Review

Automated extraction of body composition metrics from abdominal CT or MR imaging: A scoping review

Christopher Winder [a,b,*,1], Matthew Clark [c], Russell Frood [c,d], Lesley Smith [d], Andrew Bulpitt [b,2], Gordon Cook [d,e,2], Andrew Scarsbrook [c,e,f,2]

[a] UKRI CDT in AI for Medical Diagnosis and Care, University of Leeds, Woodhouse, LS2 9JT, Leeds, UK
[b] School of Computing, University of Leeds, Woodhouse, LS2 9JT, Leeds, UK
[c] Department of Radiology, St.James University Hospital, Beckett St, Harehills, LS9 7TF, Leeds, UK
[d] CRUK Clinical Trials Unit, Leeds Institute of Clinical Trial Research, University of Leeds, Woodhouse, LS2 9JT, Leeds, UK
[e] Leeds Cancer Centre, St. James University Hospital, Beckett St, Harehills, LS9 7TF, Leeds, UK
[f] Leeds Institute of Medical Research, University of Leeds, Woodhouse, LS2 9JT, Leeds, UK

A B S T R A C T

Purpose: To review methodological approaches for automated segmentation of subcutaneous adipose tissue, visceral adipose tissue, and skeletal muscle from abdominal cross-sectional imaging for body composition analysis.
Method: Four databases were searched for publications describing automated segmentation of subcutaneous adipose tissue, visceral adipose tissue, and/or skeletal muscle from abdominal CT or MR imaging between 2019 and 2023. Included reports were evaluated to assess how imaging modality, cohort size, vertebral level, model dimensionality, and use of a volume or single slice affected segmentation accuracy and/or clinical utility. Exclusion criteria included reports not in English language, manual or semi-automated segmentation methods, reports prior to 2019 or solely of paediatric patients, and those not describing the use of abdominal CT or MR.
Results: After exclusions, 172 reports were included in the review. CT imaging was utilised approximately four times as often as MRI, and segmentation accuracy did not significantly differ between the two modalities. Cohort size had no significant effect on segmentation accuracy. There was little evidence to refute the current practice of extracting body composition metrics from the third lumbar vertebral level. There was no clear benefit of using a 3D model to perform segmentation over a 2D approach.
Conclusion: Automated segmentation of intra-abdominal soft tissues for body composition analysis is an intense area of research activity. Segmentation accuracy is not affected by cross-sectional imaging modality. Extracting metrics from a single slice at the third lumbar vertebral level is a common approach, however, extracting metrics from a volumetric slab surrounding this level may increase the resilience of the technique, which is important for clinical translation. A paucity of publicly available datasets led to most reports using different data sources, preventing direct comparison of segmentation techniques. Future efforts should prioritise creating a standardised dataset to facilitate benchmarking of different algorithms and subsequent clinical adoption.

1. Introduction

Body composition (BC) has been shown to affect both drug efficacy and numerous patient outcomes, including overall survival [1–3]. Accurate measurement of BC may help provide non-invasive assessment of frailty and guide more personalised treatment decisions. Various metrics have been used to quantify differences in BC, the vast majority of which

involve the measurement of subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and/or skeletal muscle (SM) volumes or areas. Although this task can be performed manually, it is extremely time-consuming, which has been a barrier to its translation into clinical practice. The automation of this process can be accomplished by a computer vision process known as semantic segmentation, whereby every pixel/voxel of an image is classified. In the case of BC analysis,

each voxel of a scan could be classified as SAT, VAT, SM, or background/other tissue, allowing for measurement of the volumes, cross-sectional areas, and/or densities of these tissues.

Many decisions must be made when extracting BC metrics, one of which is measurement location. Ideally, BC metrics would be extracted from the entire body, as this would be the most comprehensive method. However, due to the laborious nature of manual segmentation, these measures are typically extracted from a single axial slice at the third lumbar (L3) vertebral level, as this has been shown to correlate with the composition of the entire body [4]. One potential flaw with a single slice method is that both varying spinal curvature and observer variation mean that the location of any single slice is likely inconsistent between subjects. Additionally, the positioning and distension of the bowel are likely to vary between subjects and time points. Consequently, two slices extracted from the same location but taken from different subjects or time points would likely contain differing proportions of bowel. A large amount of bowel visible in a slice would reduce the amount of VAT present in that slice, but this reduction would not apply to the entire body. In reality, subjects are all different and have varied body shapes and sizes. It therefore seems intuitive that a multi-slice section (volumetric slab) would better estimate individual body habitus in comparison to any single slice. Measuring a volume effectively increases the sample-size of the measurement, and potentially averages over any measurement error or anatomical variation. However, this may reduce the number of scans that can be segmented, as they would need to cover a larger area of the torso.

Another key decision is the dimensionality of the segmentation model. Although it seems intuitive to use a 2D model to segment a single slice and a 3D model to segment a volume, this is not necessary, as multiple 2D segmentations can be stacked to segment a volume. Conversely, provided that adjacent slices are available, a 3D model can be used to predict the composition of a single slice. In the case of convolutional neural networks, a commonly used method for semantic segmentation, 2D models require far fewer parameters than their 3D counterparts. This means that a 2D model is less prone to overfitting than its 3D equivalent, while requiring less computational power to train. However, 2D models do not allow for any knowledge-gain from the 3D structure. For example, if a voxel is surrounded by neighbours with attenuation values matching adipose tissue (AT), it is more likely that the voxel itself is AT rather than noise. By examining the data in three dimensions rather than two, the number of neighbouring voxels increases. This, in turn, may increase the confidence in determining whether the voxel belongs to any given class.

Previous work has compared the various imaging techniques which can be used to analyse BC [1,5,6]. Two reports found that CT and MR images provide the most accurate, specific, and comprehensive data. A third found that dual-energy X-ray absorptiometry (DXA) was preferred for muscle measurement. However, as DXA scans produce 2D images, they cannot be used for SAT or VAT segmentation, leaving CT and MRI as the only accurate methods for the extraction of SAT, VAT, and SM volumes. Additionally, these two modalities are routine in many clinical pathways, and as such BC metrics can be extracted without requiring additional scans.

This work aims to determine how the above factors affect both segmentation accuracy and clinical utility. Specifically, this review will aim to answer the following questions:

- Which tissues are most commonly segmented, and are any tissues more difficult to segment than others?
- Are BC metrics more commonly extracted from CT imaging or MRI, and does the modality affect segmentation accuracy?
- How much data is needed to train an accurate segmentation model?
- Which vertebral level should BC metrics be extracted from?
- Is it beneficial to extract BC metrics from a volume rather than a single slice?

- Is it worthwhile to use a 3D model for segmentation despite the added complexity compared to a 2D model?

2. Methodology

2.1. Information Source

Four databases were searched: Embase, Ovid Medline, Scopus, and Web of Science. These were chosen to include publications from both computing and medical fields. All databases were searched from 1st January 2018 to 23rd May 2023, providing a near four and a half year time window for review. Additionally, the reference lists of any literature reviews returned by the search were scrutinised to identify any missed reports, which were subsequently added to the main set.

2.2. Search Strategy and Exclusion Criteria

The list of search terms contained three broad topics: segmentation, CT/MRI, and BC. The segmentation topic included synonyms such as *'semantic labelling'* and common methods that may be used to perform the task, such as *'neural network'* and *'u net'*. The CT/MRI topic ensured that the segmentation was performed on either MR or CT images, containing various terms for the two imaging modalities. Finally, the BC topic ensures that SAT, VAT, and/or SM within the abdomen are segmented. This topic was more complex than the segmentation or CT/MRI topics, and hence required two approaches. Firstly, location-ambiguous terms such as *'adipose tissue'* or *'body composition'* were required to appear along with a location such as *'abdomen'* or *'L3'*. These were then combined with terms denoting fat or muscle that are located within the abdomen, such as *'psoas'*. The list of search terms contained within each topic and the rules used to combine them are presented in Fig. 1.

Additionally, two of the databases, Embase and Ovid Medline, contained trees of subject headings. These subject headings were matched to the existing search terms and added to the searches of their respective databases. Two terms were limited or excluded from the search. The term *'algorithm'* was excluded from the segmentation topic if it appeared in the context *'reconstruction algorithm'* as all CT scans go through this process. The term 'computer assisted diagnosis' was removed from the segmentation topic as its Embase subject heading contained *'computer assisted tomography'* as a narrower term, which duplicates terms from the CT/MRI topic, bypassing the necessity of the segmentation topic. When searching Embase and Ovid Medline, the keyword search field was used which combines multiple fields including title, abstract, and heading word. For Scopus, the title, abstract, and keywords were searched. For Web of Science, the topic search field was used which searched publications by title, abstract, author keywords, and keywords plus. The exact queries used to search each database are included in Appendix A.

To be included in this review, a study needed to automatically segment SAT, VAT, and/or SM in the abdomen using CT or MR images. The eight exclusion criteria which were utilised are summarised in Table 1 and were applied by a single author. However, any uncertainties were discussed with multiple authors.

