Hindawi
BioMed Research International
Volume 2022, Article ID 6911246, 9 pages
https://doi.org/10.1155/2022/6911246

Research Article

Radiomics Models Based on Magnetic Resonance Imaging for Prediction of the Response to Bortezomib-Based Therapy in Patients with Multiple Myeloma

Yang Li,[1] Ping Yin,[1] Yang Liu,[2] Chuanxi Hao,[1] Lei Chen,[1] Chao Sun,[1] Sicong Wang,[3] and Nan Hong[1]

[1]*Department of Radiology, Peking University People's Hospital, Beijing, China*
[2]*Peking University Institute of Hematology, Peking University People's Hospital, Beijing, China*
[3]*Pharmaceutical Diagnostics, GE Healthcare, Shanghai, China*

Correspondence should be addressed to Nan Hong; hongnan1968@163.com

Received 23 April 2022; Revised 4 August 2022; Accepted 20 August 2022; Published 5 September 2022

Academic Editor: Shiteng Suo

Purpose. To identify significant radiomics features based on MRI and establish effective models for predicting the response to bortezomib-based regimens. *Materials and Methods*. In total, 95 MM patients treated with bortezomib-based therapy were enrolled, including 77 with bortezomib, cyclophosphamide, and dexamethasone (BCD) and 18 with bortezomib, lenalidomide, and dexamethasone (VRD). Based on T1-weighted imaging (T1WI) and T2-weighted imaging with fat suppression (T2WI-fs), radiomics features were extracted and then selected. The random forest (RF), *k*-nearest neighbor, support vector machine, logistic regression, decision tree, and Bayes models were built using the selected features. The predictive power of six models for response to BCD and VRD regimens were evaluated. The correlation between the selected features and progression-free survival (PFS) was also analyzed. *Results*. Four wavelet features were correlated with BCD treatment response. The six models all showed predictive power for BCD regimen (AUC: 0.84-0.896 in the training set, 0.801-0.885 in the validation set), and RF performed relatively better than others. Nevertheless, all the BCD-based models were incapable of predicting the VRD treatment response. The wavelet-HLH_firstorder_kurtosis was also associated with PFS (log-rank $P = 0.019$). *Conclusion*. The four wavelet features were valuable biomarkers for predicting the response to BCD regimen. The six models based on these features showed predictive power, and RF was the best. One wavelet feature was also a survival-related biomarker. MRI-based radiomics had the potential to guide clinicians in MM management.

1. Introduction

Multiple myeloma (MM) is a malignancy of plasma cells originating from the bone marrow, and most commonly present with hypercalcemia, renal failure, anemia, and bone lesions, leading to significant impairment in quality of life and placing an immense burden both on patients and society [1, 2]. In the 1980s, high-dose chemotherapy and stem-cell rescue (ASCT) was introduced as an effective treatment modality of MM, and the treatment continues to evolve rapidly with the arrival of new classes of antimyeloma drugs such as immunomodulatory drugs and proteasome inhibitors [3, 4]. Since most patients with MM ultimately relapse and become unresponsive to currently available treatment options, thus resulting in shorter survival, durable and deep remission is the key objective of MM therapy [5].

Bortezomib, a typical proteasome inhibitor, is widely used in the induction, consolidation, and maintenance therapy of MM [6, 7]. Combining bortezomib with other agents such as IMIDs, alkylating agents/doxorubicin, and dexamethasone is the backbone for doublet/triplet regimens [7]. A systematic review assessed the prognosis effects of bortezomib and demonstrated its benefit in terms of survival and response rate of MM [8]. However, part of MM patients

achieved suboptimal or no response to bortezomib and may even suffer notable side effects like peripheral neuropathy and thrombocytopenia [9]. Therefore, identifying predictive biomarkers prior to MM treatment would help to avoid ineffective therapy and further optimize clinical patient management.

Some disease, host, and therapy-specific features such as performance status, tumor burden, and cytogenetic abnormalities have been reported to provide prognostic information [10, 11]. Nevertheless, all these factors and existing risk stratification systems help with prognosis for survival, and clinicians still lack reliable predictive biomarkers of treatment response [11]. In addition to clinical assessment, imaging evaluation is also an essential part of MM diagnosis and treatment, and MRI has been widely accepted as the optimum imaging modality. Since traditional imaging has difficulty detecting the treatment-mediated changes, functional imaging was recommended by many researchers [12]. However, previous studies demonstrated the ability of functional parameters for monitoring MM treatment response, their predictive capacity was limited, and response prediction remains a challenge [13, 14].

Radiomics-derived features quantify phenotypic characteristics of medical imaging, which contribute to produce accurate and robust predictions in survival, treatment response, and other clinical outcomes [15]. Many published prediction models confirmed the excellent predictive ability of CT or MRI radiomics for treatment response in cancer research, and these radiomics-based models either outperformed the existing predictive modalities or filled the gap of response prediction that has not been achieved in the clinic [16–19]. Considering previous promising results of radiomics and the advantages of MRI for bone marrow assessment in patients with MM, it is reasonable to speculate that MRI-based radiomics may have potential predictive powers for MM treatment.

In the present study, we aimed to explore the value of MRI-based radiomics for response prediction in MM patients treated with bortezomib-based therapy.

