

# (Semi-)Automatic Segmentation of Vertebral Bone Marrow Lesions on Magnetic Resonance and Dual Energy Computed Tomography Images: a Literature Review

T.O.M. Konings<sup>1,2</sup>

<sup>1</sup>Image Sciences Institute, Utrecht University

<sup>2</sup>Department of Radiology, University Medical Center Utrecht

**Abstract**—Bone marrow lesion syndrome (BMLS) describes the phenomenon of strong signal intensity changes in the bone marrow in fluid-sensitive magnetic resonance imaging (MRI). BMLs have taken up a central presence in many different diseases affecting the musculoskeletal system, including the spine. As the condition of the spine is of strong influence to the overall body's health, detection and characterization of BMLs could help in preventing the progression of spine related diseases. With the increased clinical availability of dual energy computed tomography (DECT), an alternative method for imaging of BMLs has presented itself through virtual non-calcium (VNCa) techniques. In this paper we explore the current state of both imaging modalities for (semi-)automatic segmentation of BMLs in the spine, as well as a variety of approaches to the segmentation of individual vertebrae that could be used to develop a fully automated process for BML segmentation. We found that for both MRI and DECT, automatic BML segmentation is still in its infancy, with few studies attempting the task. Additionally, manual segmentations of BMLs, which act as the ground truth for evaluating automated approaches, generally show significant inter- and intraobserver variability. However, with CT being the modality of choice in the traumatic setting, its potential for further developments in material decomposition and various other advantages that DECT poses over MRI, DECT could become the future modality of choice in both qualitative and quantitative imaging of BMLS.

## I. INTRODUCTION

**T**HE structure of the spine and its individual vertebrae is complicated and its condition is of strong influence to the overall body's health.[1] Correct identification of pathologies affecting the vertebrae therefore not only helps in preventing the progression of spine-related diseases, but also in providing doctors with necessary information to better design a therapeutic schedule.[1] One such disease, affecting the structure of bone marrow specifically, is known as bone marrow edema (BME).

BME describes the phenomena of strong signal intensity changes in fluid-sensitive magnetic resonance imaging (MRI) acquisitions of bone marrow.[2] Regular bone marrow is filled by adipocytes and therefore rich in lipid content. In BME, immune cells and microvessels accumulate in the bone marrow where they replace the adipocytes, leading to an increase in

water and reduction in fat content.[3] Since the introduction of the term BME in 1988, further research has revealed a wide range of edema-like signal changes that can be attributed to other causes such as bone marrow necrosis, bone marrow fibrosis, and trabecular abnormalities.[4] As these abnormalities are rather non-characteristic for any one cause, the overarching term of bone marrow lesion (BML) has become preferred to describe these signal intensity changes, and is widely used.[5]

BMLs have taken up a central presence in many different diseases that affect the musculoskeletal system, in particular being associated with many in- and noninflammatory rheumatologic conditions.[6] The actual histopathological processes behind BMLs, however, still remain largely unknown.[7]

Pain is the main symptom of BMLs, however, the large variability in causes and underlying disease, as well as in treatment and prognosis make the management of BMLs an incredibly difficult and complex challenge.[8] On the whole, studies on BML syndrome (BMLS) generally remain inconclusive, often not meeting the diagnostic accuracy needed to properly diagnose BMLS.[8] Etiology based classifications of BML related diseases have been proposed by Starr et al. and Eriksen and Ringe. BMLS has also been categorized by means of its underlying cause, into mechanical, reactive and ischemic.[9][6][2][10][8] Another categorization one often encounters in literature is that of traumatic (or acute) and non-traumatic BMLs.[11][12] In posttraumatic settings, BMLs can serve as a key indicator for non-displaced fractures that might otherwise be missed in standard radiological images, as even in the absence of fractures BMLs are often present.[11]

MRI has since the first descriptions of BMLS been the de-facto standard in the detection of and research into BMLS.[13] However, in everyday clinical practice, computed tomography (CT) has always the modality of choice for inspecting and assessing bony structures in the body.[14] Unlike in MR images, conventional CT does not allow for the visualization of bone marrow, as the trabecular structure surrounding bone marrow can not be completely resolved.[15] Recent advancements in the development and availability of dual-energy or spectral CT (DECT) have seemed to turn the tides, as through the use of

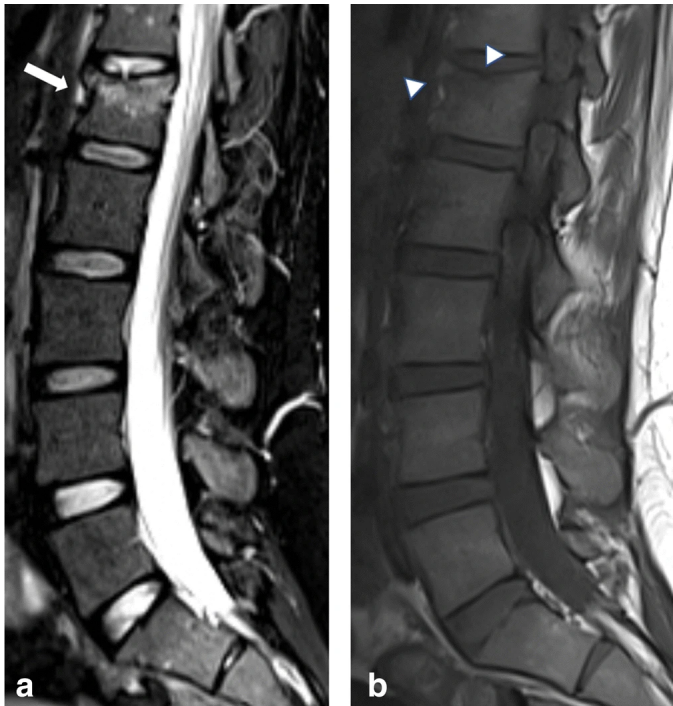


Figure 1: Two sagittal MR images of the lumbar spine of a twenty-six-year-old woman with spinal trauma after motorcycle accident. (a) is a turbo inversion recovery magnitude (TIRM) series showing a BML in the upper two quadrants of the L1 vertebrae (indicated by the arrow). (b) shows a spin-echo (SE)  $T_1$ -weighted image demonstrating two distinct acute fracture lines that affect the anterior and upper (indicated by arrowheads) cortical surfaces. The images illustrate a typical case of BMLS developed along acute fracture lines, as well as BMLs lack of sharp boundaries or anatomical edges. Images from Cavallaro et al..[16]

DECT, images can be acquired that allow the bone marrow to be resolved. This development has led to a wide body of research since, dedicated to the uses and potential of DECT for imaging BMLS.

