Radiography 31 (2025) 103010

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

The impact of intensity discretisation and filtration on the performance of the radiomic and machine learning models in brain metastasis patients treated with gamma knife radiosurgery

A. Umaru [a,b], H.A. Manan [c], R.K.A. Kumar [d], S.K. Hamsan [d], N. Yahya [a,*]

[a] *Diagnostic Imaging and Radiotherapy, CODTIS, Faculty of Health Sciences, National University of Malaysia, Jalan Raja Muda Aziz, 50300, Kuala Lumpur, Malaysia*

[b] *Department of Medical Radiography, Ahmadu Bello University Zaria, Nigeria*

[c] *Department of Radiology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, Kuala Lumpur, Malaysia*

[d] *Gamma Knife Centre, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia*

A R T I C L E I N F O

Article history:
Received 5 December 2024
Received in revised form
13 May 2025
Accepted 5 June 2025
Available online xxx

Keywords:
Intensity discretisation
Filtration
Radiomic
Machine learning
Brain metastasis

A B S T R A C T

Introduction: Image preprocessing is crucial for optimizing radiomics feature extraction, however, inconsistencies in the implementation process and a lack of universally accepted methods lead to diverse approaches. This study evaluates the impact of radiomics and machine learning (ML) performance in brain metastasis.

Methods: In this retrospective study, 108 lesions (regions of interest, ROI) were analysed. Using contrast-enhanced T1-weighted (T1W) images resampled to $1 \times 1x1mm^3$, radiomics features were extracted using various fixed bin sizes (8, 16, 32, 64, 128, 256) and bin numbers (1, 5, 10, 25, 50) with relative ROI (max-min) intensity rescaling. Intensity rescaling methods, including 64 bins with relative mean ROI $\pm$3SD and relative min–max rescaling, were applied. Additionally, four filters (Laplacian of Gaussian (LOG), Wavelet, Laws, and Mean) were tested. Feature selection was conducted using correlation analysis and Least Absolute Shrinkage and Selection Operator (LASSO) regression, yielding 11–18 features. K-nearest neighbors (KNN) was used for classification, and Synthetic Minority Over-sampling Technique (SMOTE) addressed class imbalance.

Results: The 32-bin model achieved the highest accuracy (70 %) and AUC (0.70; 95 % CI: 0.51–0.87) among the bin sizes, while the 10-bin model performed best among bin numbers, with accuracy of 79 % and AUC of 0.69 (95 % CI: 0.45–0.89). No significant difference was found in accuracy ($p = 0.278$) and AUC ($p = 0.288$) between bin sizes and numbers. Mean relative ROI $\pm$3SD rescaling improved accuracy (73 % vs 61 %) and AUC (0.74 vs 0.60). The LOG and mean filters demonstrated superior performance.

Conclusion: Image preprocessing choices significantly influence radiomics features and ML performance. Standardized preprocessing is essential for enhancing reproducibility.

Implications for practice: Standardizing preprocessing methods can improve the reliability and generalizability of radiomics models, with potential applications in clinical decision-making for brain metastasis treatment.

Introduction

Radiomics is a non-invasive technique that extracts large amounts of quantitative data from medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and ultrasound.[1] This emerging tool provides valuable insights into tumor phenotypes, potentially serving as a “virtual biopsy”.[2,3] Among these modalities, MRI is particularly favored for diagnosing brain tumors and planning Gamma Knife radiosurgery, owing to its superior soft tissue contrast.[4–7] Numerous studies have demonstrated that integrating radiomics with machine learning (ML), can assist in predicting treatment outcomes for brain metastasis.[8–14]

A critical step in radiomics analysis is image preprocessing, which often includes normalization, filtering, intensity rescaling, and intensity discretisation.[3,15] The appropriate selection of image discretisation parameters can harmonise the process, reducing

* Corresponding author.
E-mail address: azrulyahya@ukm.edu.my (N. Yahya).

https://doi.org/10.1016/j.radi.2025.103010
1078-8174/

non-biological variability introduced by differences in image acquisition parameter.[16] The Imaging Biomarker Standardisation Initiative (IBSI) recommends resampling images and discretising intensities using different bin size and numbers. This standardisation aligns voxel sizes, simplifies intensity ranges, and minimise image noise.[17] Filters are also recommended to improve intensity resolution and highlight specific image properties.[3] However, inconsistencies in the implementation process and a lack of universally accepted methods leads to diverse approach to image pre-processing.[18]

Previous studies have explored the impact of intensity discretisation on the number of features, reproducibility, and stability in CT[19] and PET[20,21] and, to a lesser extent, in MRI.[22–25] However, the number of features does not necessarily correlate with model performance in radiomics. To our knowledge, no study has specifically investigated the impact of intensity discretisation, intensity rescaling, and filters on radiomics model performance in MRI-based studies. In this context, the present study aims to assess the impact of varying intensity discretisation parameter, intensity rescaling and filters on the radiomics and ML performance in brain metastasis using contrast enhanced T1-weighted images (CE-T1WI).