2. Material and Methods

2.1. Patients. The Institutional Ethics Committee of our hospital approved this retrospective study, and the requirement for informed consent was waived. The 357 MM patients who underwent lumbar MRI at the initial diagnosis in our hospital between January 2015 and January 2021 were preliminarily included. Inclusion criteria are as follows: (1) Patients were newly diagnosed with MM and had no previous systematic chemotherapy or radiotherapy. (2) MRI examination included sagittal T1-weighted imaging (T1WI) and T2-weighted imaging with fat suppression (T2WI-fs). (3) All patients were treated with bortezomib-based induction therapy with 3-4 cycles, and efficacy was evaluated every 2 or 3 cycles. Patients who were combined with other malignant diseases ($n = 3$), were treated with other chemotherapy regimens ($n = 82$), had previous systematic chemotherapy or radiotherapy ($n = 56$), were without complete T1WI and T2WI-fs examination ($n = 31$), had less than three cycles of induction chemotherapy ($n = 57$), and were without regular treatment response assessment at every 2 or 3 cycles ($n = 17$) were excluded. Finally, 111 patients met the criteria. Of these, 77 patients were treated with bortezomib, cyclophosphamide, and dexamethasone (BCD); 18 with bortezomib, lenalidomide, and dexamethasone (VRD); 9 with bortezomib, thalidomide, and dexamethasone (BTD); and 7 with bortezomib, doxorubicin, and dexamethasone (PAD). Considering that the number of patients treated with BTD and PAD were small, and these two regimens were less widely used in clinical practice, we included only the patients treated with BCD and VRD for further study.

Patients were followed until May 2021. PFS was defined as the time from diagnosis to the date of disease progression, death from any cause, or the latest follow-up.

2.2. Treatment and Response Assessment. For the BCD regimen, patients received bortezomib 1.3 mg/m^2 on days 1, 4, 7, and 10; cyclophosphamide 300 mg/day on days 1-5; and dexamethasone 20-40 mg/day on days 1–4, 7, and 10. Each cycle of induction therapy was 21 days. For the VRD regimen, patients received lenalidomide 25 mg every other day; bortezomib 1.3 mg/m^2 on days 1, 4, 7, and 10; and dexamethasone 20-40 mg/day on days 1–4, 7, and 10. Each cycle of induction therapy was 28 days.

In accordance with International Myeloma Working Group (IMWG) guidelines that was based on monoclonal protein level in serum and urine, the treatment responses of MM were categorized as complete remission (CR), very good partial remission (VGPR), partial remission (PR), stable disease (SD), or progressive disease (PD) [20]. Our study defined the CR and VGPR as good response and PR, SD, and PD as poor response.

2.3. MRI Protocol. Baseline MR images were performed on a 1.5 T scanner (Signa Excite, GE Medical Systems), 3.0 T scanner (Discovery 750, GE Medical Systems), and 3.0 T scanner (Discovery 750w, GE Medical Systems). The scan parameters were described as follows: sagittal T1WI: repetition time (TR) = 405 − 843 msec, echo time (TE) = 7.1 − 8.1 msec, slice thickness = 4 − 5 mm, matrix = 300 × 256, and field of view (FOV) = 32 × 32 cm; sagittal T2WI-FS: TR = 2500 − 3000 msec, TE = 85.3 − 125.3 msec, slice thickness = 4 − 5 mm, matrix = 300 × 256, and FOV = 32 × 32 cm.

2.4. Evaluation of Conventional MRI Patterns for MM. The bone marrow infiltration patterns of each patient were interpreted in consensus by two radiologists with 5 years and 13 years of experience, respectively. When there was discordance between the readers, a senior radiologist with 25 years' experience made the final decision.

The five recognized infiltration patterns are listed as follows: (1) normal appearance of bone marrow despite minor microscopic plasma cell infiltration, (2) focal involvement, (3) homogeneous diffuse infiltration, (4) combined diffuse and focal infiltration, and (5) "salt-and-pepper" pattern [12, 21]. For the quick and complete assessment of all patterns, a combination of a T1WI and T2WI with fat suppression should be employed [21].

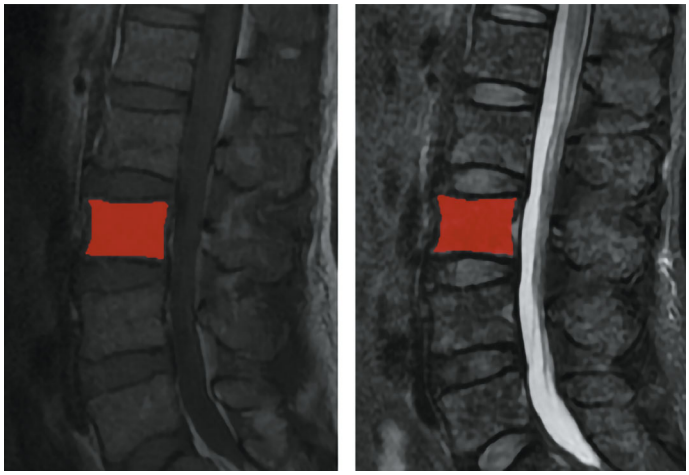

Figure 1: ROI segmentation.

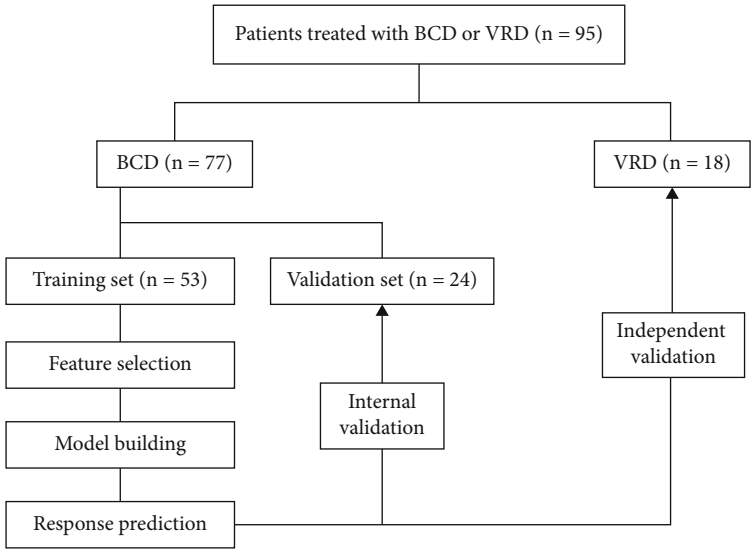

Figure 2: The process for MM response prediction of bortezomib, cyclophosphamide, and dexamethasone (BCD) and bortezomib, lenalidomide, and dexamethasone (VRD).

