Skeletal Radiology (2025) 54:267–273
https://doi.org/10.1007/s00256-024-04733-0

SCIENTIFIC ARTICLE

Whole-body low-dose computed tomography in patients with newly diagnosed multiple myeloma predicts cytogenetic risk: a deep learning radiogenomics study

Shahriar Faghani[1] · Mana Moassefi[1] · Udit Yadav[2] · Francis K. Buadi[3] · Shaji K. Kumar[3] · Bradley J. Erickson[1] · Wilson I. Gonsalves[3] · Francis I. Baffour[1]

Received: 4 April 2024 / Revised: 31 May 2024 / Accepted: 12 June 2024 / Published online: 27 June 2024

Abstract

Objective To develop a whole-body low-dose CT (WBLDCT) deep learning model and determine its accuracy in predicting the presence of cytogenetic abnormalities in multiple myeloma (MM).

Materials and methods WBLDCTs of MM patients performed within a year of diagnosis were included. Cytogenetic assessments of clonal plasma cells via fluorescent in situ hybridization (FISH) were used to risk-stratify patients as high-risk (HR) or standard-risk (SR). Presence of any of del(17p), t(14;16), t(4;14), and t(14;20) on FISH was defined as HR. The dataset was evenly divided into five groups (folds) at the individual patient level for model training. Mean and standard deviation (SD) of the area under the receiver operating curve (AUROC) across the folds were recorded.

Results One hundred fifty-one patients with MM were included in the study. The model performed best for t(4;14), mean (SD) AUROC of 0.874 (0.073). The lowest AUROC was observed for trisomies: AUROC of 0.717 (0.058). Two- and 5-year survival rates for HR cytogenetics were 87% and 71%, respectively, compared to 91% and 79% for SR cytogenetics. Survival predictions by the WBLDCT deep learning model revealed 2- and 5-year survival rates for patients with HR cytogenetics as 87% and 71%, respectively, compared to 92% and 81% for SR cytogenetics.

Conclusion A deep learning model trained on WBLDCT scans predicted the presence of cytogenetic abnormalities used for risk stratification in MM. Assessment of the model's performance revealed good to excellent classification of the various cytogenetic abnormalities.

Keywords Multiple myeloma · Cytogenetic risk · CT skeletal survey

Introduction

Multiple myeloma (MM) is a hematologic malignancy accounting for approximately 1–2% of all cancers in the United States [1]. Although the disease is considered a single entity, there are distinct cytogenetic profiles that reflect heterogeneity in the disease biology that alter the disease course, influence the response to therapy, and dictate prognosis [2]. Patients with MM whose clonal plasma cells contain high-risk cytogenetic abnormalities such as t(4;14), t(14;16), t(14;20), del(17p)/monosomy 17, and gain or amplification (1q) experience worse overall outcomes compared to the remainder of MM patients without these abnormalities [3, 4]. Conventionally, the cytogenetic assessment is performed via fluorescent in situ hybridization (FISH) on clonal plasma cells obtained from the staging aspirate of the bone marrow from the posterior iliac crest.

Despite this advancement in risk stratification via FISH testing, the spatial heterogeneity of clonal plasma cells throughout the body implies that sampling only the clonal plasma cells from the posterior iliac crest may not uncover the presence of clonal plasma cells with differing cytogenetic profiles located in other anatomical regions of a patient

Portions of this paper are presented at ARRS 2024.

✉ Francis I. Baffour
baffour.francis@mayo.edu

[1] Department of Radiology, Mayo Clinic, 200 1st St SW, Rochester, MN 55905, USA

[2] Division of Hematology, Mayo Clinic, 13400 E. Shea Blvd, Scottsdale, AZ 85259, USA

[3] Division of Hematology, Mayo Clinic, 200 1st St SW, Rochester, MN 55905, USA

[5]. This is especially important for patients with conditions such as solitary bone plasmacytomas characterized by a single mass of clonal plasma cells arising from the bone in areas other than the posterior iliac crest that may or may not contain minimal amounts of clonal plasma cells [6]. Additionally, in some newly diagnosed MM patients with macrofocal disease, the bone marrow aspirates obtained from the posterior iliac crest may not always yield sufficient clonal plasma cells that can be utilized for cytogenetics analysis by FISH [7].

