of the ExtraTrees model. The solid black line (Treat all) corresponds to the net benefit of universal intervention, and the dashed line (Treat none) corresponds to the net benefit of no intervention. The x-axis indicates threshold probability, and the y-axis indicates net benefit.

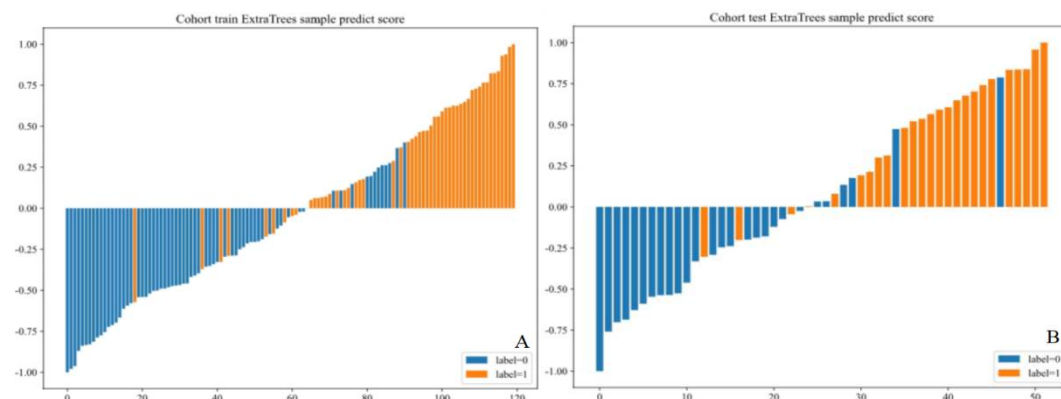


Figure 10. Waterfall plot demonstrating the classification results of the ExtraTrees model for patients with severe and non-severe primary lower limb lymphedema. Each vertical bar corresponds to one individual patient. Patients in the severe group are indicated by red bars, and those in the non-severe group by blue bars.

Figure A and Figure B show the results for the training cohort and the test sets, respectively.

Table 1: 2023 International Society of Lymphology Clinical Classification criterion for Lymphedema.

Grade	(Affected limb volume—Unaffected limb volume) / Unaffected limb volume
Normal Range	<5%
Mild	5%-20% (Includes lower bound, excludes upper bound)
Moderate	20%-40% (Includes lower bound, excludes upper bound)
Severe	>40%

Table 2: Baseline Clinical Characteristics.

Clinical features	Training set				Test set			
	Total (n=120)	Non-severe group (n=68)	Severe group (n=52)	<i>P</i>	Total (n=52)	Non-severe group (n=27)	Severe group (n=25)	<i>P</i>
Diagnosis age (years)	20.27±17.26‡	19.84±15.19‡	20.85±19.78‡	0.736	18.79±13.94‡	18.59±13.14‡	19.00±15.03‡	0.927
Disease duration (years)	10.90±10.79‡	8.71±9.12‡	13.76±12.14‡	0.025*	10.04±10.91‡	6.22±7.57‡	14.16±12.52‡	0.006*
Admission age (years)	31.07±17.75‡	28.40±14.88‡	34.58±20.56‡	0.104	28.29±15.16‡	24.81±13.62‡	32.04±16.09‡	0.101
Gender				0.357				0.630
Female	60 (50.0)	31 (45.6)	29 (55.8)		38 (73.1)	21 (77.8)	17 (68.0)	
Male	60 (50.0)	37 (54.4)	23 (44.2)		14 (26.9)	6 (22.2)	8 (32.0)	
Onset at birth	18 (15.0)	7 (10.3)	11 (21.2)	0.164	5 (9.6)	2 (7.4)	3 (12.0)	0.928
Non-pitting edema	24 (20.0)	13 (19.1)	11 (21.2)	0.963	5 (9.6)	1 (3.7)	4 (16.0)	0.302
Elevated skin temperature	23 (19.2)	9 (13.2)	14 (26.9)	0.098	8 (15.4)	1 (3.7)	7 (28.0)	0.041*
Cutaneous erythema	32 (26.7)	9 (13.2)	23 (44.2)	<0.001*	13 (25.0)	4 (14.8)	9 (36.0)	0.149
Reduced skin elasticity	85 (70.8)	40 (58.8)	45 (86.5)	0.002*	35 (67.3)	19 (70.4)	16 (64.0)	0.847