2.5. Radiomics Analysis

2.5.1. Image Preprocessing, ROI Segmentation, and Feature Extraction. All the eligible images were imported to the Artificial Intelligence Kit software version 3.3.0 (AK, GE Healthcare) for preprocessing, including resampling the image into $1 \times 1 \times 1\ mm^3$, bias field correction, signal smoothing by a Gaussian filter with the standard deviation of 0.5, and intensity standardization by z-score normalization [22].

The regions of interest (ROIs) were segmented by using the ITK-SNAP software v.3.6.0 (http://www.itksnap.org/) [23]. In our study, the entire bone marrow of the second lumbar vertebra was designated as the target region to avoid the discrepancy of the anatomical structure and ensure the stability of results [14]. The ROI was manually segmented by a radiologist with five years' experience; meanwhile, the cortical bone and degenerative changes were carefully avoided. Figure 1 shows the ROI segmentation. Then, all the segmented ROIs were validated by a senior radiologist with 13 years of experience.

Based on AK software, there were a total of 1316 radiomics features extracted from the second lumbar vertebral body, including 18 first-order histogram features, 14 shape features, 75 texture features, and 1209 second-order features generated from the derived images via wavelet transformation, local binary pattern transformation, and Laplacian of Gaussian transformation.

2.5.2. Feature Selection and Radiomics Model Construction. In the present study, 77 patients treated with BCD were used for the feature selection, model establishment, and internal validation, and 18 patients with VRD regimen were independently validated by the BCD-based model. The process of MM response prediction is illustrated in Figure 2.

Seventy-seven patients treated with BCD were randomly divided into a training set ($n = 53$) and a validation set

($n = 24$) at a ratio of 7 : 3. All the extracted features were firstly normalized before feature selection. The outliers were replaced with the median of the particular variance vector. Moreover, the data were standardized by z-score transformation. The standardized formula is as follows: $(\text{fi} - u)/\text{std}$, where fi represents a single characteristic data, u is the average value of the data column, and std pertains to the standard deviation of the data column. In order to simplify the model and increase the interpretability and stability of the model, we used a combination of three commonly used methods to select features, including Pearson correlation analysis (threshold, 0.7), variance threshold method (threshold, 1) and the least absolute shrinkage and selection operator (LASSO).

Based on the selected features, the logistic regression (LR), support vector machine (SVM), Bayes, k-nearest neighbor (KNN), decision tree (DT), and random forest (RF) models were built for MM response prediction. All the models were trained by applying the repeated fivefold cross-validation technique in the training set; then, the performance of models was evaluated in the validation set. The predictive ability was compared by using the DeLong test. Finally, the VRD regimen group was independently validated by the models above.

2.5.3. The Potential Association between the Radiomics Features and PFS. For each of the selected radiomics features, the optimum cutoff value was determined by the Youden index in ROC analysis. Then, the patients were assigned into two groups by using the cutoff values. The PFS was compared between the two groups in each of the selected features.

2.6. Statistical Analysis. The clinical characteristics and MRI patterns of MM were evaluated by the chi-square test. The receiver operating characteristic (ROC) analysis was conducted to assess the model's predictive power, and the area under the curve (AUC), sensitivity, specificity, and accuracy were all calculated. The DeLong test was applied for model comparison. Kaplan-Meier curve analysis and log-rank tests were used to analyze the correlation between radiomics features and PFS. All statistical analyses were performed using R software (version 3.5.1) and SPSS (version 24.0). A two-sided P value < 0.05 was considered significant.

3. Results

3.1. Patients. Among the 95 selected patients, 36 patients were IgG type, 33 were IgA type, 4 were IgD type, 20 were light chain type, and 2 were nonsecretory type. For treatment response, 53 patients were classified as good responders and 42 as poor responders. All the clinical characters and MRI patterns had no significant differences between the good and poor response groups, as summarized in Table 1.

3.2. The Selected Radiomics Feature and Models' Predictive Ability. Based on the training set of the BCD regimen, four significant radiomics features were identified. Two of them were extracted from T1WI and the other two from T2WI, and all these features were wavelet transformed. The odds ratio (OR) of each feature was calculated by multivariate logistic regression. The details are presented in Table 2.

Table 1: Clinical characteristics and MRI patterns of patients.

Variable	Good responders ($n = 53$)	Poor responders ($n = 42$)	P value
Age ≥ 65 (years)	14	15	0.328
Sex			
Female	18	16	0.676
Male	35	26	
BMPC ≥ 60%	11	9	0.566
Treatment			
BCD	43	34	0.593
VRD	10	8	
D-S staging			
II	6	5	0.588
III	47	37	
ISS staging			
I	12	7	0.643
II	18	13	
III	23	22	
R-ISS staging			
I	10	7	0.834
II	33	25	
III	10	10	
MRI pattern			
Normal	9	7	0.191
Focal	4	6	
Diffuse	31	18	
Focal and diffuse	4	9	
Salt-and-pepper	5	2	

BMPC: bone marrow plasma cells; D-S: Durie-Salmon staging system; ISS: International Staging System; R-ISS: Revised International Staging System.

The LR, SVM, Bayes, KNN, DT, and RF models were constructed. The ROC curves are drawn in Figure 3, and the RF model showed the higher AUC value both in the training (0.896) and validation set (0.885) when compared with others, but the differences did not reach statistical significance ($P > 0.05$). Meanwhile, the accuracy, sensitivity, and specificity of each model were also calculated (Table 3). The confusion matrices of different models were shown in supplementary materials (available here).

3.3. Independent Validation of VRD Regimen. The response to the VRD regimen was independently validated by the six models constructed based on the BCD regimen. The AUC range was 0.500-0.719, and the ACC range was 0.556-0.722 (Figure 4).

Table 2: Selected radiomics features and their odds ratio.