Whole-body low-dose computed tomography (WBLDCT) is commonly used to detect osteolytic bone lesions, which is one of the primary causes of morbidity from MM, and it has been proven to be more sensitive than conventional skeletal X-rays [8]. Since the scan volume includes the whole body, analysis of CT images may provide further information about the disease process throughout the whole body rather than from a single biopsied site. Furthermore, recent advances in deep learning techniques have allowed the development of artificial intelligence models to incorporate these image-based features into predictive models. To this end, we hypothesize that a deep learning model may detect differences in CT features among a heterogeneous cohort of MM patients that could predict the presence of specific cytogenetic abnormalities typically detected in clinical practice by FISH performed on the clonal plasma cells obtained from the bone marrow aspirate derived from the posterior iliac crest. Thus, this study aimed to develop such a non-invasive imaging-based deep learning model and determine its accuracy in predicting the presence of various cytogenetic abnormalities determined by conventional FISH assessments.

Materials and methods

Study cohort

The Mayo Clinic electronic medical record system was queried to identify all patients with a new diagnosis of MM as defined by the 2014 International Myeloma Working Group (IMWG) criteria [9] who had a WBLDCT exam performed within a year of their diagnosis for potential inclusion in this study. Patients with newly diagnosed MM for whom cytogenetic data by FISH was not available or not performed and/or did not have a WBLDCT were excluded from this analysis. Patients with monoclonal gammopathy of undetermined significance (MGUS) and smoldering MM were not included in this study. Clinical, laboratory, and pathologic data regarding these patients were extracted from prospectively maintained databases and a review of their electronic medical records. All patients had consented to use their medical records, and the study was conducted in accordance with the institutional guidelines with the approval of the institutional review board and in accordance with the principles of the Helsinki Declaration.

Whole-body low-dose CT scan parameters

The WBLDCTs included in this study were from a single institution. The scan range was a single acquisition from above the raised elbows through the knees. CT acquisition and reconstruction parameters were as follows: tube potential of 120 kV; 128 × 0.6 or 192 × 0.6 mm collimation; rotation time of 0.5 s with a pitch of 1; quality reference mAs of 100; CT dose index if 4.72 mGy; slice increment of 1 mm; 2-mm or 3-mm section thickness; 512 × 512 matrix; a smooth kernel series for bone marrow and soft tissue evaluation; and a sharp kernel series to evaluate the bones. The entire volume of the axial smooth kernel series was used for model development.

Risk stratification for multiple myeloma by FISH assessments of clonal plasma cells

Cytogenetic assessments of clonal plasma cells via FISH were used to risk-stratify patients as having either high-risk (HR) or non-HR (standard risk—SR) cytogenetics as per the Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART) guidelines [10]. The following FISH probes were utilized routinely to detect the presence of trisomies (3, 5, 7, 9, 11, 13, 15, and 17), del(17p), t(11;14), t(14;16), t(4;14), and t(14;20) which were defined as HR. Patients whose MM cells had the presence of any of the following FISH abnormalities, such as del(17p) or monosomy 17, t(14;16), t(4;14), and t(14;20), were classified as having HR FISH. The FISH assay was performed on the sorted MM cells obtained from the bone marrow aspiration from the posterior iliac crest, and a minimum of 50 nuclei were analyzed per probe set. Probe sites with less than 15 interphase nuclei were deemed insufficient for analysis. FISH analysis was performed by two qualified clinical cytogenetic technologists and interpreted by a board-certified clinical cytogeneticist. The level of detection required to identify abnormalities for individual probe sets was as follows: a minimum of 3 of 50 cells (6%) displaying fusion signals in the setting of dual fusion probes, a minimum of 5 of 50 cells (10%) with disrupted separated signals in the setting of BAP probes, a minimum 5 of 50 cells (10%) for tetraploid clones.

Dataset splitting

The dataset was evenly divided into five groups (folds) at the individual patient level using the GroupKfold function from the sci-kit-learn library, version 1.2.0. To determine the reliability of the model outcomes, the development

utilized a fivefold cross-validation. The model was trained on four folds for each development cycle and validated on the remaining one in each iteration [11]. With this fivefold cross-validation approach, each fold serves as a validation for the remaining 4-folds such that 80% of the data was committed to training and 20% to testing.

