

Original Paper

A Scalable Radiomics- and Natural Language Processing–Based Machine Learning Pipeline to Distinguish Between Painful and Painless Thoracic Spinal Bone Metastases: Retrospective Algorithm Development and Validation Study

Hossein Naseri¹, MSc; Sonia Skamene², MD; Marwan Tolba², MD; Mame Daro Faye², MD; Paul Ramia², MD; Julia Khriouian², MD; Marc David², MD; John Kildea¹, PhD

¹Medical Physics Unit, McGill University Health Centre, Montreal, QC, Canada

²Division of Radiation Oncology, McGill University Health Centre, Montreal, QC, Canada

Corresponding Author:

Hossein Naseri, MSc
Medical Physics Unit
McGill University Health Centre
Cedars Cancer Centre
1001 boul Décarie Montréal
Montreal, QC, H4A 3J1
Canada
Phone: 1 514 934 1934 ext 44158
Email: 3naseri@gmail.com

Abstract

Background: The identification of objective pain biomarkers can contribute to an improved understanding of pain, as well as its prognosis and better management. Hence, it has the potential to improve the quality of life of patients with cancer. Artificial intelligence can aid in the extraction of objective pain biomarkers for patients with cancer with bone metastases (BMs).

Objective: This study aimed to develop and evaluate a scalable natural language processing (NLP)– and radiomics-based machine learning pipeline to differentiate between painful and painless BM lesions in simulation computed tomography (CT) images using imaging features (biomarkers) extracted from lesion center point–based regions of interest (ROIs).

Methods: Patients treated at our comprehensive cancer center who received palliative radiotherapy for thoracic spine BM between January 2016 and September 2019 were included in this retrospective study. Physician-reported pain scores were extracted automatically from radiation oncology consultation notes using an NLP pipeline. BM center points were manually pinpointed on CT images by radiation oncologists. Nested ROIs with various diameters were automatically delineated around these expert-identified BM center points, and radiomics features were extracted from each ROI. Synthetic Minority Oversampling Technique resampling, the Least Absolute Shrinkage And Selection Operator feature selection method, and various machine learning classifiers were evaluated using precision, recall, F_1 -score, and area under the receiver operating characteristic curve.

Results: Radiation therapy consultation notes and simulation CT images of 176 patients (mean age 66, SD 14 years; 95 males) with thoracic spine BM were included in this study. After BM center point identification, 107 radiomics features were extracted from each spherical ROI using pyradiomics. Data were divided into 70% and 30% training and hold-out test sets, respectively. In the test set, the accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve of our best performing model (neural network classifier on an ensemble ROI) were 0.82 (132/163), 0.59 (16/27), 0.85 (116/136), and 0.83, respectively.

Conclusions: Our NLP- and radiomics-based machine learning pipeline was successful in differentiating between painful and painless BM lesions. It is intrinsically scalable by using NLP to extract pain scores from clinical notes and by requiring only center points to identify BM lesions in CT images.

(JMIR AI 2023;2:e44779) doi: [10.2196/44779](https://doi.org/10.2196/44779)

KEYWORDS

cancer; pain; palliative care; radiotherapy; bone metastases; radiomics; natural language processing; machine learning; artificial intelligent; radiation therapy

Introduction

Overview

Most patients with cancer with bone metastasis (BM) experience pain [1] and most receive radiotherapy to control it [2]. But, it has been shown that due to the subjective and qualitative nature of the pain, clinicians often underestimate pain [3]. As a result, many patients with BM receive radiotherapy after their pain has already become debilitating [4].

Although patient-reported outcomes can be used to obtain pain scores directly from patients themselves, the efficacy of these pain scores is limited due to the fact that these ratings are highly qualitative and subjective [5]. Because of this, it is desirable to have pain scoring systems that are more objective. The goal of this study was to explore ways to automatically and objectively quantify pain associated with BMs using computed tomography (CT) images.

We hypothesized that tumor features extracted from CT images of BMs contain imaging biomarkers that may be used to objectively identify BM-associated pain. These pain biomarkers may provide the opportunity to develop objective pain scoring tools to aid in the diagnosis, treatment, understanding, and prognosis of BM pain.

Background

The search for imaging and nonimaging pain biomarkers has been the focus of numerous studies [5-12]. Various studies [13-21] have shown how artificial intelligence (AI), including machine learning and radiomics, can be used to understand and quantify pain. For example, Mashayekhi et al [22] showed that radiomic features extracted from the CT images of the pancreas can help to identify functional abdominal pain in patients. Vedantam et al [23] explored the viability of using radiomics features extracted from magnetic resonance images to detect pain following percutaneous cordotomy. At least 1 study [13] has reported using radiomics to identify painful metastatic lesions in radiographic images. However, we found no reports in the literature of a scalable approach that can be used efficiently on a large set of unlabeled patient data. To the best of our knowledge, our work is the first to combine natural language processing (NLP) and radiomics to enable an efficient and scalable pain identification pipeline using unstructured data.

A fundamental challenge in developing any AI model for use in medicine is the need to obtain sufficient patient data for training and testing. For example, the data set used by Wakabayashi et al in the study that we mentioned earlier [13], was limited to 69 patients. One limiting factor is obtaining standard patient-reported pain scores for use as ground-truth data, and another limiting factor is obtaining segmented images from which to extract tumor biomarkers. For the work reported in this paper, we overcame the data set size limitation by using 2 novel strategies. First, by combining NLP with radiomics, we quickly mined pain scores from clinical notes and used these

NLP-extracted scores to label our radiomics features for supervised learning. Second, by asking our clinical colleagues to pinpoint only the center points of BM lesions in radiotherapy simulation CT images, we maximized the number of lesions identified in the time available. In the medical field, NLP has shown promising results in extracting biomedical information and clinical outcomes such as pain from unstructured text data [24-26]. Moreover, as we reported previously [21], by automatically delineating geometrical regions around BM lesion center points, it is possible to successfully extract radiomics features for robust BM lesion detection. In this study, we report how our combined radiomics-NLP machine learning pipeline can successfully identify pain in radiotherapy simulation CT images of patients with cancer with BMs.

Methods

Ethical Considerations

This retrospective study was approved by the research ethics board of the McGill University Health Centre (2020-5899) with the waiver of informed consent. We confirm that the entire research was performed in accordance with research ethics board's guidelines and regulations.

Data Selection

Our patient-selection process is outlined in Figure 1. The initial number of 200 pairs of radiation oncology consultation notes and CT images of patients with spinal BM were included in this study based on the minimum sample size calculation as explained in Section A.1 in Multimedia Appendix 1 [27]. In total, 120 of the notes and all 200 of the CT images from this study were independently used in 2 studies we previously reported on [21-25]. The first [25] of these studies showed the feasibility of extracting pain from consultation notes of patients with cancer, using NLP. The second [21] demonstrated the feasibility of using lesion center point-based radiomics models to differentiate healthy and metastatic bone lesions in CT scans of patients with BMs. This study combined the data and results from these 2 prior studies and expanded upon them to build an NLP- and radiomics-based model to detect pain using the CT scans of patients.

