Machine learning for image analysis in the cervical spine: Systematic review of the available models and methods

C.M.W. Goedmakers a,b,*, 1, L.M. Pereboom c,1, J.W. Schoones d, M.L. de Leeuw den Bouter e, R.F. Remis f, M. Staring g,h, C.L.A. Vleggeert-Lankamp a,i,j

a Department of Neurosurgery, Leiden University Medical Center, Leiden, the Netherlands
b Computational Neuroscience Outcomes Center, Department of Neurosurgery, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA
c Faculty of Mechanical, Maritime and Materials Engineering (3mE), Delft University of Technology, Delft, the Netherlands
d Walaeus Library, Leiden University Medical Center, Leiden, the Netherlands
e Circuits and Systems Group, Microelectronics Department, Delft University of Technology, Delft, the Netherlands
f Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands
g Intelligent Systems Department, Delft University of Technology, Delft, the Netherlands
h Department of Neurosurgery Haaglanden Medical Centre and HAGA Teaching Hospitals, The Hague, the Netherlands
i Department of Neurosurgery, Spaarne Gasthuis Haarlem/Hoofddorp, the Netherlands
j Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands

ARTICLE INFO

Keywords:
Cervical spine
Computer aided diagnostics
Radiological imaging
Machine learning
Image analysis

1. Introduction

Neck pain is the number four cause of physical disability worldwide, and it can be an important symptom in identifying degenerative pathologies of the cervical spine. In most cases, acute neck pain resolves without invasive treatment, but in nearly 50% of patients, the pain returns or develops a chronic nature. With the current ageing population and the relatively high prevalence of neck pain and spine disease, there is increasing demand on radiological image analysis in healthcare (Cohen, 2015). However, the analysis of those visualizations is time-consuming and is subject to significant interobserver variability (Urrutia et al., 2017). Automating parts of the radiological image analysis process can support clinicians in providing a more accurate and consistent image assessment with increased time efficiency.

Over the last decade, the application of artificial intelligence (AI) in medical research has become increasingly popular. Machine Learning (ML) techniques show promise in computer aided diagnostics (CAD), specifically for clinical tasks related to detection and segmentation, as well as classification and prediction (Esteva et al., 2017; Tabibu et al., 2019; Tiulpin et al., 2018; Xu et al., 2019; Zhang et al., 2020). A ML algorithm is able to “learn,” which means in this context that the algorithm can improve performance by previous experience or provided data to give a valid result for never-before-seen data, without being explicitly programmed to do so (Jakubicek et al., 2020).

The majority of the available literature on image analysis concentrates on the thoracic and lumbar spine, while the cervical spine is studied less often. The difference can be partly attributed to the lower incidence of neck pain in the general population, compared to (lower) back pain (Sinnott et al., 2017). Nevertheless, the neck is an essential part of the body with several vital anatomical structures whose functioning can be visualized using radiological imaging. Additionally, considering the relative novelty of the subject matter no systematic reviews have been published, while this could significantly improve the quality of future research on this topic.

Therefore, we aim to create the first overview of the available Machine Learning methods for image analysis of the cervical spine, while weighing and discussing their risks and benefits and providing
recommendations for future research in this field. We will divide the systematic review into two sections, one focusing on ML for segmentation and the other on applying ML to automate the study different properties, such as segment mobility and curvature, of the cervical spine on radiological imaging. The overview provided in this systematic review may function as a reference for all authors conducting research on computer aided diagnostics of cervical spine disease.

2. Methods

2.1. Literature search

The initial literature search was performed in PubMed, EMBASE and Web of Science, on December 18th, 2020. Two of the authors (CG, LP) separately evaluated the articles by title, abstract and full text, when necessary, to select the studies that met the predefined selection criteria.

(Performed: 20-12-2020)

Fig. 1. The search strategy used to perform the systematic search in the medical databases.
As the topic of this review touches both the medical, and the technical research field, both points of view had to be highlighted. Therefore, an additional literature search was performed in the Google Scholar, Scopus, SPIE Digital Library and IEEE Explore databases, to obtain as many articles as possible from both medical and technical journals. The search strategies used in the different databases were based on the search string as shown in Fig. 1.