Both MRI and CT are widely available, adapted and established modalities in both the clinical and research setting, each offering their own strengths and weaknesses. This review article will attempt to provide a relevant overview of some available work on semi-automatic vertebrae segmentation and the current state of (semi-)automatic BMLS detection for both MRI and CT. Through these findings it will be discussed what modalities and approaches might prove most fertile in realizing a (semi-)automated method for segmenting vertebral BMLS in the future.

## II. SEGMENTATION OF THE VERTEBRAE

Proper identification of BMLS in the spine is only possible with accurate and correct segmentation of the individual vertebrae. As BMLS may appear even in the outer regions of the bone marrow, accurate delineation of vertebra's borders

is essential. Although manual segmentation by experts may prove reliable and accurate, the complex shapes of individual vertebrae and long length of the spine make manual slice-for-slice segmentation of vertebrae an exceptionally labour-intensive task. It is therefore clear that a reliable (semi-)automated approach to segmentation is the only sustainable option for processing larger numbers of patients for research into BMLS of the spine.

Of course a good segmentation can only be performed on well-acquired images. This means that although the data may be acquired using different parameters or sequences, a certain base level of spatial resolution, temporal resolution and noise is required. A good segmentation approach would ideally be able to identify and distinguish between individual vertebrae. Additionally it is required for the method to function adequately for both healthy vertebrae as well as those affected by pathologies such as minor fractures, metastases or other factors that may impact the vertebra's morphology in the image.

The segmentation quality of an approach is an important factor in comparing different methods. A widely standardized evaluation metric for segmentation quality is unfortunately not in use, resulting in a large variety of metrics being used across different studies.[17] This can complicate the process of comparing study outcomes significantly. During this study it was observed that one of the most widely used metrics was the Sørensen–Dice coefficient (DICE). The DICE assumes comparison between a binary (usually manually segmented) ground truth and segmentation, being defined between 0 and 1 as two times the overlap divided by the cumulative amount of voxels of the segmentation and ground truth. Additionally the DICE can also be used as a measure of reproducibility of manual segmentation by comparing between segmentations of different experts.[17]

Another often calculated metric is the Hausdorff distance (HD). The HD provides the smallest euclidean distance between a point on the ground truth and any point on the performed segmentation. This measure can be an indication of the severity of deviation between two sets and alert for deviations that would be muddled out when judging purely on the DICE, or other metrics such as average absolute distance (AAD), mean symmetric surface distance (MSSD) or mean radial Euclidean distance (MRED). The HD is generally susceptible to outliers.[17]

Sensitivity (or recall) and specificity may also be mentioned throughout various articles. They respectively represent the parts of positive and negative voxels in the ground truth that are correctly identified as such by the segmentation. The problem with this metric is that it doesn't scale well to larger segmentation sizes, which are often the case in three-dimensional medical images.[17]

### A. MRI

In the aim of minimizing the risk of cancer due to radiation exposure in X-ray and CT, MRI has become a widely adapted and indispensable pillar of clinical decision making. Although

a wide range of different MRI sequences may be applied, it is important for a segmentation approach to be able to function adequately regardless of sequence, especially with the lack of standardized measurement units in MRI.[18][19]

On both  $T_1$ - and  $T_2$ -weighted MRI sequences, bone structures emit similar signals, although these may be affected due to various pathologies.[2] For the spine, intensity gradients are concentrated on the edges of the vertebral bodies.[20] In order to shorten acquisition times, MR images often feature anisotropic voxels, amplifying partial volume effects.[21]

A 2002 study by Davatzikos et al. proposed a method for spatial normalization and segmenting the spines using a pre-constructed deformable model of the spine that was subsequently transformed onto images.[21] The model was based on the manual segmentation of the spine from a typical  $T_1$ -weighted image of a single patient. The pre-constructed nature of this model generally prevents arbitrary deformations of the segmentation, however unrealistic deformations due to close-proximity edges of structure can pose a problem. Additionally, the relative rigidity of the model allows less for the accurate modeling of more subtle morphological defects of the vertebrae. On a relatively small dataset of  $T_1$ -weighted images of 14 healthy individuals, a DICE of 0.815 with a standard deviation (SD) of 0.036 was achieved. Davatzikos et al. argued that the relatively large overlap error was due to the fact that for shapes as complex as the whole spine, even small disagreements due to for example pixelation, can lead to low overlap scoring.

A 2011 study by Štern et al. presented a similar approach utilizing deformable shapes, but based in a more abstract superquadric model. [20] The model requires an initial seed point and is applied for individual vertebrae. The vertebral bodies are modelled along 25 parameters for shape (representing hand-picked clinically meaningful transformations) and 6 parameters for pose. Through adjusting these parameters an initially elliptical cylinder is displaced and oriented, after which it is deformed and aligned to best fit the vertebrae by maximization of a similarity measure. This approach was applied on a set of 75  $T_2$ -weighted images of vertebrae, both healthy and suffering from pathologies, in the thoracolumbar part of the spine. To evaluate the performance of the approach the MRED of the model was compared to 16 manually placed ground truth points for each vertebra. An MRED was measured of  $1.85 \pm 0.47$  mm. Processing times could be rather lengthy ranging from 1 to 15 minutes per vertebrae, however, in addition to computer processing speed improvements since, the study also noted that the program design and code showed room for optimization.

Gaonkar et al. in a 2017 study used a superpixel based multi-parameter ensemble model trained on manually segmented  $T_2$  images of the lumbar. [19] Although the algorithm was only trained on 6  $T_2$ -weighted images, segmentations were performed on both 48  $T_2$ - and 15  $T_1$ -weighted images yielding a mean DICE of  $0.83 \pm 0.06$  for  $T_2$  and expected lower results for  $T_1$ -weighted images (mean DICE of 0.75). The exact number of vertebrae was not specified. For the

small amount of training data used the results were found to be quite satisfactory, showing potential of better performance through more training data. The method was however rather computationally expensive, taking between 12 to 14 hours to train on a high-spec desktop computer.

An approach that coupled deformable models to convolutional neural networks (CNN) was proposed in 2016 by Korez et al..[22] On a set of 23 publicly available  $T_2$  weighted images used for both training and testing a DICE of  $0.934 \pm 0.017$ , MSSD of  $0.54 \pm 0.14$  mm and HD of  $3.83 \pm 1.04$  mm were achieved. No pathologies were noted to be present within the data that consisted of the lumbar and part of the thoracic vertebrae. Initial centroids for the vertebrae were coarsely determined manually, suggesting that the replacement of this by a fully automated method would not affect the outcome.