We searched our institution's Oncology Information System for the radiotherapy plans of patients diagnosed with a "secondary malignant neoplasm of bone" between January 2016 and September 2019. From the retrieved list, we selected those who were treated for thoracic spinal BM. Then, we retrieved the corresponding consultation notes and simulation CT images. A note-image pair was included if (1) the note was in English, (2) pain was documented, (3) the simulation CT image was taken up to 10 days post consultation, and (4) simulation CT revealed BM lesions in the thoracic spine. Patients with multiple but nonoverlapping note-image pairs were considered independent samples. We only considered the same patients as new participants if they had CT scans and associated

consultation notes for BM lesions in different areas of their spines. As a result, each BM lesion was included only once in our study. Also, it should be noted that palliative patients normally have their simulation CT scan (for treatment planning) on the same day or within a few days after the consultation, and radiotherapy is delivered on the same day or within a few days after treatment planning. To assure that there is no change in the BM lesion structure or pain status, we did not allow more than a 10-day gap between the two. Figure A1 in [Multimedia](#)

[Appendix 1](#) displays the distribution of the time interval between the radiotherapy consultation and CT acquisition dates.

We randomly assigned note-image pairs to the training or cross-validation set (approximately 70%) or the holdout test set (approximately 30%). We used stratified randomization to preserve the original sample ratio between pain labels in each sample set. In addition, we performed a paired *t* test and a chi-square analysis [28] to ensure that there was no systematic bias in any of our sample sets regarding gender, age, or primary cancer type. Patient demographics are presented in [Table 1](#).

Figure 1. The patient selection criteria used to obtain the radiotherapy consultation notes and simulation computed tomography (CT) images that formed our training and test data sets. The initial number of 200 note-image pairs included in this study was based on the minimum sample size calculation as explained in Section A.1 in [Multimedia Appendix 1](#). BM: bone metastases; DICOM: Digital Imaging and Communications in Medicine; RT: radiotherapy; T-spine: thoracic spine. *Four patients had pairs in both the training and test sets.

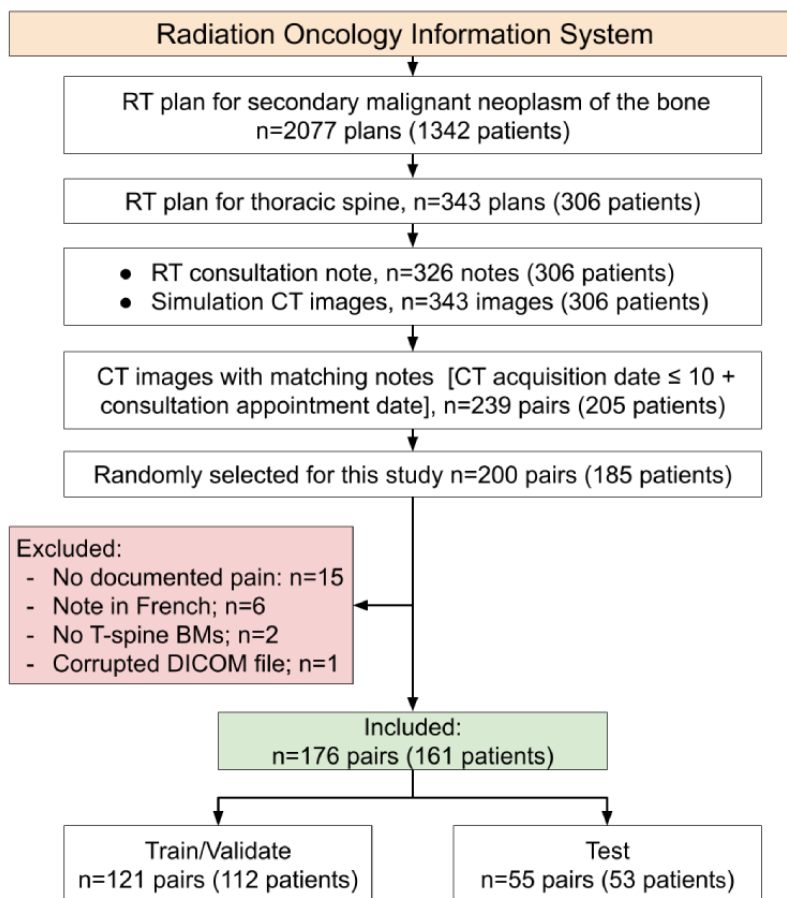


Table 1. Patient demographics in the training and test sets.

Characteristics	Training and validation set (n=121)	Test set (n=55)	<i>P</i> value ^a
Gender, n (%)			N/A ^b
Female	56 (46)	25 (45)	
Male	65 (54)	30 (55)	
Age (years), mean (SD)^c			N/A
Female	63 (14)	64 (12)	.99
Male	67 (14)	64 (13)	.72
Primary cancer type, n (%)			.06
Lung	32 (26)	20 (36)	
Breast	23 (19)	11 (20)	
Prostate	19 (16)	5 (9)	
Multiple myeloma	8 (7)	6 (11)	
Renal cell carcinoma	7 (6)	2 (4)	
Other and unknown	64 (53)	31 (56)	
Bone metastasis lesions, n (%)			.42
Lytic	220 (52)	76 (47)	
Blastic	122 (29)	57 (35)	
Mix	81 (19)	30 (18)	
Pain label, n (%)			N/A
Pain	357 (84)	136 (83)	
No pain	66 (16)	27 (17)	

^a*P* values for numerical values (age) and categorical features (primary cancer site and bone metastasis lesion type) were calculated using a 2-tailed heteroscedastic *t* test and a chi-square test, respectively.

^bN/A: not applicable.

^cThe *P* value for the age difference between males and females was .20 for the training and validation set and .50 for the test set.

NLP-Extracted Pain Labels

Due to the absence of patient-reported pain scores in our Oncology Information System, we extracted physician-reported pain scores from patients' radiation oncology consultation notes using our previously reported NLP pipeline [25]. While pain scores were typically reported as part of the "history of the present illness" in our hospital, for the sake of generalizability, we extracted pain scores from the entire note.

Our NLP pipeline first processed the text with MetaMap [29] and mapped it to the UMLS (ie, Unified Medical Language System) Metathesaurus [30] in order to identify pain terminologies and their severity scores. Next, it applied rules to filter out hypothetical, conditional, and historical references to pain in order to focus solely on references to pain at the time of the consultation. Then, it calculated the average pain intensity (API) in each note by averaging the pain scores therein. Finally, it assigned each note a "verbally declared pain" (VDP) label, as VDP="no pain" (if API 0), and VDP="pain" (if API>0). These pain labels were used to train, validate, and test our radiomics model.

Expert-Extracted Pain Scores

To evaluate the effect of NLP-extracted pain labels on the performance of our pipeline, we also generated best-available ground-truth pain labels using expert-annotated pain scores. To do so, our radiation oncologists used the texTRACTOR [31] pain labeling application to manually read consultation notes and label valid pain scores in our training and test data sets using a 4-grade verbal rating scale (no pain, mild, moderate, and severe). A mention of pain was regarded as valid if it reflected the status of pain at the metastatic sites for which treatment was planned at the time of the consultation. Table A1 in [Multimedia Appendix 1](#) contains all the NLP- and expert-extracted pain scores, and Figure A2 in [Multimedia Appendix 1](#) illustrates the level of agreement between them. Due to the quality of the documented pain scores and lack of interrater agreement among experts (Fleiss $\kappa=0.43$), as explained by Naseri et al [25], we subsequently defined a binary pain score as "no pain" and "pain" in order to establish satisfactory interrater agreement ($\kappa=0.66$) [25]. To create binary ground-truth pain labels comparable to the NLP-extracted labels, we assigned notes scored as "no pain" to "no pain" and notes scored as "mild," "moderate," and "severe" pain to "pain." These

expert-extracted pain scores were used to measure how well the NLP pipeline works.