The review process followed the steps outlined in the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [7]. Firstly, the results from each database were combined and semi-automatically deduplicated. Next, titles and abstracts were screened to eliminate irrelevant results and literature reviews. Reports were then sought for retrieval, excluding unavailable reports or records that were not full reports, such as conference posters. The retrieved reports were assessed for eligibility using the previously discussed exclusion criteria. This process was repeated for the reference lists of any literature reviews identified in the search, with the relevant results added to the main set of reports.
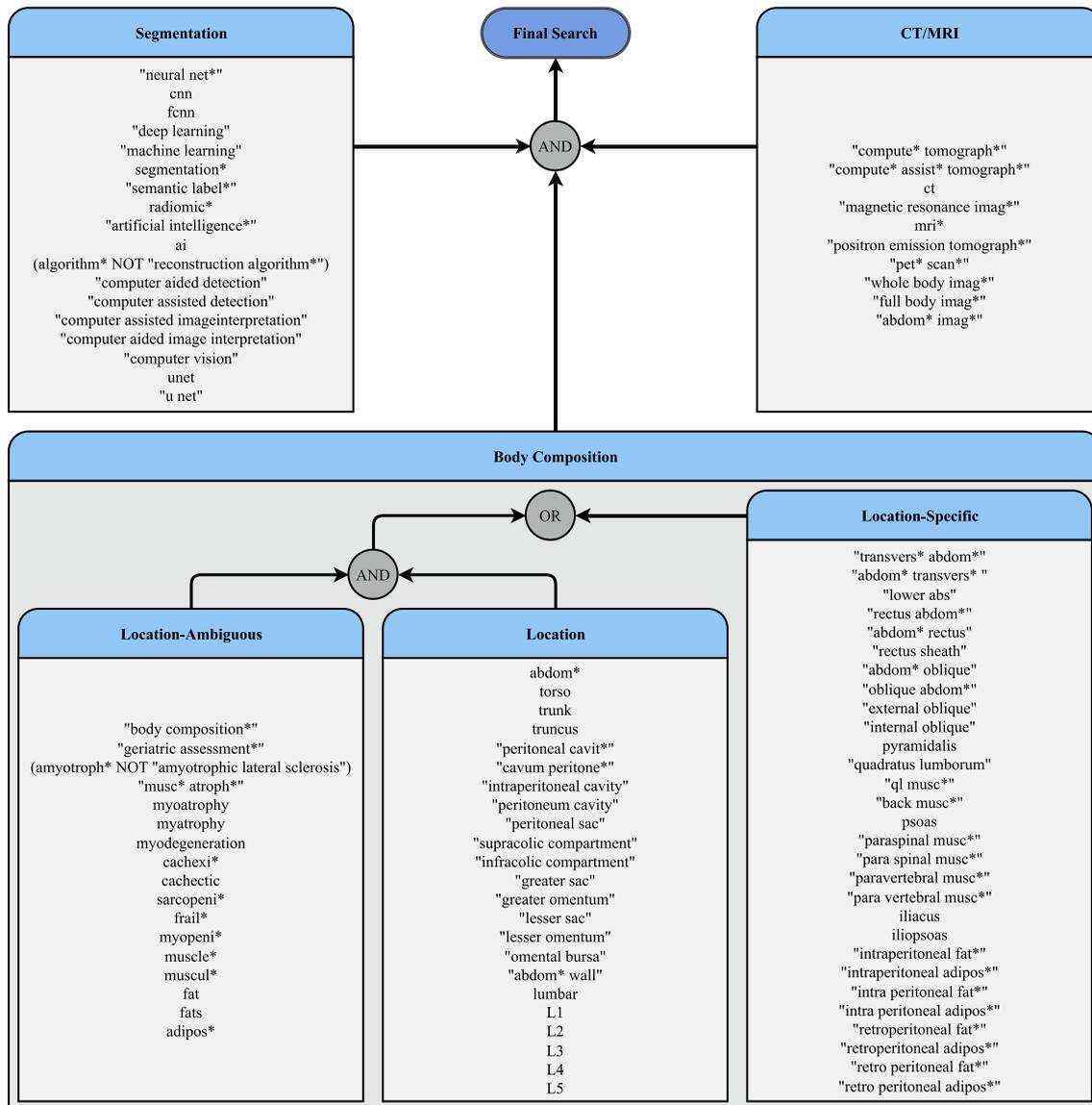
Fig. 1. Search terms used for this literature review, and the rules used to combine them.

2.3. Data collection process and analysis

Data was extracted from the selected reports by a single author and entered into a table for further analysis. The data items collected from each report are shown in Table 2. Categorical data items, such as segmented tissue types and image modalities, were plotted against the number of reports to identify the most frequent values. When applicable, the distributions of Dice similarity coefficients (DSCs) were compared both visually using box-plots and statistically using Mann–Whitney U tests to explore potential relationships with segmentation accuracy. For numerical data items, such as cohort size, histograms were used to represent their distribution. To evaluate any potential relationships between numerical data items and segmentation accuracy, scatter plots and Pearson correlation coefficients (PCCs) were analysed.

Additionally, reports were reviewed to identify any findings that directly address the specific questions this review aims to answer. Any such findings were summarised and included in this review.

3. Results

3.1. Paper selection

The search yielded 2,975 records, of which 30 were immediately removed using database filters as they were not written in English. Semi-automated duplicate detection eliminated 1,654 records, leaving 1,291 for title/abstract screening, which excluded a further 855. Of the remaining 436 records, 102 were either not available or only the abstract was available. In the latter case, it was usually either a poster or presentation abstract. Finally, the remaining 334 reports were assessed for eligibility. 127 were removed as they used either manual or semi-automated segmentation, and the focus of this search was on automated techniques. 26 reports did not segment SAT, VAT, or SM, and one used DXA as its imaging modality instead of CT or MRI and was hence excluded. 11 reports performed segmentation in areas other than the abdomen, resulting in a total of 169 reports. Additionally, the search returned ten literature reviews which were assessed, revealing three missed reports. This step is expanded upon in Appendix B. These three reports were added to the existing search records, yielding a final total of 172 reports. The PRISMA diagram showing this process is illustrated in

Table 1

The exclusion criteria utilised in this review, along with the reason for this exclusion and any clarifications.

Exclusion criteria	Exclusion reason and clarifications
Reports published prior to 2019	These were excluded as this review had a broad scope and covered many reports. 2019 was chosen as a cutoff in order to focus on state-of-the-art techniques.
Manual or semi-automated segmentation	As this review is focused on techniques used to extract body composition (BC) metrics rather than their application, any studies extracting BC metrics with manual intervention were excluded. This did not include the selection of the slice(s) of the image from which to extract BC metrics; there were no limitations on how studies perform this step.
No relevant segmentation	This review covers the measurement of subcutaneous adipose tissue, visceral adipose tissue, and skeletal muscle. Therefore, any studies which did not segment at least one of these tissues were excluded. Studies which segmented a subset of these tissues, such as the psoas muscle (a subset of skeletal muscle), were included.
Not CT/MRI	As discussed previously, CT and MR images are the ideal choices for the extraction of BC metrics.
Not abdominal	BC metrics may be extracted from areas other than the abdomen, however this review excludes any work outside this region. Studies which utilised larger areas were included if the abdomen was part of this region.
Paediatrics	Due to vast physiological variations, conclusions drawn from paediatric studies (subjects under 18 years old) may not apply to an adult population and vice versa. As BC analysis of adult populations is far more common, any studies utilising a paediatric population were excluded.
Literature reviews	Any records which were literature reviews were excluded. However, any reports included in these reviews were assessed for eligibility in this review.
Reports not written in English	Out of necessity, any reports not written in the English language were removed.

Table 2

Data items collected from each report along with a description of each item.

Data item	Description
Location of segmentation	The anatomical location from which BC metrics were extracted, recorded as free text.
Vertebral level(s)	Whether each of the first to fifth lumbar (L1-L5) vertebral levels were included in the segmentation. Each vertebra is only flagged if explicitly confirmed, or a volume overlapping these vertebrae is segmented.
Imaging modality	Whether CT or MR images were utilised.
Method	The method used to perform segmentation.
Supervision	The type of supervision used to train the model, including whether the model was pre-trained.
Cohort size	The number of subjects in the study.
Training cohort size	The number of subjects used to train the model.
Testing cohort size	The number of subjects used to assess the performance of the model.
Model dimensionality	The dimensionality of the model used to perform segmentation.
Tissue(s) Segmented	Whether subcutaneous adipose tissue, visceral adipose tissue, and/or skeletal muscle were segmented along with any other relevant tissues.
Dice Similarity Coefficient (DSC)	The DSC of each segmented tissue.

Fig. 2, and the complete list of reports and extracted data is available in the supplementary material.

3.2. DSC Reporting

Of the 172 reports included in this review, 80 (47%) included the DSC(s) of their segmentation(s). The remaining 92 (53%) did not include any DSCs, with 64 of these utilising pre-trained models.

3.3. Tissues Segmented

The number of reports segmenting each tissue are shown in Fig. 3, showing that SAT, VAT, and SM were the most commonly segmented tissues by far, utilised in 68%, 66%, and 59% of reports, respectively. The additional segmented tissues were the psoas/iliopsoas muscle, inter/intramuscular adipose tissue, paraspinal muscles, other muscles, and fat.