The very same set of  $T_2$ -weighted images was tackled a year earlier in 2015 by Chu et al..[23] Here the vertebrae were localized using random forest regression, also resulting in a probability map. This probability is used to perform a random forest classification on the ROI's and then combined to calculate the borders of the vertebrae. This yielded a DICE of  $0.887 \pm 0.029$  with a 3D HD of  $6.4 \pm 1.2$  mm and 3D AAD  $1.5 \pm 0.2$  mm. On a relatively high-end but still desktop-spec computer, segmentation times were on average around 2.0 minutes with training times around 7.4 minutes.

## B. CT

For inspecting and assessing bony structures in the body, CT is the modality of choice.[14] Unlike in MR images CT does not allow for the visualization of bone marrow, as the trabecular structure surrounding bone marrow can not be completely resolved.[15] DECT, although first envisioned and put into scientific practice in the 1970s, has only recently become a clinically available tool.[24][25] In DECT, two images of the same patient are acquired at different distinct energy levels (ideally simultaneously, to minimize patient movement). This allows for the images to be combined to generate a variety of new data.

DECT is clinically mainly used for the ability to generate virtual mono energetic (monoE) images, which simulate scan data as if it were acquired at a single desired energy level. Certain energy levels generate contrasts more suited for certain applications, such as optimal contrasts between lesions and tissues in CTA (45-55 keV), soft-tissue evaluation (60-75 keV) or the reduction of artefacts from metal implants (95-140 keV).[25] This study revealed no available literature on using any DECT specific methods for vertebrae segmentation, however, the flexibility in imaging contrasts may potentially improve the performance of already available (semi-)automatic segmentation approaches.

In a 2013 study Huang et al. used an improved version of the traditional level-set method by combining both region- and edge-based level set methods, to successfully segment vertebrae in a set of 56 images showing intervertebral disc protrusions. Due to the strong absorption of bone materials in CT images, a relatively simple approach such as Otsu

thresholding can be used to quickly generate an initial level set, also affecting computational times beneficially. This approach yielded a DICE of  $0.94 \pm 0.02$  and HD of  $10.06 \pm 1.71$  mm over 293 segmented vertebrae. Compared to the widely known local-binary-fitting model the approach compares favourably by being less sensitive to initial contours as well as much more computationally efficient. Processing times on a mid-spec computer ranged between 0.8 and 28 seconds.

In a 2021 study, Cheng et al. employed a two-stage Dense-U-Net for fully automated segmentation. The first stage uses a 2D-Dense-U-Net to localise and label the vertebrae centroids. Around the centroids an ROI is cropped and some pre-processing is applied, after which in the second stage a 3D-Dense-U-Net was trained and tested on. The method was tested and evaluated on the datasets from the CSI 2014 Workshop challenge and the 2017 xVertSeg challenge which also includes fractured vertebrae. This resulted in a DICE of  $0.953 \pm 0.014$  with HD of  $4.013 \pm 2.128$  mm, and DICE of  $0.877 \pm 0.035$  with no given HD respectively.[1]

The 2015 approach by Chu et al. that was described earlier for MRI images, was also evaluated for CT. Using a slightly smaller set of 10 CT images for both training and testing, yielding a DICE of  $0.910 \pm 0.070$  along with a mean HD of  $7.3 \pm 2.2$  mm and AAD of  $0.9 \pm 0.3$  mm. Segmentations took around 2.3 minutes per image on average and training 9.3 minutes.[23]

The 2011 superquadric model used by Štern et al. was also applied on a set of 75 CT acquisitions of thoracolumbar vertebrae from different institutes and machines, extracted from both normal and pathological spines.[20] The measured MRED between the final 3D models and the ground truth points was determined at  $1.17 \pm 0.33$  mm. It was noted that although the system is based on 25 parameters, it could be expanded to more parameters to represent more deformations that could occur and potentially increase fitting performance.

### III. BME IDENTIFICATION AND QUANTIFICATION

Having achieved successful segmentations of the vertebrae, the next step is to investigate the bone marrow for the presence of BMLs. The spine is the largest store of bone marrow in the body and its composition changes continually through aging and different states of health. [27] Both red and yellow marrow can be found, composed of different ratios of lipids, water and proteins. Awareness of the various bone marrow changes in the body is essential for radiologists and will avoid overrating normal bone marrow patterns as pathologic states.[27]

#### A. MRI

BMLs show signals higher than muscle or disc on  $T_1$ -weighted images and a slightly higher signal on  $T_2$ -weighted images. On fat suppressed, short-tau inversion recovery (STIR) and contrast enhanced sequences, BMLs show hyperintense compared to regular bone marrow.[9] Due to lower cost and no risk of a patient's reaction to a contrast agent, STIR sequences are generally advised.[28]  $T_1$ ,  $T_2$  and STIR sequences provide gross morphological data. Non-routine MRI sequences such

as chemical-shift or diffusion-weighted imaging may improve on MRI's ability to distinguish between different types of heterogeneities in the bone marrow. These have however not been as widely explored.[27]

On MRI images the appearance of BMLs should be somewhat homogeneous but will also display a lack of sharp boundaries or anatomical edges (see Figure 1).[6][9] Within the manual assessment of lesion boundaries this is a cause of substantial subjectivity and leads to high inter-observer variability.[29]

Although MRI shows excellent diagnostic performance and holds the position as standard for (qualitative) detection and imaging of BMLs, there are several barriers to its application, especially in the emergent setting. These can include the associated cost of operation and installation, limited access, rigorous screening requirements, relatively long examinations and the incompatibility with certain pacemakers or other implants.[13][30] Additionally, MRI signal is highly sensitive to radiofrequency inhomogeneities and small positional changes of the patient within the coil. This means that even when performing an acquisition of the same patient twice using the same parameters, a slightly different distribution of gray scale values will turn out.[31]

Kucybała et al. performed a study in 2020 to create an efficient tool for semi-automated detection of BMLs in the sacroiliac joints of patients with axial spondyloarthritis.[32] This approach attempts to automatically identify a reference signal level that exists a user set distance from the segmentation's border, after which various ROIs are defined (which are again split into quadrants) to be searched for inflammatory changes between  $T_1$ -weighted and STIR images. 22 MR images of sacroiliac joints of patients with confirmed axial spondyloarthritis were included into the study. The pixel-by-pixel comparison between manual and automated findings yielded a Spearman's correlation coefficient yielded 0.87 while the per-quadrant comparison yielded 0.83. As the manual segmentations were performed twice by different radiologists their correlation was also determined, yielding 0.91 for the pixel-by pixel and 0.88 for the quadrant-wise comparison. This is indicative of the levels of inter-observer variability in manual BML segmentations on MRI. Average computational times were  $0.64 \pm 0.30$  s per slice with patient data ranging between 18 and 24 slices.