Pain	6 (5.0)	2 (2.9)	4 (7.7)	0.447	1 (1.9)	0 (0)	1 (4.0)	0.969
------	---------	---------	---------	-------	---------	-------	---------	-------

Note: Brackets denote percentages; ‡: mean ± standard deviation; **P* <0.05

Table 3: Comparison of in phase imaging biomarker indices across different models

Model	AUC	95% CI	Precision	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy	Recall	F1	group
LR	0.882	0.817–0.947	0.817	0.846	0.794	0.759	0.871	0.759	0.846	0.800	training set
LR	0.822	0.704–0.940	0.808	0.680	0.926	0.895	0.758	0.895	0.680	0.773	test set
SVM	0.929	0.879–0.979	0.867	0.846	0.882	0.846	0.882	0.846	0.846	0.846	training set
SVM	0.830	0.713–0.946	0.750	0.720	0.778	0.750	0.750	0.750	0.720	0.735	test set
RF	0.930	0.882–0.978	0.858	0.865	0.853	0.818	0.892	0.818	0.865	0.841	training set
RF	0.825	0.710–0.940	0.750	0.600	0.889	0.833	0.706	0.833	0.600	0.698	test set
ET	0.896	0.842–0.951	0.825	0.712	0.912	0.860	0.805	0.860	0.712	0.779	training set
ET	0.864	0.762–0.965	0.808	0.760	0.852	0.826	0.793	0.826	0.760	0.792	test set
LightGBM	0.926	0.879–0.973	0.842	0.923	0.779	0.762	0.930	0.762	0.923	0.835	training set
LightGBM	0.876	0.782–0.970	0.788	0.560	1.000	1.000	0.711	1.000	0.560	0.718	test set

LR: Logistic Regression; SVM: Support Vector Machine; LightGBM: Lightweight Gradient Boosting Machine; RF: Random Forest; ET: ExtraTrees; CI: Confidence Interval

Table 4: Comparison of out of phase imaging-based biomarker indices across different models

Model	AUC	95% CI	Precision	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy	Recall	F1	group
LR	0.844	0.773–0.915	0.775	0.712	0.824	0.755	0.789	0.755	0.712	0.733	training set
LR	0.841	0.724–0.959	0.808	0.720	0.889	0.857	0.774	0.857	0.720	0.783	test set
SVM	0.898	0.838–0.957	0.825	0.904	0.765	0.746	0.912	0.746	0.904	0.817	training set
SVM	0.816	0.695–0.937	0.769	0.680	0.852	0.810	0.742	0.810	0.680	0.739	test set
RF	0.925	0.877–0.973	0.858	0.923	0.809	0.787	0.932	0.787	0.923	0.850	training set
RF	0.847	0.736–0.957	0.788	0.680	0.889	0.850	0.750	0.850	0.680	0.756	test set
ET	0.875	0.812–0.937	0.817	0.673	0.926	0.875	0.787	0.875	0.673	0.761	training set
ET	0.846	0.730–0.962	0.808	0.720	0.889	0.857	0.774	0.857	0.720	0.783	test set
LightGBM	0.923	0.874–0.972	0.858	0.923	0.809	0.787	0.932	0.787	0.923	0.850	training set
LightGBM	0.826	0.714–0.938	0.731	0.520	0.926	0.867	0.676	0.867	0.520	0.650	test set

LR: Logistic Regression; SVM: Support Vector Machine; LightGBM: Lightweight Gradient Boosting Machine; RF: Random Forest; ET: ExtraTrees; CI: Confidence Interval

Table 5: Comparison of water-only-phase imaging biomarker indices across different models