Fig. 4 shows the distribution of each tissue with five or more reported DSCs. The DSC distributions resulting from the segmentation of SAT, VAT, and SM have median DSCs of 0.961, 0.943, and 0.949, respectively. The Mann–Whitney U tests, comparing the DSC distribution of SAT to those of the other tissues, produce significant p values at the 5% level. As the distributions share a similar shape, these findings suggest that the median DSC of SAT is significantly higher than that of the other tissues. The median DSC of VAT and SM did not significantly differ at the 5% level.

3.4. Image Modality

The vast majority of reports (137, 80%) utilised CT images and 36 (21%) used MR images. There was a single report which utilised both CT and MR images. Fig. 5 illustrates the distribution of the reported DSC by modality and tissue. Only a single study utilising MR images reported the DSC of SM segmentation, however the reported DSC of 0.83 was lower than any DSC resulting from SM segmentation of CT images. The median DSC resulting from SAT segmentation of MR images is lower than those utilising CT images, but the opposite is true for VAT segmentation. The Mann–Whitney U tests show that neither of these differences are significant at the 5% level.

3.5. Cohort Size

Fig. 6 demonstrates the number of subjects utilised in each report, ranging from 10 to 19,766. It shows that 78 reports utilised fewer than 150 subjects, and 71 included more. The remaining 23 reports did not include the number of subjects used, instead usually reporting the number of scans. The term *'scans'* was used differently across reports; many used it to refer to a single slice, meaning that multiple scans could be taken from a single 3D volume, whereas others used the term without definition. Due to this ambiguity, only the number of subjects is analysed, as the term is unambiguous.

Many of the studies utilising large cohorts used pre-trained models, with no assessment of segmentation accuracy on their dataset. As one aim of this review was to assess the number of subjects needed to train an accurate segmentation model, such studies are discarded from this analysis. Fig. 7 shows only the number of subjects used to train models, showing far smaller populations than those presented in Fig. 6.

Fig. 8 shows the DSC for each tissue segmented against the number of subjects used to train the model. The Pearson correlation coefficient for each tissue ranges between 0.09 and 0.29, demonstrating no significant correlations at the 5% level.

3.6. Segmentation Location

A plot showing the number of reports extracting BC metrics from each vertebral level is presented in Fig. 9, showing that the vast majority of reports (81%) utilise the L3 vertebral level. The second most common was the fourth lumbar (L4) vertebral level used in 39% of reports, with the remaining lumbar vertebral levels having approximately equal usage.

Additionally, three reports were found comparing differing vertebral levels. A study by Hong et al. aimed to compare the correlation between volumes of SAT, VAT, and SM extracted at a single slice and those measured using the entire body [8]. Their results are shown in Table 3, showing that although the best single slice for SM and VAT was
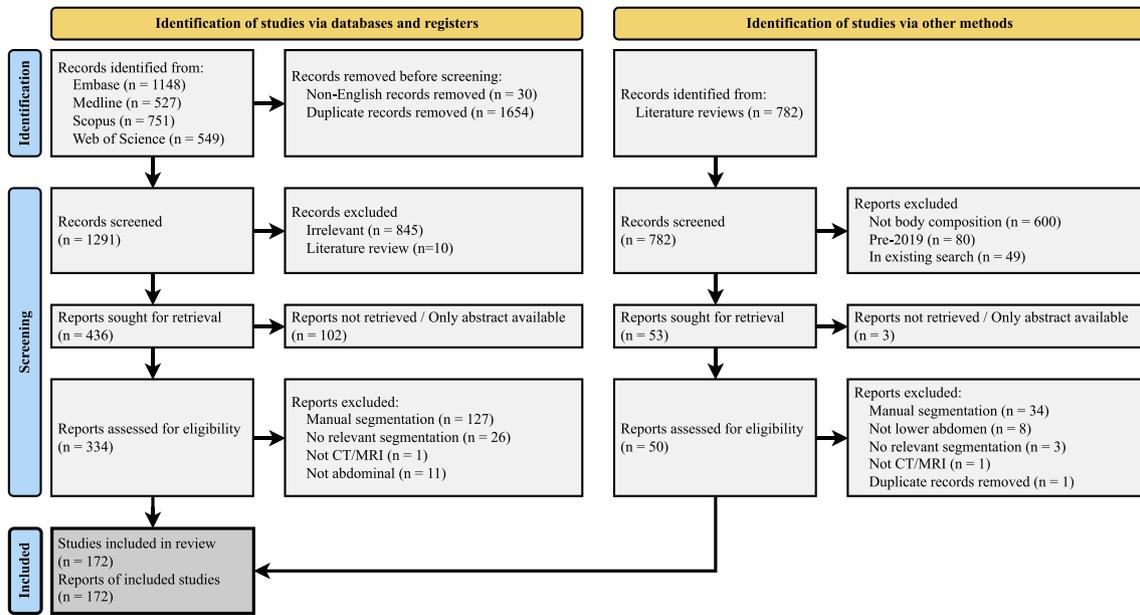
Fig. 2. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 2020 flowchart showing the screening process of the literature review.
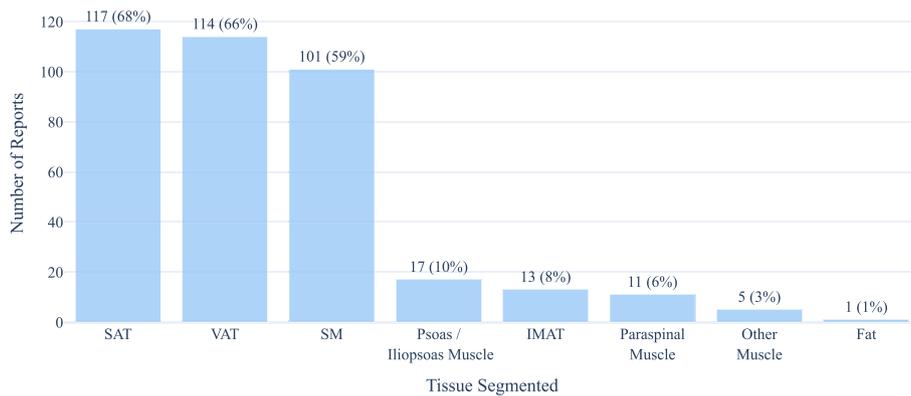
Fig. 3. Bar chart showing the number of reports segmenting each tissue. The tissues included are subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), skeletal muscle (SM), psoas/iliopsoas muscle, intermuscular adipose tissue (IMAT), paraspinal muscle, other muscle, and fat.
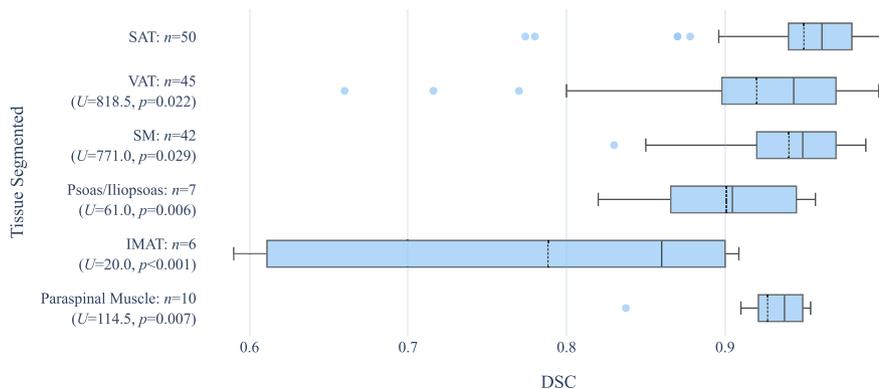
Fig. 4. Box plots showing the distribution of the Dice similarity coefficient (DSC) of each tissue segmented. The tissues included are subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), skeletal muscle (SM), psoas/iliopsoas muscle, intermuscular adipose tissue (IMAT), and paraspinal muscle. The median DSC is shown by the solid vertical line within each box, and the mean is shown by the dashed vertical line. Only tissues with five or more reported DSCs are shown. Additionally, the Mann–Whitney U test statistics (U) and associated p values are shown on the left, comparing the DSC distribution of each tissue to that of SAT. Although not shown, an additional Mann–Whitney U test was conducted to compare the DSC distributions of VAT and SM, showing that the two distributions did not significantly differ at the 5% level ($U = 896.5$, $p = 0.683$).
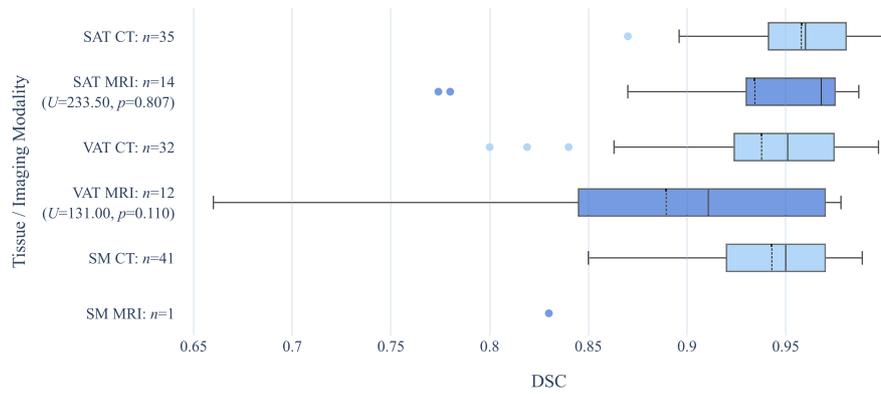
Fig. 5. Box plots showing the distribution of the Dice similarity coefficient (DSC) of each commonly segmented tissue by imaging modality. The tissues included are subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and skeletal muscle (SM). The median DSC is shown by the solid vertical line within each box, and the mean is shown by the dashed vertical line. Additionally, the Mann–Whitney U test statistics (U) and associated p values are shown on the left, comparing the median DSC resulting from segmentation of both CT and MR images. Please note that only a single DSC for SM segmentation via MR imaging was reported. Hence, this is shown as a single point rather than a box plot and the Mann–Whitney U test is not performed.
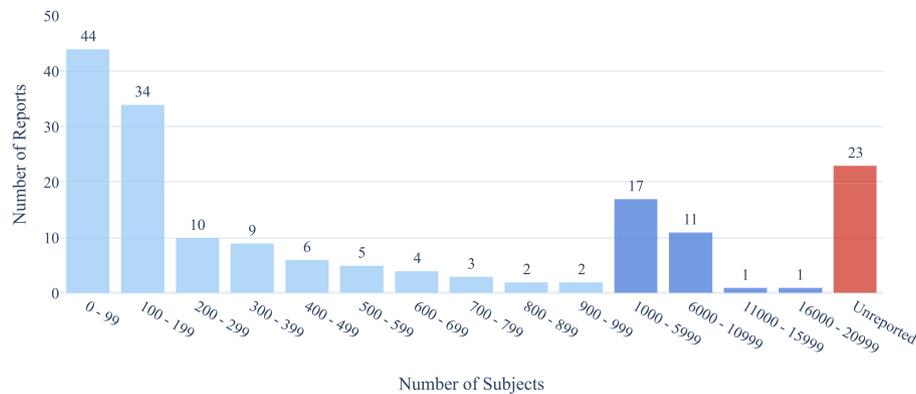
Fig. 6. Histogram showing the number of subjects in each report. Bins which are 100 subjects wide are shown in light blue, whereas those shown in dark blue are 5000 subjects wide. The number of reports which did not specify the number of subjects used is shown in red.
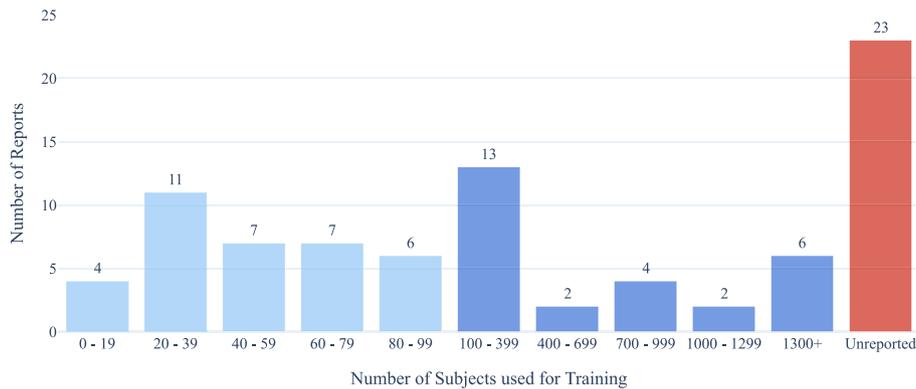
Fig. 7. Histogram showing the number of subjects used to train each model. Bins which are 20 subjects wide are shown in light blue, whereas those shown in dark blue are 300 subjects wide, including an overflow bin. The number of reports which did not specify the number of subjects used is shown in red.