Data preprocessing and model development

2D axial CT images were converted to 3D volumes with appropriate normalization, resizing, and data augmentation to prevent overfitting [12]. A 3D-DenseNet-121 classifier, optimized with AdamW and employing a CosineAnnealing learning rate scheduler, was used to train on the preprocessed images [13–18]. To address dataset imbalances, a weighted cross-entropy loss function was applied [19]. This approach incorporated occlusion interpretation maps to identify critical regions within the images influencing model predictions. Model training leveraged a GPU cluster with NVIDIA A100 GPUs, utilizing PyTorch and MONAI frameworks. A comprehensive description of the data preprocessing, model development techniques, and occlusion interpretation approach are included as Figs. 1 and 2 and in Supplementary Material. The code for the algorithm can be found at [20].

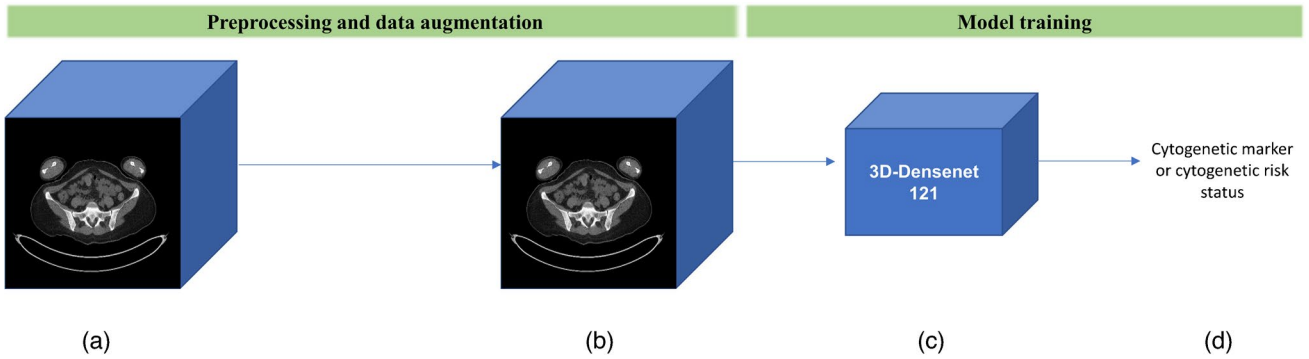

Fig. 1 Schematic illustration of the preprocessing, data augmentation, and model training pipeline. **a** 3D low-dose whole-body CT volume. **b** Preprocessed and augmented low-dose whole-body CT to mitigate overfitting. **c** 3D-DenseNet model. **d** Prediction of cytogenetic marker status