Feature	Sequences	OR (95% CI)
Wavelet-HLH_firstorder_kurtosis	T1WI	0.448 (0.229 -0.877)
Wavelet-HLH_glcm_correlation	T1WI	3.784 (1.818 7.876)
Wavelet-HHL_firstorder_kurtosis	T2WI	2.842 (1.328 6.081)
Wavelet-LLH_firstorder_mean	T2WI	4.340 (1.538 12.244)

glcm: gray level cooccurrence matrix; OR: odds ratio; CI: confidence interval.

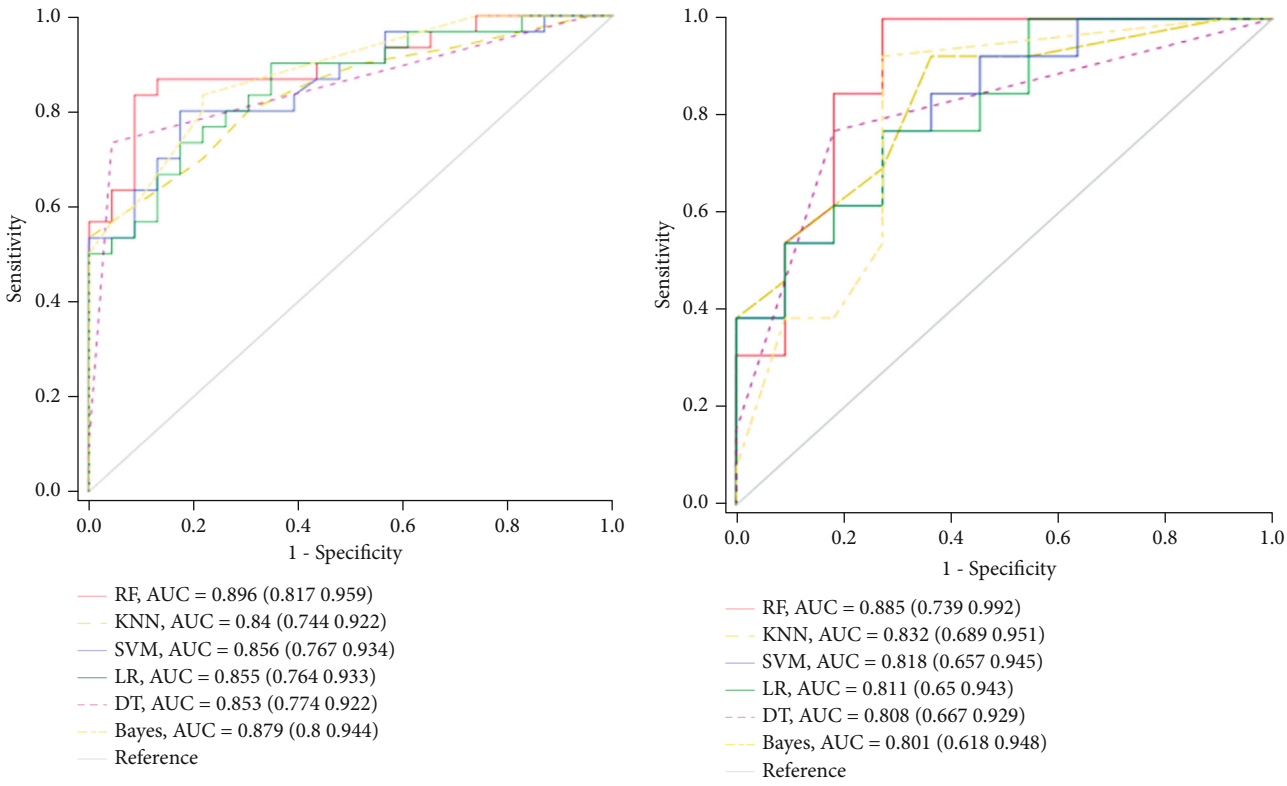

Figure 3: Receiver operating characteristic (ROC) curves of random forest (RF), *k*-nearest neighbor (KNN), support vector machine (SVM), the logistic regression (LR), decision tree (DT), and Bayes models for BCD response prediction. (a) The ROC curves in the training set. (b) The ROC curves in the validation set.

Table 3: The comparison of different models.

Model	Training set				Validation set			
	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE
RF	0.896	0.868	0.867	0.870	0.885	0.833	0.923	0.727
KNN	0.840	0.755	0.800	0.696	0.832	0.792	0.923	0.636
SVM	0.856	0.792	0.800	0.783	0.818	0.708	0.769	0.636
LR	0.855	0.755	0.700	0.826	0.811	0.708	0.692	0.727
DT	0.853	0.830	0.733	0.957	0.808	0.792	0.769	0.818
Bayes	0.879	0.792	0.800	0.783	0.801	0.833	0.923	0.727

AUC: area under the curve; ACC: accuracy; SEN: sensitivity; SPE: specificity; RF: random forest; KNN: *k*-nearest neighbor; SVM: support vector machine; LR: logistic regression; DT: decision tree.

3.4. The Relationship between the Selected Features and PFS. For patients treated with BCD regimen, median PFS was 29.72 months. According to the Youden index, the optimum cutoff values of the four selected radiomics features (wavelet-HLH_firstorder_kurtosis, wavelet-HLH_glcm_correlation, wavelet-HHL_firstorder_kurtosis, and wavelet-LLH_firstorder_mean) were 0.320, 0.297, 0.105, and 0.288, respectively. For wavelet-HLH_firstorder_kurtosis, the PFS was significantly different between the two groups (Figure 5(a)), and the other three features indicated no difference between the two groups (Figures 5(b)–5(d)).