A 2021 feasibility study by Rzecki et al. (also featuring Kucybała as second author) investigated a fully automatic BML detection method on MRI images of 30 patients also suffering from axial spondyloarthritis (axSpA). Initial bone segmentations were performed on  $T_1$  and STIR images using a deep learning approach, after which the ROIs (in the subchondral bone and central part of the sacral bone) were extracted.[29][32] Manual segmentations of BME in the STIR images were made by three independent radiologists and majority voted to create a ground truth. An U-net deep learning based lesion classification was trained and applied after which its performance was evaluated. As the amount of available data was moderate, the U-net architecture was

Table I: Summarized vertebrae segmentation results

Study	Modality	Method	Results	Data	Computational time
Davatzikos et al.[21]	MRI	Deformable shape model registration	DICE: 0.865±0.036	14 lumbar spines	None specified
Štern et al.[20]	MRI CT	Superquadric model	MRI: MRED: 1.85±0.47 mm CT: MRED: 1.17±0.33 mm	75 thoracolumbar vertebrae	1-15 minutes per vertebra
Gaonkar et al.[19]	MRI	Superpixel multi-parameter ensemble model	$T_2$ DICE: 0.83±0.06 $T_1$ DICE: 0.75	6 $T_2$ lumbar training images 48 $T_2$ lumbar images 15 $T_1$ lumbar images	12-14 hours training
Korez et al.[22]	MRI	CNN coupled 3D deformable model	DICE: 0.934±0.017 MSSD: 0.54±0.14 mm HD: 3.83±1.04 mm	161 thoracic and lumbar vertebrae	None specified
Chu et al.[23]	MRI CT	Random forests	MRI: DICE: 0.887±0.029 3D HD: 6.4±1.2 mm 3D AAD 1.5±0.2 mm CT: DICE: 0.91±0.07 3D HD: 7.3±2.2 mm AAD: 0.9±0.3 mm	MRI: 161 thoracic and lumbar vertebrae CT: 50 thoracic and lumbar vertebrae	MRI: 7.4 minutes training 2.0 minutes per image CT: 9.3 minutes training 2.3 minutes per image
Huang et al.[26]	CT	Region and edge based level-set	DICE 0.94±0.02 HD: 10.06±1.71 mm	293 lumbar vertebrae	0.8-28 seconds per vertebra
Cheng et al.[1]	CT	Two stage U-net	CSI 2014: DICE:0.953±0.014 HD: 4.013±2.128 xVertSeg 2017: DICE:0.877±0.035	CSI 2014 Workshop data xVertSeg challenge data	10 hours training 50 seconds per image

AAD: Average Absolute Distance; MSSD: Mean Symetric Surface Distance; MRED: Mean Radial Euclidian Distance; HD: Hausdorff Distance; DICE: Sørensen–Dice coefficient

selected a priori and not specifically optimized for the data. It was observed that some regions clearly showing increased STIR signal were still denoted as negative in the majority voting. It was therefore suspected that some positive regions had been missed during manual segmentation, leading to the false positive regions identified by the deep learning network being reviewed. This significantly increased the evaluated specificity and sensitivity (0.96 and 0.95 respectively) of the deep learning approach, however, it again is strongly indicative of the inter- and intraobserver variability in BML detection. The Spearman’s coefficient of correlation between automated and manual measurements was determined at 0.866 (95% CI from 0.735 to 0.934).

A 2017 study by Aizenberg et al. explored the possibilities for a fully automated quantitative approach by examining the carpal bone in the wrists of 485 early arthritis patients for BMLs. Using atlas based image registration, the carpal bones were located and segmented on  $T_1$ -weighted frequency-selective fat saturation images ( $T_1 - Gd$ ). The presence of BMLS was quantified through intensity based fuzzy C-means clustering, where two clusters are assumed (BMLs and normal bone marrow). This results in two probability maps, in which each voxel corresponds to the probability of that voxel belonging to the respective cluster. Using an optimized threshold, the total fraction between 0 and 1 of the segmented voxels determined to exhibit BMLS represent the quantitative BML score.

A Pearson correlation score between the sum of visual BML scores across all carpal bones and the sum of the quantitative method across all carpal bones was determined at 0.83. As the correlation scoring here only takes into account a cumulative grading it doesn’t provide insight on the performance of the approach on actual BML segmentation quality.

### B. DECT

The effectiveness of DECT for qualitative assessment of BMLs has been widely studied and proven. A meta-analysis by Suh et al. pooled together 12 different studies, evaluating DECT performance in detecting BMLS over 1091 BMLs in 450 patients. A pooled a specificity of 0.85 (95% CI, 0.78-0.90) and sensitivity of 0.97 (95% CI, 0.92-0.98) was determined. 7 studies included the detection of BMLs in the spine where a pooled sensitivity for detecting BMLs of 0.84 (95% CI, 0.72–0.92) and specificity of 0.98 (95% CI, 0.95–0.99) were found. All studies included in the meta-analysis use the technique of virtual non-calcium (VNCA) whereby present calcium can be subtracted from anatomical structures, allowing for the visibility of features or pathological conditions that might be obscured on standard CT.[35] Using this technique to observe within the bone marrow, the presence of BML is detectable and directly visualised, as bone marrow attenuation increases when its fatty content is replaced by edema and micro-hemorrhages.[35][36] Use of CT then poses several advantages over the application of MRI such as

shorter investigation times, higher spatial resolution and lower cost.[36]

Analysis indicated that the image plane number, slice thickness and evaluation method were all sources of heterogeneity between studies. The pooled sensitivity was shown to be higher in studies with two or more image planes, thinner than 1 mm slice thickness, and those applying a binary evaluation method. Decrease in slice thickness however did lead to decreases in specificity, most likely related to increases in noise. As such the use of slices thinner than 0.5mm is not recommended.[34]

Another, more recent, meta-analysis by BÄCKER et al. collected studies specifically looking at detection of BMLs (and disc edema) in patients with spine fractures. 13 studies (showing some overlap with the 2018 Suh et al. study) spanning 515 patients, 3335 vertebrae, and 926 acute fractures defined by MRI were included. The overall sensitivity was found to be 86.2% with a specificity of 91.2% and accuracy of 89.3%. Heterogeneity between the studies was found to be rather considerable, additionally, significant inter observer differences were again reported.

