Strahlentherapie und Onkologie

Non-invasive Prediction of Secondary Enucleation Risk in Uveal Melanoma Based on Pretreatment CT and MR Imaging Prior to Stereotactic Radiotherapy

--Manuscript Draft--

Manuscript Number:	SUON-D-24-00281R1
Full Title:	Non-invasive Prediction of Secondary Enucleation Risk in Uveal Melanoma Based on Pretreatment CT and MR Imaging Prior to Stereotactic Radiotherapy
Article Type:	Original Article
Corresponding Author:	Yagiz Yedekci Hacettepe University: Hacettepe Universitesi Ankara, Ankara TÜRKIYE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Hacettepe University: Hacettepe Universitesi
Corresponding Author's Secondary Institution:	
First Author:	Yagiz Yedekci
First Author Secondary Information:	
Order of Authors:	Yagiz Yedekci
	Hidetaka Arimura
	Yu Jin
	Melek Tugce Yilmaz
	Takumi Kodama
	Gokhan Ozyigit
	Gozde Yazici
Order of Authors Secondary Information:	
Funding Information:	Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (2219) — Dr. Yagiz Yedekci
Abstract:	Purpose: To develop a radiomic model to non-invasively predict the risk of secondary enucleation (SE) in patients with uveal melanoma (UM) prior to stereotactic radiotherapy using pretreatment computed tomography (CT) and magnetic resonance (MR) images. Materials and Methods: This retrospective study encompasses a cohort of 308 patients diagnosed with UM who underwent stereotactic radiosurgery (SRS) or fractionated stereotactic radiotherapy (FSRT) using the CyberKnife system between 2007 and 2018. Each patient received comprehensive ophthalmologic evaluations, including assessments of visual acuity, anterior segment, fundus examinations, and ultrasonography. All patients were followed up for a minimum of five years. The cohort was composed of 65 patients who underwent SE (SE+) and 243 who did not (SE-). Radiomic features were extracted from pretreatment computed tomography (CT) images as well as magnetic resonance (MR) images. To develop a robust predictive model, four different machine learning algorithms were evaluated using these features. Results: The stacking model utilizing CT+MR radiomic features achieved the highest predictive performance, with an AUC of 0.90, accuracy of 0.86, sensitivity of 0.81, and specificity of 0.90. The feature of Robust Mean Absolute Deviation derived from Laplacian of Gaussian-filtered MR images was identified as the most significant predictor, demonstrating a statistically significant difference between SE+ and SE- cases (p=0.005). Conclusion: Radiomic analysis of pretreatment CT and MR images can non-invasively

	predict the risk of SE in UM patients undergoing SRS/FSRT. The combined CT+MR radiomic model may inform more personalized therapeutic decisions, thereby reducing unnecessary radiation exposure potentially and improving patient outcomes.
Response to Reviewers:	Reviewer #2 Comment1: The setting and the data are interesting. The major issue is the specific setting. Validation on an external data set would be very important for general applicability. Pipeline code sharing is mandatory to ascertain reproducibility. Response1: We sincerely thank the reviewer for highlighting the importance of validation on an independent external dataset to support generalizability. While we fully agree that external validation is a critical step, in this retrospective, single-institution study, access to an external dataset with comparable imaging protocols and outcome annotations was not feasible. To mitigate the risk of overfitting and to simulate external testing conditions as closely as possible, we separated a portion of the dataset at the very beginning of the study and did not include it in any stage of model training. This held-out dataset was used exclusively for final performance evaluation and was treated as an independent test set. We acknowledge this limitation and have discussed it more explicitly in the revised manuscript (in the last paragraph of the discussion). We also consider this a key direction for future work, including efforts to validate the model on multicenter or external cohorts. In addition, to promote transparency and reproducibility, we have shared the full pipeline code as Supplementary 2. Comment2: Although radiomics may predict the probability of secondary enucleation to a certain degree, classical parameters such as size, tumor volume and location should be controlled. 6.4% with T4 stage tumors were included and, on the other hand, 22.6% were small tumors. How might preselection of tumors with a volume of </> 1mL impact on the role of radiomics (cf. DOI: 10.1136/bjo-2022-322750)? Thus, how does radiomics perform in comparison with classical risk factors? Response2: Thank you for this thoughtful and important comment. We fully acknowledge the prognostic value of classical parameters such as tumor size, volume, and location, which have been consistently associated with secondary enucleation in previous studies. However, our study was designed to explore whether radiomic features could provide complementary or even superior predictive value, particularly in cases where classical markers might not fully capture the risk. As shown in Figure 1, secondary enucleation was observed even among patients with low COMS stages, suggesting that risk is not confined to patients with large or advanced tumors. In our dataset, 9.7% of the tumors were classified as small and 24% as large, based on the COMS classification. SE rate was 13% for the small tumor group, 18% for the medium tumor group, and 29% for the large tumor group. Regarding tumor location, the SE rate was 47% for anteriorly located tumors, 22% for tumors located in the middle region, and 15% for posteriorly located tumors. We included the results into the result section. And we discussed it in discussion part. Thank you again for your recommendation. To avoid introducing bias and to reflect the heterogeneity of real-world clinical populations, we deliberately chose not to apply any preselection based on tumor volume, such as a 1 mL threshold. Our goal was to develop a model that would be generalizable across the full clinical spectrum of patients referred for radiotherapy, regardless of tumor size, volume, or stage. Regarding your question on comparative performance: while our study primarily focused on developing a radiomics-based model, we agree that direct comparison with classical clinical risk factors is important. In response to your suggestion, we have now expanded the Discussion section to include commentary on this point and cited the relevant reference (DOI: 10.1136/bjo-2022-322750). Future work including combined models integrating both radiomic and classical features may offer even greater predictive accuracy and will be a natural next step in our research. Comment3: Methods, page 2, lines 50ff: Add information on fractionation schedules used. Response3:

	Thank you for this suggestion. Information on the fractionation schedules has now been added to the Methods section (page 2, lines 50ff). We specified the number of fractions and the total prescribed dose., consistent with the clinical protocols used during treatment planning and delivery. Comment4: Page 6, lines 31f: Elaborate on the split use for hyperparameter selection. Response4: Thank you for pointing this out. We have now elaborated on the data split procedure used for hyperparameter selection in the Methods section (page 6, lines 31f). Specifically, we clarified that hyperparameter tuning was performed using a stratified k-fold cross-validation on the training/validation set only, ensuring that the independent test set remained completely untouched during model development and selection. Comment5: Page 6, line 4: (fig suppl 1B) move to page 5 second paragraph. Response5: Thank you for the helpful editorial comment. We have now relocated the reference to Supplementary Figure 1B from page 6, line 4, to the second paragraph on page 5, as suggested. This improves the flow and consistency of the text. Comment6: The current standard of SE is below 10%; Discuss the potential impact of a more performing treatment method with a lower SE rate on the usefulness of your method to predict SE. Responce6: We sincerely thank the reviewer for this insightful comment. In response, we have added a paragraph to the Discussion section addressing the potential implications of declining SE rates on the clinical utility of our prediction model. While we acknowledge that more effective treatment modalities may result in SE rates falling below 10%, we believe our model remains valuable. In such a low-incidence context, identifying high-risk patients becomes even more critical to avoid unnecessary radiotherapy, optimize follow-up strategies, and support personalized treatment planning. Moreover, the model can serve as a complementary tool in evaluating the efficacy of new treatment methods. We have elaborated on this perspective in the revised manuscript as follows: “While the SE rate in our study cohort was 21%, the current standard SE rate in centers with more advanced treatment techniques may be below 10%... [paragraph as above].” We hope this addition addresses the reviewer’s concern appropriately and strengthens the clinical relevance of our work.