extracted somewhere in the level of the second to fourth lumbar vertebrae, they did not differ significantly from those extracted at the L3 vertebral level. This is not the case for SAT, where metrics extracted from the fifth lumbar vertebral level had significantly higher correlation with the entire body than those extracted from the L3 vertebral level.

In a study by Liu et al., the predictive value of SAT/VAT volumes extracted at the first lumbar (L1) and third lumbar(L3) vertebral levels were compared by assessing mortality in 9066 asymptomatic adults [9]. They reported that volumes extracted at either level were comparable,

but the predictive power of the ratio of visceral to subcutaneous AT was significantly higher at the L3 vertebral level. In subsequent work by some of the same authors, Pickhardt et al. compared the ability of SM volume and mean attenuation to predict hip fractures and mortality in an asymptomatic population of 9,223 adults, evaluating metrics extracted at the L1 and L3 vertebral levels [10]. Their findings indicated that metrics from both locations performed similarly for most outcomes. However, SM volumes from the L3 vertebral level demonstrated a significantly higher area under the receiver operating characteristic
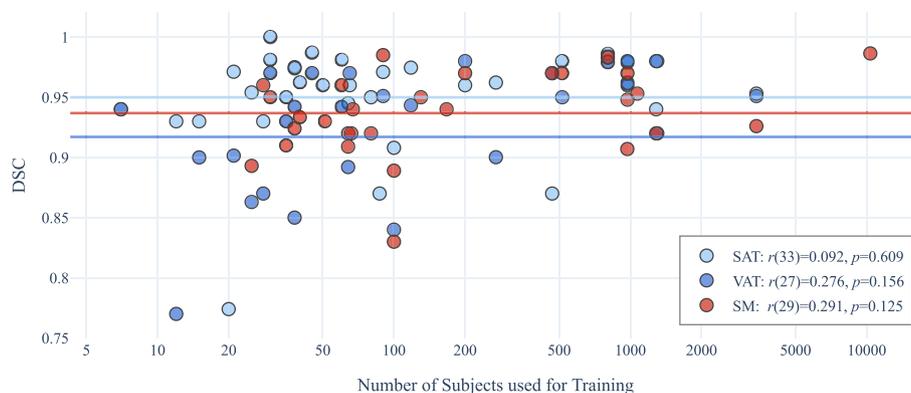
Fig. 8. Scatter plot showing the Dice similarity coefficient (DSC) against training size by tissue. The tissues included are subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and skeletal muscle (SM). The Pearson correlation coefficients (r) and associated p values are shown on the bottom right, showing no significant correlations. The mean DSC for each tissue is shown by the horizontal lines. Please note that a single outlier has been removed for clarity, a study segmenting VAT with a reported DSC of 0.66.
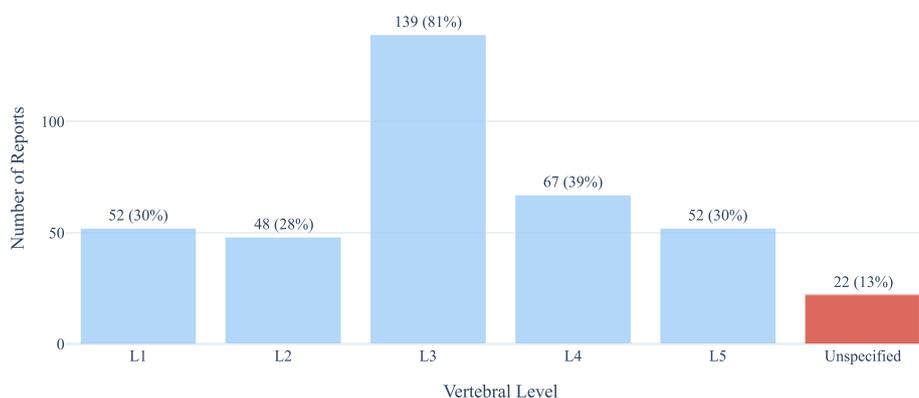
Fig. 9. Bar chart showing the number of reports extracting body composition metrics from each of the first to fifth lumbar (L1-L5) vertebral levels. Please note that any given paper may utilise multiple vertebral levels, and hence the total number of reports exceeds the number returned by the search.

Table 3
Comparison of correlation between subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and skeletal muscle (SM) volumes extracted from single/multiple slices versus the entire body. Measurements were taken from the twelfth thoracic (T12) to the first lumbar (L1) vertebral levels. Values in the L3 vertebra column were measured at the centre of the vertebra, whereas the best single slice(s) column includes measurements from any location within the vertebra. The highest Pearson correlation coefficient (PCC) per tissue/sex is highlighted in bold, along with the associated confidence interval (CI). Results are taken from Hong et al., 2023 [8].

Tissue	Sex	L3 Vertebra		Best Single Slice(s)			Multi-Slice (T12-L1)	
		PCC	CI	Vertebra(e)	PCC	CI	PCC	CI
SAT	F	0.84	0.78–0.89	L5	**0.91**	0.87–0.94	**0.91**	0.88–0.94
	M	0.89	0.84–0.93	L5	**0.94**	0.90–0.96	0.91	0.85–0.94
VAT	F	0.97	0.95–0.98	L2-L3	0.97	0.96–0.98	**0.98**	0.97–0.99
	M	0.97	0.96–0.98	L2	**0.98**	0.96–0.99	0.97	0.95–0.98
SM	F	0.85	0.78–0.90	L3	0.86	0.80–0.91	**0.88**	0.82–0.92
	M	0.86	0.78–0.92	L4	0.87	0.81–0.91	**0.91**	0.87–0.93
All	-	0.90	0.87–0.92	L2-L3	0.90	0.88–0.92	**0.92**	0.92–0.95

curve compared to those from the L1 vertebral level in predicting hip fractures within ten years of the scan. Mean SM attenuation outperformed SM volume measured at either location, with mean attenuation extracted from both levels being nearly identical.