Studies were included when they reported on a form of automated radiologic image analysis focusing on the human cervical spine or whole spine including the cervical vertebrae.

Studies were excluded if they met any of the following criteria: (1) Publications not written in English; (2) Conference abstracts; (3) Narrative reviews; (4) Cadaver studies without proven clinical application; (5) Phantom studies without proven clinical application; (6) Studies that describe a protocol without any form of analysis; (7) Studies on the thoracic or lumbar spine; (8) Studies on radiation dose, artifact reduction, sequence analysis or robotic surgery; (9) Studies on image processing without segmentation, landmarking or any other measurement on the spine involved.

Any discrepancy in selection between the reviewers was resolved in open discussion (CG, LP), and, if needed, a third reviewer was asked to make a final decision (CVL). Reference screening and citation tracking were performed on the identified articles. This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement (Moher et al., 2009).

### 2.2. Quality assessment

The methodological quality of all studies was assessed separately by two reviewers (CG, LP), using a version of the Modified New Castle – Ottawa Quality assessment scale for Cohort Studies (Wells et al., 2020). If there was no consensus about the assessment, a third reviewer (CVL) was consulted. The New Castle – Ottawa scale was manually adjusted to better fit human to model comparison studies with a technical nature.

The items reviewed in the assessment were: 1.1 Representativeness of cohort; 1.2 Model selection, development and implementation; 1.3 Comparison made; 1.4 Ground truth assessment and Data extraction; 2. Applicability and Generalizability (data variability, semi-/fully-automatic, different modalities); 3.1 Outcome Assessment (clear split, ground truth objectified); 3.2 Outcome reporting (different outcome measures, uncertainty metrics reported); 3.3 Sharing (data or code sharing). All items could be awarded a maximum of 1 point, except for ‘Applicability and Generalizability’, for which a maximum of 2 points could be given. Studies could maximally be awarded 9 points. Studies were then divided into low (7–9 points), intermediate (5–6 points) or high (4 or less points) risk of bias.

### 2.3. Data extraction

Data extraction for all included articles was performed by two reviewers separately (CG, LP) and any controversies were resolved by a third reviewer (CVL). From each article the following information was...
collected: year of publication, image modality, spine region, model description, degree of automation, number of images included, train to test set distribution and description of how the ground truth was acquired. The determination of the ground truth can be done by either one or more clinical experts and can be provided in different formats; i.e. bounding boxes, vertebra centers, or complete pixel-wise segmentations. Only outcomes that were mentioned in the text or tables of a publication were included into the analysis, as extracting outcomes from graphs was deemed too imprecise and time-intensive.

In order to compare model performance, commonly reported outcome measures were extracted from each publication. Outcomes were divided in either the internal comparison group; when the model’s performance was compared to the ground truth, or the external comparison group; when the model’s performance was compared to model performance from previous publications.

Outcomes of articles in the segmentation category were reported in five major groups:

**Accuracy:** Accuracy, Identification Rate (IR), Detection Rate (DR)

**Error (mm):** Localization error (LE), Mean Distance Closest Point (MDCP),

Mean Absolute Surface Distance (MASD), Point-to-Surface error, Hausdorff Distance (HD), Center of Gravity (COG)

**Overlap:** Dice Overlap (DO), Dice Coefficient (DC), Dice Index (DI)

**Time:** Runtime, Efficiency.

**Other:** Precision, Sensitivity, Specificity.

The aims of studies included into the second category, cervical spine analysis, can be divided up into five broad categories: 1. Biomechanical analysis; 2. CVM stage; 3. Clinical prognosis/prediction; 4. Image registration/Planning; 5. Clinical/Radiological Feature Detection. Additional variables collected for the second category articles were aim, included vertebrae and key points.