Methods

Patients and image acquisition

The study included 108 brain metastases from 30 patients who underwent stereotactic radiosurgery (SRS) using Gamma Knife at the specialist centre in Hospital Canselor Tuanku Muhriz (HCTM), Kuala Lumpur. Only patients who met the inclusion criteria were included: 1) brain metastasis patients treated with Gamma knife; 2) availability of follow-up MRI data; and 3) availability of Contrast enhanced T1 sequences. MRI data, including pretreatment and follow-up scans, were acquired on a Siemens Magnetom Avanto 1.5 T system with a dedicated Tx/Rx CP head coil and gadolinium contrast agent. Pretreatment images had a resolution of 288 $\times$ 288 pixels, and a slice thickness of 1.5 mm. These were processed using various discretisation techniques and filtering methods for radiomics features extraction. Follow-up images were utilized to assess treatment response. Tumors were classified using the Response Assessment in Neuro-Oncology Brain metastasis (RANO-BM) criteria. We classified the complete response and partial response groups as “good response” while the stable disease and progressive disease groups were classified as “poor response”.[26,27]

ROI segmentation

Regions of interest (ROIs) were manually segmented slice by slice by an experienced neurosurgeon using the Gamma Knife planning system. These were verified by an experienced neuroradiologist. The DICOM data and ROIs were uploaded to LifeX radiomics software for feature extraction. Fig. 1 shows the example of tumor segmentation. Following IBSI guidelines, images were resampled to a voxel size of 1 $\times$ 1x1 mm^3 to harmonise voxel size across scans.[17] The DICOM data and segmented ROI were imported into the LifeX radiomics software (v7.6.0) for features extraction.[28] LifeX calculates texture features only on ROIs containing 64 or more voxels to ensure reliable texture computation.

Features extraction

Intensity discretisation

Radiomics features were extracted from all 108 tumors, each containing more than 64 voxels, which is sufficient for texture feature calculations. Various fixed bin sizes (8, 16, 32, 64, 128, and 256) were tested to identify the most suitable bin size for capturing relevant texture within the ROI. In addition to testing different bin sizes, a range of fixed bin numbers (1, 5, 10, 25, and 50) were also tested to assess their impact on model performance. This method divides intensity values into fixed numbers across the entire dataset.

A relative ROI (max-min) intensity rescaling was employed to normalize pixel values, The maximum and minimum intensities within each ROI were identified, and pixel intensities were rescaled accordingly. At this stage, no filters were applied, ensuring that the variation in performance was not influenced by any filter.

Intensity rescaling

Additionally, two different intensity rescaling methods were tested to assess the impact of intensity rescaling on radiomic features extraction and subsequently on model performance. Two different intensity rescaling methods tested were:

1. Relative ROI (min–max) rescaling: This method ensures the intensity values within the ROI are distributed between the minimum and maximum value of the intensities in the ROI.
2. Relative ROI (mean $\pm$ 3 SD) rescaling: this method rescales intensity values using the mean and 3 standard deviations above and below the mean of each ROI to limit the influence of extreme values or outliers.

Both intensity recalling methods were independently applied, with a fixed bin size of 64, allowing for an exploration of different normalization strategies on radiomics feature performance.

Filters

To further explore the impact of filters on radiomics, four filters were applied before feature extraction, using a fixed bin size of 64 and relative ROI $\pm$3 SD intensity rescaling:

1. Laplacian of Gaussian (LOG): Enhance edges and detects fine details by highlighting regions with rapid intensity changes, helping to identify boundaries within the ROI that may be associated with significant clinical outcomes.
2. Wavelet filter: Decomposed images into multiple frequency components, allowing us to capture both fine and coarse image patterns in the BM region, providing a comprehensive assessment of texture across multiple scale.
3. Laws: Focuses on capturing texture patterns such as roughness and smoothness within the ROI, which are often not apparent through visual inspection.
4. Mean filter: Smooths intensity variation and reduces noise, preserving larger structures within the ROI and emphasizing global texture features rather than pixel-level noise.

All radiomic features were extracted using LifeX (v7.6.0) radiomics software.[28] A total of 179 features were extracted for each model, including 56 *Texture features, 50 Morphological features and 73 intensity features*.

The original MRI and segmentation data, as well as the clean extracted radiomic dataset supporting these findings, are available upon request to the corresponding author, subject to institutional and ethical guidelines.

Features selection

To mitigate overfitting and ensure the models generalizability, several strategies were implemented. First, pairwise correlation analysis was performed among to eliminate highly correlated features (with a threshold of 0.8).[29] This step reduced the