3.7. Single Slice vs 3D Volume

Of the 172 reviewed reports, approximately one third (58) segmented a 3D volume, whereas 112 did not. Additionally, two reports did not specify whether they segmented a 3D volume.

Two reports directly compared metrics extracted from a single slice with those extracted from a larger volume. The previously discussed

analysis by Hong et al. shown in Table 3 compares the correlation between metrics extracted from single and multi-slice volumes with the composition of the entire body [8]. The multi-slice metrics consistently had higher correlation with the composition of the entire body than those extracted at the L3 vertebral level. Additionally, they usually outperformed the best single slice, only failing to do so in two cases.

Anyene et al. compared BC metrics measured from the L3 slice to those extracted from a volume spanning from the twelfth thoracic (T12) to the fifth lumbar (L5) vertebral levels [11], finding a high degree of correlation for all tissues. However, they found that the two metrics differed significantly in 5.5–6.4% of the population, varying by tissue. They noted that these outliers had a higher average BMI in comparison

to non-outliers. They additionally used these metrics to analyse all-cause mortality in colorectal cancer patients, finding no significant differences between single-slice and multi-slice metrics.

3.8. Model Dimensionality

2D models were over five times more common than 3D models, as shown in Fig. 10. Additionally, six reports utilised 2.5D models, i.e., models which utilised two or three anatomical planes simultaneously.

Fig. 11 compares the distribution of the reported DSCs of SAT, VAT, and SM segmentation using both 2D and 3D models. The Mann–Whitney U tests show no significant difference in the segmentation of SAT when using the two methods. For VAT and SM this is not the case, with both Mann–Whitney U tests producing significant results at the 5% level. Combined with the plotted distributions, this suggests that the median DSC achieved by 2D models is higher than that achieved by 3D models.

The literature search identified three reports comparing the use of 2D and 3D models. In the first paper, Liu. et al. compared 2D and 3D implementations of their model for the prediction of SM, SAT, and VAT volumes in the entire torso [12]. The second paper, by Lee et al., performed the same task but with segmentation of the entire body [13]. The third paper, by Wesselink et al., aimed to segment six paraspinal muscles [14]. Their results are summarised in Table 4, illustrating that two of the three reports found that 3D models outperformed their 2D counterparts.

4. Discussion

4.1. Search coverage

This work presents an extensive scoping review capturing an intense area of research at the interface of medical imaging and scientific computing. It covers a broad scope and hence evaluated many studies. Additionally, the analysis of existing literature reviews revealed only three reports which were not identified using the search strategy presented in Fig. 1. This small number suggests that the search was comprehensive and likely missed very few relevant reports.

4.2. DSC reporting

DSCs are widely used to compare the agreement between predicted segmentations and ground truths, but surprisingly, this measure was only reported in approximately 50% of the reports. Although other measures of segmentation accuracy such as the Jaccard index were occasionally used, this only occurred in a handful of reports as the DSC is the most commonly used measure. Most reports which did not include the DSC utilised pre-trained models with no quantitative assessment of segmentation accuracy. In some cases, the model was either trained on or previously evaluated on their exact dataset, but this was rare. The remainder assumed that previous results would apply to their data,

which may be problematic, as minor variations in scanning parameters or patient characteristics, not present in the original training data, could lead to inaccurate segmentations. Although a qualitative assessment of results was likely performed, this may be insufficient, as it can be hard to spot minor mistakes when checking numerous scans. When utilising pre-trained models, a small subset of the data should be manually labelled to quantitatively assess the model's performance and compare it to existing results.

4.3. Tissues segmented

SAT, VAT, and SM were the most commonly segmented tissues by far, segmented in over six times as many reports as the psoas, the next most common tissue. Comparing the distributions using Mann–Whitney U tests showed that the median DSC of SAT segmentation is higher than those resulting from VAT and SM, suggesting that it is the easiest tissue to segment. This is unsurprising as it is the most geometrically simple shape, loosely consisting of a cylinder surrounding the abdomen. By comparison, SM and VAT are more complicated. SM is a similar shape to SAT, but with the added complexity of the paraspinal musculature. VAT is the most geometrically complex, as it contains multiple cavities where other tissues such as the bowel may lie. This suggests that extra care should be taken when segmenting VAT and SM in comparison to SAT.

4.4. Image modality

This review found that BC metrics were far more commonly extracted from CT imaging in comparison to MRI. This is likely because CT is more widely performed than MRI due to a greater number of scanners and faster acquisition/throughput. Overall, the choice of imaging modality is likely to be determined by the clinical pathway, as many will involve routine diagnostic imaging. Even if there were a preferred modality which could be used to acquire significantly more accurate BC metrics, it would be more efficient to use routine scans rather than expending extra resources to acquire a second image. Fortunately, there appears to be no significant difference in the accuracy of segmentations of CT or MR images, and as such there is no reason to expend these extra resources. One exception to this is the segmentation of SM, as there was only a single study performing this task, which performed poorly in comparison to those utilising CT images. Developing a toolkit capable of extracting a full suite of accurate BC metrics from MR images would be beneficial, as many patients undergo MRI scans.

4.5. Cohort size

When first examining the results presented in Fig. 8, it was surprising to see no correlation between cohort size and DSC as it seems intuitive that more training data would result in a more accurate model. One possible explanation is that smaller cohorts likely contain less variation,
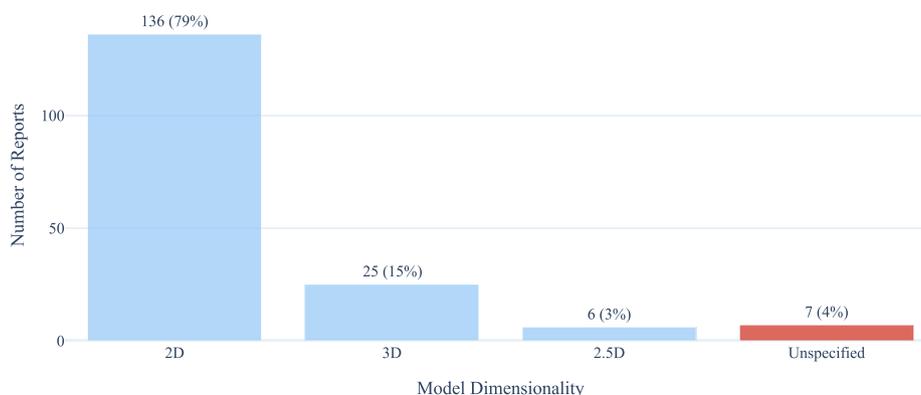
Fig. 10. Bar chart showing the number of reports extracting body composition metrics using 2D, 3D, or 2.5D models.
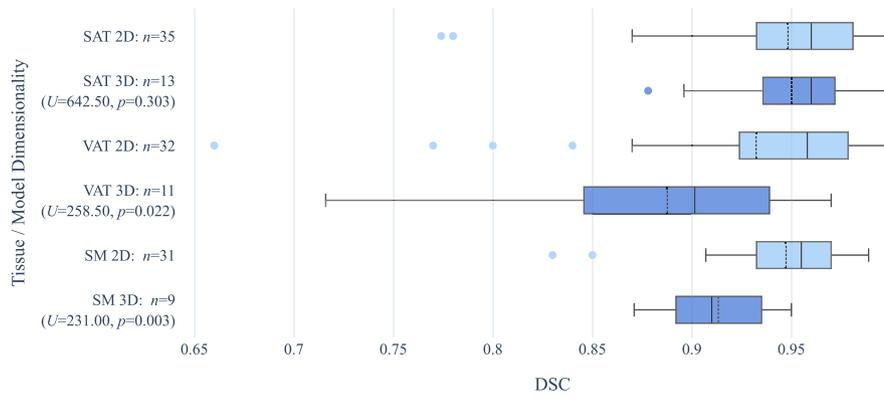
Fig. 11. Box plots showing the distribution of the Dice similarity coefficient (DSC) of each commonly segmented tissue by model dimensionality. The tissues included are subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and skeletal muscle (SM). The median DSC is shown by the solid vertical line within each box and the mean is shown by the dashed vertical line. Additionally, the Mann–Whitney U test statistics (*U*) and associated *p* values are shown on the left, comparing the median DSC resulting from segmentation using both 2D and 3D models.

Table 4
A comparison of three reports extracting body composition metrics in 3D using both 2D and 3D methods. Two reports segmented subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and skeletal muscle (SM), while the third report segmented six paraspinal muscles. For brevity, the mean Dice similarity coefficient (DSC) of these paraspinal muscles is reported and hence no standard deviations (SDs) are shown. Two reports employed U-Net models, whereas the third used a dual-pathway deep dilated convolutional neural network (DDD-CNN). The best performing dimensionality for each model is highlighted in bold. Metrics reported for Liu et al. are taken from their internal dataset to allow for a fair comparison, as the other two reports did not use external datasets.