3. Results

3.1. Article selection

Through searching PubMed, EMBASE and Web of Science, using the predefined search strategy, 956 records could be identified. 654 remained after duplicates were removed. An added search in Google Scholar, Scopus, SPIE Digital Library and IEEE Explore yielded an additional 28 publications. The 682 unique records were screened for title and abstract, after which a total of 506 articles could be excluded. The full-texts of the remaining 176 articles were screened, and 125 did not fit all in- and exclusion criteria and were therefore removed. The remaining 51 articles were included in this systematic review and, based on their primary aim, divided into the two main categories; 1. Segmentation (n = 32) and 2. Cervical Spine Analysis (n = 19). The first category was then divided into two subcategories; 1.1 Conventional Machine Learning Segmentation (n = 20) and 1.2 Deep Learning Segmentation (n = 12) (Fig. 2).

![Fig. 3. Number of publications plotted per year.](image)

![Fig. 4. Number of publications per subcategory per year.](image)
Where articles in the first subcategory focus more on the conventional Machine Learning methods for segmentation, studies in the second category deploy the relatively newer, neural networks. In the second category studies were included that did not necessarily focus on segmentation but in some other way analyzed the cervical spine and its radiologic characteristics.

The increasing popularity of Machine Learning for image analysis of the cervical spine is clearly illustrated when the number of included publications in this study is plotted against the year of publication in total and per subcategory (Fig. 3, Fig. 4). The majority of the included articles (n = 39) is published within the last 5 years.

| Author (Year) | Search | 1.1 | 1.2 | 1.3 | 1.4 | 2   | 3.1 | 3.2 | 3.3 | Total |
|---------------|--------|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| Zamora (2003) |        | 1   | 1   | 1   | 1   | 1   | 0   | 0   |     | 6     |
| Burnett (2004)| PubMed | 1   | 1   | 1   | 1   | 1   | 1   | 0   | 0   | 6     |
| Weiss (2006)  | PubMed | 1   | 1   | 1   | 1   | 0   | 1   | 0   | 0   | 5     |
| Schmidt (2007)| PubMed | 0   | 1   | 1   | 1   | 2   | 1   | 1   | 0   | 7     |
| Klinder (2009)| PubMed | 1   | 1   | 1   | 1   | 2   | 1   | 0   | 0   | 7     |
| Huang (2009)  | PubMed | 1   | 1   | 1   | 1   | 2   | 0   | 0   | 0   | 6     |
| Banik (2010)  | PubMed | 1   | 1   | 1   | 1   | 0   | 1   | 1   | 0   | 6     |
| Giuliani (2011)| PubMed | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 0   | 6     |
| Chen (2012)   | PubMed | 0   | 0   | 1   | 1   | 0   | 1   | 0   | 0   | 3     |
| Glocke (2012) | PubMed | 1   | 1   | 1   | 1   | 2   | 0   | 1   | 0   | 7     |
| Xu (2012)     | Additional | 1   | 1   | 1   | 1   | 1   | 1   | 0   | 0   | 6     |
| Glocke (2013) | PubMed | 1   | 1   | 1   | 1   | 2   | 0   | 1   | 1   | 8     |
| Mirzaalian (2013)| PubMed | 1   | 1   | 1   | 1   | 1   | 0   | 1   | 0   | 6     |
| Lahrham (2014)| PubMed | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 0   | 7     |
| Daenzer (2014)| PubMed | 1   | 1   | 1   | 1   | 2   | 1   | 1   | 1   | 9     |
| De Leener (2015)| PubMed | 1   | 1   | 1   | 1   | 2   | 1   | 1   | 0   | 8     |
| Clogenson (2015)| Embase | 0   | 1   | 1   | 1   | 0   | 0   | 1   | 1   | 5     |
| Al Arif (2016)| Additional | 1   | 1   | 1   | 1   | 2   | 1   | 1   | 0   | 8     |
| Mehmood (2017)| Additional | 1   | 1   | 1   | 1   | 1   | 0   | 0   | 0   | 5     |
| Hanaoka (2017)| PubMed | 0   | 1   | 0   | 1   | 2   | 1   | 0   | 0   | 5     |
3.2. Quality assessment