4. Discussion

In the present study, four wavelet-transformed features extracted from the MRI were confirmed as predictive factors for BCD treatment response in MM. Based on the selected features, six predictive models were built, and the RF model holds relatively more predictive power than the other models in both training and validation groups, though not

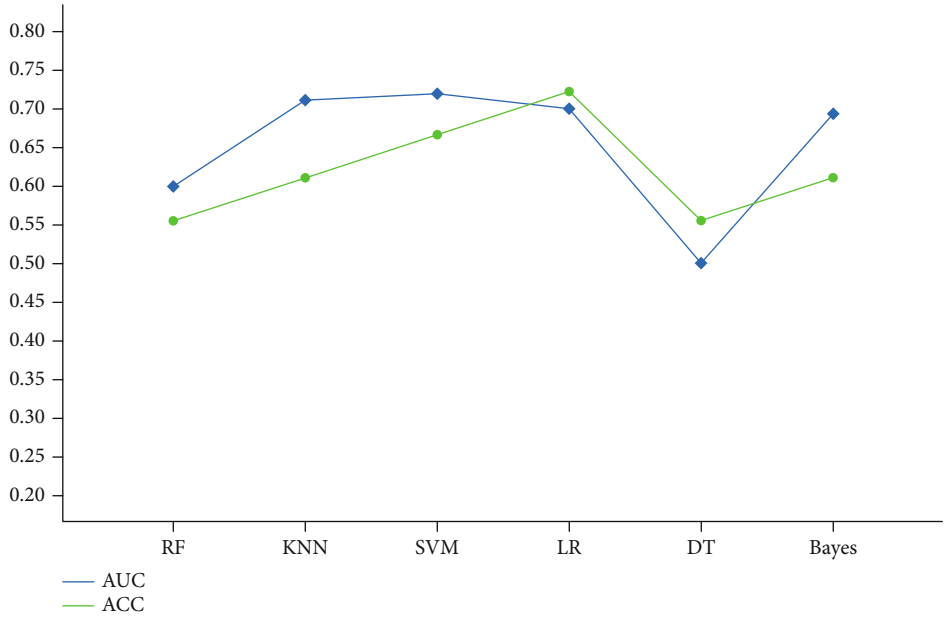

Figure 4: Scatterplots for the area under the curve (AUC) and accuracy (ACC) depiction of RF, KNN, SVM, LR, DT, and Bayes models to predict VRD response.

statistically significant. Meanwhile, all the models lost their predictive capacity when the VRD regimen set was independently validated. Finally, the wavelet-HLH_firstorder_kurtosis was correlated with the PFS, which may be regarded as a relevant biomarker for both the BCD response and survival in patients with MM.

Although achieved PR or better represents therapeutic effectivity, our study defined CR and VGPR as good response and PR, SD, and PD as poor response for further analysis [24]. With the application of new drugs such as bortezomib and lenalidomide, MM patients usually achieved at least PR [25]. Moreover, a previous study showed that at least VGPR after induction could improve progression-free survival [26]. Achieving only PR may not satisfy the clinical expectations entirely. Some studies noted that the quality of response was not a validated effective surrogate for overall survival, but undertreatment could lead to failure of the potential cure and further progression with loss of disease control [25, 27]. Furthermore, a study demonstrated that the lenalidomide maintenance improved the prognosis, particularly among the patients who did not achieve VGPR [28]. Therefore, the achievement of VGPR has a significant implication for clinical practice.

There is an urgent need to predict the efficacy of treatment, but no promising biomarkers are universally accepted at present. Although some studies confirmed the predictive power of gene expression profiles, the results remain controversial [11, 29, 30]. The Durie-Salmon staging system, International Staging System (ISS), and Revised-ISS were commonly used for risk stratification and treatment guidance, but all these staging systems performed predictive ability for survival but were not predictive of response to therapy [10, 31]. In our study, we also found no association of staging systems between the good and poor response groups, and bone marrow plasma cell percentage $\geq 60\%$ that reflects high tumor burden also had no ability to distinguish the quality of response. In addition, our study showed that the MRI patterns had no relationship with response to therapy. This result was reasonable, for the previous meta-analysis demonstrated that the MRI patterns were associated with survival [32], but no study indicated a correlation between MRI patterns and treatment response.

In recent years, several studies have explored the role of CT or MRI radiomics on MM treatment evaluation, the results showed good performance of radiomics, but a limitation was mentioned that patients who underwent different treatment regimens were not comparable [33, 34]. In our study, the commonly used BCD and VRD regimens were selected and analyzed separately in order to provide more reliable results. For MRI-based radiomics, Ekert et al. [33] explored the logistic model in MM treatment response assessment and reported AUC values of 0.60-0.84 based on different sequences. Our study explored and compared six predictive models both in the training and validation set; the results confirmed their predictive power, which may serve as a more comprehensive reference for future research. In addition, previous studies confirmed the efficiency of diffusion-weighted imaging (DWI) and dynamic contrasted enhanced- (DCE-) MRI [13, 14]. Therefore, the functional imaging-based radiomics could be explored in future research. However, the resolution of DWI was limited, and the DCE-MRI was rarely used due to the renal impairment in MM. The routine T1WI and T2WI were still the most frequently used for MM examination, and the radiomics based on routine sequences in our study may have the potential for better clinical practicability and higher clinical application value.