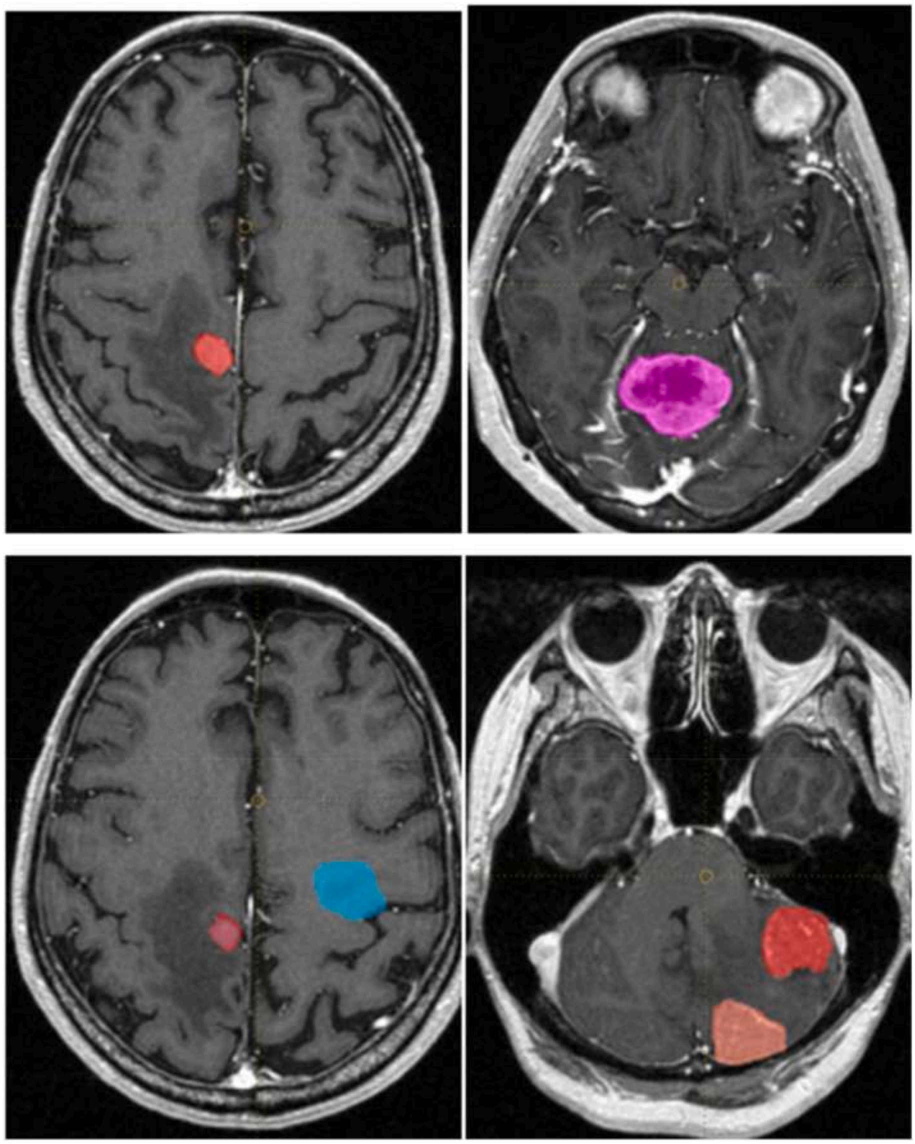

Figure 1. CE-T1WI showing segmentation of six ROIs in different colors in four patients.

complexity of the model by removing redundant features and retaining only unique, informative ones.

Next, LASSO (Least Absolute Shrinkage and Selection Operator) regression was applied as a regularization technique. LASSO shrinks the coefficients of less important features toward zero, effectively eliminating them and further reducing the feature set. Only the most significant features, with non-zero coefficients, were retained.

Additionally, SMOTE (Synthetic Minority Over-sampling Technique) was applied to address class imbalance, by generating synthetic data for minority class. To assess the importance of each feature and prevent overfitting, permutation feature importance was also used, identify retaining only the significantly contributed to the model's performance. All of these steps were taken to mitigate this risk of bias, reduce overfitting, and improve the model's ability to generalize across different datasets.

Machine learning model classifier

The K-nearest neighbours (KNN) algorithm was selected as the classifier due to its simplicity and effectiveness in handling complex, multi-dimensional data. The model was trained on 70 % of the dataset, with the remaining 30 % used for testing and validating the model's performance. A K-value of 5 and 1000 bootstraps were applied to ensure robust performance, with the goal to predict treatment response in brain metastasis patients.

KNN was chosen for this study due to its simplicity, interpretability, and non-parametric nature. This study was exploratory, aimed at assessing how varying preprocessing steps affect model performance rather than optimizing the classifier. KNN provided an accessible baseline for comparison, allowing us to focus on the impact of preprocessing without the complexity of more advanced models. $K = 5$ was selected as it offers a balance between bias and variance, providing a reasonable trade-off for generalization without overfitting to noise. Additionally, 1000 bootstraps were used to assess model stability, providing more reliable estimates of performance metrics and helping quantify the variability of the model due to the small sample size.

Addressing class imbalance using SMOTE

To mitigate bias toward the minority class, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. which generates synthetic data for the minority class by interpolating existing samples, ensuring a more balance class distribution in the training set.[30,31]

Model performance evaluation

The KNN model's performance was assessed using accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC) metrics. To further evaluate the role of individual features, permutation feature importance was applied, which involved randomly shuffling each feature's values and measuring the resulting decrease in model performance, thus identifying influential features.

All machine learning workflows, including feature selection, model training, and performance evaluation, were implemented using Python's Scikit-learn library (V3.12.1). The radiomic workflow is illustrated in Fig. 2.

Results

A total of 108 brain metastasis lesions from 30 patients (13 male and 17 female), who met the inclusion criteria, were analysed. Table 1 provides a summary of the patient and tumor characteristics. Radiomics features were extracted using varying bin sizes, bin numbers, intensity rescaling methods, and filters to evaluate their effect on model performance.

Table 1
Patient demographics and tumor characteristics.

	N	Minimum	Maximum	Mean	Std. Deviation
Age (years)	30	38	88	58.9	13.68
Follow-up time (months)	30	1	33	6.82	6.27
Pre-treatment volume (cm^3)	108	0.010	45.334	2.966	5.790
Post-treatment volume (cm^3)	108	0.000	39.455	2.294	6.636
% Volume changes (cm^3)	108	−909.201	100.00	13.890	138.079

The selected features for the best model (Relative ROI: mean ± 3 SD rescaling) includes: 5 *Morphological* (Compacity, Compactness2, Centre of Mass Shift, Radius Sphere Norm-Max Intensity Coor-Roi Centroid Coor-Dist, Radius Sphere Norm-Max Intensity Coor-PerimeterCoor-2DCoronalSmallestDist), 5 *Intensity-based*