Non-invasive Prediction of Secondary Enucleation Risk in Uveal Melanoma Based on Pretreatment CT and MR Imaging Prior to Stereotactic Radiotherapy

Yagiz Yedekci[1,2], Hidetaka Arimura[2], Yu Jin[3], Melek Tugce Yilmaz[1], Takumi Kodama[3], Gokhan Ozyigit[1], Gozde Yazici[1]

[1]*Department of Radiation Oncology, Faculty of Medicine, Hacettepe University, Ankara 06100, Turkey*

[2]*Division of Medical Quantum Science, Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan*

[3]*Division of Medical Quantum Science, Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan*

Corresponding Author:

Yagiz Yedekci, PhD

Department of Radiation Oncology, Faculty of Medicine, Hacettepe University

Sihhiye, Ankara 06100, Turkey

Phone: +90 (312) 305 2900; e-mail: yagiz.yedekci@hacettepe.edu.tr

Non-invasive Prediction of Secondary Enucleation Risk in Uveal Melanoma Based on Pretreatment CT and MR Imaging Prior to Stereotactic Radiotherapy

Abstract

Purpose: To develop a radiomic model to non-invasively predict the risk of secondary enucleation (SE) in patients with uveal melanoma (UM) prior to stereotactic radiotherapy using pretreatment computed tomography (CT) and magnetic resonance (MR) images.

Materials and Methods: This retrospective study encompasses a cohort of 308 patients diagnosed with UM who underwent stereotactic radiosurgery (SRS) or fractionated stereotactic radiotherapy (FSRT) using the CyberKnife system between 2007 and 2018. Each patient received comprehensive ophthalmologic evaluations, including assessments of visual acuity, anterior segment, fundus examinations, and ultrasonography. All patients were followed up for a minimum of five years. The cohort was composed of 65 patients who underwent SE (SE+) and 243 who did not (SE-). Radiomic features were extracted from pretreatment computed tomography (CT) images as well as magnetic resonance (MR) images. To develop a robust predictive model, four different machine learning algorithms were evaluated using these features.

Results: The stacking model utilizing CT+MR radiomic features achieved the highest predictive performance, with an AUC of 0.90, accuracy of 0.86, sensitivity of 0.81, and specificity of 0.90. The feature of Robust Mean Absolute Deviation derived from Laplacian of Gaussian-filtered MR images was identified as the most significant predictor, demonstrating a statistically significant difference between SE+ and SE- cases ($p=0.005$).

Conclusion: Radiomic analysis of pretreatment CT and MR images can non-invasively predict the risk of SE in UM patients undergoing SRS/FSRT. The combined CT+MR radiomic model may inform more personalized therapeutic decisions, thereby reducing unnecessary radiation exposure potentially and improving patient outcomes.

Keywords: Uveal Melanoma, Radiomics, Secondary Enucleation, Stereotactic Radiosurgery, Fractionated Stereotactic Radiotherapy, Predictive Model, Machine Learning

Non-invasive Prediction of Secondary Enucleation Risk in Uveal Melanoma Based on Pretreatment CT and MR Imaging Prior to Stereotactic Radiotherapy

Introduction

Uveal melanoma (UM) is the most common primary intraocular malignancy in adults. Historically, enucleation has been the standard treatment for UM. However, radiation therapy (RT) has gained prominence in recent years due to its potential to preserve the globe and retain vision [1-3]. Enucleation remains necessary in cases of large tumor size, or when the eye causes intolerable pain or discomfort [4]. Although RT is an alternative to primary enucleation, some patients still require secondary enucleation (SE) following RT. The SE is often caused by tumor progression or adverse effects associated with RT [5, 6].

RT modalities for the treatment of UM include brachytherapy, stereotactic radiosurgery (SRS), fractionated stereotactic radiotherapy (FSRT), and proton therapy [7-9]. The incidence of SE following SRS/FSRT can reach up to 38% [8, 10]. In contrast, the incidence of SE is reported to be 15% and 22.6% following brachytherapy and proton therapy, respectively [6, 11]. Previous studies have identified significant risk factors for SE, including increased tumor height and the development of neovascular glaucoma post-RT [12, 13]. Non-invasive techniques are frequently employed in the diagnosis of uveal melanoma. These include ophthalmoscopy, ultrasonography, fluorescein angiography, optical coherence tomography and magnetic resonance imaging (MRI)[14]. SE risk prediction has been generally attempted to be made according to the results of these methods. To a lesser extent, biopsy is typically used in diagnosis when other diagnostic methods are inconclusive or in ambiguous cases. Notably, Bosch et al.[12] reported that biopsy results contain valuable information for SE risk estimation. However, eye biopsies carry risks, including complications, potential tumor spread, diagnostic reliability issues, and vision loss [15]. In contrast, radiomics could offer a non-invasive and reliable alternative approach. To date, there is no study employing radiomics to predict SE risk.

Radiomics is an evolving discipline that enables the extraction of quantitative features from radiological images, providing predictive insights. By leveraging mathematical and statistical techniques, radiomics facilitates a thorough analysis of medical imaging data, often incorporating machine learning (ML) and advanced image analysis methods [16-18][14-16]. This

comprehensive approach allows for personalized assessments [19][17]. We hypothesized that radiomics could predict the risk of secondary effects prior to administering RT in patients with UM.

The objective of this study was to develop a predictive model that could forecast the risk of SE in UM patients undergoing SRS/FSRT. This approach integrates radiomics to achieve a comprehensive understanding of treatment outcomes and equip clinicians with essential tools for making personalized therapeutic decisions.

Materials and Methods

Patient cohort

This retrospective study was approved by the institutional ethics committee and was performed in accordance with the Declaration of Helsinki. The study is based on the data from 308 patients diagnosed with UM who were treated in our department between 2007 and 2018. The diagnosis of UM was confirmed through ophthalmoscopy. Each patient underwent a comprehensive ophthalmologic assessment, including best-corrected visual acuity testing, anterior segment examination, fundus examination, and fundus fluorescein angiography. These assessments aimed to evaluate the impact of the tumor on vision and ocular structures. Tumor thickness was measured using A- and B-mode ocular ultrasonography, providing accurate quantification of the tumor's dimensions. Prior to treatment, each patient underwent a comprehensive pretreatment staging process. This process included a detailed review of the patient’s medical history, a physical examination, blood biochemistry analysis, and imaging studies (abdominal CT and/or PET/CT) to detect possible metastasis. This comprehensive staging was critical for evaluating the extent of the disease and planning appropriate treatment strategies.