In the Conventional Machine Learning Segmentation group there was one study included with a high risk of bias, eleven studies with an intermediate risk of bias and eight with a low risk of bias (Table 1a). In the Deep Learning Segmentation group three studies showed intermediate risk of bias and nine a low risk of bias, while there were no studies included with a high risk of bias (Table 1b). Lastly, in the Cervical Spine Analysis group there was one study with a high, 14 with an intermediate and 4 with a low risk of bias (Table 1c).

In general it can be observed that the more recently studies were published, the more likely they were to have a decreased risk of bias. Therefore, the percentage of low risk of bias studies in the Deep Learning Segmentation group is higher than in the Conventional Machine Learning Segmentation group, as the latter includes more recent studies. The same pattern - a decreased risk of bias over time - can be observed in the Cervical Spine Analysis group.

3.3. Qualitative synthesis

3.3.1. Conventional Machine Learning Segmentation techniques

The total number of included studies involving Conventional Machine Learning segmentation techniques is 20, of which 6 studies focused on X-ray images, 6 on MR imaging and 8 studies on CT imaging. The major part, consisting of 14 studies, involved two-dimensional models. The remaining 6 studies used three-dimensional models, of which 4 studies used CT imaging and 2 studies used MRI. The number of images used per study varied widely by image modality. The range of the number of included X-ray images was between 66 (Larhmam et al., 2014) and 10024 (Xi et al., 2012). The range of included MR images and CTs was diffusely reported, as publications did not only use different numbers of scans but also different numbers of slices, sometimes differentiating per spine region. The number of studies with semi- or fully-automated methods was the same (n = 10) (Table 2a).

The highest accuracy for MR imaging were reported by Weiss et al. (2006); 96% for the initial model and 100% for the modified model, for the whole spine and the cervical spine, respectively. The study included the entire spine; the vertebrae and intervertebral disks and the ground truth consisted of ‘independent assignments’ of neurologists. In total, 50 MR images were included, 27 were used for the initial model and 23 MR images were used for the modified model. Image volumes are enhanced with a tophat filter, the program assigns the threshold values and applies a median spatial filter to the search regions. Voxels exceeding threshold values are then subjected to additional constraints and the centroids of these voxel clusters are then connected. 3D linear interpolation and Gaussian filters were applied, the longest disc chains were then analyzed in clusters, which obtained the above mentioned accuracies (De Leener et al., 2015).

The best performing methods are VoHOG for MR images Daenzer et al. (2014), Modified GHT and K-means clustering with the use of X-ray

| Author (Year)       | Search    | 1.1 | 1.2 | 1.3 | 1.4 | 2   | 3.1 | 3.2 | 3.3 | Total |
|---------------------|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| Suzani (2015)       | Additional| 1   | 1   | 1   | 0   | 2   | 1   | 1   | 0   | 7     |
| Chen (2015)         | Additional| 1   | 1   | 1   | 0   | 2   | 1   | 1   | 0   | 7     |
| Cai (2016)          | PubMed    | 1   | 1   | 1   | 0   | 2   | 1   | 0   | 0   | 6     |
| Forsberg (2017)     | PubMed    | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 0   | 7     |
| Liu (2018)          | PubMed    | 1   | 1   | 1   | 1   | 2   | 1   | 1   | 0   | 8     |
| Al Arif (2018)      | PubMed    | 1   | 1   | 1   | 0   | 2   | 1   | 1   | 0   | 7     |
| Jakubicek (2019)    | PubMed    | 1   | 1   | 0   | 1   | 1   | 1   | 1   | 0   | 6     |
| Bae (2019)          | Embase    | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 0   | 7     |
| Rak (2019)          | PubMed    | 1   | 1   | 1   | 0   | 1   | 1   | 1   | 0   | 6     |
| Wang (2019)         | PubMed    | 1   | 1   | 1   | 1   | 2   | 1   | 1   | 0   | 8     |
| Chen (2020)         | Embase    | 1   | 1   | 1   | 0   | 2   | 1   | 1   | 0   | 7     |
| Jakubicek (2020)    | PubMed    | 1   | 1   | 1   | 1   | 2   | 1   | 1   | 0   | 8     |
imaging Larhmam et al. (2014) and a statistical and Gaussian shape model in combination with a principal component in combination with CT imaging Clogenson et al. (2015). The research of Daenzer et al. (2014) approached the cervical vertebra detection with a proposed novel machine learning method based on new radiological features, combined with a linear SVM. An accuracy of 98.1% was achieved with the baseline model and improved to 99.1% with the VolHOG. In addition, various levels of artificial noise are used during the performance analysis of the algorithm.