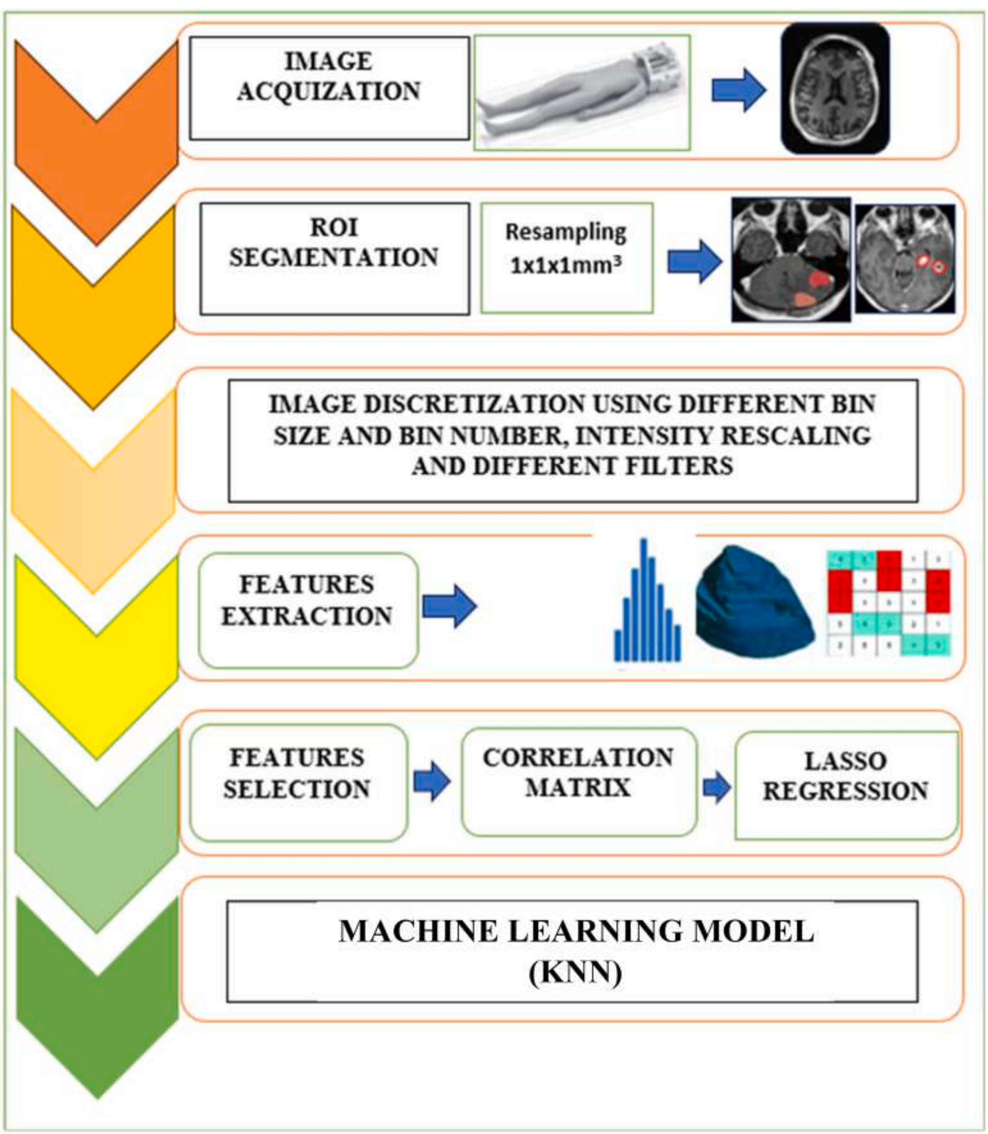

Figure 2. Radiomic workflow: starting from image acquisition, ROI segmentation, intensity discretisation and feature extraction, followed by features selection and machine learning model generation.

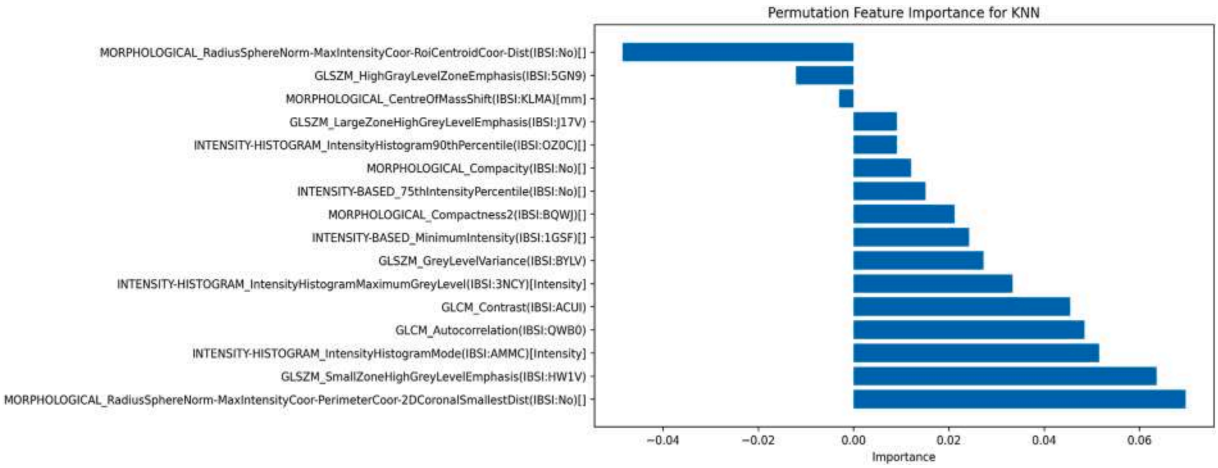

Figure 3. Final selected features for the best performing model (Relative ROI Mean ± 3SD) and their Importance.

(Minimum Intensity, 75th Intensity Percentile, Intensity Histogram Mode, Intensity Histogram 90th Percentile, Intensity Histogram Maximum Grey Level), and 6 *Texture* (including 2 GLCM (Contrast, Auto correlation), and 4 GLSZM (High Gray Level Zone Emphasis, Small Zone High Grey Level Emphasis, Large Zone High Grey Level Emphasis, Grey Level Variance) (see Fig. 3).

Details of these features, including their contribution to model performance as calculated using permutation feature importance, are presented in the supplementary material (Supplementary Figs. S1–S17). These figures outline the importance of individual features in model prediction, helping highlight the most influential features for improving predictive accuracy.

Bin size analysis

Six bin sizes (8, 16, 32, 64, 128, and 256) were tested for their impact on model performance. Models using bin size of 32, 64, and 128 demonstrated the best performance, with the 32-bin model achieving the highest accuracy of 70 %, and AUC of 0.70 (95 % CI: 0.51–0.87). The model trained using the 8-bin size performed the worst, with an accuracy of 52 % and AUC of 0.53 (0.33–0.73). Models trained with bin sizes of 16 and 256 also performed sub optimally, with accuracies of 58 % and 55 %, and AUCs of 0.57 (0.30–0.82), respectively.