All patients underwent SRS/FSRT via the Cyberknife system (Accuray, Sunnyvale, California). The prescribed fractionation schemes were as follows: 60 Gy in 3 fractions (n=101), 54 Gy in 3 fractions (n=69), 45 Gy in 3 fractions (n=96), 20 Gy in 3 fractions (n=32), and 20 Gy in a single fraction (n=10). Patients with distant metastasis, missing planning MR images, or a follow-up duration of less than five years were excluded from the study. Following treatment, patients were examined at the following intervals: every three months within the initial two years post-treatment,

followed by twice-yearly examination over the subsequent three years, and subsequently, annual examinations after that. The data for this study were sourced from a variety of channels, including departmental charts, hospital records, reports from referring physicians, and records from the General Directorate of Population and Citizenship Affairs. The dataset comprises 65 patients who have experienced SE. Patients with SE are denoted as SE+, while those without it are labeled as SE-.

Pretreatment imaging

CT images were acquired using the Brightspeed CT simulator (GE Medical System, Milwaukee, WI, USA). MR scans were conducted with an imaging protocol of structural 3D T1-weighted sequences on a 1.5T MRI system, the Signa HDx (GE Healthcare, Milwaukee, WI, USA), and a 1.5T Achieva scanner (Philips Healthcare, Amsterdam, Netherlands). All scans followed the same protocol for all patients (see Supplementary, Table1)

T1-weighted MR images, CT images, and the gross tumor volume (GTV) contours were obtained from the treatment planning system. Figure 1 illustrates examples of CT, GTV, and MR images from three SE+ and SE- patients.

Preprocessing

All image preprocessing and feature extraction calculations were conducted using in-house developed software based on Python (version 3.9.18). The detailed overall workflow is illustrated in Figure 2. Before radiomic feature extraction, systematic preprocessing steps were applied as follows. As for CT images, isotropic resampling was performed at a voxel size of 0.8 x 0.8 x 0.8 mm^3 using a b-spline interpolation. Regarding MR images, preprocessing was conducted, including re-quantization, N4 bias field correction [20] and isotropic resampling to a voxel size of 0.8 x 0.8 x 0.8 mm^3.

Feature extraction

The risk of SE was predicted by employing machine learning models with radiomic features derived from CT and MR images. For this purpose, three types of feature sets were utilized: CT radiomic, MR radiomic, and a combined CT+MR radiomic feature sets.

Radiomic features were extracted from the pretreatment images within the GTVs using the PyRadiomics (version 3.1.0) open-source package[21]. The feature extraction encompassed first-order statistical features, shape-based features, and texture features derived from both the original and filtered images. To enhance the features, various filters were applied, including a three-dimensional Laplacian-of-Gaussian (LoG) filter with sigma values of 1,2,3,4, and 5, and a Wavelet filter. Texture features were computed using several types of matrices: the Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Gray Level Dependence Matrix (GLDM), and Neighboring Gray Tone Difference Matrix (NGTDM).

After feature extraction, the dataset was divided into training and test sets. The training dataset consisted of 266 patients, while the test dataset comprised 42. The test dataset demonstrated an equal distribution of SE- and SE+. Within the training dataset, only 44 patients were classified as SE+, indicating an inherent class imbalance. To address the imbalance, we implemented an imbalance adjustment strategy based on the approach described by Schiller et al.[22]. As illustrated in Supplementary Figure 1A, the original imbalanced training dataset was divided into five balanced subsets. Each subset contained 44 SE+ cases and 44 or 45 randomly selected SE- cases. Subsequently, Z-score normalization was performed to mitigate the potential dominance of features with larger scales in the analysis. This normalization process facilitates the standardization of feature values, thereby ensuring the accurate contribution of each feature to the analysis[23]. Following Z-score normalization, a harmonization technique was applied to the radiomic data derived from MR images. This harmonization process utilized Python's neuroCombat package (version 0.2.12), aiming to reduce systematic disparities between radiomic features extracted from different scanners [24]. The harmonization was not applied to CT radiomic features as only one scanner was used.

Feature selection

Our study employed a feature selection algorithm based on the random forest (RF) algorithm to identify the most pertinent features to find the appropriate input dimensionality to machine learning models, thereby avoiding the over- and under-fitting problems, and improving the generalization of the developed model. Feature selection with RF improves the model accuracy by

identifying the most relevant features. Feature importance scores generated by RF indicated the degree of the feature's contribution to the model's predictive power, aiding in understanding and optimizing model performance[25, 26]. Additionally, we optimized the hyperparameters using grid search, which systematically tests model hyperparameters within specified value ranges to determine the best parameters for the optimal model performance [27, 28]. Eight out of the 1316 features were selected for each fold based on their importance scores. These eight features exhibited the highest importance scores, signifying their significant contribution to the model.

Construction and evaluation of predictive models

We constructed four ML classifiers: RF, Support Vector Machine (SVM), Gradient Boosting Decision Tree (GBDT), and a stacking model. The stacking model is an ensemble learning approach that combines RF, SVM, and GBDT as base learners, with Logistic Regression (LR) serving as the meta-learner. By utilizing the stacking approach, the predictive performance of base learners can be enhanced through the training of a meta-learner. A visual representation of the stacking model architecture is provided in the Supplementary Figure A.1. A total of 15 models were used to train the meta-learner.

All computational implementations were carried out in the Python environment. Hyperparameter tuning for each of the five machine learning models —RF, SVM, GBDT, SM, and LR—was performed individually using Bayesian optimization. This approach allowed efficient and model-specific exploration of the hyperparameter space to identify the optimal configuration for each classifier. As described in detail by Wu et al. [29], the training dataset was further divided into inner training and validation subsets using a stratified 5-fold cross-validation scheme. Hyperparameter selection was based solely on performance across these internal validation folds. To prevent data leakage and ensure unbiased model evaluation, the independent test dataset was not used at any stage of the model development or tuning process.

Each dataset was divided into five subsets, which were then used sequentially for model training, and subsequently evaluated on an independent test cohort (see Supplementary, Figure 1B).

To assess the predictive capability of the classifiers, various performance metrics were computed. The metrics included the area under the receiver operating characteristic curve (AUC), accuracy,

sensitivity, and specificity. These metrics provided the models' effectiveness in distinguishing between SE+ and SE- cases across the different datasets.

Statistical analysis

Statistical analyses were performed using SPSS version 23 software (SPSS, Inc.). For the demographic distributions of patients (Table 1), Mann-Whitney U tests were employed to assess significant differences for continuous variables, which did not follow the normal distribution. Chi-squared tests were used for the other categorical variables. As a result of the tests, $p < 0.05$ was considered statistically significant.

Results

The clinical characteristics of the 308 patients included in this study are detailed in Table 1, which also shows their distribution across both the training and test datasets. The training dataset consisted of 266 cases, with 44 SE+ and 222 SE- patients. The test dataset included 42 cases, with an equal number of patients in each category (21 SE+ and 21 SE-). The results of the statistical analysis demonstrated that there were no significant differences between the training and test datasets.

A total of 65 patients (21%) underwent SE during the follow-up period. Among the cases, 38 (58%) underwent SE due to tumor progression, while 27 (42%) underwent SE due to side effects related to RT. SE rate was 13% for the small tumor group, 18% for the medium tumor group, and 29% for the large tumor group. Regarding tumor location, the SE rate was 47% for anteriorly located tumors, 22% for tumors located in the middle region, and 15% for posteriorly located tumors.