The ground truth in these studies is based on manually determined datapoints by (clinical) experts. All studies reported an internal comparison, comparing the performance of their model to the ground truth.
| Author (Year) | Modality | Spine Region | 2D/3D | Semi-/Fully-Automatic | N ~ Splits | Ground Truth | Comparison |
|--------------|----------|--------------|-------|-----------------------|------------|--------------|------------|
| Zamarra (2023) [66] | X-ray | Cervical Spine | 2D | Semi | + Unclear | Morphometric points placed by three expert radiologists | + Stages: * GHT | - |
| Barrett (2004) [16] | CT | Whole Spinal canal | 2D | Fully | + 5 CF’s total | * Contours drawn by six dosimetrist | - |
| Wein (2006) [62] | MRI | Whole Spine | 2D | Semi | + 59 MRI total | * Two radiation oncologists | - |
| Schmidt (2007) [51] | MRI | Whole Spine | 2D | Fully | + 28 MRI total | Annotated ground truth data by expert | - |
| Kinsler (2009) [16] | CT | Whole spine | 3D | Fully | + 74 CT total | * Corrected mesh ground truth & Voxel based | - |
| Hung (2009) [51] | MRI | Cervical Spine | 2D | Fully | + 22 MRI total | Manually marked pixelwise ground truth | - |
| Banik (2010) [5] | CT | Whole Spine and Spinal Canal | 2D | Semi | + 39 CT total | Contours manually annotated by an expert radiologist | - |
| Guaitelli (2011) [25] | MRI (EPI) | Spinal Canal | 2D | Semi | + Unclear + 7 patients total | Manually annotated by two experienced operators both trained in spinal segmentation | - |
| Chen (2012) [15] | X-ray | Cervical Spine | 2D | Semi | + Unclear + 120 X-rays total | Orthodontic Novices and Orthodontic experts (ON and OE), compared rather than setting GT | - |
| Glocker (2012) [26] | CT | Cervical Spine and Whole Spine | 2D | Semi | + Unclear + 10024 Cervical X-rays total | * Performance per stage reported (Stage 1: Regression Forest, Stage 2: BMM) | - |
| Xu (2012) [51] | X-ray | Cervical Spine | 2D | Fully | + Unclear + 10024 Cervical X-rays total | Coordinates control points obtained from segmentation results compared to points marked manually by doctors | - |
| Glocker (2013) [27] | CT | Cervical Spine and Whole Spine | 2D | Semi | + Unclear + 200 CT scans total | Manual annotations, not further specified by whom, Glocker (2012) by expert radiologist | - |

Table 2a: Conventional Machine Learning Segmentation articles overview.
and 9 studies additionally reported some form of external comparison, with earlier publications, published in the years 2003–2009. Of all Conventional Machine Learning segmentation studies, 7 studies reported segmentation results for only the whole spine, while 8 reported results for specifically the cervical spine. In 4 studies, the segmentation results were reported for both the cervical spine specifically and the whole spine. Clogenson et al. (2015) is an exception, just focusing on vertebra C2, which decreases external validity as compared to the other included studies (Table 3a).