Center Point Identification of BM Lesions

BM lesion center points were identified by a team comprising a staff radiation oncologist (SS) with 10 years' experience, a radiation oncology fellow (MT), and 3 third-year radiation oncology residents (J Khriguian, PR, and MF). Simulation CT DICOM (ie, Digital Imaging and Communications in Medicine) files were exported from the radiotherapy treatment planning software and deidentified. Then, the CT images were randomly divided into 5 sets and loaded into the diCOMBINE [32] application for BM lesion center point identification. Our experts were blinded to patients' pain statuses and identities. We requested each expert to label center points for all visually identifiable BM lesions in all CT images within 1 of the 5 sets, and another expert was assigned to validate their labels. A key benefit of this radiomics pipeline [21] is that it does not require full lesion segmentation, making it feasible to engage busy clinicians.

Segmentation of Regions of Interest

Using our previously reported methodology [21], we automatically segmented lesion center point-based nested

spherical (SP) regions of interest (ROIs). To do this, we first delineated nested spherical ROIs around the identified BM lesion center points (see [Textbox 1](#), top panel). ROI diameters ranged from 7 mm (3×3 voxels) to 50 mm (average size of the vertebral body) [33]. Then, in addition to what was reported by Naseri et al [21], we used Hounsfield units thresholding to exclude fat and air regions from the delineated ROIs. For this, motivated by Deglint et al [34] and Ulano et al [35], we applied a threshold to remove voxels with negative Hounsfield units from our ROIs. Hounsfield units of <0 are associated with fat and air [34]. We used OpenCV [36] (version 4.4.0) for Hounsfield units thresholding and applied a Gaussian filter to reduce noise. Then, we used pynrrd [37] (version 0.4.2) to export each ROI as a 3D binary mask and store it as a.nrrd [38] file. Finally, we aggregated these nested ROI masks to form ensemble ROIs. In this study, we examined 2 contrasting ensemble (EN) ROIs as shown in [Textbox 1](#) (bottom panel): one with small size and 3 layers (EN3) and the other with large size and 6 layers (EN6). Wakabayashi et al [13] and Naseri et al [21] have shown that radiomics-based machine learning models trained on ensemble ROIs have better classification performance than single ROI-based models.

Textbox 1. The characteristics of the spherical and ensemble regions of interest (ROIs) used in this study.

Nested spherical (SP) ROIs with Hounsfield units (HUs) intensity thresholds (HU>0):

- SP7 (diameter 7 mm)
- SP10 (diameter 10 mm)
- SP15 (diameter 15 mm)
- SP20 (diameter 20 mm)
- SP30 (diameter 30 mm)
- SP50 (diameter 50 mm)

Ensemble (EN) ROIs:

- EN3 (ROI SP7+SP10+SP15)
- EN6 (ROI SP7+SP10+SP15+SP20+SP30+SP50)

Radiomics Models

Our radiomics pipeline is illustrated in [Figure 2](#). We essentially used our previously reported pipeline [21] but with our NLP- and expert-extracted pain labels to train and test it. We made one improvement to the pipeline by incorporating Imbalanced-learn [39] (version 0.7.0) as a resampling step to account for imbalance (see below).

Radiomics features were extracted from each CT image using masks composed of the ensemble ROIs listed in [Textbox 1](#). Then, the feature space was scaled using *z* score normalization [40], and the associated NLP-extracted binary pain labels (pain=1, no pain=0) were incorporated. A single NLP-extracted pain score was assigned to all the lesions extracted from a given paired CT image.

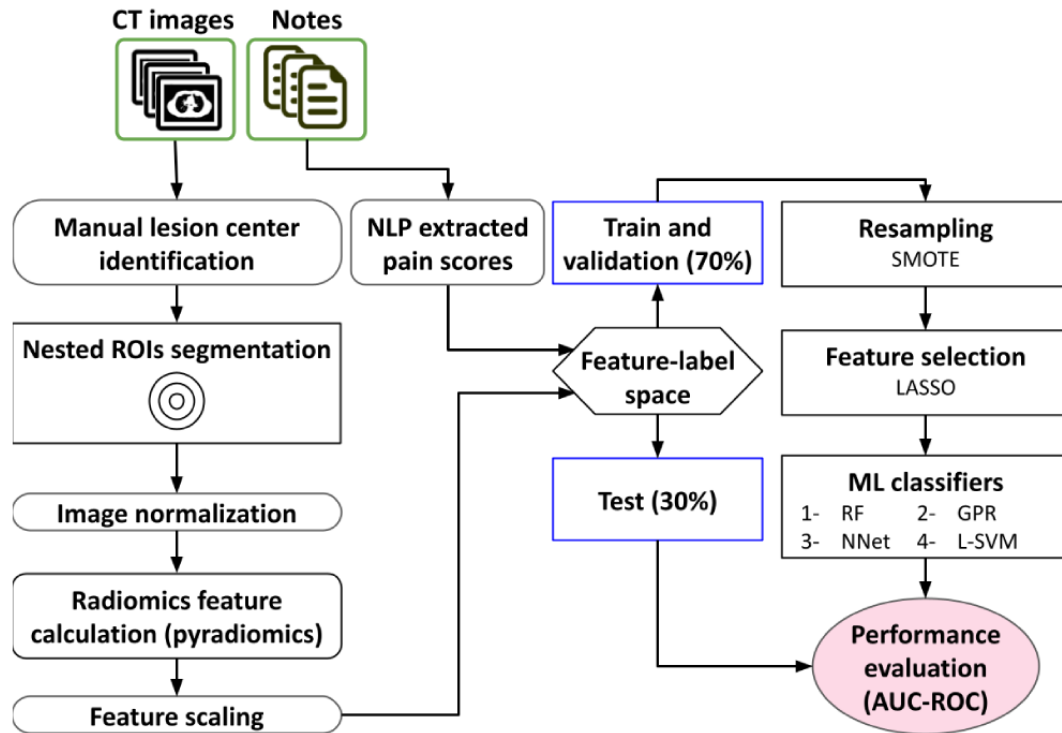
Due to the nature of BM pain [41], there was a large imbalance between the number of painful and painless lesions (493 pain,

93 no pain). Therefore, we used the Synthetic Minority Oversampling Technique (SMOTE) [42] in the training phase as it has been shown to be the best-performing resampling method for radiomics [43]. We did not apply resampling to our test set in order to maintain the original sample imbalance. Then, the Least Absolute Shrinkage And Selection Operator [44] feature selection method was applied to the feature space to remove noninformative features. Least Absolute Shrinkage And Selection Operator is a commonly used feature selection method in radiomics studies [45,46]. Finally, we examined the Gaussian process regression, linear support vector machine, random forest, and neural networks classifiers, as they were the best performing machine learning classifiers in our previous work. We evaluated the performance of our models on the training set using 5-fold cross-validation. Final evaluation was performed on the test set. The receiver operating characteristic (ROC) [47] curve, area under the ROC curve (AUC), precision, sensitivity, specificity, and F_1 -score metrics were used to report the performance of

our models on the training and test sets. We also trained and tested our best performing pipeline using the expert-extracted

pain scores (best-available ground-truth) to evaluate the impact of NLP-extracted pain labels.

Figure 2. The radiomics-based pipeline that we used to select and train a machine learning model to separate painful and painless bone metastasis lesions. Our pipeline is the same as that published by Naseri et al [21] but using NLP-extracted pain labels and modified to account for sample imbalance. AUC-ROC: area under the receiver operating characteristic curve-receiver operating characteristic; CT: computed tomography; GPR: Gaussian process regression; LASSO: Least Absolute Shrinkage And Selection Operator; L-SVM: linear support vector machine; ML: machine learning; NLP: natural language processing; NNet: neural network; RF: random forest; ROI: region of interest; SMOTE: Synthetic Minority Oversampling Technique.