Interestingly, the four radiomics features selected from the T1WI and T2WI were all wavelet transformed high-dimensional features, which indicated that the radiomics

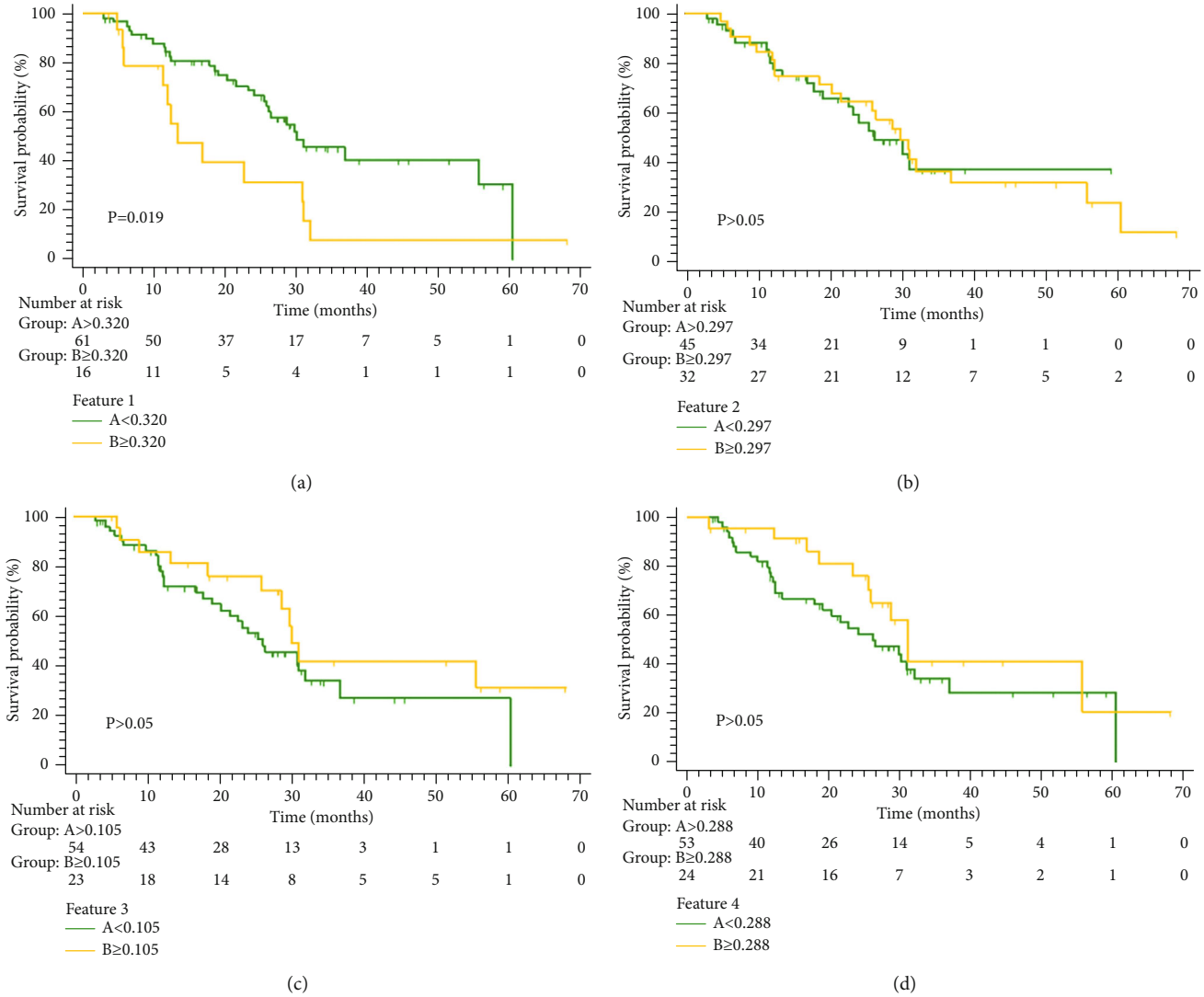

Figure 5: Kaplan-Meier progression-free survival (PFS) analysis of four selected wavelet features. (a) The PFS analysis of feature 1. The PFS of groups A (<0.320) and B (≥0.320) showed a significant difference. (b–d) The PFS analysis of features 2, 3, and 4. The PFS showed no significant difference between the A and B groups of each feature. Features 1, 2, 3, and 4 were wavelet-HLH_firstorder_kurtosis, wavelet-HLH_glcm_correlation, wavelet-HHL_firstorder_kurtosis, and wavelet-LLH_firstorder_mean, respectively.

features for MM response prediction might be more complex, and wavelet features may play a crucial role. In a previous study for rectal cancer chemotherapy response prediction, the majority of the predictive features were wavelet transformed, and authors also cited many studies with other purposes such as prediction for lymph node metastasis to stress the importance of wavelet features [16]. In a recent study for predicting the response of osteosarcoma, the wavelet transformed features also accounted for most of the proportion, and researchers speculated that the wavelet features might be more sensitive for treatment prediction [19]. To our knowledge, apart from ours, there was only one previous study that utilized the MRI radiomics for MM treatment assessment [33], but the wavelet transformed features were not included for analysis, and in-depth explorations were needed in future studies.

As shown in some studies, different machine learning models could influence the predictive performance of radiomics [35, 36]. Nevertheless, there was no consensus on which would be better. Our study constructed six predictive models, including RF, KNN, SVM, LR, DT, and Bayes, and all of them were confirmed as effective methods in many previous radiomics reports [17–19, 36]. As the results showed, all these models had the predictive ability since the AUC values in each of them were greater than 0.80 in both training and validation groups. Additionally, the RF model outperformed the other five models, although this result was not statistically significant. The radiomics models were recommended for BCD response prediction. Moreover, other effective methods such as deep learning should be explored with a large sample in the future.

Though the established models exhibited some predictive capacity for response to the BCD regimen, all these models were incapable of predicting the treatment response of the VRD regimen. To some extent, it was interpretable that some studies indicated that the effect of VRD was

superior to the BCD regimen [37, 38]. Clinically, there were many patients who achieved poor response with BCD, but the replacement with VRD improved the efficacy of treatment. Therefore, our study analyzed the two regimens separately, despite both of them commonly using bortezomib-based therapy in clinical practice. In addition, the result may also be affected by the smaller number size in the VRD group. It is necessary to continue collecting data for VRD-specific predictive model building in the future. Ultimately, other regimens such as BTD and PAD were not studied due to the small sample size and the limited practical clinical application.