Bone sclerosis may locally increase the measured HU values and as such lead to false positive identification of BMLs. Additionally, excessive subtraction processes may hinder BML detection.[11] In most studies, color maps overlayed on top of the regular images are generated based on the DECT's three material decomposition, where after VNCA, differences among voxels mainly reflect water and fat content in the marrow (see Figure 2). This allows for both easier visual/qualitative assessment as well as quantization in ROIs expressed in Hounsfield units (HU).[35][37] The presence of the standardized HU in DECT is an advantage in areas of less obvious BML, however, thresholds can depend on many factors such as anatomical area being evaluated, age of the patient or the imaging parameters used, resulting in a wide range of values found from research to research.[11] In differentiating between regular marrow and BMLs, Cavallaro et al. found an effective cutoff around -0.43HU, Petritsch et al. of -47HU, -80HU for Wang et al. and Bierry et al. discerned between thoracic at 35 HU and 6.5HU for lumbar vertebrae.[16][36][38][15]

A 2016 study by Biondi et al. investigated a variation on the VNCA approach through performing a virtual non-cortical bone (VNCB) technique. On a set of both MR and DECT images of 8 patients, VNCA and VNCB were applied on the same ROIs corresponding to regular bone marrow. Threshold values were individually determined per image. Agreement between the color mapped DECT and MR image of the entire acquisition volume was scored by two radiologists on a grading scale ranging from 1 (completely disagree) to 5 (complete agreement). It was found that on all images the agreement scores obtained between MRI and VNCB were significantly higher than for the VNCA approach ( $4.12 \pm 0.83$  compared to  $2.25 \pm 0.46$  respectively)[37] Although the study has some severe limitations such as its small sample size, it is indicative of the versatility and possibility for further methodological improvements in using DECT for BML imaging.

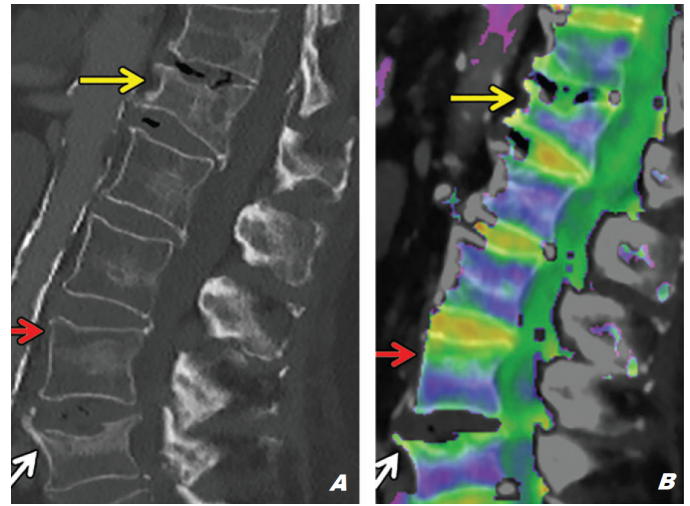


Figure 2: Saggital images of osteoporotic compression fractures in a 97-year-old man with back pain. (a) shows a nonenhanced CT image of the lumbar spine with compressions of the superior endplate of T12 (indicated by the yellow arrow) and L4 (indicated by the white arrow) vertebrae, and a nearly occult fracture of the superior endplate of L3 (indicated by the red arrow). (b) shows a VNCA overlay made with DECT, the attenuation within the bone marrow is displayed by a color map in purple reflects the presence of fat, and higher HU values are displayed in green. The color map indicates the presence of BMLs in the L4 and L3 vertebra, however not the T12. Images from Gosangi et al..[13]

With the clinical availability of DECT scanner being rather recent, ongoing developments on the manufacturer's side of DECT scanners (the current offering is the third generation) promise even better acquisitions in the future. Research by Petritsch et al. comparing the quality of second generation to third generation DECT for quantitative analysis confirmed significant increases in sensitivity and specificity.[36]

#### IV. DISCUSSION

Both for MRI and CT, a wide range of vertebral body segmentation methods are available that all produce good to excellent results. The more homogeneous nature and inherently high signal of bone material might suggest CT images to be more suitable for "traditional" segmentation approaches such as level-sets than MRI images. The Huang et al. study using level-sets presents one of the higher DICE scores, however, a high mean HD suggests susceptibility of the method to outliers or unrealistic deformations.

Segmentation methods that were applied and evaluated for both CT and MRI showed slightly better performance on CT, however, observed differences were minor.[23][20] Additionally, direct comparison between CT and MRI outcomes in the Chu et al. study is difficult, as training set size was significantly smaller for CT. Recent research on deep learning based segmentation for both MRI and CT, Korez et al. and Cheng et al. respectively, have yielded even better results

that only promise more robustness and reliability with more training data. Segmentation of individual vertebrae on both CT and MRI seems yield the best results, although computationally expensive, through using deep learning approaches. With prospective of computational power only becoming cheaper as time goes on, this however does not seem to pose a limitation for deep learning approaches.

A 2021 study by Cavallaro et al. directly compared the diagnostic performance of a third generation DECT to a 3.0 T MRI scanner in assessment of acute vertebral fractures. 88 patients were taken up in the study and imaged with both modalities within a maximum interval of a week. This resulted in 730 included vertebral bodies of the thoracic or lumbar spine. As expected, mean acquisition times were significantly shorter for DECT than MRI at 21 s (range, 18–32 s) compared to 1103 s (range, 902–1306 s) respectively.

Reference standard assessments were made by two highly experienced radiologists where, in consensus reading sessions, the presence of traumatic BMLs and fracture lines was rated on a 4-point scale. Additionally the vertebrae were split up into quadrants each representing 25% of the total volume, and received a scoring of the extent of BML based on the number of quadrants containing BMLs (ranging between 1 to 4). Afterwards, evaluations were performed on all images (both DECT and MRI) by 5 different radiologists using the same rating system along with assessments of image quality, image noise, and diagnostic confidence on a 5-point Likert scale (1 = unacceptable to 5 = excellent).

The study found for both DECT and MRI to provide similar levels of diagnostic confidence, image noise and image quality in assessment of acute vertebral fractures. It was also found that DECT yielded high overall diagnostic accuracy in depicting the presence of BMLs through application of colored VNCa overlays compared to MRI. These findings are directly in line with that of other studies and meta-analyses on the diagnostic potential of DECT for BMLS, such as Foti et al., Suh et al. and BÄCKER et al..