Results

Patient Demographics

A total of 176 pairs of radiotherapy consultation notes and simulation CT images of patients with thoracic spinal BM were included in this study. As summarized in Table 1, a total of 121 sample pairs (mean patient age 63, SD 14 years; males: n=65, mean age 67, SD 14 years; $P=.20$) were included for training and cross-validation, and 55 sample pairs (mean patient age 64, SD 12 years; males: n=25, mean age 64, SD 13 years; females: mean age 64, SD 23 years; $P=.50$) were included in the test set. The sample selection procedure and data quantities are presented in Figure 1. The demographics of the patients in the training and test sets are presented in Table 1. The most common primary

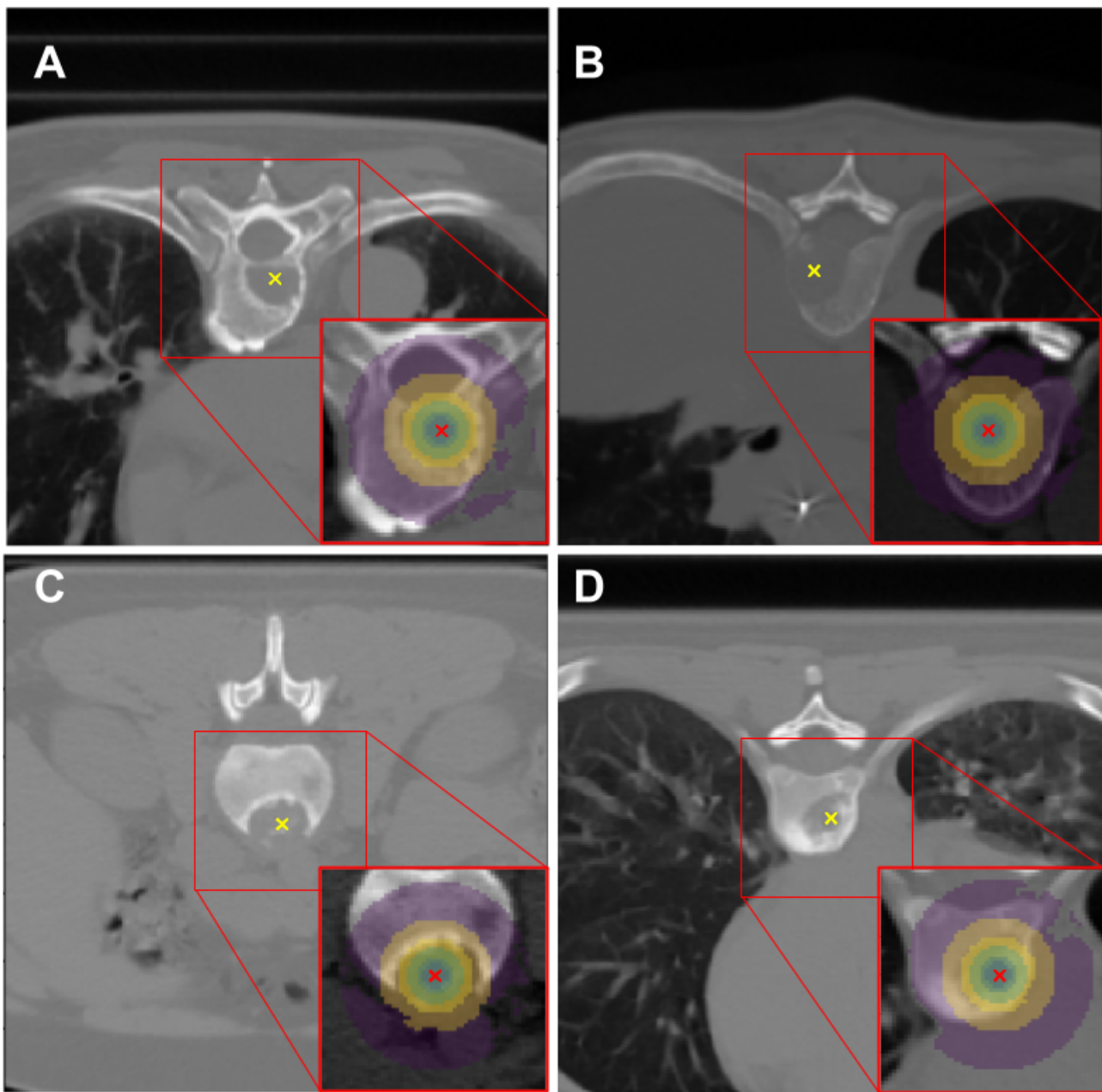
cancer sites were the lungs (n=52), breasts (n=34), and prostate (n=24).

A total of 586 BM center points were identified by our experts on the training (n=423 lesions) and test (n=163 lesions) data sets. In the training set, 357 (84%) lesions were labeled by the NLP pipeline as painful and 66 lesions were labeled as painless. In the test set, 136 (83%) lesions were identified by the NLP pipeline as painful, and 27 lesions were labeled as painless. This represented a significant but equal imbalance in our training and test sets.

Segmented ROIs

Examples of segmented ROIs with the Hounsfield units threshold applied are presented in Figure 3 for painful and painless BMs.

Figure 3. Examples of segmented nested spherical regions of interest (ROIs) with the Hounsfield units threshold applied on computed tomography images of patients with painful (A, B) and painless (C, D) bone metastases lesions. Nested ROIs with diameters of 50, 30, 20, 15, 10, and 7 mm are shown in the insets as different hues.



Testing Our Radiomics Models

In total, 107 radiomics features were extracted from each of the 6 nested ROIs. Then, they were aggregated to form feature spaces for the EN3 (with 321 features) and EN6 (with 642 features) ensemble ROIs. Figure 4 shows the ROC curve of each model in the training (black lines) and test (red squares) data sets using the EN3 and EN6 ROIs. On the training set, the gray range represents the mean (SD) AUC of the 5-fold

cross-validation. The AUC and F_1 -score grids are presented in Table 2.

The precision, accuracy, sensitivity, specificity, F_1 -score, and AUC values of our best-performing pipeline (neural networks with the EN6 ROI) are presented in Table 3. The performance of this pipeline (trained and tested) on the data set of expert-extracted pain labels (best-available ground-truth) is provided as a quality measurement. The performance of the model described previously by Wakabayashi et al [13] is also provided for comparison.

Figure 4. Receiver operating characteristic curves for our classifiers using 3-layer ensemble (EN3) (top row) and 6-layer ensemble (EN6) (bottom row) lesion center point–based ensemble regions of interest in training (black lines) and test (dark red squares) data sets. AUC: area under the receiver operating characteristic curve; GPR: Gaussian process regression; L-SVM: linear support vector machine; NNet: neural network; RF: random forest.