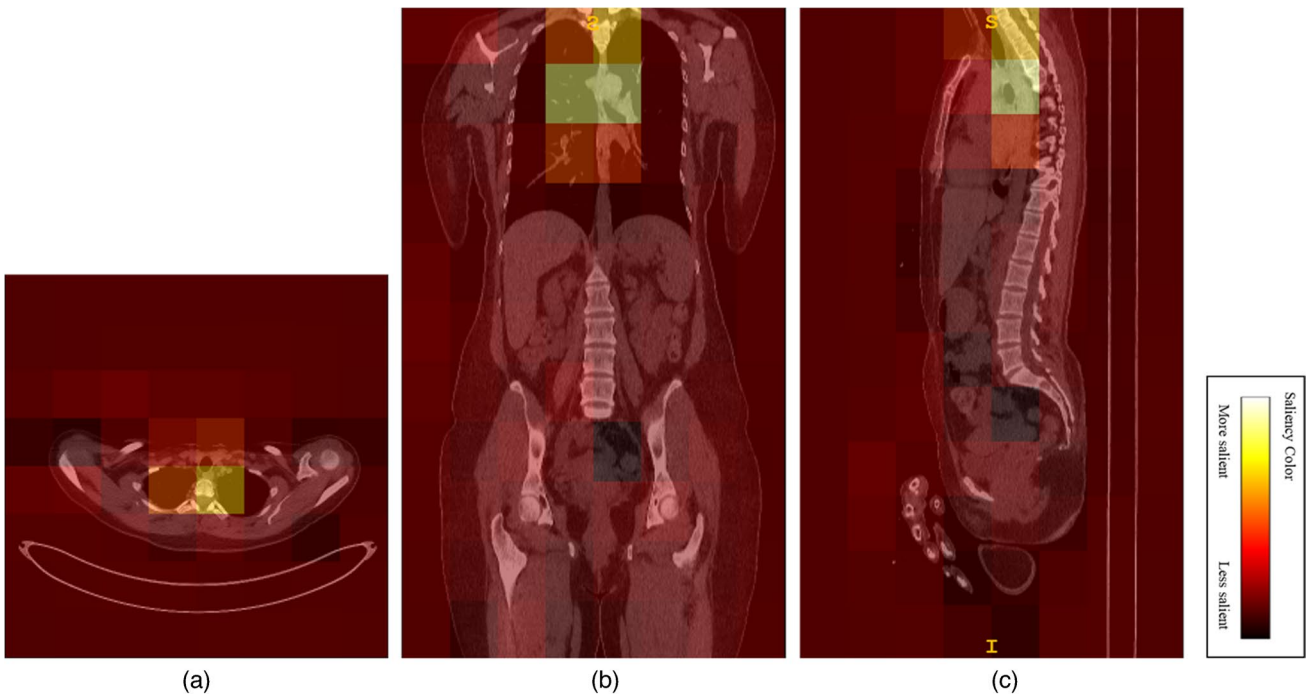

Fig. 2 Illustration of key regions in the model's decision-making for predicting the t(11, 14) biomarker. Axial (**a**), coronal (**b**), and sagittal (**c**) views of a whole-body CT scan (WBCT) with an overlay of the occlusion mask

Statistical methods

Participant characteristics were summarized using the median and interquartile range for age and frequency for categorical variables. For cytogenetic risk and each cytogenetic marker in the study, mean and standard deviation (SD) of the area under the receiver operating curve (AUROC) across the folds were reported to showcase the discriminative ability of our model [21].

Overall survival (OS) was measured from the day of diagnosis of MM to death from any cause, with censoring performed at the date of last contact. Kaplan–Meier analysis was used to analyze and create the OS curves, and log-rank test was used to compare survival curves.

Results

A total of 151 patients with MM were included in the study. The baseline demographics are summarized in Table 1. The median age of the participants was 63.4 years, with an interquartile range of 11.4 years. 21.2%, 38.4%, and 40.4% of patients were ISS stage I, II, and III, respectively. Table 1 summarizes the cytogenic risk of the patients and the distribution of each cytogenetic abnormality in the patient cohort. Overall, 43 patients (28%) were classified as having HR cytogenetics and 108 (72%) as SR cytogenetics.

Table 1 Participant characteristics, international staging system status, and cytogenetic markers

Subject characteristics	Frequency (unless otherwise stated)
Median age in years (IQR) (range in years)	63.4 (11.4) (33–83)
Gender (*N*, %)	
Female	52
Male	99
International staging system (*N*)	
Stage I	32
Stage II	58
Stage III	61
FISH risk (*N*)	
Standard risk	108
High risk *	43
Distribution of FISH abnormality	
(*N*)**	19
t(4:14) present	36
t(11:14) present	7
t(14:16) present	1
t(6;14) present	1
t(14;20) present	23
Deletion 17p/monosomy 17	65
Any trisomy present	29
Normal FISH	

FISH, fluorescent in situ hybridization; *IQR*, interquartile range

*High-risk FISH includes t(4;14), t(14;16), t(14;20), and monosomoy 17/del 17p

**Some patients may have more than 1 FISH abnormality; thus, the total sum is not equal to 151

Model performance—accuracy

Model performance is reported as average AUROCs across all 5-folds and is summarized in Table 2. For the specific cytogenetic markers, the model performed best for t(4;14) with an average (SD) AUROC of 0.874 (0.073). The lowest accuracy was observed for trisomies with an average (SD) AUROC of 0.717 (0.058). The median overall survival of patients based on high vs. standard risk cytogenetics was not reached for both groups (Fig. 3a). However, the 2- and 5-year survival rates for patients with high-risk cytogenetics were 87% and 71%, respectively, compared to 91% and 79% observed in patients with standard-risk cytogenetics. For six patients with more than one HR cytogenetic abnormality, the 2- and 5-year survival rates were 77% and 53%, respectively. When assessing the overall survival based on cytogenetic predictions by the AI model generated from the WBLDCT images from the same patients, the median overall survival of patients was again not reached for either group (Fig. 3b). However, the 2- and 5-year survival rates for patients with high-risk cytogenetics were also 87% and 71%, respectively, compared to 92% and 81% observed in patients with standard-risk cytogenetics.