Aside from the prediction of response to bortezomib-based therapy, our study also analyzed the correlation between the four selected wavelet features and PFS in patients with MM and confirmed that the wavelet-HLH_firstorder_kurtosis was associated with PFS. Our research was not the first to explore the connection between treatment response-related radiomics and survival. As reported in some other cancer studies, the radiomics score calculated based on selected radiomics features for response prediction was also survival-associated [18, 39]. In contrast to these studies, we explored the relationship between each of the selected features and PFS and then identified that one feature was associated with survival, while the other three were only treatment response-related features with no survival significance. The confirmed feature that correlated with both treatment response and survival may allow clinicians to develop more individualized treatment strategies for MM patients.

There were several limitations in the present study. First, the inevitable selection bias of retrospective design. Second, the sample size was small, and multicenter studies with a large sample size were required. Third, all the ROIs were manually delineated. Further study should explore the method for automatic segmentation to avoid this laborious and time-consuming process. Lastly, the parameters of the models were optimized according to experience or experimental adjustment, which may not be the most effective.

In conclusion, our study provided six effective models based on four significant wavelet features for predicting the response to the BCD regimen, which was of great value as none of the clinical characters or conventional MRI patterns had the predictive ability. Meanwhile, one wavelet feature was also found to correlate with MM survival. MRI-based radiomics had the potential to guide clinicians in MM management.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' Contributions

Study management and guidance was handled by Nan Hong. Study design was handled by Nan Hong and Yang Li. Clinical data acquisition and analysis were handled by Yang Li, Ping Yin, and Yang Liu. Imaging data acquisition and analysis were handled by Yang Li, Ping Yin, and Chao Sun. Statistical analysis was handled by Yang Li, Sicong Wang, and Chuanxi Hao. Manuscript preparation was handled by Yang Li, Ping Yin, and Lei Chen. Manuscript review was handled by Nan Hong. Yang Li and Ping Yin are co-first authors on the paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 81971575).

Supplementary Materials

The supplementary material shows the confusion matrices for the different models. *(Supplementary Materials)*

References

[1] D. E. Joshua, C. Bryant, C. Dix, J. Gibson, and J. Ho, "Biology and therapy of multiple myeloma," *The Medical Journal of Australia*, vol. 210, no. 8, pp. 375–380, 2019.

[2] H. Ludwig, P. Moreau, M. A. Dimopoulos et al., "Health-related quality of life in the ENDEAVOR study: carfilzomib-dexamethasone vs bortezomib-dexamethasone in relapsed/refractory multiple myeloma," *Blood Cancer Journal*, vol. 9, no. 3, p. 23, 2019.

[3] S. K. Kumar, S. V. Rajkumar, A. Dispenzieri et al., "Improved survival in multiple myeloma and the impact of novel therapies," *Blood*, vol. 111, no. 5, pp. 2516–2520, 2008.

[4] S. V. Rajkumar and S. Kumar, "Multiple myeloma current treatment algorithms," *Blood Cancer Journal*, vol. 10, no. 9, p. 94, 2020.

[5] E. Grywalska, B. Sosnowska-Pasiarska, J. Smok-Kalwat, M. Pasiarski, P. Niedzwiedzka-Rystwej, and J. Rolinski, "Paving the way toward successful multiple myeloma treatment: chimeric antigen receptor T-cell therapy," *Cells-Basel*, vol. 9, no. 4, p. 983, 2020.

[6] Y. Xu, L. Xing, J. Su, X. Zhang, and W. Qiu, "Model-based clustering for identifying disease-associated SNPs in case- control genome-wide association studies," *Scientific Reports*, vol. 9, no. 1, article 13686, 2019.

[7] S. Gandolfi, J. P. Laubach, T. Hideshima, D. Chauhan, K. C. Anderson, and P. G. Richardson, "The proteasome and proteasome inhibitors in multiple myeloma," *Cancer Metastasis Reviews*, vol. 36, no. 4, pp. 561–584, 2017.

[8] K. Scott, P. J. Hayden, A. Will, K. Wheatley, I. Coyne, and Cochrane Haematological Malignancies Group, "Bortezomib for the treatment of multiple myeloma," *Cochrane Database of Systematic Reviews*, vol. 4, article D10816, 2016.

[9] S. G. Cengiz and M. Beksac, "The safety of bortezomib for the treatment of multiple myeloma," *Expert Opinion on Drug Safety*, vol. 17, no. 9, pp. 953–962, 2018.

[10] D. Smith and K. Yong, "Advances in understanding prognosis in myeloma," *British Journal of Haematology*, vol. 175, no. 3, pp. 367–380, 2016.

[11] S. B. Amin, W. K. Yip, S. Minvielle et al., "Gene expression profile alone is inadequate in predicting complete response in multiple myeloma," *Leukemia*, vol. 28, no. 11, pp. 2229–2234, 2014.

[12] E. Zamagni, P. Tacchetti, and M. Cavo, "Imaging in multiple myeloma: how? When?," *Blood*, vol. 133, no. 7, pp. 644–651, 2019.

[13] M. Horger, K. Weisel, W. Horger, A. Mroue, M. Fenchel, and M. Lichy, "Whole-body diffusion-weighted MRI with apparent diffusion coefficient mapping for early response monitoring in multiple myeloma: preliminary results," *AJR. American Journal of Roentgenology*, vol. 196, no. 6, pp. W790–W795, 2011.

[14] M. Merz, J. Ritsch, C. Kunz et al., "Dynamic contrast-enhanced magnetic resonance imaging for assessment of anti-angiogenic treatment effects in multiple myeloma," *Clinical Cancer Research*, vol. 21, no. 1, pp. 106–112, 2015.

[15] A. Zwanenburg, M. Vallieres, M. A. Abdalah et al., "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.

[16] B. Petresc, A. Lebovici, C. Caraiani, D. S. Feier, F. Graur, and M. M. Buruian, "Pre-treatment T2-WI based radiomics features for prediction of locally advanced rectal cancer non-response to neoadjuvant chemoradiotherapy: a preliminary study," *Cancers (Basel)*, vol. 12, no. 7, p. 1894, 2020.