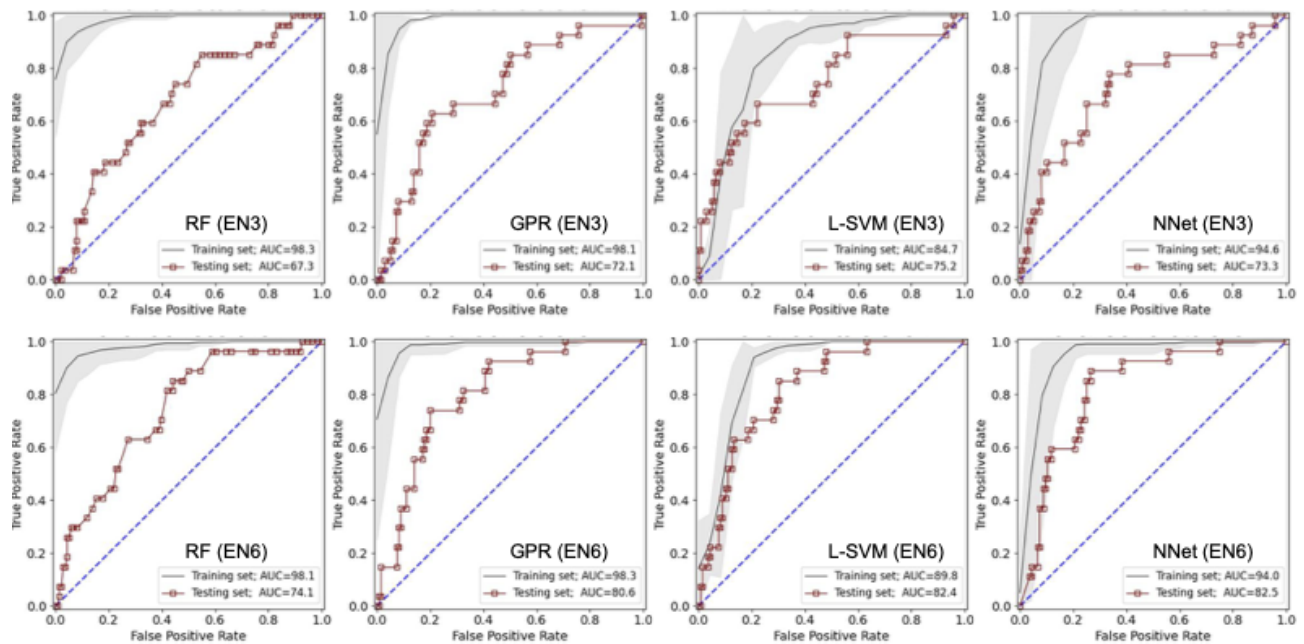


Table 2. The area under the receiver operating characteristic curves (AUCs) and F_1 -scores of our machine learning classifiers in the training and test data sets using the ensemble (EN) regions of interest EN3 and EN6 for each of the RF (random forest), GPR (Gaussian process regression), L-SVM (linear support vector machine), and NNet (neural networks) classifiers.

Region of interest	Training set				Test set			
	RF	GPR	L-SVM	NNet	RF	GPR	L-SVM	NNet
Areas under the receiver operating characteristic curve								
EN3	98.3	98.1	84.7	94.6	67.3	72.1	75.2	73.3
EN6	98.1	98.3	89.8	94.0	74.1	80.6	82.4	82.5
F_1-scores								
EN3	90.0	89.9	79.4	90.5	60.9	64.7	65.4	63.6
EN6	93.0	93.0	84.7	91.6	63.8	66.9	67.4	69.5

Table 3. The performance of our best-performing natural language processing (NLP)–radiomics pipeline (neural networks with the ensemble 6 region of interest) on the training and test sets using NLP and manually extracted pain labels, together with the results from a prior study by Wakabayashi et al [13].

	Accuracy	Precision	Sensitivity	Specificity	F_1 -score	AUC ^a
This study (training set)	92.4	93.2	92.4	86.4	91.6	94.0
This study (test set)	81.0	67.9	59.2	85.3	69.5	82.5
This study (training set); using manual pain scores	94.2	94.8	98.7	89.7	94.4	98.1
This study (test set); using manual pain scores	83.5	64.9	64.7	85.7	68.0	82.3
Wakabayashi et al [13] (training test only)	73.9	— ^b	71.0	86.0	—	82.0

^aAUC: area under the receiver operating characteristic curve.

^bNot determined.

Discussion

Underestimation and undertreatment of cancer pain can significantly diminish the quality of life of patients with cancer.

Accordingly, systems that can objectively measure cancer pain have the potential to improve quality of life. In this study, we created a scalable NLP-radiomics pain identification pipeline. Our pipeline is designed for palliative treatment for patients

with cancer undergoing radiotherapy therapy, for whom there are typically just 2 contemporaneous sources of relevant medical information at the time of the treatment: consultation notes and simulation CT images. We used an NLP pipeline to extract physician-reported pain scores from radiotherapy consultation notes. NLP-extracted pain scores are appropriate, when structured patient-reported pain scores are unavailable (as is the case for at least 25% to 35% of all patients with cancer [13,48] and for all patients with cancer receiving palliative care who are treated with radiotherapy at our institution at the time the data were used in this study). Our lesion center point-based spherical ROI delineation method significantly sped up the ROI segmentation procedure, enabling us to rapidly delineate BM center points in 176 images in this study. For comparison, the radiomics pipeline that was developed by Wakabayashi et al [13] required full 3D segmentation of each ROI (69 images).

Due to the unbalanced nature of BM pain, our data set contained significantly fewer “no pain” samples. In order to better train our models, we applied SMOTE resampling to the training set to balance the number of samples with the NLP-extracted “pain” and “no pain” labels. We did not apply any resampling techniques to our test (hold out) set to maintain the original sample imbalance. Therefore, while our training set was balanced, our test set had 5 times more “pain” cases than “no pain” cases (136 pain versus 27 no pain cases). This caused a significant change in the pipeline’s performance between the training and test sets. It has been shown that oversampling improves the overall performance of machine learning models, but the effect is stronger on the training set due to the inclusion of replicated samples in the cross-validation subsets [49]. Moreover, the imbalance in our test set led to high specificity (ability to properly identify pain instances) and low sensitivity (ability to correctly identify no pain cases) in the performance evaluation. For comparison, the sample imbalance reported by Wakabayashi et al [13] was 2:1, resulting in a more balanced relationship between the sensitivity and specificity of their model.

The performance of our pipeline did not improve much when we trained and tested it using expert-extracted pain labels (best-available ground-truth). This might be the case because, in the first experiment, we both trained and tested our pipeline using NLP-extracted pain labels, and in the second experiment, we both trained and tested our pipeline using expert-extracted pain labels. Consequently, after being trained with one set of labels (NLP- or expert-extracted), our pipeline performed well on the test set that was labeled using the same method (NLP or expert). We also demonstrated that our pipeline’s performance is comparable to that of Wakabayashi et al [13], who achieved their results using patient-reported pain labels.

Our pipeline performed significantly better on the EN6 ROIs than on the EN3 ROIs. This could be the case because in comparison to EN3, our EN6 ROIs include additional ROIs with sizes of 20, 30, and 50 mm. From visual inspection, we suspect that, in addition to the characteristics of the BM lesion itself, its location (eg, its proximity to the spinal cord) may be a significant contributor to the BM pain. As a result, larger ROIs enable our algorithm to extract characteristics from outside the

BM lesion. Wakabayashi et al [13] also demonstrated the effectiveness of using ROIs outside of the BM lesion.

We are unable to offer a convincing explanation as to why neural networks outperformed random forest and support vector machine classifiers in our analysis. Notwithstanding, it has been demonstrated that neural network classifiers perform better when applied to more difficult problems and larger data sets, while random forest and support vector machine classifiers typically perform well with smaller data sets [46,50,51].

Our pipeline was successful in extracting radiomics biomarkers capable of distinguishing between painful and painless BM lesions. These biomarkers potentially provide the opportunity to objectively identify clinical pain-related indicators that may aid in the diagnosis, treatment, and understanding of BM pain.