### 3.3.2. Deep Learning Segmentation techniques

There was a total number of 12 studies included that proposed Deep Learning segmentation techniques, of which two studies focused on MR imaging, 8 studies focused on CT imaging, and just one study used X-ray imaging. The study of Cai et al. (2016) involved both MRI and CT imaging. The majority, consisting of 7 studies, involved three-dimensional models, of which one study Jakubicek et al. (2019) combined 2D and 3D. The remaining 4 studies used two-dimensional models. The number of images used per study varied again widely per image modality, comparable to the Conventional Machine Learning segmentation studies. The range of the number of included CT images was between 41 Bae et al., 2019 and 392 (Jakubicek et al., 2019). The range of used MR images was slightly smaller but comparable; from 60 (Cai et al., 2016) to 245 MRI images (Forsberg et al., 2017). However, the interpretation of this range is difficult as publications, like in the conventional Machine Learning group, did not only use different numbers of scans but also different numbers of slices, sometimes differentiating per spine region.

Almost all studies deployed fully automated methods. Only one study used a semi-automated approach Forsberg et al. (2017), which then also achieves highest detection accuracies (98.8–99.8%). Forsberg et al. (2017) focused on both the cervical and lumbar spine, creating two separated training and configuration pipelines, both having the same CNN setup. The CNN uses fully connected layers, drop-out rate of 0.5, a categorical cross-entropy cost function and Nesterov momentum accelerated Stochastic gradient descent (SGD). The included MR images, together with the annotated spine labels, are focused on either the lumbar or cervical part of the spine. The dataset was originated from an image archive. The missed detections were mainly concerning partly visible vertebrae on the available images. This research showed promising results for labeling and detection by a CNN, focusing on both the cervical and lumbar spine (Glocker et al., 2012) (Table 2b).

The highest segmentation accuracy was achieved by the SpineCNN from Jakubicek et al. (2019) (93.3%). Thereby, the best performing methods are CNN based methods for both CT and MR images. The study presents a fully automated approach based on 130 CT scans, which includes two CNNs and a spine tracing algorithm, among which a
| Author (Year) | Modality | Spine Region | 2D/3D | Semi-Fully-Automatic | N ~ Split | Ground Truth | Comparison |
|---------------|----------|--------------|-------|---------------------|----------|--------------|------------|
| Suzuki (2015) [36] | CT | Cervical Spine and Whole Spine | 3D | Fully | 224 CT total * 50:50 split * 112 images each * Two-fold cross-validation | Expert annotations | Detection rates and localization errors for different regions of the vertebral column * Compared with detection rates and localization errors of Glockler (2013) and Glockler (2012) |
| Chen (2015) [24] | CT | Cervical Spine and Whole Spine | 3D | Fully | 302 CT total * 242 CT training * 60 CT testing | Manually, not further specified by whom | J-CNN compared with ‘standard’ CNN and GT * Compared with IR, Mean error and Std Glockler (2013) |
| Cai (2016) [13] | CT, MRI | Cervical Spine and Whole Spine | 2D | Fully | 60 MRI total * 90 CT total * 6 MRI Training * 4 CT Training * 54 MRI Testing * 56 CT Testing | Manually selected and labeled by one-click annotation | Proposed vs. Baseline MR * Proposed vs. Baseline CT |
| Forberg (2017) [24] | MRI | Cervical Spine | 2D | Semi | 245 MRI total * 223/221 T1-T2 * 232/ spine labels * Random split * Training : Validation : Test | Navigation support annotations, manual quality assurance step performed | T1 vs. T2 compared for sensitivity, precision accuracy, localization error, labeling accuracy * DR for Huang (2009), Klings (2009), Zhan (2015) compared * Extensive comparison table (including localization error and labeling rate) available in full publication. |
| Liu (2018) [40] | CT | Cervical Spine | 3D | Fully | 60 CT total * 40 Training set * 20 Test set | Manually labelled by ‘professional physicians’ | Proposed method DI, Mean absolute surface distance, JID * Proposed method vs. GT per vertebra, see full publication |
| Al Arif (2018) [11] | X-ray | Cervical Spine | 2D | Fully | 296, train 124, test 172 * Dataset 138 total | Manual annotation of vertebral boundaries, * Global Localization * Center Localization |
| Jakobek (2019) [13] | CT | Whole Spine | 2D/3D | Fully | 392 CT total * 4:2:3 split * Training : Validation : Test | Positions determined by experts | The algorithm is compared in centerline detection to the GT * The performance of CordCNN and SpineCNN were compared |
| Jie (2019) [6] | CT | Cervical Spine | 2D | Fully | 41 CT total * 80:20 split * Train:Validation, (n=17 set) 14 train, 3 validation * (n=24 healthy controls set) 19 train, set, 5 validation * Validation performed on both sets for different models * Two different sets from two different hospitals | Provided by 2 human experts, kappa calculated | Intra-validation (which seems like validation) and extra-validation (which seems like test set performance) * Models functioned once as training and once as test set * Comparison with lumber model * Extra-validation results reported in Outcome as this seems test set performance |
| Rah (2019) [19] | MRI | Whole Spine | 3D | Fully | 64 MRI total * 4:1 = 51 : 13 | Falsely detected vertebrae were corrected manually, i.e. user-specified vertebra center was used, not further specified by whom | Model performance on different 1 (whole spine images) T1 vs. T2 weighted MRI images |
| Wang (2019) [61] | CT | Cervical Spine and Whole Spine | 3D | Fully | 58 CT total * 50:50 split * Two sets ~40 * Randomly selecting 31 from 63 normal | Centerline annotated by two experts | Cervical vs. whole spine Localization error for SRF step and refinement step * Cervical vs. whole spine Identification Rate |