Quantitative MRI measurements with segmentation of BMLs have attracted relatively little interest in addition to being difficult due to the lack of physical units in MRI.[31][33] Manual approaches generally do not attempt direct quantification of bone lesion volume but instead grade MRI images slice by slice based on the extent lesion within a this slice making these methods more semi-quantitative.[29]

Attempts at automating the segmentation of BML in MR images have been met with mixed results. Aizenberg et al. demonstrated that the resulting probability maps lead to grading that showed good correlation to manual/visual grading. Although the method outputs probability maps that in theory describe lesions on a pixel-by-pixel basis, no manual segmentations were performed to evaluate this aspect of the performance. Kucybała et al. showed similar pixel based quantification of BMLS to be comparable in reliability to manual assessment, however, was only assessed in the Sacroiliac joint and not in vertebrae. Additionally, the statistical approach used to achieve these results is on a per-pixel basis, meaning

that it holds no regard for the anatomical shape of the lesions. This leads to the segmentation as scattered pixels rather than as a map of connected regions. Rzecki et al. showed that the automated segmentation of BMLs through deep learning in MRI is possible, however, noted the small volume of lesions compared to the bone region as a serious difficulty in training the system. As the U-net employed was not optimized for the data (the available data set was deemed too small to optimize the U-net architecture) there is significant potential to further improve on the study results, both through optimization of the U-net architecture as well as by supplying more training data.

The inherent use of the standardized HU in CT images poses an advantage in performing quantitative analysis with DECT over MRI. The construction of color-coded maps for visual analysis has been widely used in many studies, however, this research revealed no studies that truly make an attempt at further automating the analysis process in the vertebrae. Standardized HU cutoffs could be an easy approach to automatically identifying (at least on an individual pixel basis) BMLs, but the high heterogeneity seen in effective cutoff values (ranging from 35 to -80 HU) suggests further research is necessary to establish some form of reference standard. This variability might be explained by the variation in bone marrow composition in different vertebrae and age groups.

DECT's unique abilities in performing material decomposition might be used to further expand on its performance and uses, such as through use of the Biondi et al. VNCB approach. Another possibility might be in distinguishing between different types of BMLs through characteristic HU levels.

## V. CONCLUSION

Both on MRI and DECT it seems that automatic BML segmentation (in the vertebrae) is still in its infancy. In order to fully automate the process from imaged vertebrae to segmented BML a robust segmentation approach for both is required. A wide array of vertebrae segmentation approaches have been developed and tested for both modalities. Current deep learning approaches seemingly yield the best results, assuming enough accurate training data is available. Attempts at automatic segmentation of BMLs on MRI images has so far been met with mixed results, showing that although feasible, much further work is necessary. Additionally, manual segmentations of BMLs seem to show significant interobserver variability. The standardized HU in CT suggests an easy means to segmentation through utilizing HU cutoffs, however, a large heterogeneity in found cutoff values for different studies underlines the need for further research into a reference standard. Additionally, meta-studies into the application of DECT for BMLs found studies showed significant inter observer differences. Although literature still supports the use of MRI as the gold standard, the widely studied and accepted effectiveness of current generation DECT for assessing BMLs can not be ignored, showing similar levels of diagnostic confidence and image quality. With an increased adaptation of DECT in the future, inter- and intraobserver agreement will probably improve. With CT being the modality of choice in

the traumatic setting, its potential for further developments in material decomposition and various other advantages of DECT over MRI, DECT could become the future modality of choice in both qualitative and quantitative imaging of BMLS.

## REFERENCES

- [1] Pengfei Cheng, Yusheng Yang, Huiqiang Yu, and Yongyi He. Automatic vertebrae localization and segmentation in CT with a two-stage dense-u-net. *Scientific Reports*, 11(1), nov 2021. doi: 10.1038/s41598-021-01296-1.
- [2] Sam Akhavan, Stephen C. Martinkovich, Connor Kasik, and Patrick J. DeMeo. Bone marrow edema, clinical significance, and treatment options: A review. *Journal of the American Academy of Orthopaedic Surgeons*, 28(20): e888–e899, jul 2020. doi: 10.5435/jaaos-d-20-00142.
- [3] Georg Schett. Bone marrow edema. *Annals of the New York Academy of Sciences*, 1154(1):35–40, feb 2009. doi: 10.1111/j.1749-6632.2009.04383.x.
- [4] A J Wilson, W A Murphy, D C Hardy, and W G Totty. Transient osteoporosis: transient bone marrow edema? *167(3):757–760*, jun 1988. doi: 10.1148/radiology.167.3.3363136.
- [5] F.W. Roemer, R. Frobell, D.J. Hunter, M.D. Crema, W. Fischer, K. Bohndorf, and A. Guermazi. MRI-detected subchondral bone marrow signal alterations of the knee joint: terminology, imaging appearance, relevance and radiological differential diagnosis. *Osteoarthritis and Cartilage*, 17(9):1115–1131, sep 2009. doi: 10.1016/j.joca.2009.03.012.
- [6] Erik Fink Eriksen and Johan Diederich Ringe. Bone marrow lesions: a universal bone response to injury? *32(3): 575–584*, sep 2011. doi: 10.1007/s00296-011-2141-2.
- [7] M. Manara and M. Varenna. A clinical overview of bone marrow edema. *66(2):184–196*, jul 2014. doi: 10.4081/reumatismo.2014.790.
- [8] Sebastian F. Baumbach, Vanessa Pfahler, Susanne Bechtold-Dalla Pozza, Isa Feist-Pagenstert, Julian Fürmetz, Andrea Baur-Melnyk, Ulla C. Stumpf, Maximilian M. Saller, Andreas Straube, Ralf Schmidmaier, and Jan Leipe. How we manage bone marrow edema—an interdisciplinary approach. *Journal of Clinical Medicine*, 9(2):551, feb 2020. doi: 10.3390/jcm9020551.
- [9] A. M. Starr, M. A. Wessely, U. Albastaki, C. Pierre-Jerome, and N. W. Kettner. Bone marrow edema: pathophysiology, differential diagnosis, and imaging. *49(7): 771–786*, sep 2008. doi: 10.1080/02841850802161023.
- [10] Siegfried Hofmann, Josef Kramer, Anosheh Vakil-Adli, Nicolas Aigner, and Martin Breitscheher. Painful bone marrow edema of the knee: differential diagnosis and therapeutic concepts. *35(3):321–333*, jul 2004. doi: 10.1016/j.ocl.2004.04.005.
- [11] Giovanni Foti, Gerardo Serra, Venanzio Iacono, and Claudio Zorzi. Identification of traumatic bone marrow oedema: The pearls and pitfalls of dual-energy CT (DECT). *7(3):424–433*, sep 2021. doi: 10.3390/tomography7030037.
- [12] Giovanni Foti, Gerardo Serra, Venanzio Iacono, Stefania Marocco, Giulia Bertoli, Stefania Gori, and Claudio Zorzi. Identification of non-traumatic bone mar-