Paper	Model	Dim	SAT		VAT		SM		Paraspinal Muscles	
			DSC	SD	DSC	SD	DSC	SD	DSC	SD
Liu 2019 [12]	DDD-CNN	2D	0.898	0.147	0.732	0.143	0.824	0.061	-	-
		3D	**0.954**	0.039	**0.863**	0.062	**0.893**	0.036	-	-
Lee 2021 [13]	U-Net	2D	0.961	0.031	0.905	0.078	0.972	0.010	-	-
		3D	**0.971**	0.031	**0.951**	0.049	**0.981**	0.006	-	-
Wesselink 2022 [14]	U-Net	2D	-	-	-	-	-	-	**0.921**	-
		3D	-	-	-	-	-	-	0.916	-

as larger cohorts likely require data to be collected from multiple centres, patients of differing conditions, and/or longer periods of time. These may result in a wider range of scanning parameters to be used, and more variety in the subjects being scanned. Conversely, small populations may be single centre studies of a small homogeneous population, leading to a simpler segmentation task. Analysis of cohort sizes was limited by the varied and often ambiguous ways this figure was reported. Future work should ideally report three figures, the number of subjects, volumes, and slices, as these terms are unambiguous.

4.6. Segmentation location

As expected, an overwhelming majority of reports utilised the L3 vertebral level for their segmentation, as this has been the status quo for many years. The report by Hong et al. showed that although the L3 vertebral level was not necessarily the best choice for all tissues, the improvement gained by using an alternate vertebral level was usually insignificant [8]. They showed that for work segmenting all tissues, either the second lumbar (L2) or L3 vertebral level was the optimum choice. Two other studies compared the clinical predictive value of BC metrics extracted from the L1 and L3 vertebral levels, finding that although there were usually no significant differences between the two measures, there were some instances where the metrics extracted from the L3 vertebral level were better [9,10]. Combining this information, there is little evidence to refute the practice of extracting BC metrics from the L3 vertebral level.

4.7. Single slice vs 3D volume

This review found that the vast majority of reports did not segment a volume, and instead extracted BC metrics from a single axial slice.

Reports directly comparing the two found minimal differences between BC metrics extracted in either manner. However, the point made by Anyene et al. regarding outliers is an important one [11]. The similar performance of the two measures is generally assessed across a population, but if a model were to be used for personalised recommendations, the number of outliers is likely a more important factor than mean segmentation accuracy. We advise that any studies aiming to implement BC models in clinical practice should segment a volumetric slab surrounding the L3 vertebral level, and that any outliers should be analysed and reported. However, for large population studies, a single slice is likely sufficient.

4.8. Model dimensionality

As shown in Fig. 11, employing a 3D model appears to either offer no benefits or significantly impairs segmentation accuracy in comparison to a 2D model. This finding disagrees with two of the three reports which directly compared the two methods. These three reports may provide more conclusive evidence than those presented in Fig. 11, as they directly compare the same model architectures on the same datasets. As 2D models are easier to train than their 3D counterparts, it is likely that a 2D model is the first to be used on any segmentation task. Consequently, a 3D model might only be used on more difficult datasets where a 2D model cannot produce accurate segmentations, biasing the results. Hence, any future studies should begin by using a 2D model. If this proves to be insufficient, delving into the added complexity of a 3D model may be warranted.

4.9. Issues and recommendations

The main difficulty in this review is the fact that these reports are all

utilising different datasets, preventing direct comparison of segmentation techniques except for a few reports. At the time of writing, we are aware of no public datasets containing SAT, VAT, and SM. This, along with the previously discussed issues with current work and recommendations for future work are summarised in Table 5.

4.10. Limitations

This review has several limitations. Firstly, reports published prior to 2019 were excluded for pragmatic reasons and an aim to highlight more recent advancements. Despite this limited time period, the search returned a substantial number of reports. Secondly, reports written in languages other than English were necessarily excluded, although this only accounted for approximately 1% of records. Lastly, a single author performed both the filtering of reports and data extraction, whereas it is common for two authors to perform this task independently, allowing them to discuss potential disagreements and reduce the number of erroneously excluded reports. This was a necessity, and to counteract this the results were compared to existing literature reviews with any missed reports added into the results. Additionally, any uncertainties were discussed with multiple authors. Finally, skeletal muscle density has recently emerged as a potentially superior predictor of mortality compared to muscle area/bulk, and these aspects were not considered as part of this scoping review [15].

5. Conclusion

This review identified and analysed recent literature describing automated segmentation of SAT, VAT, and/or SM, and used data extracted from these reports to quantify the effects of decisions made in the automatic acquisition of BC metrics. It has found that SAT, VAT, and SM are all equally prevalent in BC analysis, and that VAT and SM are harder to segment than SAT. There was little evidence to refute the status quo of extracting BC metrics from a single axial slice at the L3 vertebral level, as the composition of this slice is typically highly correlated with overall BC. However, there are edge cases where this relationship may not hold. Extracting BC metrics from a 3D volume surrounding this level would increase the sample size of the measurement and may reduce the frequency of such cases. This is crucial for clinical applications, where inaccurate metrics may result in adverse outcomes. In contrast, population studies are generally focused on average outcomes, and as such metrics extracted from a single slice are likely sufficient. There were no significant differences in the accuracy of segmentations of both CT and MR images, and hence the choice of modality should be based on data availability. Unfortunately, paucity of standardised datasets limited the analysis of both model dimensionality and the number of subjects required to train a model. Future work should prioritise creating a standard dataset to allow for the direct comparison of different methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 5
Common issues with present work and recommendations for future work.

Issue	Recommendation
Direct comparison of segmentation techniques is rare	Creation of a public dataset for subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), and skeletal muscle (SM) segmentation to allow for the direct comparison of techniques.
Dice similarity coefficients (DSCs) are rarely reported when utilising pre-trained models	A portion of the data should be manually labelled to quantitatively assess the performance of the pre-trained model using DSCs.
Little work automatically extracts VAT or SM volumes from MRI	Development of a model to extract a full suite of body composition (BC) metrics from MRI.
Ambiguous reporting of dataset sizes	Any work performing segmentation on medical images should report the number of subjects, volumes, and slices used for training, testing, and validation.
BC metrics are rarely extracted from a volume	BC metrics should be extracted from a volume if they are to be used in clinical practice. For large population studies, a single slice is likely sufficient.

Appendix A. Literature Search Strings

A.1. Embase

(("neural net*" or cnn or fcnn or "deep learning" or "machine learning" or segmentation* or "semantic label*" or radiomic* or "artificial intelligence*" or ai or (algorithm* not "reconstruction algorithm*") or "computer aided detection" or "computer assisted detection" or "computer assisted image interpretation" or "computer aided image interpretation" or "computer vision" or unet or "u net").mp. or exp *artificial intelligence/ or exp *machine learning/ or *radiomics/ or exp *algorithm/ or *computer vision/ or exp *image segmentation/ or automated biomarker*.mp.) and (((("body composition*" or "geriatric assessment*" or (amyotroph* not "amyotrophic lateral sclerosis") or "musc* atroph*" or myoatrophy or myatrophy or myodegeneration or cachexi* or cachectic or sarcopeni* or frail* or myopeni* or muscle* or muscul* or fat or fats or adipos*).mp. or exp *body composition/ or *geriatric assessment/ or exp *muscle atrophy/ or *cachexia/ or *frailty/ or exp *muscle/ or exp *adipose tissue/) and ((abdom* or torso or trunk or truncus or "peritoneal cavit*" or "cavum peritone*" or "intraperitoneal cavity" or "peritoneum cavity" or "peritoneal sac" or "supracolic compartment" or "infracolic compartment" or "greater sac" or "greater omentum" or "lesser sac" or "lesser omentum" or "omental bursa" or "abdom* wall" or lumbar or L1 or L2 or L3 or L4 or L5).mp. or *abdomen/ or *abdominal cavity/ or *trunk/)) or (("transvers* abdom*" or "abdom* transvers*" or "lower abs" or "rectus abdom*" or "abdom* rectus" or "rectus sheath" or "abdom* oblique" or "oblique abdom*" or "external oblique" or "internal oblique" or pyramidalis or "quadratus lumborum" or "ql musc*" or "back musc*" or psoas or "paraspinal musc*" or "para spinal musc*" or "paravertebral musc*" or "para vertebral musc*" or iliacus or iliopsoas or "intraperitoneal fat*" or "intraperitoneal adipos*" or "intra peritoneal fat*" or "intra peritoneal adipos*" or "retroperitoneal fat*" or "retroperitoneal adipos*" or "retro peritoneal fat*" or "retro peritoneal adipos*").mp. or exp *abdominal wall musculature/ or *back muscle/ or *paraspinal muscle/ or *iliacus muscle/ or exp *iliopsoas muscle/ or exp *abdominal fat/)) and (("compute* tomograph*" or "compute* assist* tomograph*" or ct or "magnetic resonance imag*" or mri* or "positron emission tomograph*" or "pet* scan*" or "whole body imag*" or "full body imag*" or "abdom* imag*").mp. or *computer assisted tomography/ or exp *X-ray computed tomography/ or *nuclear magnetic resonance imaging/ or *whole body imaging/ or exp *whole body tomography/ or *abdominal radiography/) [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword heading word, floating subheading word, candidate term word].