Our work has several limitations. First, we used data from a single center for this retrospective study. A multicenter study with a larger data set is necessary to assess the generalizability of our radiomics pipeline for pain quantification. We anticipate that the performance of our NLP-radiomics pipeline will vary based on the pain scoring systems of the cohorts tested. Second, by using lesion center point-based geometrical ROIs, we ignored lesion characteristics such as size and shape, which may be important in the context of pain. Although we used Hounsfield units intensity thresholding to preserve some tumor information, we are considering implementing deep learning-based ROI segmentation in the future as it may better account for full tumor and surrounding tissue characteristics. Lastly, we used SMOTE resampling to address the issue of class imbalance. An alternative solution might be to develop cost-sensitive machine learning classifiers that account for the cost of misclassifying minority samples [52]. However, there is no clear consensus in the literature on whether cost-sensitive learning outperforms resampling [53]. A model that can differentiate between painful and painless lesions from medical imaging is a critical component of any possible radiomics-based pain quantification pipeline. This work not only shows the feasibility of developing a pain quantification tool, but also it removes some of the barriers to its development. As a result, our future work will be to apply our pipeline to patients’ past and current CT images and consultation notes in order to develop a longitudinal model of pain. Such a model should take into account not only images (taken before, during, and after delivering radiotherapy) but also other internal and external parameters that can influence how pain evolves over time (such as primary cancer type, radiation dose, other treatments, and pain medications). Also, it will include patient-reported pain scores to provide more accurate ground-truth pain labels in order to develop a more robust deep learning-based NLP pipeline [24,54]. This, however, is beyond the scope of this investigation.

In conclusion, we demonstrated that our NLP and radiomics-based machine learning pipeline can effectively differentiate between painful and painless BM lesions in simulation CT images using ensemble lesion center point-based geometrical ROIs. Using NLP-extracted pain labels in conjunction with lesion center point-based radiomics features is time efficient. This helps to pave the way for the development of quickly trained and efficient clinical AI-based

decision-making tools that can objectively measure cancer pain. and improve the quality of life of patients with BMs. Such a tool may help alleviate the burden of pain management

Acknowledgments

This research was supported by the startup grant of J Kildea at the Research Institute of the McGill University Health Centre (RI-MUHC), the Ruth and Alex Dworkin scholarship award from the Faculty of Medicine and Health Sciences at McGill University, an RI-MUHC studentship award, a Grad Excellence Award-00293 from the Department of Physics at McGill University, Fonds de recherche du Québec - Santé (FRQS), and by the CREATE Responsible Health and Healthcare Data Science (SDRDS) grant of the Natural Sciences and Engineering Research Council. The authors would like to thank Dr Luc Galarneau for his help with statistical analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sample data.

[\[DOCX File , 103 KB-Multimedia Appendix 1\]](#)

References

1. van den Beuken-van Everdingen MH, Hochstenbach LM, Joosten EA, Tjan-Heijnen VC, Janssen DJ. Update on prevalence of pain in patients with cancer: systematic review and meta-analysis. *J Pain Symptom Manage* 2016 Jun;51(6):1070-1090.e9 [FREE Full text] [doi: [10.1016/j.jpainsymman.2015.12.340](https://doi.org/10.1016/j.jpainsymman.2015.12.340)] [Medline: [27112310](https://pubmed.ncbi.nlm.nih.gov/27112310/)]
2. McQuay HJ, Collins SL, Carroll D, Moore RA, Derry S. WITHDRAWN: radiotherapy for the palliation of painful bone metastases. *Cochrane Database Syst Rev* 2013 Nov 22;2013(11):CD001793 [FREE Full text] [doi: [10.1002/14651858.CD001793.pub2](https://doi.org/10.1002/14651858.CD001793.pub2)] [Medline: [24271498](https://pubmed.ncbi.nlm.nih.gov/24271498/)]
3. Grossman SA. Undertreatment of cancer pain: barriers and remedies. *Support Care Cancer* 1993 Mar;1(2):74-78. [doi: [10.1007/bf00366899](https://doi.org/10.1007/bf00366899)]
4. Cleeland CS, Janjan NA, Scott CB, Seiferheld WF, Curran WJ. Cancer pain management by radiotherapists: a survey of radiation therapy oncology group physicians. *Int J Radiat Oncol Biol Phys* 2000 Apr 01;47(1):203-208. [doi: [10.1016/s0360-3016\(99\)00276-x](https://doi.org/10.1016/s0360-3016(99)00276-x)] [Medline: [10758325](https://pubmed.ncbi.nlm.nih.gov/10758325/)]
5. Tracey I, Woolf CJ, Andrews NA. Composite pain biomarker signatures for objective assessment and effective treatment. *Neuron* 2019 Mar 06;101(5):783-800 [FREE Full text] [doi: [10.1016/j.neuron.2019.02.019](https://doi.org/10.1016/j.neuron.2019.02.019)] [Medline: [30844399](https://pubmed.ncbi.nlm.nih.gov/30844399/)]
6. Xu X, Huang Y. Objective pain assessment: a key for the management of chronic pain. *F1000Res* 2020 Jan 23;9:35 [FREE Full text] [doi: [10.12688/f1000research.20441.1](https://doi.org/10.12688/f1000research.20441.1)] [Medline: [32047606](https://pubmed.ncbi.nlm.nih.gov/32047606/)]
7. Niculescu AB, Le-Niculescu H, Levey DF, Roseberry K, Soe KC, Rogers J, et al. Towards precision medicine for pain: diagnostic biomarkers and repurposed drugs. *Mol Psychiatry* 2019 Apr;24(4):501-522 [FREE Full text] [doi: [10.1038/s41380-018-0345-5](https://doi.org/10.1038/s41380-018-0345-5)] [Medline: [30755720](https://pubmed.ncbi.nlm.nih.gov/30755720/)]
8. Diaz MM, Caylor J, Strigo I, Lerman I, Henry B, Lopez E, et al. Toward composite pain biomarkers of neuropathic pain-focus on peripheral neuropathic pain. *Front Pain Res (Lausanne)* 2022 May 11;3:869215 [FREE Full text] [doi: [10.3389/fpain.2022.869215](https://doi.org/10.3389/fpain.2022.869215)] [Medline: [35634449](https://pubmed.ncbi.nlm.nih.gov/35634449/)]
9. Furfari A, Wan BA, Ding K, Wong A, Zhu L, Bezjak A, et al. Genetic biomarkers associated with pain flare and dexamethasone response following palliative radiotherapy in patients with painful bone metastases. *Ann Palliat Med* 2017 Dec;6(Suppl 2):S240-S247. [doi: [10.21037/apm.2017.09.04](https://doi.org/10.21037/apm.2017.09.04)] [Medline: [29156912](https://pubmed.ncbi.nlm.nih.gov/29156912/)]
10. Gunn J, Hill MM, Cotten BM, Deer TR. An analysis of biomarkers in patients with chronic pain. *Pain Physician* 2020 Jan;23(1):E41-E49 [FREE Full text] [Medline: [32013287](https://pubmed.ncbi.nlm.nih.gov/32013287/)]
11. Marchi A, Vellucci R, Mameli S, Rita Piredda A, Finco G. Pain biomarkers. *Clinical Drug Investigation* 2009;29(Supplement 1):41-46. [doi: [10.2165/0044011-200929001-00006](https://doi.org/10.2165/0044011-200929001-00006)]
12. Ota Y, Connolly M, Srinivasan A, Kim J, Capizzano AA, Moritani T. Mechanisms and origins of spinal pain: from molecules to anatomy, with diagnostic clues and imaging findings. *Radiographics* 2020 Jul;40(4):1163-1181. [doi: [10.1148/rg.2020190185](https://doi.org/10.1148/rg.2020190185)] [Medline: [32501739](https://pubmed.ncbi.nlm.nih.gov/32501739/)]
13. Wakabayashi K, Koide Y, Aoyama T, Shimizu H, Miyauchi R, Tanaka H, et al. A predictive model for pain response following radiotherapy for treatment of spinal metastases. *Sci Rep* 2021 Jun 18;11(1):12908 [FREE Full text] [doi: [10.1038/s41598-021-92363-0](https://doi.org/10.1038/s41598-021-92363-0)] [Medline: [34145367](https://pubmed.ncbi.nlm.nih.gov/34145367/)]
14. Carlson LA, Hooten WM. Pain-linguistics and natural language processing. *Mayo Clin Proc Innov Qual Outcomes* 2020 Jun;4(3):346-347 [FREE Full text] [doi: [10.1016/j.mayocpiqo.2020.01.005](https://doi.org/10.1016/j.mayocpiqo.2020.01.005)] [Medline: [32542226](https://pubmed.ncbi.nlm.nih.gov/32542226/)]
15. Dave AD, Ruano G, Kost J, Wang X. Automated extraction of pain symptoms: a natural language approach using electronic health records. *Pain Physician* 2022 Mar;25(2):E245-E254. [Medline: [35322976](https://pubmed.ncbi.nlm.nih.gov/35322976/)]