Bin number analysis

Five bin numbers (1, 5, 10, 25 and 50) were evaluated for their effect on model performance. Models using bin numbers of 5, 10 and 25 showed comparable results, with the model using bin number of 10 achieving the best overall performance: 79 %, and AUC of 0.69 (95 % CI: 0.45–0.89). While variations in individual model performance were observed, no statistically significant differences were found in accuracy ($P = 0.278$) or AUC ($P = 0.288$) among models with different bin numbers and sizes.

Effect of intensity rescaling

The model trained with a bin size of 64 combined with relative mean ROI ± 3SD intensity rescaling outperformed those using relative ROI (min–max) rescaling. The mean ROI ± 3SD method achieved an accuracy of 73 %, sensitivity of 71 %, specificity of 71 %, and AUC of 0.74 (95 % CI: 0.55–0.91), compared to the relative ROI (min–max) intensity recalling method, which yielded an accuracy of 61 %, sensitivity of 59 %, specificity of 59 %, and an AUC of 0.60 (95 % CI: 0.36–0.81).

Filter application

The impact of four filters (LOG, Law, Mean and Wavelet) was examined. The Laplacian of Gaussian (LOG) and Mean filters produced the best results, with the LOG filter achieving 73 % accuracy and an AUC of 0.66 (0.40–0.89), and the Mean filter achieving 76 % accuracy and an AUC of 0.64 (0.43–0.85). In contrast, the Wavelet filter underperformed, with an accuracy of 48 % and AUC of 0.60 (0.34–0.84). Table 2 provides a detailed comparison of model performance, while Fig. 4 presents a bar chart summarizing the accuracy and AUC across models.

Discussion

Intensity discretisation is an essential preprocessing step in radiomics, offering noise-suppressing properties and optimizing texture calculations.[17] While previous studies have evaluated the number of robust features derived from various discretisation methods,[18] the quantity of features alone does not guarantee better model performance. This study explored the impact of intensity discretisation, rescaling, and filters on radiomics based ML models in predicting treatment response in brain metastasis. Our findings confirm that grey level and intensity discretisation significantly affect predictive performance.

The analysis of bin sizes revealed that models using 32, 64, and 128 demonstrated the best performance. The model using the 32-bin size achieved the highest performance with an accuracy of 70 % and an AUC of 0.70. In contrast, the model trained using the 8-bin size performed the worst. Models trained with bin sizes of 16 and 156 also performed sub optimally. These findings emphasize the importance of optimizing bin size to improve model efficacy, allowing the model to capture essential patterns without overfitting. Moderate bin sizes (32, 54, and 128) strike an optimal balance between features details and noise, allowing the model to capture essential patterns without overfitting. These findings align with Zhao et al.,[24] who compared different models against a 64-bin and found no difference in performance between 64 and 32 bins but reported fewer reproducible features for both high (256) and low[16] bins sizes. Similar results have been supported by Khodabakhshi et al.,[23] and Molina et al.,[22,32] contradicting the notion that increasing bin numbers and sizes inherently enhances

Table 2
Features selection for the various discretisation methods, intensity rescaling and filters used. The performance of each model is given in terms of accuracy, sensitivity, specificity, and AUC. at 95 % CI. * Represent the best performing model for each group.

	number of features selected by correlation matrix	number of features selected by LASSO	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (95 % CI)
Bin size						
8	32	15	52	52	52	0.53 (0.33–0.73)
16	40	13	58	55	55	0.57 (0.34–0.78)
32	41	13	70	67.3	67.3	0.70 (0.51–0.87)*
64	44	14	61	59	59	0.60 (0.36–0.81)
128	49	14	61	72	72	0.62 (0.42–0.80)
256	45	16	55	53	53	0.57 (0.30–0.82)
Bin number						
1	86	11	55	50	50	0.55 (0.35–0.77)
5	53	18	64	40	40	0.37 (0.15–0.61)
10	49	14	79	63	63	0.69 (0.45–0.89)*
25	44	16	58	55	55	0.61 (0.41–0.79)
50	46	17	52	55	55	0.54 (0.31–0.75)
Intensity rescaling						
Relative ROI (max-min)	44	14	61	59	59	0.60 (0.36–0.81)
Relative ROI mean ± 3SD	51	17	73	71	71	0.74 (0.55–0.91)*
Filters						
LOG	50	16	73	69	69	0.66 (0.40–0.89)*
Law	56	17	58	59	59	0.56 (0.32–0.79)
Mean	52	15	76	62.5	62.5	0.64 (0.43–0.85)
Wavelet	32	17	48	58	58	0.60 (0.34–0.84)