limit 1 to (english language and yr="2019–2023")

A.2. Ovid Medline

(("neural net*" or cnn or fcnn or "deep learning" or "machine learning" or segmentation* or "semantic label*" or radiomic* or "artificial intelligence*" or ai or (algorithm* not "reconstruction algorithm*") or "computer aided detection" or "computer assisted detection" or "computer assisted image interpretation" or "computer aided image interpretation" or "computer vision" or unet or "u net").mp. or exp *neural networks, computer/ or *algorithms/ or *artificial intelligence/ or exp *machine learning/) and (((("body composition*" or "geriatric assessment*" or (amyotroph* not "amyotrophic lateral sclerosis") or "musc* atroph*" or myoatrophy or myatrophy or myodegeneration or cachexi* or cachectic or sarcopeni* or frail* or myopeni* or muscle* or muscul* or fat or fats or adipos*).mp. or exp *body composition/ or *geriatric assessment/ or *frailty/ or *muscles/ or *muscle, skeletal/ or exp *Adipose Tissue/) and ((abdom* or torso or trunk or truncus or "peritoneal cavit*" or "cavum peritone*" or "intraperitoneal cavity" or "peritoneum cavity" or "peritoneal sac" or "supracolic compartment" or "infracolic compartment" or "greater sac" or "greater omentum" or "lesser sac" or "lesser omentum" or "omental bursa" or "abdom* wall" or lumbar or L1 or L2 or L3 or L4 or L5).mp. or exp *abdomen/ or *trunk/)) or (("transvers* abdom*" or "abdom* transvers* " or "lower abs" or "rectus abdom*" or "abdom* rectus" or "rectus sheath" or "abdom* oblique" or "oblique abdom*" or "external oblique" or "internal oblique" or pyramidalis or "quadratus lumborum" or "ql musc*" or "back musc*" or psoas or "paraspinal musc*" or "para spinal musc*" or "paravertebral musc*" or "para vertebral musc*" or iliacus or iliopsoas or "intraperitoneal fat*" or "intraperitoneal adipos*" or "intra peritoneal fat*" or "intra peritoneal adipos*" or "retroperitoneal fat*" or "retroperitoneal adipos*" or "retro peritoneal fat*" or "retro peritoneal adipos*").mp. or exp *abdominal muscles/ or *paraspinal muscles/ or *psoas muscles/ or exp *abdominal fat/)) and (("compute* tomograph*" or "compute* assist* tomograph*" or ct or "magnetic resonance imag*" or mri* or "positron emission tomograph*" or "pet* scan*" or "whole body imag*" or "full body imag*" or "abdom* imag*").mp. or exp *Tomography, X-ray Computed/ or exp *Magnetic Resonance Imaging/ or exp *Whole Body Imaging/) [mp = title, book title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms, population supplementary concept word, anatomy supplementary concept word].

limit 1 to (english language and yr="2019–2023")

A.3. Scopus

TITLE-ABS-KEY (("neural net*" OR cnn OR fcnn OR "deep learning" OR "machine learning" OR segmentation* OR "semantic label*" OR radiomic* OR "artificial intelligence*" OR ai OR (algorithm* AND not "reconstruction algorithm*") OR "computer aided detection" OR "computer assisted detection" OR "computer assisted image interpretation" OR "computer aided image interpretation" OR "computer vision" OR unet OR "u net") AND (((("body composition*" OR "geriatric assessment*" OR (amyotroph* AND not "amyotrophic lateral sclerosis") OR "musc* atroph*" OR myoatrophy OR myatrophy OR myodegeneration OR cachexi* OR cachectic OR sarcopeni* OR frail* OR myopeni* OR muscle* OR muscul* OR fat OR fats OR adipos*) AND (abdom* OR torso OR trunk OR truncus OR "peritoneal cavit*" OR "cavum peritone*" OR "intraperitoneal cavity" OR "peritoneum cavity" OR "peritoneal sac" OR "supracolic compartment" OR "infracolic compartment" OR "greater sac" OR "greater omentum" OR "lesser sac" OR "lesser omentum" OR "omental bursa" OR "abdom* wall" OR lumbar OR l1 OR l2 OR l3 OR l4 OR l5)) OR ("transvers* abdom*" OR "abdom* transvers* " OR "lower abs" OR "rectus abdom*" OR "abdom* rectus" OR "rectus sheath" OR "abdom* oblique" OR "oblique abdom*" OR "external oblique" OR "internal oblique" OR pyramidalis OR "quadratus lumborum" OR "ql musc*" OR "back musc*" OR psoas OR "paraspinal musc*" OR "para spinal musc*" OR "paravertebral musc*" OR "para vertebral musc*" OR iliacus OR iliopsoas OR "intraperitoneal fat*" OR "intraperitoneal adipos*" OR "intra peritoneal fat*" OR "intra peritoneal adipos*" OR "retroperitoneal fat*" OR "retroperitoneal adipos*" OR "retro peritoneal fat*" OR "retro peritoneal adipos*")) AND ("compute* tomograph*" OR "compute* assist* tomograph*" OR ct OR "magnetic resonance imag*" OR mri* OR "positron emission tomograph*" OR "pet* scan*" OR "whole body imag*" OR "full body imag*" OR "abdom* imag*")) AND (LIMIT-TO (PUBYEAR, 2023) OR LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019)) AND (LIMIT-TO

(LANGUAGE, "english")).
A.4. *Web of science*

("neural net*" OR cnn OR fcnn OR "deep learning" OR "machine learning" OR segmentation* OR "semantic label*" OR radiomic* OR "artificial intelligence*" OR ai OR (algorithm* NOT "reconstruction algorithm*") OR "computer aided detection" OR "computer assisted detection" OR "computer assisted image interpretation" OR "computer aided image interpretation" OR "computer vision" OR unet OR "u net") AND ((("body composition*" OR "geriatric assessment*" OR (amyotroph* NOT "amyotrophic lateral sclerosis") OR "musc* atroph*" OR myoatrophy OR myatrophy OR myodegeneration OR cachexi* OR cachectic OR sarcopeni* OR frail* OR myopeni* OR muscle* OR muscul* OR fat OR fats OR adipos*) AND (abdom* OR torso OR trunk OR truncus OR "peritoneal cavit*" OR "cavum peritone*" OR "intraperitoneal cavity" OR "peritoneum cavity" OR "peritoneal sac" OR "supracolic compartment" OR "infracolic compartment" OR "greater sac" OR "greater omentum" OR "lesser sac" OR "lesser omentum" OR "omental bursa" OR "abdom* wall" OR lumbar OR L1 OR L2 OR L3 OR L4 OR L5)) OR ("transvers* abdom*" OR "abdom* transvers* " OR "lower abs" OR "rectus abdom*" OR "abdom* rectus" OR "rectus sheath" OR "abdom* oblique" OR "oblique abdom*" OR "external oblique" OR "internal oblique" OR pyramidalis OR "quadratus lumborum" OR "ql musc*" OR "back musc*" OR psoas OR "paraspinal musc*" OR "para spinal musc*" OR "paravertebral musc*" OR "para vertebral musc*" OR iliacus OR iliopsoas OR "intraperitoneal fat*" OR "intraperitoneal adipos*" OR "intra peritoneal fat*" OR "intra peritoneal adipos*" OR "retroperitoneal fat*" OR "retroperitoneal adipos*" OR "retro peritoneal fat*" OR "retro peritoneal adipos*")) AND ("compute* tomograph*" OR "compute* assist* tomograph*" OR ct OR "magnetic resonance imag*" OR mri* OR "positron emission tomograph*" OR "pet* scan*" OR "whole body imag*" OR "full body imag*" OR "abdom* imag*") (Topic) and English (Languages).
Timespan: 2019–01–01 to 2023–12–31 (Publication Date).

Appendix B. Comparison with Existing Literature Reviews

The literature search returned ten literature reviews, summarised in Table B.1, whose results were compared with the previously found 169 reports to assess coverage and identify any missed reports. Most of these reviews explored either broader or narrower scopes than the work presented here, and hence had minimal overlap. Many of the reports included in these reviews were published prior to 2019, leading to their exclusion from this review. Of the 102 reports covering body composition post-2018, 49 were already included in this review and 49 were excluded or not accessed. The four remaining reports included a duplicate, meaning that there were three missed reports.

Table B.1
Existing reviews covering body composition.

Author(s)	Year	Review Topic	Papers			
			Body composition (post-2018)	In this search	Excluded or not retrieved	Missed by this search
Greco, F and Mallio, C A [16]	2021	Abdominal adipose tissue analysis	12	11	1	0
D'Antoni, F et al. [17]	2021	Lower back pain	2	1	0	1
Bedrikovetski, S et al. [18]	2022	Body composition	21	18	2	1
Lenchik, L et al. [19]	2019	Tissue segmentation	0	0	0	0
Paris, M [20]	2019	CT body composition	3	3	0	0
Elfanagely, O et al. [21]	2020	CT analysis in abdominal wall reconstruction	4	0	4	0
Meyer, H-J et al. [22]	2023	Fat as a prognostic marker in gastric adenocarcinoma	5	0	5	0
Tolonen, A et al. [23]	2021	CT body composition	36	0	35	1
Zhao, Y et al. [24]	2023	Multi-task deep learning	1	0	1	0
Huang, Y-T et al. [25]	2021	Sarcopenia	18	16	1	1

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrad.2024.111764.