* 10: Graphs CNN based and RF+HMM based approaches error compared per vertebra.
fine-tuned AlexNet and a VGG-16 R-CNN. A population approach was used to increase robustness. The novel combination of the CNN and the tracing, results in almost 90% of correctly identified spinal centerlines within 20 s of computing time (Forsberg et al., 2017).

The only study focusing on X-ray imaging Al Arif et al. (2018) used a 6-layered FCN, with an accuracy of center localization of 93.7%.

Similar to the Conventional Machine Learning segmentation studies, the ground truth in the Deep Learning segmentation publications is based on manually determined ground truth by (clinical) experts. The majority of the included studies regarding Deep Learning segmentation methods used both internal and external comparison of their results (n = 9). Results were reported for the whole spine and cervical spine, in 2 and 4 studies, respectively. In 6 studies, half of the total number of studies related to Deep Learning segmentation methods included, the results were reported for both the cervical part and the whole spine (Table 3b).

### 3.3.3. Cervical Spine Analysis

There was a total number of 19 studies included involving cervical spine analysis, of which four studies focused on MR imaging, two studies focused on CT imaging, and the majority of the studies used X-ray imaging (n = 11). The aims of the included studies could be further divided into five subgroups: 1. Biomechanical analysis (n = 7), 2. CVM stage (n = 3), 3. Clinical prognosis/prediction (n = 2), 4. Image registration/Planning (n = 4), 5. Clinical/Radiological Feature detection (n = 3).

Two studies included both CT and MR imaging, of which du Bois d’Aische et al. (2007) included PET imaging as well. The use of two-dimensional and three-dimensional visualizations were equal (n = 9), and the remaining study of Kage et al. (2020) (Kage et al., 2020) used a combination of 2D and 3D imaging. Most studies included vertebrae C2–C6, of which several expanded with inclusion of the vertebrae C1, C7 or T1. Other studies used a smaller area of the spine, vertebrae C2–C4, which had the aim to determine the CVM stage Kök et al. (2019) and Amasya et al. (2020). The study by Dzyubachyk et al. (2013) was the only one to include the entire spine in the analysis model, with the aim to create an automated reconstruction of the complete spine, based on multistation 7T MR images. The authors applied intensity inhomogeneity correction and used coherent local intensity clustering (CLIC) and fuzzy-c-means-clustering. The performance of the model by Dzyubachyk et al. (2013) was validated based on 18 different datasets, which showed a mean registration error of 0.53 mm, which was lower than the MR image pixel size and showed thereby sufficient accuracy.