Model	AUC	95% CI	Precision	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy	Recall	F1	group
LR	0.917	0.866–0.968	0.867	0.865	0.868	0.833	0.894	0.833	0.865	0.849	training set
LR	0.893	0.803–0.984	0.827	0.880	0.778	0.786	0.875	0.786	0.880	0.830	test set
SVM	0.945	0.903–0.987	0.900	0.885	0.912	0.885	0.912	0.885	0.885	0.885	training set
SVM	0.890	0.795–0.985	0.827	0.840	0.815	0.808	0.846	0.808	0.840	0.824	test set
RF	0.961	0.931–0.991	0.908	0.942	0.882	0.860	0.952	0.860	0.942	0.899	training set
RF	0.893	0.810–0.977	0.808	0.760	0.852	0.826	0.793	0.826	0.760	0.792	test set
ET	0.903	0.846–0.960	0.850	0.827	0.868	0.827	0.868	0.827	0.827	0.827	training set
ET	0.891	0.801–0.981	0.827	0.840	0.815	0.808	0.846	0.808	0.840	0.824	test set
LightGBM	0.945	0.906–0.983	0.867	0.808	0.912	0.875	0.861	0.875	0.808	0.840	training set
LightGBM	0.875	0.776–0.974	0.808	0.680	0.926	0.895	0.758	0.895	0.680	0.773	test set

LR: Logistic Regression; SVM: Support Vector Machine; LightGBM: Lightweight Gradient Boosting Machine; RF: Random Forest; ET: ExtraTrees; CI: Confidence Interval

Table 6: Comparison of fat-only-phase imaging biomarker indices across different models

Model	AUC	95% CI	Precision	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy	Recall	F1	group
LR	0.905	0.849–0.961	0.850	0.769	0.912	0.870	0.838	0.870	0.769	0.816	training set
LR	0.810	0.695–0.926	0.731	0.880	0.593	0.667	0.842	0.667	0.880	0.759	test set
SVM	0.940	0.893–0.986	0.883	0.865	0.897	0.865	0.897	0.865	0.865	0.865	training set
SVM	0.870	0.775–0.964	0.769	0.800	0.741	0.741	0.800	0.741	0.800	0.769	test set
RF	0.954	0.919–0.990	0.892	0.865	0.912	0.882	0.899	0.882	0.865	0.874	training set
RF	0.854	0.744–0.964	0.827	0.800	0.852	0.833	0.821	0.833	0.800	0.816	test set
ET	0.920	0.870–0.970	0.850	0.885	0.824	0.793	0.903	0.793	0.885	0.836	training set
ET	0.846	0.744–0.948	0.750	0.680	0.815	0.773	0.733	0.773	0.680	0.723	test set
LightGBM	0.942	0.898–0.986	0.892	0.865	0.912	0.882	0.899	0.882	0.865	0.874	training set
LightGBM	0.864	0.766–0.961	0.750	0.520	0.963	0.929	0.684	0.929	0.520	0.667	test set

LR: Logistic Regression; SVM: Support Vector Machine; LightGBM: Lightweight Gradient Boosting Machine; RF: Random Forest; ET: ExtraTrees; CI: Confidence Interval

Table 7: Comparison of fusion-phase imaging-based biomarker indices across different models

Model	AUC	95% CI	Precision	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy	Recall	F1	group
LR	0.926	0.877–0.975	0.867	0.769	0.941	0.909	0.842	0.909	0.769	0.833	training set
LR	0.893	0.802–0.984	0.846	0.920	0.778	0.793	0.913	0.793	0.920	0.852	test set
SVM	0.957	0.922–0.992	0.892	0.923	0.868	0.842	0.937	0.842	0.923	0.881	training set
SVM	0.890	0.805–0.975	0.808	0.800	0.815	0.800	0.815	0.800	0.800	0.800	test set
RF	0.978	0.956–1.000	0.942	0.942	0.941	0.925	0.955	0.925	0.942	0.933	training set
RF	0.876	0.785–0.966	0.788	0.920	0.667	0.719	0.900	0.719	0.920	0.807	test set
ET	0.899	0.846–0.953	0.817	0.808	0.824	0.778	0.848	0.778	0.808	0.792	training set
ET	0.907	0.824–0.990	0.846	0.760	0.926	0.905	0.806	0.905	0.760	0.826	test set
LightGBM	0.949	0.912–0.987	0.883	0.788	0.956	0.932	0.855	0.932	0.788	0.854	training set
LightGBM	0.839	0.735–0.943	0.750	0.560	0.926	0.875	0.694	0.875	0.560	0.683	test set

LR: Logistic Regression; SVM: Support Vector Machine; LightGBM: Lightweight Gradient Boosting Machine; RF: Random Forest; ET: ExtraTrees; CI: Confidence Interval