[17] J. T. Antunes, A. Ofshteyn, K. Bera et al., "Radiomic features of primary rectal cancers on baseline T2-weighted MRI are associated with pathologic complete response to neoadjuvant chemoradiation: a multisite study," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 5, pp. 1531–1541, 2020.

[18] L. Zhao, J. Gong, Y. Xi et al., "MRI-based radiomics nomogram may predict the response to induction chemotherapy and survival in locally advanced nasopharyngeal carcinoma," *European Radiology*, vol. 30, no. 1, pp. 537–546, 2020.

[19] H. Chen, X. Zhang, X. Wang et al., "MRI-based radiomics signature for pretreatment prediction of pathological response to neoadjuvant chemotherapy in osteosarcoma: a multicenter study," *European Radiology*, vol. 31, no. 10, pp. 7913–7924, 2021.

[20] S. Kumar, B. Paiva, K. C. Anderson et al., "International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma," *The Lancet Oncology*, vol. 17, no. 8, pp. e328–e346, 2016.

[21] A. Baur-Melnyk, S. Buhmann, H. R. Durr, and M. Reiser, "Role of MRI for the diagnosis and prognosis of multiple myeloma," *European Journal of Radiology*, vol. 55, no. 1, pp. 56–63, 2005.

[22] M. Bologna, V. Corino, and L. Mainardi, "Technical note: virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain," *Medical Physics*, vol. 46, no. 11, pp. 5116–5123, 2019.

[23] P. A. Yushkevich, J. Piven, H. C. Hazlett et al., "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroiamge*, vol. 31, no. 3, pp. 1116–1128, 2006.

[24] A. Latifoltojar, M. Hall-Craggs, N. Rabin et al., "Whole body magnetic resonance imaging in newly diagnosed multiple myeloma: early changes in lesional signal fat fraction predict disease response," *British Journal of Haematology*, vol. 176, no. 2, pp. 222–233, 2017.

[25] S. Lonial and K. C. Anderson, "Association of response endpoints with survival outcomes in multiple myeloma," *Leukemia*, vol. 28, no. 2, pp. 258–268, 2014.

[26] P. Moreau, M. Attal, B. Pegourie et al., "Achievement of VGPR to induction therapy is an important prognostic factor for longer PFS in the IFM 2005-01 trial," *Blood*, vol. 117, no. 11, pp. 3041–3044, 2011.

[27] K. C. Anderson, R. A. Kyle, S. V. Rajkumar et al., "Clinically relevant end points and new drug approvals for myeloma," *Leukemia*, vol. 22, no. 2, pp. 231–239, 2008.

[28] M. Attal, V. Lauwers-Cances, G. Marit et al., "Lenalidomide maintenance after stem-cell transplantation for multiple myeloma," *The New England Journal of Medicine*, vol. 366, no. 19, pp. 1782–1791, 2012.

[29] A. J. Vangsted, S. Helm-Petersen, J. B. Cowland et al., "Drug response prediction in high-risk multiple myeloma," *Gene*, vol. 644, pp. 80–86, 2018.

[30] E. A. Punnoose, J. D. Leverson, F. Peale et al., "Expression profile of BCL-2, BCL-XL, and MCL-1 predicts pharmacological response to the BCL-2 selective antagonist venetoclax in multiple myeloma models," *Molecular Cancer Therapeutics*, vol. 15, no. 5, pp. 1132–1144, 2016.

[31] A. Silva, M. C. Silva, P. Sudalagunta et al., "An ex vivo platform for the prediction of clinical response in multiple myeloma," *Cancer Research*, vol. 77, no. 12, pp. 3336–3351, 2017.

[32] S. Y. Lee, H. J. Kim, Y. R. Shin, H. J. Park, Y. G. Lee, and S. J. Oh, "Prognostic significance of focal lesions and diffuse infiltration on MRI for multiple myeloma: a meta-analysis," *European Radiology*, vol. 27, no. 6, pp. 2333–2347, 2017.

[33] K. Ekert, C. Hinterleitner, K. Baumgartner, J. Fritz, and M. Horger, "Extended texture analysis of non-enhanced whole-body MRI image data for response assessment in multiple myeloma patients undergoing systemic therapy," *Cancers (Basel)*, vol. 12, no. 3, p. 761, 2020.

[34] C. P. Reinert, E. M. Krieg, H. Bosmuller, and M. Horger, "Mid-term response assessment in multiple myeloma using a texture analysis approach on dual energy-CT-derived bone marrow images – a proof of principle study," *European Journal of Radiology*, vol. 131, article 109214, 2020.

[35] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. Aerts, "Machine learning methods for quantitative radiomic biomarkers," *Scientific Reports*, vol. 5, no. 1, p. 13087, 2015.

[36] S. Bernatz, J. Ackermann, P. Mandel et al., "Comparison of machine learning algorithms to predict clinically significant prostate cancer of the peripheral zone with multiparametric MRI using clinical assessment categories and radiomic features," *European Radiology*, vol. 30, no. 12, pp. 6757–6769, 2020.

[37] P. Moreau, M. J. San, P. Sonneveld et al., "Multiple myeloma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†," *Annals of Oncology*, vol. 28, pp. iv52–iv61, 2017.

[38] K. Uttervall, B. J. Borg, C. Gran et al., "Upfront bortezomib, lenalidomide, and dexamethasone compared to bortezomib, cyclophosphamide, and dexamethasone in multiple myeloma," *European Journal of Haematology*, vol. 103, no. 3, pp. 247–254, 2019.

[39] K. J. Park, J. L. Lee, S. K. Yoon, C. Heo, B. W. Park, and J. K. Kim, "Radiomics-based prediction model for outcomes of PD-1/PD-L1 immunotherapy in metastatic urothelial carcinoma," *European Radiology*, vol. 30, no. 10, pp. 5392–5403, 2020.