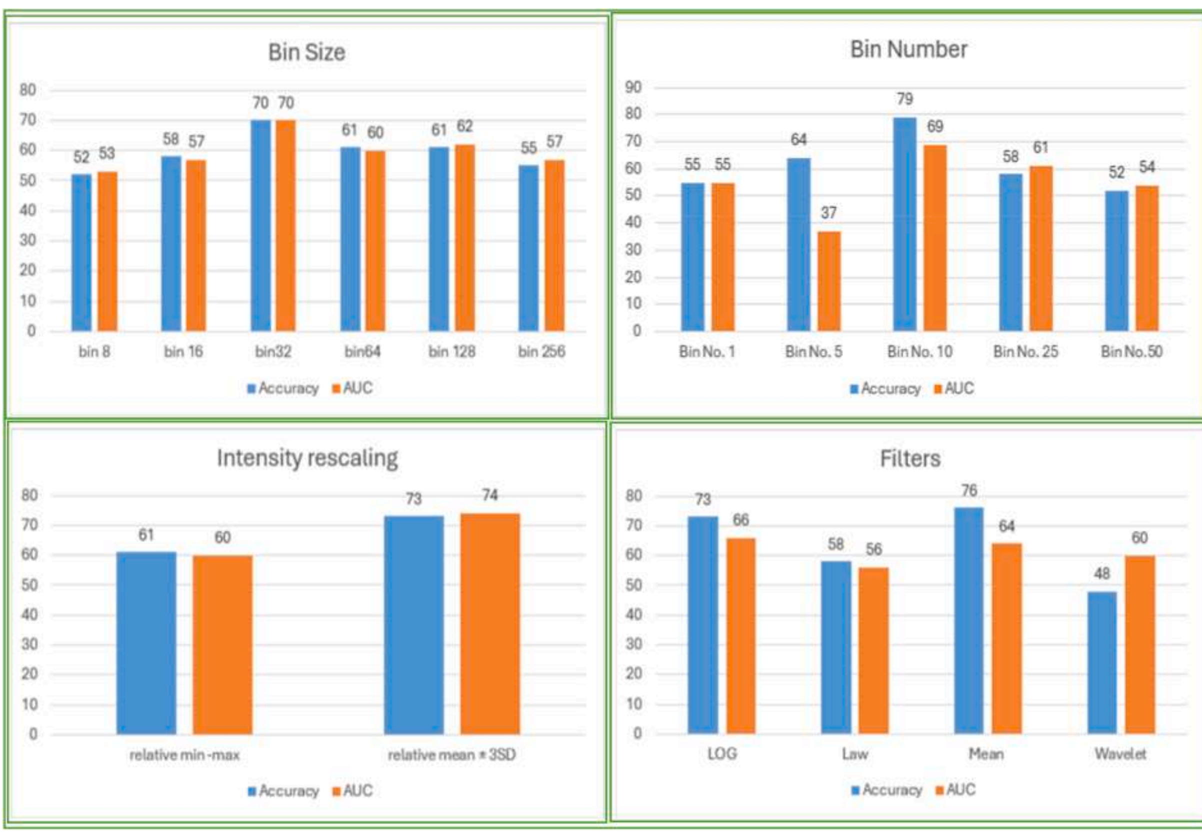

Figure 4. Performance comparison of accuracy and AUC between models.

model stability.[23,25] While higher bin counts may increase the number of features, they do not necessarily improve model performance. This finding is further supported by studies that use 32 bins for feature extraction due to its performance.[16]

Analysis of fixed bin numbers revealed that a bin number of 10 yielded the best performance, with an accuracy of 79 % and an AUC of 0.69. Although these results are promising, the **exploratory** nature of the study suggests that there is still room for improvement in terms of clinical applicability. Further work with larger, multi-center datasets and more advanced classifiers will likely improve the model's clinical utility. In clinical practice, a higher **AUC** is desired for making reliable predictions regarding treatment response.

While variation in individual model performance was observed, no statistically significant difference was found in accuracy

($p = 0.278$) or AUC ($p = 0.288$) among the combined bin numbers and sizes. Models based on fixed bin sizes (FBS) marginally outperformed fixed bin numbers (FBN), demonstrating slightly better performance overall. Similar findings were reported by Duron et al.,[18] who observed that FBS models generally outperform FBN models, likely due to their ability to better preserve original intensity values, which may enhance model performance.

The lack of statistically significant differences between bin sizes and bin numbers reflects the comparison across all models as a group, rather than evaluating individual model performance in isolation. Although some models were selected based on their overall performance in the group, the choice was made empirically to find the best combination of model performance metrics, including accuracy and AUC, despite the absence of statistical significance, this approach aimed at understanding the impact of various of various preprocessing techniques.

Intensity rescaling techniques also significantly influenced model performance. Relative intensity rescaling (mean ROI ± 3SD) achieved higher accuracy (73 %) and AUC (0.74, 0.55–0.91), compared to relative intensity ROI (min – max) method (accuracy: 61 %, AUC 0.60 95 % 0.36–0.81). This underscore the critical role of rescaling in radiomic preprocessing, consistent with Leijenaar et al.,[21] who demonstrated the dependency of texture features on intensity resolution. The mean ROI ± 3SD rescaling method is preferred for improving model accuracy and robustness because it limits the influence of extreme values or outliers, ensuring that the intensity values are normalised in a way that retains important features while minimizing noise, thereby improving reproducibility of texture features across datasets.

Furthermore, studies have shown that rescaling techniques that normalize intensity values within a consistent range improve the predictive power,[33] demonstrating that normalization methods improve texture features stability of texture features. The min–max rescaling method, while simple and effective in some context, tend to scale the data to a fixed range, potentially amplifying the influence of outliers. This may explain why mean ROI ± 3SD rescaling method outperformed min–max rescaling in our study. In clinical radiomics, where the aim is to ensure that features extracted are robust and reproducible across different scanners, protocols, and patient population, mean ROI ± 3SD is considered more reliable.

The evaluation of filters revealed that the Laplacian of Gaussian (LOG) and mean filters provided superior performance. LOG yielded an accuracy of 73 % and an AUC of 0.66 (95 % CI: 0.40–0.89), while the mean filter achieved 76 % accuracy and an AUC of 0.64 (95 % CI: 0.43–0.85). In contrast, the wavelet filter underperformed, suggesting it may be unsuitable for this specific application in brain metastasis. The LOG filter is a well-known edge-detection filter that enhances fine details by emphasizing regions with rapid intensity changes. This characteristic is particularly important when analysing tumor boundaries or fine structural differences in medical imaging. Bologna et al.,[34] noted that while Gaussian filters, such as LOG, often provide modest improvements in model performance, they are particularly useful for enhancing the clarity of image edges, which helps the model better distinguish between tumor and normal tissue. Our finding suggest that LOG and mean filters are more suitable for improving model performance in this specific application. These filters preserve relevant structural information while reducing noise, leading to more reliable radiomics features for machine learning models.