A total of twelve distinct predictive models were developed, each utilizing a different feature set (CT, MR, and CT+MR) and a distinct ML algorithm (RF, SVM, GBDT, and SM). Table 2 presents the performance metrics, including accuracy, sensitivity, specificity, and AUC, for each model. The test AUC for RF, SVM, GBDT, and stacking models with the CT image features were 0.52, 0.57, 0.54, and 0.80, respectively. The test AUC values for RF, SVM, GBDT, and stacking models with the MR image features were 0.78, 0.75, 0.79, and 0.83, respectively. The test AUC values for RF, SVM, GBDT, and stacking models with CT+MR combined features were 0.74, 0.68, 0.69, and

0.90, respectively. The stacking model achieved the highest AUC among the models with the three feature types (CT, MR, and CT+MR). The combined CT+MR image features enabled the stacking model to reach the highest AUC (0.90), with an accuracy of 0.86, sensitivity of 0.81, and specificity of 0.90 (Figure 3).

A total of 1316 features were extracted from each image type of CT and MR for each patient. These extracted features were subsequently utilized to create the CT+MR feature set. The eight features with the highest importance scores for each set (CT, MR, and CT+MR) are presented in Supplementary Figure 2. For the CT, MR, and CT+MR, the most significant features were identified as wavelet-HHL_glcm_Imc1, wavelet-HHH_glszm_ZoneEntropy, and log-sigma-5-0-mm-3D_firstorder_RobustMeanAbsoluteDeviation, respectively. Figure 4 presents violin plots illustrating the distribution of these features for SE- and SE+ patients. The difference in the distribution of the wavelet-HHL_glcm_Imc1 feature scores, which have the highest importance score for CT images, is not statistically significant between SE- and SE+ patients (p=0.457). However, the wavelet-HHH_glszm_ZoneEntropy feature scores from MR images demonstrate a statistically significant difference between SE- and SE+ patients (p=0.026). Furthermore, the log-sigma-5-0-mm-3D_firstorder_RobustMeanAbsoluteDeviation feature from the combined CT+MR set also demonstrates a statistically significant difference between SE- and SE+ patients, with the lowest p-value among the features (p=0.005).

Discussion

This study aimed to investigate the potential of radiomic features derived from pretreatment CT and MR images to predict the risk of SE in patients with UM undergoing SRS/FSRT. The findings demonstrated that radiomic predictive models, particularly the stacking model applying to CT+MR features, exhibited satisfactory performance in classifying patients who would need SE following RT. The highest AUC of 0.90 in the test cohort was observed with the CT+MR feature set using stacking model, indicating that this model exhibited superior predictive capability.

The Robust Mean Absolute Deviation (RMAD) feature for the Laplacian of Gaussian (LoG)-filtered MR images emerged as the most significant feature in the CT+MR model. The LoG filter enhances and detects edges within the image by first applying Gaussian smoothing to reduce noise, followed by the Laplacian operator to accentuate edge information. RMAD quantifies the

variability in voxel intensities by calculating the mean absolute difference between each voxel's intensity and the median intensity of the entire image or region of interest (ROI). In our analysis, RMAD demonstrated the lowest p-value in discriminating between SE- and SE+ cases, indicating a statistically significant distinction between these groups. This indicates that RMAD is capable of capturing critical variability in intensity patterns, which may provide valuable insights into the tumor's characteristics or its response to treatment. The significance of RMAD in the CT+MR model highlights its potential to enhance the predictive power of radiomic features in clinical settings.

It is evident that accurately predicting the risk of SE prior to administering SRS/FSRT offers substantial clinical benefits. The SRS/FSRT process involves numerous preparatory steps, such as immobilizing the eye before each treatment, acquiring treatment images, contouring and planning [30, 31]. In the event that SE becomes necessary after SRS/FSRT, the initial choice of RT may be unwarranted. This results in the waste of clinical resources and exposes patients to potential RT complications [32-34]. Anticipating the risk for SE may prevent unnecessary radiation exposure and thus reduce the risks associated with secondary cancer development and RT-related complications. Furthermore, performing SE after RT presents technical challenges and is associated with a higher incidence of complications compared to primary enucleation, as reported by Chau et al. [35]. Consequently, our predictive model, which exhibits a high AUC of 0.90 for prediction of the risk of SE, may significantly mitigate such risks by facilitating more informed decision-making. The proposed model has the potential to enable clinicians to more accurately identify patients who will benefit from RT, thereby improving clinical workflows and avoiding unnecessary treatments.

In our study, the SE rate was 21% among 310 patients. In comparison, Bosch et al. [12] reported an SE rate of 16% in 118 patients treated with FSRT. Similarly, Fabian et al. [13] observed a 19% SE rate among 515 patients treated with plaque RT, proton therapy, and transpupillary thermotherapy, while Wang et al. [6] observed a 15% SE rate in 880 patients undergoing plaque RT. Regarding the reasons for SE, Bosch et al. [12] indicated that 31% of SE cases were due to treatment failure, with the remaining cases attributed to complications. In contrast, Wang et al. [6] found that 39.4% of SE cases resulted from complications and 46.9% from treatment failure. Fabian et al. [13] reported that 21% of SE cases were due to complications, with the remaining

79% attributed to treatment failure. In our study, treatment failure emerged as the predominant cause of SE, accounting for 58% of SE cases. With respect to risk factors for SE, Bosch et al. [12] observed that a large tumor height was associated with an increased risk of SE, highlighting high mitotic count and the gain of chromosome 8q as common features in SE cases. Wang et al. [6] identified neovascular glaucoma as the primary cause of SE, whereas Fabian et al. [13] emphasized a high-risk group exhibiting radioresistance to treatment. These findings underscore that while treatment failure and complications are the main contributors to SE, their specific rates vary across studies [6, 12, 13, 36]. Moreover, the clinical delineation of SE risk factors remains incomplete. Our study contributes to the existing literature by validating the role of radiomics in predicting the risk of SE, thereby enhancing our understanding of its predictors.

The SE rate varied notably with both tumor size and location. Patients with large tumors exhibited the highest SE rate (29%), followed by those with medium-sized (18%) and small tumors (13%), indicating that tumor volume is a significant prognostic factor. This aligns with the findings of Sreenivasa et al. [37], who reported that tumor volume was the most robust physical prognostic marker for local control and disease progression. In our study, tumors located in the anterior segment of the eye were associated with a substantially higher SE rate (47%) compared to those situated in the middle (22%) or posterior (15%) regions. These results support the notion that anteriorly located and larger tumors may be more susceptible to treatment-related complications or less responsive to radiotherapy, ultimately increasing the likelihood of SE.

While the SE rate in our study cohort was 21%, the current standard SE rate in centers with more advanced treatment techniques may be below 10%[37]. In such a scenario, the clinical utility of our prediction model warrants reevaluation. A lower incidence of SE would reduce the absolute number of events, potentially limiting the statistical power of predictive models. However, the ability to identify rare but high-impact adverse outcomes becomes even more valuable in a context where complications are infrequent. Accurate prediction of SE risk, even in a low-incidence setting, could help clinicians avoid unnecessary radiotherapy in patients with a high likelihood of eventual enucleation. Furthermore, the model may assist in tailoring follow-up strategies for high-risk individuals and contribute to individualized treatment decision-making. Importantly, our model could also serve as a complementary tool in assessing the effectiveness of emerging treatment modalities by identifying patients most likely to benefit. Therefore, even with decreasing SE rates,

the predictive model retains its potential clinical relevance, particularly in optimizing patient selection and minimizing overtreatment.