16. Tighe PJ, Sannapaneni B, Fillingim RB, Doyle C, Kent M, Shickel B, et al. Forty-two million ways to describe pain: topic modeling of 200,000 PubMed pain-related abstracts using natural language processing and deep learning-based text generation. *Pain Med* 2020 Nov 01;21(11):3133-3160 [FREE Full text] [doi: [10.1093/pm/pnaa061](https://doi.org/10.1093/pm/pnaa061)] [Medline: [32249306](https://pubmed.ncbi.nlm.nih.gov/32249306/)]
17. Matsangidou M, Liampas A, Pittara M, Pattichi CS, Zis P. Machine learning in pain medicine: an up-to-date systematic review. *Pain Ther* 2021 Dec 26;10(2):1067-1084 [FREE Full text] [doi: [10.1007/s40122-021-00324-2](https://doi.org/10.1007/s40122-021-00324-2)] [Medline: [34568998](https://pubmed.ncbi.nlm.nih.gov/34568998/)]
18. Neijenhuijs KI, Peeters CFW, van Weert H, Cuijpers P, Leeuw IV. Symptom clusters among cancer survivors: what can machine learning techniques tell us? *BMC Med Res Methodol* 2021 Aug 16;21(1):166 [FREE Full text] [doi: [10.1186/s12874-021-01352-4](https://doi.org/10.1186/s12874-021-01352-4)] [Medline: [34399698](https://pubmed.ncbi.nlm.nih.gov/34399698/)]
19. Hong JH, Jung J, Jo A, Nam Y, Pak S, Lee S, et al. Development and validation of a radiomics model for differentiating bone islands and osteoblastic bone metastases at abdominal CT. *Radiology* 2021 Jun;299(3):626-632. [doi: [10.1148/radiol.2021203783](https://doi.org/10.1148/radiol.2021203783)] [Medline: [33787335](https://pubmed.ncbi.nlm.nih.gov/33787335/)]
20. Sun W, Liu S, Guo J, Liu S, Hao D, Hou F, et al. A CT-based radiomics nomogram for distinguishing between benign and malignant bone tumours. *Cancer Imaging* 2021 Feb 06;21(1):20 [FREE Full text] [doi: [10.1186/s40644-021-00387-6](https://doi.org/10.1186/s40644-021-00387-6)] [Medline: [33549151](https://pubmed.ncbi.nlm.nih.gov/33549151/)]
21. Naseri H, Skamene S, Tolba M, Faye MD, Ramia P, Khriuguian J, et al. Radiomics-based machine learning models to distinguish between metastatic and healthy bone using lesion-center-based geometric regions of interest. *Sci Rep* 2022 Jun 14;12(1):9866 [FREE Full text] [doi: [10.1038/s41598-022-13379-8](https://doi.org/10.1038/s41598-022-13379-8)] [Medline: [35701461](https://pubmed.ncbi.nlm.nih.gov/35701461/)]
22. Mashayekhi R, Parekh VS, Faghih M, Singh VK, Jacobs MA, Zaheer A. Radiomic features of the pancreas on CT imaging accurately differentiate functional abdominal pain, recurrent acute pancreatitis, and chronic pancreatitis. *Eur J Radiol* 2020 Feb;123:108778 [FREE Full text] [doi: [10.1016/j.ejrad.2019.108778](https://doi.org/10.1016/j.ejrad.2019.108778)] [Medline: [31846864](https://pubmed.ncbi.nlm.nih.gov/31846864/)]
23. Vedantam A, Hassan I, Kotrotsou A, Hassan A, Zinn PO, Viswanathan A, et al. Magnetic resonance-based radiomic analysis of radiofrequency lesion predicts outcomes after percutaneous cordotomy: a feasibility study. *Oper Neurosurg (Hagerstown)* 2020 Jun 01;18(6):721-727. [doi: [10.1093/ons/opa288](https://doi.org/10.1093/ons/opa288)] [Medline: [31665446](https://pubmed.ncbi.nlm.nih.gov/31665446/)]
24. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online October 11, 2018. [doi: [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805)]
25. Naseri H, Kafi K, Skamene S, Tolba M, Faye MD, Ramia P, et al. Development of a generalizable natural language processing pipeline to extract physician-reported pain from clinical reports: Generated using publicly-available datasets and tested on institutional clinical reports for cancer patients with bone metastases. *J Biomed Inform* 2021 Aug;120:103864 [FREE Full text] [doi: [10.1016/j.jbi.2021.103864](https://doi.org/10.1016/j.jbi.2021.103864)] [Medline: [34265451](https://pubmed.ncbi.nlm.nih.gov/34265451/)]
26. Elbattah M, Arnaud É, Gignon M, Dequen G. The role of text analytics in healthcare: a review of recent developments and applications. 2021 Presented at: 14th International Joint Conference on Biomedical Engineering Systems and Technologies - Scale-IT-up; February 11-13, 2021; Online. [doi: [10.5220/0010414508250832](https://doi.org/10.5220/0010414508250832)]
27. Smith TMF, Cochran WG. Sampling techniques, second edition. *Applied Statistics* 1964;13(1):54. [doi: [10.2307/2985224](https://doi.org/10.2307/2985224)]
28. Freedman D, Pisani R, Purves R. *Statistics: Fourth International Student Edition*. New York, NY: W.W. Norton & Company; 2007.
29. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001:17-21 [FREE Full text] [Medline: [11825149](https://pubmed.ncbi.nlm.nih.gov/11825149/)]
30. McCray AT, Aronson AR, Browne AC, Rindflesch TC, Razi A, Srinivasan S. UMLS knowledge for biomedical language processing. *Bull Med Libr Assoc* 1993 Apr;81(2):184-194 [FREE Full text] [Medline: [8472004](https://pubmed.ncbi.nlm.nih.gov/8472004/)]
31. hn617/texTRACTOR: texTRACTOR. Zenodo. 2021. URL: <https://zenodo.org/record/4649625> [accessed 2023-04-18]
32. hn617/diCOMBINE: diCOMBINE. Zenodo. 2021. URL: <https://zenodo.org/record/5218743> [accessed 2023-04-18]
33. Busscher I, Ploegmakers JJW, Verkerke GJ, Veldhuizen AG. Comparative anatomical dimensions of the complete human and porcine spine. *Eur Spine J* 2010 Jul 26;19(7):1104-1114 [FREE Full text] [doi: [10.1007/s00586-010-1326-9](https://doi.org/10.1007/s00586-010-1326-9)] [Medline: [20186441](https://pubmed.ncbi.nlm.nih.gov/20186441/)]
34. Deglinc HJ, Rangayyan RM, Ayres FJ, Boag GS, Zuffo MK. Three-dimensional segmentation of the tumor in computed tomographic images of neuroblastoma. *J Digit Imaging* 2006 Aug 25;20(1):72-87. [doi: [10.1007/10278-006-0769-3](https://doi.org/10.1007/10278-006-0769-3)]
35. Ulano A, Bredella MA, Burke P, Chebib I, Simeone FJ, Huang AJ, et al. Distinguishing untreated osteoblastic metastases from enostoses using CT attenuation measurements. *Am J Roentgenol* 2016 Aug;207(2):362-368. [doi: [10.2214/ajr.15.15559](https://doi.org/10.2214/ajr.15.15559)]
36. Smoothing Images. OpenCV. URL: https://docs.opencv.org/4.x/d4/d13/tutorial_py_filtering.html [accessed 2023-09-18]
37. mhe/pynrrd: v0.4.3 Released. Zenodo. 2022. URL: <https://zenodo.org/record/6501810> [accessed 2023-04-18]
38. Nearly Raw Raster Data. URL: <http://teem.sourceforge.net/nrrd/index.html> [accessed 2022-09-04]
39. Lemaitre G, Nogueira F, Aridas CK. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J Mach Learn Res* 2017;18:1-5 [FREE Full text]
40. Low E. Review of Understanding Basic Statistics. *Am Stat* 1998;52(2):198. [doi: [10.2307/2685480](https://doi.org/10.2307/2685480)]
41. Torvik K, Hølen J, Kaasa S, Kirkevold ?, Holtan A, Kongsgaard U, et al. Pain in elderly hospitalized cancer patients with bone metastases in Norway. *Int J Palliat Nurs* 2008 May;14(5):238-245. [doi: [10.12968/ijpn.2008.14.5.29491](https://doi.org/10.12968/ijpn.2008.14.5.29491)] [Medline: [18563017](https://pubmed.ncbi.nlm.nih.gov/18563017/)]
42. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002 Jun 01;16:321-357. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]