The supplementary material provides a detailed breakdown of permutation features importance across different preprocessing technique (bin sizes, intensity rescaling, and filters). Certain features, such as compacity and compacness2, showed greater impact with higher bin sizes (e.g., 128 and 256), while intensity histogram mode and minimum intensity were more influential under relative mean ROI ± 3SD rescaling. The LOG and mean filters notably enhanced GLCM contrast and GLSZM features, aligning with our results that these preprocessing have different impact on extracted features and model performance.

Although this single centre study minimized variability, the result highlights the substantial impact of preprocessing methods on radiomics features and ML model performance. Thess findings offer valuable guidance for researchers and clinicians in selecting appropriate preprocessing parameters.

Given the small sample size in this study, overfitting was a potential concern. However, several techniques were used to reduce this risk. We performed feature selection through pairwise correlation analysis to eliminate highly correlated features, and LASSO regression, a regularization technique, was applied to shrink the coefficients of less important features toward zero. To address potential class imbalance, we used SMOTE, which generates synthetic samples for the minority class. Additionally, permutation feature importance was used to assess the stability and relevance of the features, ensuring that only the most significant features were retained. These methods collectively ensure the model performs reliably despite the limited dataset, providing a better estimate of its generalizability.

While the study provides valuable insights into the impact of intensity discretization, rescaling, and filtering of radiomic and machine learning models, there are several limitations that should be considered. The relatively small sample size consisting of 30 patients and 108 lesions, may limits the generalizability of the result. A larger cohort would help strengthen the robustness of the findings and reduce risk of overfitting.

Additionally, the study relied on a single pulse sequence (CE-T1WI), which, while commonly used for tumor delineation, treatment planning and most radiomics study, it may not capture all relevant tumor characteristics. While the main aim of this study was to assess the impact of preprocessing, the inclusion of multiple MRI sequences, such as T2 and FLAIR, could provide complementary information on tumor heterogeneity, improving model performance. Including multiple MRI sequences would allow for a more comprehensive assessment of tumor's structural and functional probertites, which could further enhance model performance.

Furthermore, the use of a single center dataset may limit the applicability of the results to broader populations. While multi-center data would help assess the robustness of preprocessing methods across different institutions, imaging protocols, and patient populations, the focus of this study was specifically to explore how different preprocessing steps affect model performance within a controlled, consistent dataset. Introducing multi-center could add variations, such as differences in scanner type and protocols, that may not be directly relevant to the primary objective of evaluating preprocessing technique. By minimize confounding variables, this study aimed to isolate the impact of preprocessing methods. However, future studies incorporating multi-sequence MRI data and multi-center datasets could enhance the external validity of the finding by allowing the models to be tested on more varied data, improving the generalizability of radiomic features across different clinical contexts.

A critical next step in improving the generalizability and clinical applicability of our findings is to perform external validation using multi-center datasets. This would allow for the evaluation of the model's robustness across different institutions, imaging protocols, and scanner types. Moreover, the integration of multiple MRI sequences would enable the model to capture a broader range of tumor characteristics, improving its robustness for clinical decision-making.

Conclusion

The study highlights the significant impact of intensity discretisation, intensity rescaling, and filters on radiomics and ML performance in predicting treatment response for brain metastasis. Standardising preprocessing parameters is crucial for enhancing reproducibility and ensuring and ensuring the clinical applicability of radiomics-based models. Future studies should focus on external validation as a key step in establishing the generalizability and clinical applicability of our model in real-world settings. By incorporating diverse datasets, including various MRI sequences, and testing across multiple centres, we can ensure that our findings are both robust and reproducible across different clinical environments.

Ethical statement

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional/regional/national ethics/committee/ethics board of ethical committee of Universiti Kebangsaan Malaysia (UKM.FSK.PNI.800-2/27/9 (JEP-2022-524)) and individual consent for this retrospective analysis was waived.

Contribution of authors

AU, HAM, RKAK, SKH and NY conceptualised the idea. AU, NY and HAM developed the methodology. AU, RKAK and SKH collected the data; AU, HAM and NY analysed the data. AU, and NY wrote the original draft. HAM, RKAK and SKH reviewed and edited the manuscript. All authors read and agreed to the final draft for publication.

Funding

No funds, grants, or other support was received.

Conflict of interest statement

All authors have no conflicts of interest to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radi.2025.103010.