It is important to acknowledge the limitations of our study. Firstly, as the research was conducted within a single department, the findings may be influenced by specific institutional practices, equipment, and patient demographics. This context-specific nature may constrain the generalizability of our results to other clinical settings or broader populations. Secondly, the dataset exhibits a notable imbalance, with only 65 cases of SE+ compared to 243 cases with SE-. This imbalance presents challenges for model training and performance, as the models may be biased towards the majority class SE-. Although imbalance adjustment strategies were implemented, the relatively low number of SE+ cases could impact on the robustness and generalizability of our predictive models, potentially leading to overfitting or under-representation of SE+ cases. Furthermore, while all CT scans were acquired using the same machine, MR images were obtained from different scanners, introducing variability. Despite efforts to harmonize radiomic features using neuroCombat, systematic differences due to varying imaging protocols and equipment could still influence feature extraction and, consequently, model performance. The heterogeneity in MR imaging may introduce biases that affect the reliability of the features derived from these images. Additionally, our study did not consider potential variables beyond radiomic features that might influence the risk of SE, such as genetic factors or specific dosimetric parameters. These unaccounted factors could play a significant role in SE risk and might provide additional predictive value.

Conclusion

This study demonstrates the potential of radiomic features derived from pretreatment CT and MR images to predict the risk of SE in UM patients undergoing SRS/FSRT. Notably, stacking model, particularly the combined CT+MR model, achieved a high AUC of 0.90, demonstrating superior predictive accuracy compared to traditional clinical models. The RMAD feature, identified as the most significant in distinguishing SE- from SE+ cases, appears to capture critical aspects of tumor characteristics and treatment responses. This highlights its utility in refining patient selection for RT, potentially leading to more targeted and effective treatment strategies. The promising results of our radiomics-based approach represent a substantial advancement in personalized treatment

strategies for UM, offering a valuable tool for enhancing clinical decision-making and improving patient outcomes.

Acknowledgements

The first author acknowledges support from The Scientific and Technological Research Council of ****** through the 2219 - International Postdoctoral Research Fellowship Program.

References

1. **Semeniuk, O., E. Yu, and M.J. Rivard, *Current and Emerging Radiotherapy Options for Uveal Melanoma.* Cancers (Basel), 2024. 16(5).**

2. **Reichstein, D.A. and A.L. Brock, *Radiation therapy for uveal melanoma: a review of treatment methods available in 2021.* Current Opinion in Ophthalmology, 2021. 32(3): p. 183-190.**

3. **Kosydar, S., et al., *Systematic Review and Meta-Analysis on the Use of Photon-based Stereotactic Radiosurgery Versus Fractionated Stereotactic Radiotherapy for the Treatment of Uveal Melanoma.* Am J Clin Oncol, 2021. 44(1): p. 32-42.**

4. **Bai, H., J.J. Bosch, and L.M. Heindl, *Current management of uveal melanoma: A review.* Clin Exp Ophthalmol, 2023. 51(5): p. 484-494.**

5. **van Beek, J.G.M., et al., *Fractionated stereotactic radiotherapy for uveal melanoma: Long-term outcome and control rates.* Acta Ophthalmol, 2022. 100(5): p. 511-519.**

6. **Wang, H., et al., *Retrospective analysis of secondary enucleation for uveal melanoma after plaque radiotherapy.* BMC Ophthalmol, 2022. 22(1): p. 163.**

7. **Buonanno, F., et al., *Local tumor control and treatment related toxicity after plaque brachytherapy for uveal melanoma: A systematic review and a data pooled analysis.* Radiotherapy and Oncology, 2022. 166: p. 15-25.**

8. **Yazici, G., et al., *Every other day stereotactic radiation therapy for the treatment of uveal melanoma decreases toxicity.* Radiotherapy and Oncology, 2022. 176: p. 39-45.**

9. **Gollrad, J., et al., *Quality of life and treatment-related burden during ocular proton therapy: a prospective trial of 131 patients with uveal melanoma.* Radiation Oncology, 2021. 16: p. 1-11.**

10. **Furdova, A., et al., *Clinical experience of stereotactic radiosurgery at a linear accelerator for intraocular melanoma.* Melanoma Research, 2017. 27(5): p. 463-468.**

11. **Marinkovic, M., et al., *Clinical Outcomes after International Referral of Uveal Melanoma Patients for Proton Therapy.* Cancers (Basel), 2021. 13(24).**

12. **van den Bosch, T., et al., *Risk factors associated with secondary enucleation after fractionated stereotactic radiotherapy in uveal melanoma.* Acta Ophthalmologica, 2015. 93(6): p. 555-560.**

13. **Fabian, I.D., et al., *Secondary Enucleations for Uveal Melanoma: A 7-Year Retrospective Analysis.* Am J Ophthalmol, 2015. 160(6): p. 1104-1110.e1.**

14. **Branisteanu, D.C., et al., *Uveal melanoma diagnosis and current treatment options (Review).* Exp Ther Med, 2021. 22(6): p. 1428.**

15. Frizziero, L., et al., *Uveal Melanoma Biopsy: A Review.* Cancers (Basel), 2019. 11(8).

16. Arimura, H., et al., *Potentials of radiomics for cancer diagnosis and treatment in comparison with computer-aided diagnosis.* Radiological physics and technology, 2018. 11: p. 365-374.

17. Tomaszewski, M.R. and R.J. Gillies, *The biological meaning of radiomic features.* Radiology, 2021. 298(3): p. 505-516.

18. Reginelli, A., et al., *Radiomics as a new frontier of imaging for cancer prognosis: a narrative review.* Diagnostics, 2021. 11(10): p. 1796.

19. Ger, R.B., et al. *The promise and future of radiomics for personalized radiotherapy dosing and adaptation*. in *Seminars in Radiation Oncology*. 2023. Elsevier.

20. Dovrou, A., et al., *A segmentation-based method improving the performance of N4 bias field correction on T2weighted MR imaging data of the prostate.* Magnetic Resonance Imaging, 2023. 101: p. 1-12.

21. Van Griethuysen, J.J., et al., *Computational radiomics system to decode the radiographic phenotype.* Cancer research, 2017. 77(21): p. e104-e107.

22. Schiller, T.W., et al., *Modeling radiation-induced lung injury risk with an ensemble of support vector machines.* Neurocomputing, 2010. 73(10-12): p. 1861-1867.

23. Demircioğlu, A., *The effect of feature normalization methods in radiomics.* Insights into Imaging, 2024. 15(1): p. 2.

24. Fortin, J.-P., et al., *Harmonization of cortical thickness measurements across scanners and sites.* Neuroimage, 2018. 167: p. 104-120.

25. Reif, D.M., et al. *Feature selection using a random forests classifier for the integrated analysis of multiple data types*. in *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*. 2006. IEEE.

26. Huljanah, M., et al. *Feature selection using random forest classifier for predicting prostate cancer*. in *IOP Conference Series: Materials Science and Engineering*. 2019. IOP Publishing.

27. Ataei, M. and M. Osanloo, *Using a combination of genetic algorithm and the grid search method to determine optimum cutoff grades of multiple metal deposits.* International Journal of Surface Mining, Reclamation and Environment, 2004. 18(1): p. 60-78.

28. Ramadhan, M.M., et al., *Parameter tuning in random forest based on grid search method for gender classification based on voice frequency.* DEStech transactions on computer science and engineering, 2017. 10(2017).