Discussion

Assessment of disease risk is critical for appropriate management of MM. It has been well established that MM comprises a group of highly heterogeneous diseases in terms of underlying genetic abnormalities, clinical presentation, and response to treatment, and these ultimately impact survival outcomes. Since a number of these genetic abnormalities

Table 2 Average area under the receiver operating curve across all 5-folds for each model

Model	Mean area under the receiving operating curve	Standard deviation
1 (t(4:14))	0.874	0.0730
2 (t(11:14))	0.805	0.0550
3 (Presence of HR FISH)	0.763	0.0280
4 (Any trisomy)	0.717	0.0584

FISH, fluorescent in situ hybridization; *HR*, high risk

High-risk FISH includes t(4;14), t(14;16), t(14;20), and monosomoy 17/del 17p

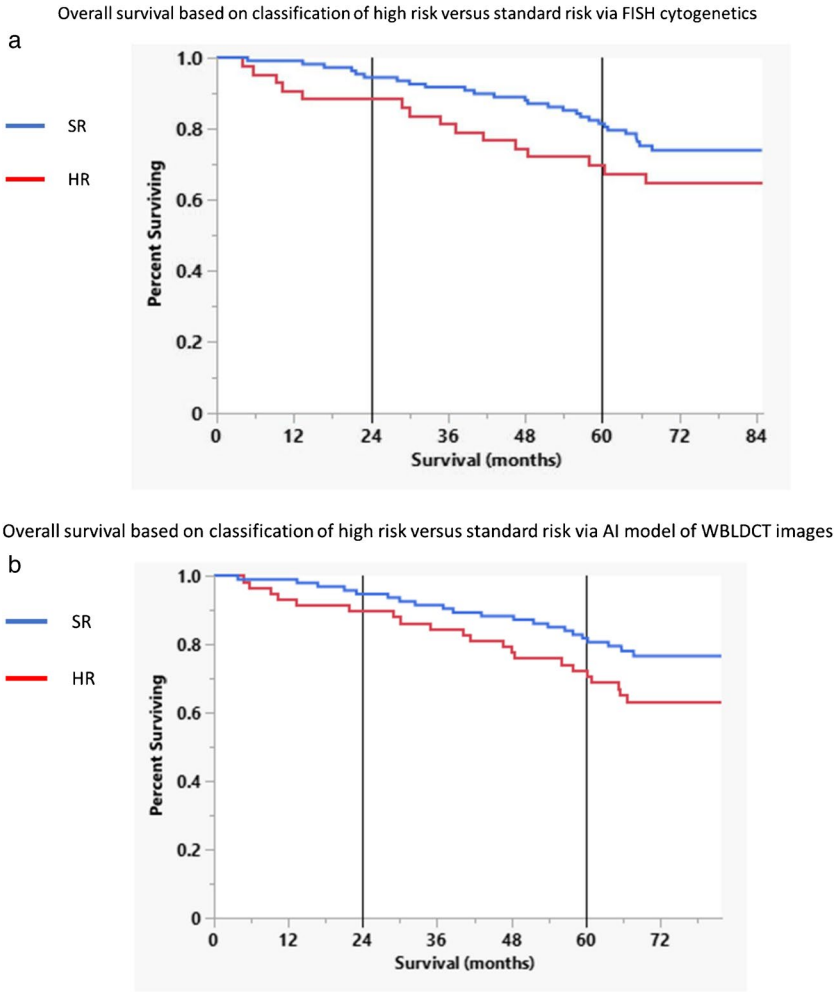

Fig. 3 **a** Overall survival based on classification of high risk versus standard risk via fluorescent in situ hybridization (FISH) cytogenetics. **b** Overall survival based on classification of high risk versus standard risk via AI model of whole-body low-dose computed tomography (WBLDCT) images