References

[1] C. Yip, C. Dinkel, A. Mahajan, M. Siddique, G.J.R. Cook, V. Goh, Imaging body composition in cancer patients: visceral obesity, sarcopenia and sarcopenic obesity may impact on clinical outcome, Insights into Imag. 6 (2015) 489–497, https://doi.org/10.1007/s13244-015-0414-0.

[2] P.J. Pickhardt, Value-added opportunistic CT screening: State of the art, Radiology 303 (2022) 241–254, https://doi.org/10.1148/radiol.211561.

[3] P.J. Pickhardt, P.M. Graffy, R. Zea, S.J. Lee, J. Liu, V. Sandfort, R.M. Summers, Automated ct biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study, The Lancet Digital Health 2 (2020) e192–e200, https://doi.org/10.1016/s2589-7500(20)30025-x.

[4] W. Shen, M. Punyanitya, Z. Wang, D. Gallagher, M.-P. St.-Onge, J. Albu, S.B. Heymsfield, S. Heshka, Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image, Journal of Applied Physiology 97 (2004) 2333–2338. doi:10.1152/japplphysiol.00744.2004, pMID: 15310748.

[5] F. Buckinx, F. Landi, M. Cesari, R.A. Fielding, M. Visser, K. Engelke, S. Maggi, E. Dennison, N.M. Al-Daghri, S. Allepaerts, J. Bauer, I. Bautmans, M.L. Brandi, O. Bruyère, T. Cederholm, F. Cerreta, A. Cherubini, C. Cooper, A. Cruz-Jentoft, E. McCloskey, B. Dawson-Hughes, J.-M. Kaufman, A. Laslop, J. Petermans, J.-Y. Reginster, R. Rizzoli, S. Robinson, Y. Rolland, R. Rueda, B. Vellas, J.A. Kanis, Pitfalls in the measurement of muscle mass: a need for a reference standard, J. Cachexia, Sarcopenia and Muscle 9 (2018) 269–278, https://doi.org/10.1002/jcsm.12268.

[6] A. Shuster, M. Patlas, J.H. Pinthus, M. Mourtzakis, The clinical importance of visceral adiposity: a critical review of methods for visceral adipose tissue analysis, The British Journal of Radiology 85 (2012) 1–10, https://doi.org/10.1259/bjr/38447238, pMID: 21937614.

[7] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The prisma 2020 statement: an updated guideline for reporting systematic reviews, BMJ 2021 (2020) n71, https://doi.org/10.1136/bmj.n71.

[8] J.H. Hong, H. Hong, Y.R. Choi, D.H. Kim, J.Y. Kim, J.-H. Yoon, S.H. Yoon, CT analysis of thoracolumbar body composition for estimating whole-body composition, Insights into Imaging 14 (2023), https://doi.org/10.1186/s13244-023-01402-z.

[9] D. Liu, J.W. Garrett, M.H. Lee, R. Zea, R.M. Summers, P.J. Pickhardt, Fully automated CT-based adiposity assessment: comparison of the l1 and l3 vertebral levels for opportunistic prediction, Abdominal, Radiology 48 (2022) 787–795, https://doi.org/10.1007/s00261-022-03728-6.

[10] P.J. Pickhardt, A.A. Perez, J.W. Garrett, P.M. Graffy, R. Zea, R.M. Summers, Fully automated deep learning tool for sarcopenia assessment on ct: L1 versus l3 vertebral level muscle measurements for opportunistic prediction of adverse clinical outcomes, Am. J. Roentgenol. 218 (2022) 124–131, https://doi.org/10.2214/AJR.21.26486, pMID: 34406056.

[11] I. Anyene, B. Caan, G.R. Williams, K. Popuri, L. Lenchik, S. Giri, V. Chow, M.F. Beg, E.M. Cespedes Feliciano, Body composition from single versus multi-slice abdominal computed tomography: Concordance and associations with colorectal cancer survival, Journal of Cachexia, Sarcopenia and Muscle 13 (2022) 2974–2984, https://doi.org/10.1002/jcsm.13080.

[12] T. Liu, P. Xu, Y. Bo, Body composition segmentation based on dual-pathway deep dilated convolutional neural network, in: Proceedings of the Third International Symposium on Image Computing and Digital Medicine, ISICDM 2019, Association for Computing Machinery, New York, NY, USA, 2019, p. 36–42. doi:10.1145/3364836.3364844.

[13] Y.S. Lee, N. Hong, J.N. Witanto, Y.R. Choi, J. Park, P. Decazes, F. Eude, C.O. Kim, H.C. Kim, J.M. Goo, Y. Rhee, S.H. Yoon, Deep neural network for automatic volumetric segmentation of whole-body CT images for body composition assessment, Clinical Nutrition 40 (2021) 5038–5046, https://doi.org/10.1016/j.clnu.2021.06.025.

[14] E.O. Wesselink, J.M. Elliott, M.W. Coppieters, M.J. Hancock, B. Cronin, A. Pool-Goudzwaard, K.A.W. II, Convolutional neural networks for the automatic segmentation of lumbar paraspinal muscles in people with low back pain, Scientific Reports 12 (2022). doi:10.1038/s41598-022-16710-5.

[15] M. Nachit, Y. Horsmans, R.M. Summers, I.A. Leclercq, P.J. Pickhardt, AI-based CT body composition identifies myosteatosis as key mortality predictor in asymptomatic adults, Radiology 307 (2023) e222008, https://doi.org/10.1148/radiol.222008.

[16] F. Greco, C.A. Mallio, Artificial intelligence and abdominal adipose tissue analysis: a literature review, Quantitative Imaging in Medicine and Surgery 11 (2021) 4461–4474, https://doi.org/10.21037/qims-21-370.

[17] F. D'Antoni, F. Russo, L. Ambrosio, G. Vollero, G. Vadalà, M. Merone, R. Papalia, V. Denaro, Artificial intelligence and computer vision in low back pain: A systematic review, International Journal of Environmental Research and Public Health 18 (2021), https://doi.org/10.3390/ijerph182010909.

[18] S. Bedrikovetski, W. Seow, H.M. Kroon, L. Traeger, J.W. Moore, T. Sammour, Artificial intelligence for body composition and sarcopenia evaluation on computed tomography: A systematic review and meta-analysis, Eur. J. Radiol. 149 (2022) 110218, https://doi.org/10.1016/j.ejrad.2022.110218.

[19] L. Lenchik, L. Heacock, A.A. Weaver, R.D. Boutin, T.S. Cook, J. Itri, C.G. Filippi, R. P. Gullapalli, J. Lee, M. Zagurovskaya, T. Retson, K. Godwin, J. Nicholson, P. A. Narayana, Automated segmentation of tissues using ct and mri: A systematic review, Academic Radiology 26 (2019) 1695–1706, https://doi.org/10.1016/j.acra.2019.07.006.

[20] M. Paris, Body Composition Analysis of Computed Tomography Scans in Clinical Populations: The Role of Deep Learning, Lifestyle, Genomics 13 (2019) 28–31, https://doi.org/10.1159/000503996.

[21] O. Elfanagely, J.A. Mellia, S. Othman, M.N. Basta, J.T. Mauch, J.P. Fischer, Computed tomography image analysis in abdominal wall reconstruction: A systematic review, Plastic and Reconstructive Surgery – Global Open 8 (2020), https://doi.org/10.1097/GOX.0000000000003307.

[22] H.-J. Meyer, A. Wienke, M. Pech, A. Surov, Computed tomography-defined fat composition as a prognostic marker in gastric adenocarcinoma: A systematic review and meta-analysis, Dig. Dis. 41 (2023) 177–186, https://doi.org/10.1159/000527532.

[23] A. Tolonen, T. Pakarinen, A. Sassi, J. Kyttä, W. Cancino, I. Rinta-Kiikka, S. Pertuz, O. Arponen, Methodology, clinical applications, and future directions of body composition analysis using computed tomography (CT) images: A review, Eur J Radiol 145 (2021) 109943, https://doi.org/10.1016/j.ejrad.2021.109943.

[24] Y. Zhao, X. Wang, T. Che, G. Bao, S. Li, Multi-task deep learning for medical image computing and analysis: A review, Comput. Biol. Med. 153 (2023) 106496, https://doi.org/10.1016/j.compbiomed.2022.106496.

[25] Y.-T. Huang, Y.-S. Tsai, P.-C. Lin, Y.-M. Yeh, Y.-T. Hsu, P.-Y. Wu, M.-R. Shen, The value of artificial Intelligence-Assisted imaging in identifying diagnostic markers of sarcopenia in patients with cancer, Dis Markers 2022 (2022) 1819841, https://doi.org/10.1155/2022/1819841.