29. **Wu, J., et al., *Hyperparameter optimization for machine learning models based on Bayesian optimization.* Journal of Electronic Science and Technology, 2019. 17(1): p. 26-40.**

30. **Yazici, G., et al., *Stereotactic radiosurgery and fractionated stereotactic radiation therapy for the treatment of uveal melanoma.* International Journal of Radiation Oncology* Biology* Physics, 2017. 98(1): p. 152-158.**

31. **Quhill, H., et al., *To suture or not to suture? Does globe immobilisation technique affect clinical outcome in stereotactic radiosurgery for uveal melanoma?* British Journal of Neurosurgery, 2021: p. 1-4.**

32. **Zemba, M., et al., *Ocular complications of radiotherapy in uveal melanoma.* Cancers, 2023. 15(2): p. 333.**

33. **Banou, L., et al., *Radiotherapy in uveal melanoma: a review of ocular complications.* Current Oncology, 2023. 30(7): p. 6374-6396.**

34. **Sekac, J., et al., *Secondary glaucoma in small versus large uveal melanoma patients treated with stereotactic radiosurgery on linear accelerator.* Bratislava Medical Journal/Bratislavske Lekarske Listy, 2019. 120(12).**

35. **Pham, C.M., P.L. Custer, and S.M. Couch, *Comparison of primary and secondary enucleation for uveal melanoma.* Orbit, 2017. 36(6): p. 422-427.**

36. **Zahorjanová, P., et al., *Enucleation after stereotactic radiosurgery in patients with uveal melanoma.* Cesk Slov Oftalmol, 2020. 76(1): p. 46-51.**

37. **Sreenivasa, S., et al., *Impact of tumour volume and treatment delay on the outcome after linear accelerator-based fractionated stereotactic radiosurgery of uveal melanoma.* British Journal of Ophthalmology, 2024. 108(3): p. 457.**

Figure

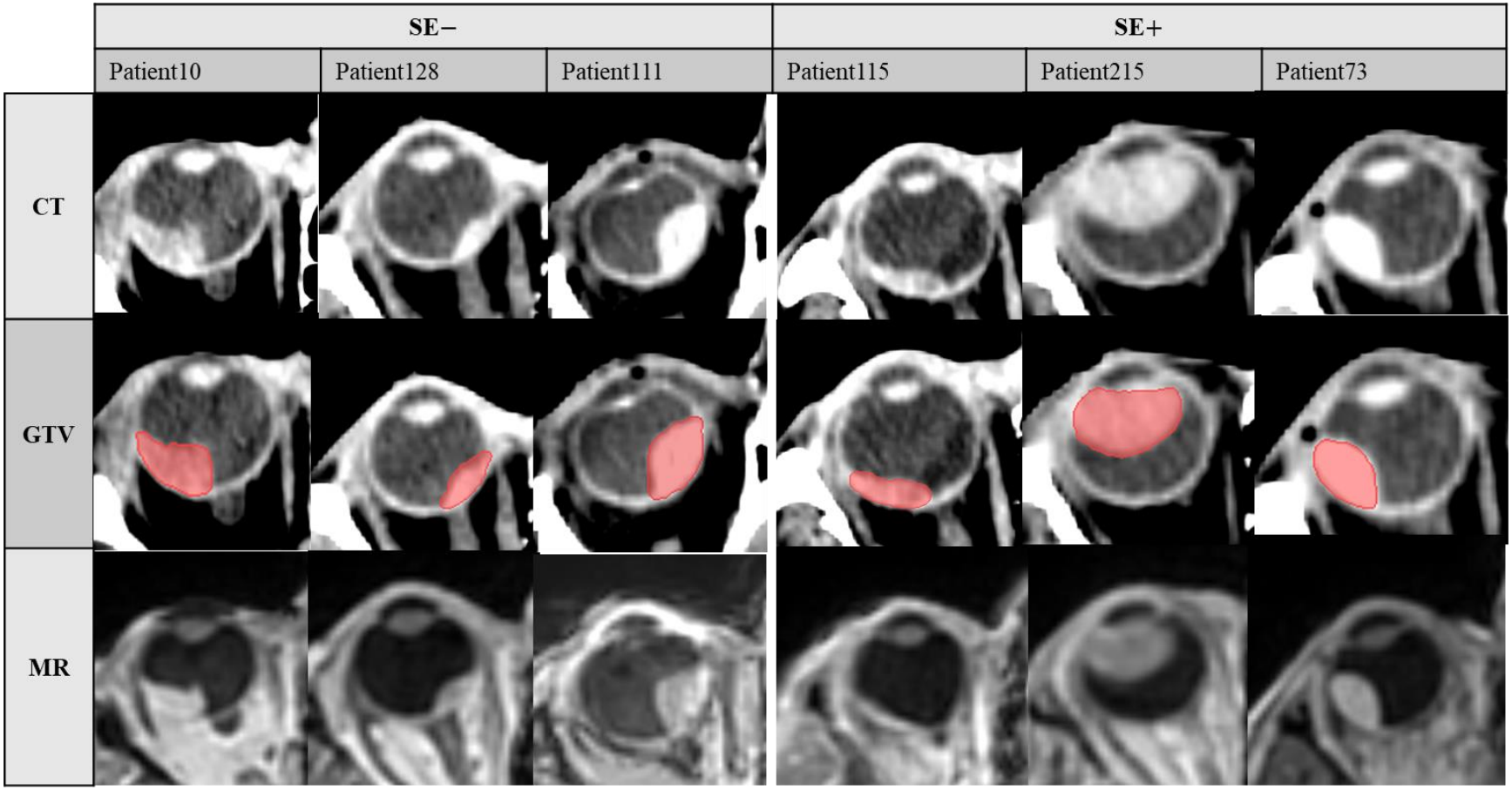

Figure 1. Treatment planning image samples of patients with and without secondary enucleation after SRS/FSRT. *Abbreviations:* SRS: stereotactic radiotherapy; FSRT= fractionated stereotactic radiotherapy; CT=computed tomography; GTV=gross tumor volume; MR= magnetic resonance; SE−= patients without secondary enucleation after radiotherapy; SE+= patients with secondary enucleation after radiotherapy