43. Xie C, Du R, Ho JW, Pang HH, Chiu KW, Lee EY, et al. Effect of machine learning re-sampling techniques for imbalanced datasets in F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *Eur J Nucl Med Mol Imaging* 2020 Nov 06;47(12):2826-2835. [doi: [10.1007/s00259-020-04756-4](https://doi.org/10.1007/s00259-020-04756-4)] [Medline: [32253486](https://pubmed.ncbi.nlm.nih.gov/32253486/)]
44. Tibshirani R. Regression shrinkage and selection via The Lasso: a retrospective. *J R Stat Soc Series B Stat Methodol* 2011;73(3):273-282. [doi: [10.1111/j.1467-9868.2011.00771.x](https://doi.org/10.1111/j.1467-9868.2011.00771.x)]
45. Yin P, Mao N, Chen H, Sun C, Wang S, Liu X, et al. Machine and deep learning based radiomics models for preoperative prediction of benign and malignant sacral tumors. *Front Oncol* 2020 Oct 16;10:564725 [FREE Full text] [doi: [10.3389/fonc.2020.564725](https://doi.org/10.3389/fonc.2020.564725)] [Medline: [33178593](https://pubmed.ncbi.nlm.nih.gov/33178593/)]
46. Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JPB, et al. Radiomics in oncology: a practical guide. *Radiographics* 2021 Oct;41(6):1717-1732. [doi: [10.1148/rg.2021210037](https://doi.org/10.1148/rg.2021210037)] [Medline: [34597235](https://pubmed.ncbi.nlm.nih.gov/34597235/)]
47. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006 Jun;27(8):861-874. [doi: [10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010)]
48. Fleischman RJ, Frazer DG, Daya M, Jui J, Newgard CD. Effectiveness and safety of fentanyl compared with morphine for out-of-hospital analgesia. *Prehosp Emerg Care* 2010 Mar 03;14(2):167-175 [FREE Full text] [doi: [10.3109/10903120903572301](https://doi.org/10.3109/10903120903572301)] [Medline: [20199230](https://pubmed.ncbi.nlm.nih.gov/20199230/)]
49. Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. Learning from imbalanced data streams. In: *Learning from Imbalanced Data Sets*. Cham: Springer; 2018.
50. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep learning vs. radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: don't forget the peritumoral region. *Front Oncol* 2020 Jan 31;10:53 [FREE Full text] [doi: [10.3389/fonc.2020.00053](https://doi.org/10.3389/fonc.2020.00053)] [Medline: [32083007](https://pubmed.ncbi.nlm.nih.gov/32083007/)]
51. Lisson CS, Lisson CG, Mezger MF, Wolf D, Schmidt SA, Thaiss WM, et al. Deep neural networks and machine learning radiomics modelling for prediction of relapse in mantle cell lymphoma. *Cancers (Basel)* 2022 Apr 15;14(8):2008 [FREE Full text] [doi: [10.3390/cancers14082008](https://doi.org/10.3390/cancers14082008)] [Medline: [35454914](https://pubmed.ncbi.nlm.nih.gov/35454914/)]
52. Thai-Nghe N, Gantner Z, Schmidt-Thieme L. Cost-sensitive learning methods for imbalanced data. 2010 Presented at: The 2010 International Joint Conference on Neural Networks (IJCNN); July 18-23, 2010; Barcelona. [doi: [10.1109/ijcnn.2010.5596486](https://doi.org/10.1109/ijcnn.2010.5596486)]
53. Liu A, Martin C, La Cour B, Ghosh J. Effects of Oversampling Versus Cost-Sensitive Learning for Bayesian and SVM Classifiers. In: Stahlbock R, Crone S, Lessmann S, editors. *Data Mining. Annals of Information Systems (volume 8)*. Boston, MA: Springer; 2010.
54. Tamang S, Humbert-Droz M, Gianfrancesco M, Izadi Z, Schmajuk G, Yazdany J. Practical considerations for developing clinical natural language processing systems for population health management and measurement. *JMIR Med Inform* 2023 Jan 03;11:e37805 [FREE Full text] [doi: [10.2196/37805](https://doi.org/10.2196/37805)] [Medline: [36595345](https://pubmed.ncbi.nlm.nih.gov/36595345/)]

Abbreviations

AI: artificial intelligence
API: average pain intensity
AUC: area under the receiver operating characteristic curve
BM: bone metastasis
CT: computed tomography
EN: ensemble
NLP: natural language processing
ROC: receiver operating characteristic
ROI: region of interest
SMOTE: Synthetic Minority Oversampling Technique
SP: spherical
VDP: verbally declared pain

Edited by K El Emam, B Malin; submitted 02.12.22; peer-reviewed by E Hmouda, SY Wang, M Elbattah; comments to author 06.02.23; revised version received 12.03.23; accepted 01.04.23; published 22.05.23

Please cite as:

Naseri H, Skamene S, Tolba M, Faye MD, Ramia P, Khriouan J, David M, Kildea J
A Scalable Radiomics- and Natural Language Processing-Based Machine Learning Pipeline to Distinguish Between Painful and Painless Thoracic Spinal Bone Metastases: Retrospective Algorithm Development and Validation Study
JMIR AI 2023;2:e44779
URL: <https://ai.jmir.org/2023/1/e44779>
doi: [10.2196/44779](https://doi.org/10.2196/44779)
PMID:

©Hossein Naseri, Sonia Skamene, Marwan Tolba, Mame Daro Faye, Paul Ramia, Julia Khriouian, Marc David, John Kildea. Originally published in JMIR AI (<https://ai.jmir.org>), 22.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.