detected by FISH are associated with poor survival, prior knowledge of this risk is of considerable importance when comparing the outcomes of patients based on different treatments. In this study, a deep learning model was trained on WBLDCT scans to predict for the presence of cytogenetic abnormalities in their clonal plasma cells that are used for risk stratification. Assessment of the model's performance revealed good to excellent classification of various cytogenetic abnormalities at the patient level with AUROC ranging between 0.717 and 0.874. While WBLDCTs have traditionally been useful mainly for osteolytic bone lesion assessment and associated sequela such as fractures, the volume of CT data comprises the entire body and provides an unrealized opportunity to detect radiogenomic features that reflect the cytogenetic diversity of clonal plasma cells present within a given patient with MM. Therefore, successful CT-based radiomic models may supplant or at least be complementary to invasive approaches that are focal and, as such, may not offer a comprehensive assessment of the entire heterogeneity of disease burden, associated cytogenetic abnormalities, and, consequently, the disease biology. This is especially important for MM cases such as those with macrofocal MM that have a small volume of clonal plasma cell burden in their posterior iliac crest, and as a result, almost 25% of such patients do not have sufficient numbers of clonal plasma cells to perform an adequate cytogenetic assessment by FISH [7]. Finally, it is possible that the worse-than-expected clinical and survival outcomes observed in certain MM patients considered not to have any high-risk cytogenetics [22] may arise from the clonal plasma cells in the posterior iliac crest being unable to reflect the presence of HR cytogenetics in clonal plasma cells present in other anatomical bone marrow regions.

Overall, the deep learning model had good or excellent performance for various cytogenetic abnormalities. While the reason for this variation is unclear, with the notably excellent performance at predicting t(4;14), the overall model prediction of HR FISH may be driven by this particular cytogenetic marker. The saliency maps of the model suggests that although no specific region is the constant "focus" of the algorithm, the spine and paraspinal soft tissue are of most relevance to the model's prediction.

It is possible that a model trained specifically on a smaller field-of-view of the spine and pelvis may demonstrate higher and more consistent performance than shown here. While there has been a few studies that have utilized radiomics-based approaches to extract features from PET/CT scans to help better prognosticate patients based on their survival outcomes [23–25], there has been only one study that utilized cross-sectional imaging of spinal MRIs to predict the presence of HR-FISH [26]. Thus, this is the first and the largest study in consecutive newly diagnosed MM patients using WBLDCT to predict the presence of specific HR-FISH cytogenetics.

While this study is novel and clinically relevant, there are a few limitations. First, since this was a pilot study, the sample size of the study cohort is small, and future validation studies with larger sample sizes are required. Second, the results of this study may not be generalizable across all populations. Not all cytogenetic markers associated with MM were present in this cohort, for instance, 1q abnormalities are not represented in this cohort. Additionally, the cases used to train the models reported here are from a single medical center, and all CT images were obtained using similar acquisition and reconstruction parameters. CT section thickness and reconstruction kernels derived from other WBLDCT scanners may impact the performance of these algorithms. Nevertheless, the training cohort comprised 2-mm and 3-mm section thicknesses, a common choice across CT practices. Finally, not all oncology centers utilize the WBLDCT as their choice of baseline cross-sectional imaging for newly diagnosed MM patients and may choose to use PET/CT scans or whole-body MRIs as the preferred imaging modality in their clinical practice. While this reduces the applicability of this model to all MM patients, this study still serves as a proof of concept that can lead to the evaluation and development of similar models trained in various other imaging modalities in the future.

In the future, larger studies incorporating radiogenomics in conjunction with artificial intelligence-based models will be needed to address the pervasive issues of genetic heterogeneity in MM to better stratify patients' risk according to their disease biology. Additionally, developing a regression model to directly predict overall survival would be beneficial.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00256-024-04733-0.

Funding The work is partly supported by grants from the National Cancer Institute of the National Institutes of Health under Award Number R01 CA254961.

Data availability Relevant code has been cited and is publicly available.

Declarations

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of interest The authors declare no competing interests.