- row oedema: The pearls and pitfalls of dual-energy CT (DECT). 7(3):387–396, aug 2021. doi: 10.3390/tomography7030034.
- [13] Babina Gosangi, Jacob C. Mandell, Michael J. Weaver, Jennifer W. Uyeda, Stacy E. Smith, Aaron D. Sodickson, and Bharti Khurana. Bone marrow edema at dual-energy CT: A game changer in the emergency department. *RadioGraphics*, 40(3):859–874, may 2020. doi: 10.1148/rg.2020190173.
- [14] Bernhard Tins. Technical aspects of CT imaging of the spine. *Insights into Imaging*, 1(5-6):349–359, oct 2010. doi: 10.1007/s13244-010-0047-2.
- [15] Chien-Kuo Wang, Jen-Ming Tsai, Ming-Tsung Chuang, Min-Tsung Wang, Kuo-Yuan Huang, and Ruey-Mo Lin. Bone marrow edema in vertebral compression fractures: Detection with dual-energy CT. *Radiology*, 269(2):525–533, nov 2013. doi: 10.1148/radiol.13122577.
- [16] Marco Cavallaro, Tommaso D’Angelo, Moritz H. Albrecht, Ibrahim Yel, Simon S. Martin, Julian L. Wichmann, Lukas Lenga, Silvio Mazziotti, Alfredo Blandino, Giorgio Ascenti, Marcello Longo, Thomas J. Vogl, and Christian Booz. Comprehensive comparison of dual-energy computed tomography and magnetic resonance imaging for the assessment of bone marrow edema and fracture lines in acute vertebral fractures. *European Radiology*, jul 2021. doi: 10.1007/s00330-021-08081-8.
- [17] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15(1), aug 2015. doi: 10.1186/s12880-015-0068-x.
- [18] Georg Hille, Sylvia Saalfeld, Steffen Serowy, and Klaus Tönnies. Vertebral body segmentation in wide range clinical routine spine MRI data. *Computer Methods and Programs in Biomedicine*, 155:93–99, mar 2018. doi: 10.1016/j.cmpb.2017.12.013.
- [19] Bilwaj Gaonkar, Yihao Xia, Diane S. Villaroman, Allison Ko, Mark Attiah, Joel S. Beckett, and Luke Macyszyn. Multi-parameter ensemble learning for automated vertebral body segmentation in heterogeneously acquired clinical MR images. *IEEE Journal of Translational Engineering in Health and Medicine*, 5:1–12, 2017. doi: 10.1109/jtehm.2017.2717982.
- [20] Darko Štern, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec. Parametric modelling and segmentation of vertebral bodies in 3d CT and MR spine images. 56(23):7505–7522, nov 2011. doi: 10.1088/0031-9155/56/23/011.
- [21] Christos Davatzikos, Dengfeng Liu, Dinggang Shen, and Edward H. Herskovits. Spatial normalization of spine MR images for statistical correlation of lesions with clinical symptoms. 224(3):919–926, sep 2002. doi: 10.1148/radiol.2243011266.
- [22] Robert Korez, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec. Model-based segmentation of vertebral bodies from MR images with 3d CNNs. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 433–441. Springer International Publishing, 2016. doi: 10.1007/978-3-319-46723-8\_50.
- [23] Chengwen Chu, Daniel L. Belavý, Gabriele Armbrecht, Martin Bansmann, Dieter Felsenberg, and Guoyan Zheng. Fully automatic localization and segmentation of 3d vertebral bodies from CT/MR images via a learning-based method. *PLOS ONE*, 10(11):e0143327, nov 2015. doi: 10.1371/journal.pone.0143327.
- [24] R. A. Rutherford, B. R. Pullan, and I. Isherwood. Measurement of effective atomic number and electron density using an EMI scanner. *Neuroradiology*, 11(1):15–21, 1976. doi: 10.1007/bf00327253.
- [25] J Grajo, M Patino, A Prochowski, and D Sahani. Dual energy ct in practice: Basic principles and applications. *Appl Radiol*, 45(7):6–12, 2016.
- [26] Juying Huang, Fengzeng Jian, Hao Wu, and Haiyun Li. An improved level set method for vertebra CT image segmentation. *BioMedical Engineering OnLine*, 12(1), may 2013. doi: 10.1186/1475-925x-12-48.
- [27] Mohamed Ragab Nouh and Ahmed Fathi Eid. Magnetic resonance imaging of the spinal marrow: Basic understanding of the normal marrow pattern and its variant. *World Journal of Radiology*, 7(12):448, 2015. doi: 10.4329/wjr.v7.i12.448.
- [28] S. EUSTACE, C. KEOGH, M. BLAKE, R.J. WARD, P.D. ODER, and M. DIMASI. MR imaging of bone oedema: Mechanisms and interpretation. *Clinical Radiology*, 56(1):4–12, jan 2001. doi: 10.1053/crad.2000.0585.
- [29] Krzysztof Rzecki, Iwona Kucybała, Daniel Gut, Aldona Jarosz, Tomasz Nabagło, Zbislaw Tabor, and Wadim Wojciechowski. Fully automated algorithm for the detection of bone marrow oedema lesions in patients with axial spondyloarthritis – feasibility study. *Biocybernetics and Biomedical Engineering*, 41(2):833–853, apr 2021. doi: 10.1016/j.bbe.2021.05.005.
- [30] HENRIK C. BÄCKER, CHIA H. WU, CARSTEN PERKA, and GERGELY PANICS. Dual-energy computed tomography in spine fractures: A systematic review and meta-analysis. *International Journal of Spine Surgery*, page 8074, may 2021. doi: 10.14444/8074.
- [31] Marius E. Mayerhoefer, Martin Breitenhofer, Siegfried Hofmann, Nicolas Aigner, Roland Meizer, Harald Siedentop, and Josef Kramer. Computer-assisted quantitative analysis of bone marrow edema of the knee: Initial experience with a new method. *American Journal of Roentgenology*, 182(6):1399–1403, jun 2004. doi: 10.2214/ajr.182.6.1821399.
- [32] Iwona Kucybała, Zbislaw Tabor, Jakub Polak, Andrzej Urbanik, and Wadim Wojciechowski. The semi-automated algorithm for the detection of bone marrow oedema lesions in patients with axial spondyloarthritis. *Rheumatology International*, 40(4):625–633, jan 2020. doi: 10.1007/s00296-020-04511-w.
- [33] Evgeni Aizenberg, Edgar A.H. Roex, Wouter P. Nieuwenhuis, Lukas Mangnus, Annette H.M. van der Helm-van Mil, Monique Reijnierse, Johan L. Bloem,