References

1. Du P, Liu X, Shen L, Wu X, Chen J, Chen L, et al. Prediction of treatment response in patients with brain metastasis receiving stereotactic radiosurgery based on pre-treatment multimodal MRI radiomics and clinical risk factors: a machine learning model. *Front Oncol* 2023;**13**.
2. Derks SH, van der Veldt AA, Smits M. Brain metastases: the role of clinical imaging. *Br J Radiol* 2022;**95**(1130):20210944.
3. Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JP, et al. Radiomics in oncology: a practical guide. *Radiographics* 2021;**41**(6):1717–32.
4. Liao C-Y, Lee C-C, Yang H-C, Chen C-J, Chung W-Y, Wu H-M, et al. Enhancement of radiosurgical treatment outcome prediction using MRI radiomics in patients with non-small cell lung cancer brain metastases. *Cancers* 2021;**13**(16):4030.
5. Tan D, Mohamad Salleh SA, Manan HA, Yahya N. Delta-radiomics-based models for toxicity prediction in radiotherapy: a systematic review and meta-analysis. *J Med Imag Radiat Oncol* 2023;**67**(5):564–79.
6. Manan HA, Yahya N, Nor Shafiza AW, Yusoff AN. *Distribution pattern of brain tumour types and location in patients who underwent MRI scans: a survey of 1240 patients in a Tertiary Malaysian University Hospital*. 2022.
7. Yahya N, Manan HA. Utilisation of diffusion tensor imaging in intracranial radiotherapy and radiosurgery planning for white matter dose optimization: a systematic review. *World Neurosurg* 2019;**130**:e188–98.
8. Chang E, Joel MZ, Chang HY, Du J, Khanna O, Omuro A, et al. Comparison of radiomic feature aggregation methods for patients with multiple tumors. *Sci Rep* 2021;**11**(1).
9. Casà C, Piras A, D'Aviero A, Preziosi F, Mariani S, Cusumano D, et al. The impact of radiomics in diagnosis and staging of pancreatic cancer. *Therap Adv Gastrointestin Endosc* 2022;**15**. 26317745221081596.
10. Guglielmo P, Marturano F, Bettinelli A, Gregianin M, Paiusco M, Evangelista L. Additional value of PET radiomic features for the initial staging of prostate cancer: a systematic review from the literature. *Cancers* 2021;**13**(23):6026.
11. Zhang Z, Yang J, Ho A, Jiang W, Logan J, Wang X, et al. A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. *Eur Radiol* 2018;**28**:2255–63.
12. Salari E, Elsamaloty H, Ray A, Hadziahmetovic M, Parsai EI. Differentiating radiation necrosis and metastatic progression in brain tumors using radiomics and machine learning. *Am J Clin Oncol* 2023;**46**(11):486–95.
13. Cao Y, Parekh VS, Lee E, Chen X, Redmond KJ, Pillai JJ, et al. A multidimensional connectomics-and radiomics-based advanced machine-learning framework to distinguish radiation necrosis from true progression in brain metastases. *Cancers* 2023;**15**(16):4113.
14. Tan D, Nasir NM, Manan HA, Yahya N. Prediction of toxicity outcomes following radiotherapy using deep learning-based models: a systematic review. *ancer Radiother* 2023;**27**(5):398–406.
15. Umaru A, Abdul Manan H, Kumar Athi Kumar R, Hamsan SK, Yahya N. Utilizing radiomics as predictive factor in brain metastasis treated with stereotactic radiosurgery: systematic review and radiomic quality assessment. *iRADIOLOGY* 2025;**3**(2):132–43.
16. Li Y, Ammari S, Balleyguier C, Lassau N, Chouzenoux E. Impact of preprocessing and harmonization methods on the removal of scanner effects in brain MRI radiomic features. *Cancers* 2021;**13**(12):3000.
17. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020;**295**(2):328–38.
18. Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-C, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS One* 2019;**14**(3):e0213459.
19. Larue RT, van Timmeren JE, de Jong EE, Feliciani G, Leijenaar RT, Schreurs WM, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol* 2017;**56**(11):1544–53.
20. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015;**60**(14):5471.
21. Leijenaar RT, Nalbantov G, Carvalho S, Van Elmpt WJ, Troost EG, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* 2015;**5**(1):11075.
22. Molina D, Pérez-Beteta J, Martínez-González A, Martino J, Velásquez C, Arana E, et al. Influence of gray level and space discretization on brain tumor heterogeneity measures obtained from magnetic resonance images. *Comput Biol Med* 2016;**78**:49–57.
23. Khodabakhshi Z, Gabrys H, Wallimann P, Guckenberger M, Andratschke N, Tanadini-Lang S. Magnetic resonance imaging radiomic features stability in brain metastases: impact of image preprocessing, image-, and feature-level harmonization. *Phys Imag Radiat Oncol* 2024;**30**:100585.
24. Zhao W, Hu Z, Kazerooni AF, Körzdörfer G, Nittka M, Davatzikos C, et al. Physics-informed discretization for reproducible and robust radiomic feature extraction using quantitative MRI. *Investig Radiol* 2024;**59**(5):359–71.
25. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep* 2020;**10**(1):12340.
26. DeVries DA, Lagerwaard F, Zindler J, Yeung TPC, Rodrigues G, Hajdok G, et al. Performance sensitivity analysis of brain metastasis stereotactic radiosurgery outcome prediction using MRI radiomics. *Sci Rep* 2022;**12**(1).
27. Du P, Liu X, Xiang R, Lv K, Chen H, Liu W, et al. Development and validation of a radiomics-based prediction pipeline for the response to stereotactic radiosurgery therapy in brain metastases. *Eur Radiol* 2023;**33**(12):8925–35.
28. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res* 2018;**78**(16):4786–9.
29. An MR radiomics framework for predicting the outcome of stereotactic radiation therapy in brain metastasis. In: Karami E, Ruschin M, Soliman H, Sahgal A, Stanisz GJ, Sadeghi-Naini A, editors. *2019 41st annual international conference of the*. IEEE: IEEE Engineering in Medicine and Biology Society (EMBC); 2019.
30. Kawahara D, Jensen A, Yuan J, Nagata Y, Watanabe Y. Predicting the BRAF mutation with pretreatment MRI radiomics features for melanoma brain metastases receiving Gamma Knife radiosurgery. *Clin Radiol* 2023;**78**(12):E934–40.
31. Yahya N, Ebert MA, Bulsara M, House MJ, Kennedy A, Joseph DJ, et al. Statistical-learning strategies generate only modestly performing predictive models

for urinary symptoms following external beam radiotherapy of the prostate: a comparison of conventional and machine-learning methods. *Med Phys* 2016;**43**(5):2040–52.
32. Molina D, Pérez-Beteta J, Martínez-González A, Martino J, Velasquez C, Arana E, et al. Lack of robustness of textural measures obtained from 3D brain tumor MRIs impose a need for standardization. *PLoS One* 2017;**12**(6).
33. Wahid KA, He R, McDonald BA, Anderson BM, Salzillo T, Mulder S, et al. Intensity standardization methods in magnetic resonance imaging of head and neck cancer. *Phys Imag Radiat Oncol* 2021;**20**:88–93.
34. Bologna M, Corino V, Mainardi L. Virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain. *Med Phys* 2019;**46**(11):5116–23.