Disclaimer The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49.
2. Kumar SK, Rajkumar SV. The multiple myelomas - current concepts in cytogenetic classification and therapy. Nat Rev Clin Oncol. 2018;15(7):409–21.
3. Abdallah NH, Binder M, Rajkumar SV, Greipp PT, Kapoor P, Dispenzieri A, et al. A simple additive staging system for newly diagnosed multiple myeloma. Blood Cancer J. 2022;12(1):21.
4. Chng WJ, Dispenzieri A, Chim CS, Fonseca R, Goldschmidt H, Lentzsch S, et al. IMWG consensus on risk stratification in multiple myeloma. Leukemia. 2014;28(2):269–77.
5. Rasche L, Chavan SS, Stephens OW, Patel PH, Tytarenko R, Ashby C, et al. Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. Nat Commun. 2017;8(1):268.
6. Yadav U, Kumar SK, Baughn LB, Dispenzieri A, Greipp P, Ketterling R, et al. Impact of cytogenetic abnormalities on the risk of disease progression in solitary bone plasmacytomas. Blood. 2023;142(22):1871–8.
7. Katodritou E, Kastritis E, Gatt M, Cohen YC, Avivi I, Pouli A, et al. Real-world data on incidence, clinical characteristics and outcome of patients with macrofocal multiple myeloma (MFMM) in the era of novel therapies: a study of the Greco-Israeli collaborative myeloma working group. Am J Hematol. 2020;95(5):465–71.
8. Moulopoulos LA, Koutoulidis V, Hillengass J, Zamagni E, Aquerreta JD, Roche CL, et al. Recommendations for acquisition, interpretation and reporting of whole body low dose CT in patients with multiple myeloma and other plasma cell disorders: a report of the IMWG Bone Working Group. Blood Cancer J. 2018;8(10):95.
9. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol. 2014;15(12):e538-548.
10. mSMART. https://www.msmart.org/.
11. Rouzrokh P, Khosravi B, Faghani S, Moassefi M, Vera Garcia DV, Singh Y, et al. Mitigating bias in radiology machine learning: 1. Data handling. Radiol Artif Intell. 2022;4(5): e210290.
12. Zhang K, Khosravi B, Vahdati S, Faghani S, Nugen F, Rassoulinejad-Mousavi SM, et al. Mitigating bias in radiology machine learning: 2. Model development. Radiol Artif Intell. 2022;4(5): e220010.
13. Faghani S, Khosravi B, Moassefi M, Conte GM, Erickson BJ. A Comparison of three different deep learning-based models to predict the MGMT promoter methylation status in glioblastoma using brain MRI. J Digit Imaging. 2023;36(3):837–46.

14. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 21–26 July 2017; 2017. p. 2261–2269.
15. Moassefi M, Faghani S, Conte GM, Kowalchuk RO, Vahdati S, Crompton DJ, et al. A deep learning model for discriminating true progression from pseudoprogression in glioblastoma patients. J Neurooncol. 2022;159(2):447–55.
16. Singh Y, Kelm ZS, Faghani S, Erickson D, Yalon T, Bancos I, et al. Deep learning approach for differentiating indeterminate adrenal masses using CT imaging. Abdom Radiol (NY). 2023;48(10):3189–94.
17. Loshchilov I, Hutter F. Decoupled weight decay regularization. International Conference on Learning Representations; 2017; 2017.
18. Gotmare A, Keskar NS, Xiong C, Socher R. A closer look at deep learning heuristics: learning rate restarts, warmup and distillation. arXiv preprint arXiv:181013243. 2018.
19. Ho Y, Wookey S. The real-world-weight cross-entropy loss function: modeling the costs of mislabeling. IEEE Access. 2020;8:4806–13.
20. Faghani S. shahriar-faghani/MM_radgen. GitHub 2024.
21. Faghani S, Khosravi B, Zhang K, Moassefi M, Jagtap JM, Nugen F, et al. Mitigating bias in radiology machine learning: 3. Performance metrics. Radiol Artif Intell. 2022;4(5): e220061.
22. Binder M, Rajkumar SV, Ketterling RP, Dispenzieri A, Lacy MQ, Gertz MA, et al. Substratification of patients with newly diagnosed standard-risk multiple myeloma. Br J Haematol. 2019;185(2):254–60.
23. Ni B, Huang G, Huang H, Wang T, Han X, Shen L, et al. Machine learning model based on optimized radiomics feature from (18) F-FDG-PET/CT and clinical characteristics predicts prognosis of multiple myeloma: a preliminary study. J Clin Med. 2023;12(6).
24. Sachpekidis C, Enqvist O, Ulén J, Kopp-Schneider A, Pan L, Mai EK, et al. Artificial intelligence-based, volumetric assessment of the bone marrow metabolic activity in [(18)F]FDG PET/CT predicts survival in multiple myeloma. Eur J Nucl Med Mol Imaging. 2024.
25. Zhong H, Huang D, Wu J, Chen X, Chen Y, Huang C. (18) F-FDG PET/CT based radiomics features improve prediction of prognosis: multiple machine learning algorithms and multimodality applications for multiple myeloma. BMC Med Imaging. 2023;23(1):87.
26. Liu J, Wang C, Guo W, Zeng P, Liu Y, Lang N, et al. A preliminary study using spinal MRI-based radiomics to predict high-risk cytogenetic abnormalities in multiple myeloma. Radiol Med. 2021;126(9):1226–35.