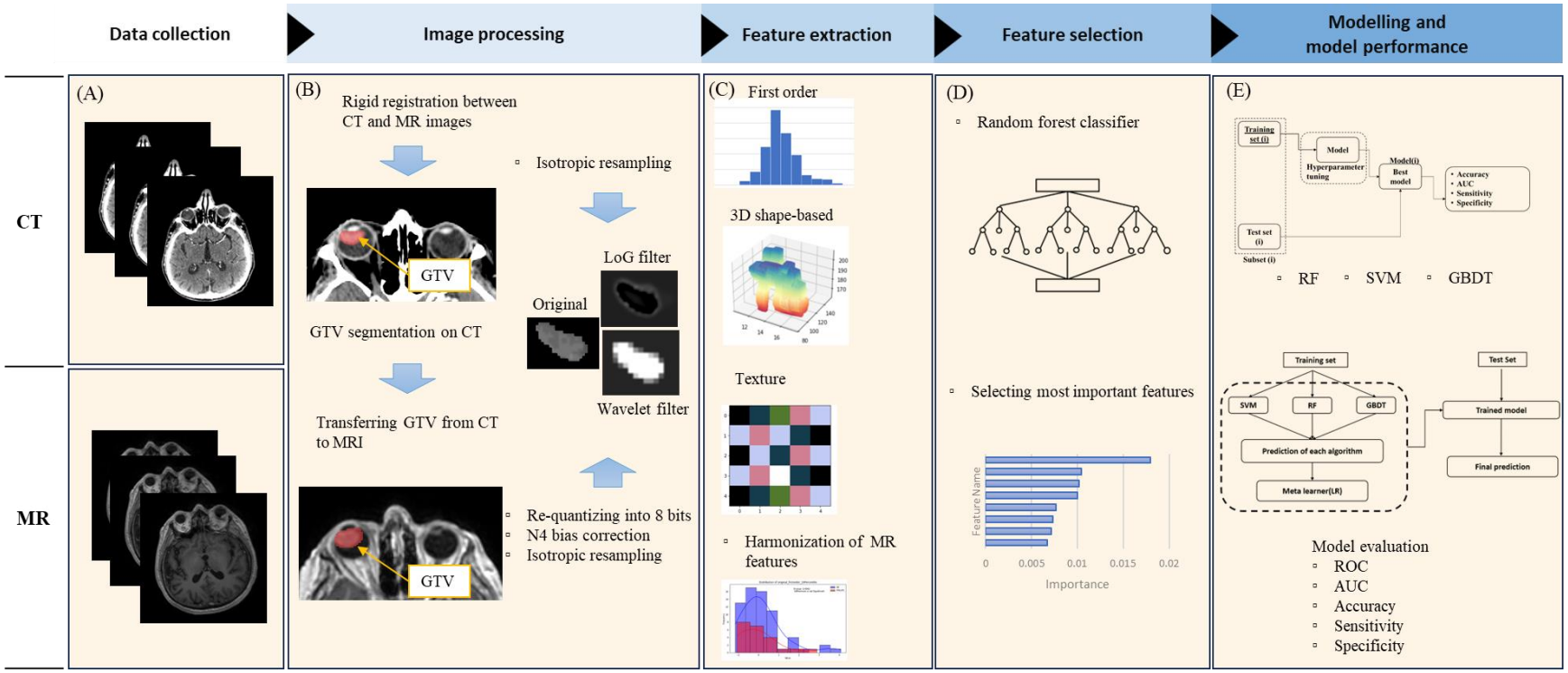

Figure 2. Study Pipeline. (A) Visualization of the collection process for pretreatment CT and MR images. (B) Illustration of the image processing framework, including the fusion of CT and MR images, GTV delineation, transferring GTVs to MR images, and the application of filters and various image processing techniques. (C) Depiction of the feature extraction and data harmonization for MR features. (D) Using of a Random Forest Classifier to identify the eight most important features. (E) Representation of the machine learning model development process, including experimental procedures and model comparisons. *Abbreviations*: CT= computed tomography: MR= magnetic resonance: GTV= gross tumor volume: RF= random forest: SVM= support vector machine: GBDT= gradient boosting decision tree: AUC= area under the curve

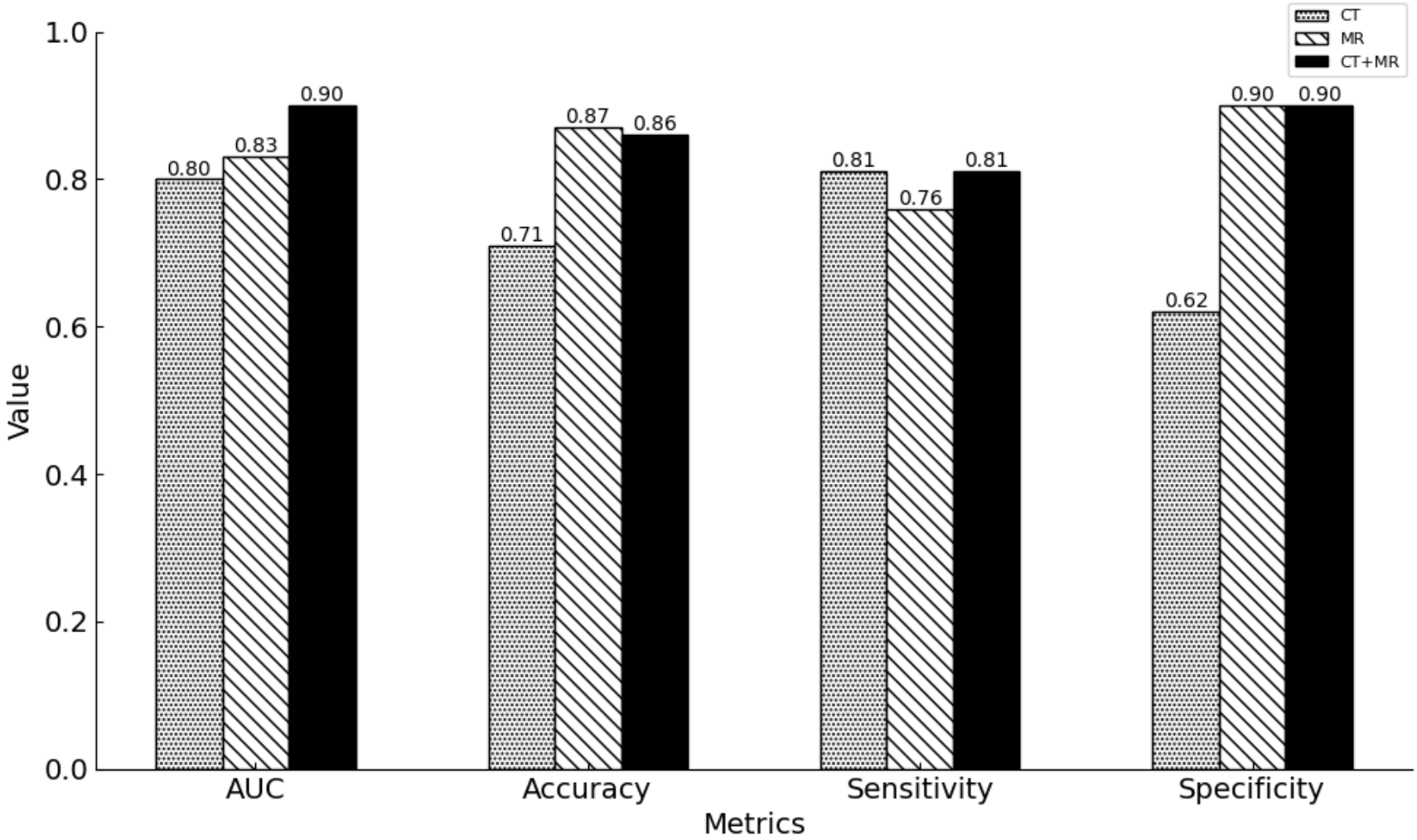

Figure 3. Comparison of performance metrics for the test cohort using the SM model across different feature sets derived from CT, MR, and combined CT+MR. *Abbreviations:* SM= stacking model; CT= computed tomography; MR=magnetic resonance