- Boudewijn P.F. Lelieveldt, and Berend C. Stoel. Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis: A feasibility study. *Magnetic Resonance in Medicine*, 79(2):1127–1134, may 2017. doi: 10.1002/mrm.26712.
- [34] Chong Hyun Suh, Seong Jong Yun, Wook Jin, Sun Hwa Lee, So Young Park, and Chang-Woo Ryu. Diagnostic performance of dual-energy CT for the detection of bone marrow oedema: a systematic review and meta-analysis. 28(10):4182–4194, apr 2018. doi: 10.1007/s00330-018-5411-5.
- [35] Tommaso D’Angelo, Moritz H. Albrecht, Danilo Caudo, Silvio Mazziotti, Thomas J. Vogl, Julian L. Wichmann, Simon Martin, Ibrahim Yel, Giorgio Ascenti, Vitali Koch, Giuseppe Cicero, Alfredo Blandino, and Christian Booz. Virtual non-calcium dual-energy CT: clinical applications. *European Radiology Experimental*, 5(1), sep 2021. doi: 10.1186/s41747-021-00228-y.
- [36] Bernhard Petritsch, Aleksander Kosmala, Andreas M. Weng, Bernhard Krauss, Anke Heidemeier, Richard Wagner, Timo M. Heintel, Tobias Gassenmaier, and Thorsten A. Bley. Vertebral compression fractures: Third-generation dual-energy CT for detection of bone marrow edema at visual and quantitative analyses. *Radiology*, 284(1):161–168, jul 2017. doi: 10.1148/radiol.2017162165.
- [37] M. Biondi, E. Vanzi, G. De Otto, F. Banci Buonamici, G.M. Belmonte, L.N. Mazzoni, A. Guasti, S.F. Carbone, M.A. Mazzei, A. La Penna, E. Foderà, D. Guerreri, A. Maiolino, and L. Volterrani. Water/cortical bone decomposition: A new approach in dual energy CT imaging for bone marrow oedema detection. a feasibility study. *Physica Medica*, 32(12):1712–1716, dec 2016. doi: 10.1016/j.ejmp.2016.08.004.
- [38] Guillaume Bierry, Aina Venkatasamy, Stéphane Kremer, Jean-Claude Dosch, and Jean-Louis Dietemann. Dual-energy CT in vertebral compression fractures: performance of visual and quantitative analysis for bone marrow edema demonstration with comparison to MRI. *Skeletal Radiology*, 43(4):485–492, jan 2014. doi: 10.1007/s00256-013-1812-3.
- [39] Giovanni Foti, Alberto Beltramello, Matteo Catania, Stefano Rigotti, Gerardo Serra, and Giovanni Carbone. Diagnostic accuracy of dual-energy CT and virtual non-calcium techniques to evaluate bone marrow edema in vertebral compression fractures. *La radiologia medica*, 124(6):487–494, feb 2019. doi: 10.1007/s11547-019-00998-x.

Magnetic Resonance Imaging (MRI) en Computed Tomography (CT) zijn beide veelgebruikte modaliteiten om non-invasief onderzoek in het menselijk lichaam te verrichten. Bone Marrow Lesion Syndrome (BMLS) beschrijft het fenomeen van verandering in signaal op vloeistofgevoelige MRI afbeeldingen in het beenmerg, en kan het gevolg zijn van een groot scala aan verschillende onderliggende aandoeningen of acute trauma. BMLS kan ook plaatsvinden in de individuele wervels in de ruggengraat. De staat van de ruggengraat en haar individuele wervels is van grote invloed op de algehele gezondheid van een patient, waardoor de juiste identificatie en characterisatie van BMLS in de wervelkolom belangrijk is. Voorheen was het nooit mogelijk om deze veranderingen in het beenmerg te observeren met CT technieken, maar door de recente opkomst van Dual Energy CT (DECT) systemen is het mogelijk ook met een andere modaliteit dan MRI, BMLS in kaart te brengen. Dit onderzoek maakt een uiteenzetting van verscheidene segmentatiemethoden voor de ruggenwervels en vervolgens individuele beenmerglaesies voor MRI en CT, om te kijken naar de mogelijkheden om het proces van de identificatie van beenmerglaesies te automatiseren.

MRI en CT verschillen fundamenteel waardoor afhankelijk van de toepassing het ene systeem een voordeel kan hebben tegenover de andere. Zo duren acquisities met MRI meestal aanzienlijk langer en zijn deze vaak van lagere resolutie. Daartegenover staat dan weer dat bij CT het contrast tussen zachte weefsels meestal lager is en er gebruik gemaakt wordt van ioniserende straling, wat de kans op de ontwikkeling van kanker verhoogt. Doordat met DECT, door het combineren van twee acquisities op verschillende energieniveaus, een virtuele subtractie van calcium uitgevoerd kan worden, is het mogelijk om ook met CT de structuur van het beenmerg te inspecteren. Laesies in het beenmerg kenmerken zich op afbeeldingen door een relatief homogeen signaal maar een gebrek aan anatomische eigenschappen of scherp gedefinieerde randen. Deze eigenschappen vermoedelijk zowel het proces van manuele als geautomatiseerde segmentatie van de laesies. Bij de manuele segmentatie van de laesies is op zowel DECT als MRI dan ook sprake van significante variabiliteit, zowel onderling tussen artsen, als tussen verschillende sessies door dezelfde arts. Variabiliteit in de manuele segmentaties bemoeilijkt ook de ontwikkeling van geautomatiseerde methoden, aangezien manuele segmentaties als gouden standaard gebruikt worden om de kwaliteit te beoordelen, en ook gebruikt worden om neurale netwerken initieel te trainen.

Concluderend is gevonden dat voor de segmentatie van de ruggenwervels, neurale netwerken de meest veelbelovende resultaten leverden, met vergelijkbare kwaliteit op zowel DECT als MRI. De techniek voor segmentatie van individuele laesies (zowel automatisch als manueel) in het beenmerg staat nog in de kinderschoenen, maar door de voordelen die DECT mogelijk biedt tegenover MRI zou het niet onrealistisch zijn dat in de toekomst DECT de voorkeur zal genieten in zowel

kwalitatief als kwantitatief onderzoek naar BMLS.