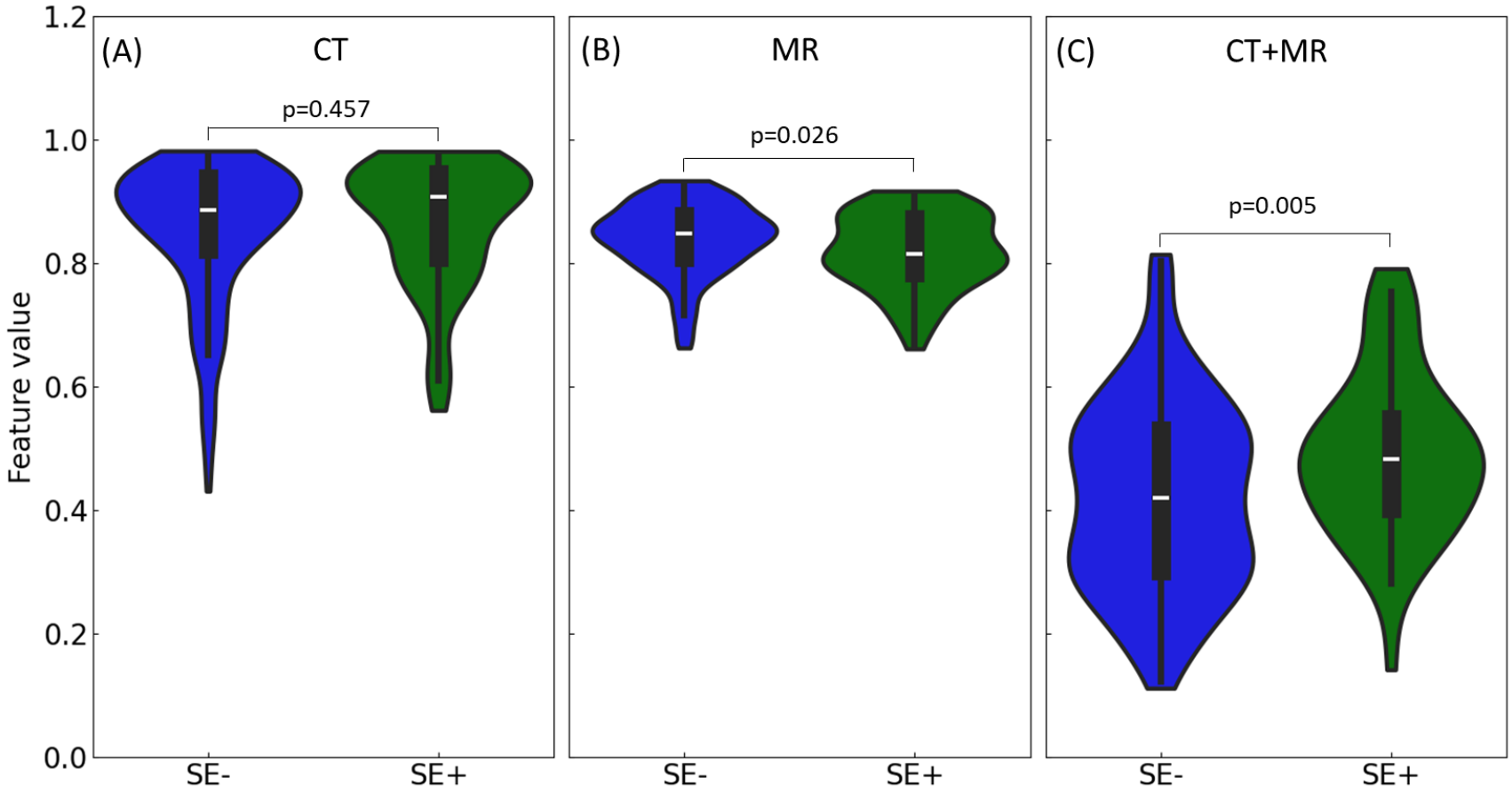

Figure 4. Violin plots illustrating the distribution of radiomic features with the highest importance scores, distinguishing between patients with and without secondary enucleation. (a) shows wavelet-HHL_glcm_Imc1, (b) represents wavelet-HHH_glszm_ZoneEntropy, and (c) demonstrates log-sigma-5-0-mm-3D_firstorder_RobustMeanAbsoluteDeviation (MR) for the CT, MR, and CT+MR, respectively. Blue represents patients without secondary enucleation, while green represents patients with secondary enucleation. Min-max normalization was performed to ensure all data was displayed on the same scale. *Abbreviations:*SE−= patients without secondary enucleation after SRS/FSRT: SE + = patients with secondary enucleation after SRS/FSRT; CT= computed tomography; MR=magnetic resonance

Table

Table1. Summary of patient characteristics

Characteristic	All cases	Training cohort	Test cohort	p-value (method)
Number of cases	308	266	42	
Age (years)	18-85 (median:56)	18-85 (median:56)	18-78 (median:53)	0.25 (Mann-Whitney U test)
Sex				0.71 (Chi-square)
Male	164	140	24	
Female	144	126	18	
Tumor diameter (mm)	2-23 (median:9)	2-23 (median:9)	3-18 (median:9)	0.31 (Mann-Whitney U test)
COMS classification				0.67 (Chi-square)
Small	30	26	4	
Medium	202	177	25	
Large	76	63	13	
T stage				0.25 (Chi-square)
T1	68	60	8	
T2	124	110	14	
T3	94	79	15	
T4	22	17	5	
Treatment dose				
BED_{3Gy} (Gy)	43-550 (median:460)	43-482 (median:460)	67-550 (median:378)	0.07 (Mann-Whitney U test)
BED_{10Gy} (Gy)	20-211(median:180)	20-188 (median:180)	38-211 (median:151)	0.07 (Mann-Whitney U test)

Abbreviations: COMS= collaborative ocular melanoma study: Large= >10 mm in apical height and >16 mm in basal diameter, Medium= 2.5–10 mm in apical height and ≤16 mm in basal diameter, Small=1.5–2.4 mm in apical height and 5–16 mm in basal diameter: BED: Biologically Effective Dose

Table 2. Sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve results of the studied models.

	Metric	RF		SVM		GBDT		Stacking model	
		Training	Test	Training	Test	Training	Test	Training	Test
CT	AUC	0.76	0.52	0.84	0.57	1.00	0.54	1.00	0.80
	Accuracy	0.75	0.55	0.93	0.60	1.00	0.56	1.00	0.71
	Sensitivity	0.80	0.50	0.95	0.52	1.00	0.50	1.00	0.81
	Specificity	0.70	0.53	0.91	0.62	1.00	0.59	1.00	0.62
MR	AUC	0.97	0.78	1.00	0.75	1.00	0.79	1.00	0.83
	Accuracy	0.97	0.82	1.00	0.82	1.00	0.84	1.00	0.87
	Sensitivity	0.98	0.80	1.00	0.80	1.00	0.81	1.00	0.76
	Specificity	0.95	0.75	1.00	0.66	1.00	0.76	1.00	0.90
CT+MR	AUC	0.81	0.74	0.96	0.68	1.00	0.69	1.00	**0.90**
	Accuracy	0.81	0.79	0.88	0.76	1.00	0.74	1.00	**0.86**
	Sensitivity	0.89	0.77	0.84	0.71	1.00	0.69	1.00	**0.81**
	Specificity	0.73	0.71	0.91	0.66	1.00	0.68	1.00	**0.90**

Abbreviations: CT= computed tomography; MR= magnetic resonance; AUC= area under the curve; RF= random forest; SVM= support vector machine; GBDT= gradient boosting decision tree