A wide range of methods was deployed. The best method for radiological feature detection is a CNN model, while the SVM model gave the best result in terms of clinical classification. The ANN approach was reported to work best for CVM stage determination and the FE model, in combination with X-ray imaging, is the most-used method for biomechanical analysis of the spine. In the included spine analysis studies, the amount of fully and semi-automated methods was 7 and 12, respectively (Table 4).

#### 3.4. Quantitative synthesis

It was considered to pool accuracy rates in the Conventional Machine Learning and Deep Learning segmentation groups, however it was found that outcomes in the included studies were too heterogeneously reported for doing so. Authors chose to report different outcome metrics and the majority did not report on uncertainty metrics (confidence intervals, standard errors, standard deviations or p-values) with their primary outcome. Pooling the data would therefore require statistical imputation for the majority of the uncertainty metrics. Subsequently, this means that heterogeneity tests, such as the $F^2$, were not performed, as data could not be pooled.

#### 4. Discussion

In this systematic review an overview was provided of the literature on the available Machine Learning techniques for automated image analysis of the cervical spine on radiological imaging. The results of the included studies show a wide variety of possibilities in Machine Learning methods, depending on the aim of the application and the available modalities. In segmentation models, Deep Learning methods show promising results with the application of (fully automatic) CNN models using X-ray, CT or MR imaging. Regarding cervical spine analysis, the biomechanical properties are most often studied using finite element models. The application of artificial neural networks and support vector machine models looks promising for other classification purposes.

Most of the published work on image analysis of the spine focusses on the (thoraco-) lumbar spine. This can be explained by the higher prevalence of lumbar spine pathology, as compared to cervical spine pathology. However, this study, focusing on the cervical spine, is the first of its kind and we therefore believe it can be used as a reference study for all researchers aiming to use radiological image analysis for the cervical spine, as well as other diseases in the neck area.

Unfortunately, results in this systematic review were too heterogeneously reported and therefore pooling the results was not possible. Reporting outcomes clearly and homogenously is an important requirement to compare performance among publications. The authors of this review want to plead for more consistent reporting of outcomes, i.e. the same set of outcome variables for every segmentation, classification or prediction study in order to increase the external validity and
| Author (Year) | Model Features | Internal | External |
|--------------|----------------|----------|----------|
| Zavros (2003) [6] | * Hierarchical * ASMs with GL/VGLM * Customized GHT * Customized DM | * GHT: 65% * GHT + ASM: 75% * GHT + ASM + DM: 75% | - | - |
| Bennett (2004) [10] | * EM * Fourier descriptors * Watershed-based Edge detection algorithm * Chamfer matching | * Automated contours: 93% vs. Manual contours: 96% * Automated successful in 91%, further editing in 7% and unsuccessful in 2% | - | - |
| Weiss (2006) [62] | * Threshold values * Median spatial filter * Three-dimensional linear interpolation * Gaussian filters * Disc contours, centroid, discs, longest chains, analyze clusters | * Initial algorithm: 96% * Modified algorithm: 100% | - | - |
| Schmidt (2007) [55] | * Probabilistic GM, object recognition framework * Fully interconnectable model * No model assumptions * Truncated Gaussian distributions, randomized classification trees, branching tests at true nodes | * Geometry: DR: 91% incl. appearance: 94% * Rotational: DR: 95% incl. appearance: 97% * Semi-automatic Rotational DR: 94% Geometry DR: 98% | - | - |
| Kinder (2009) [36] | * Geometric modelling * GHT (template matching) * Shape constrained EM | * Vertebra detection: 93% * Vertebra identification: 95% | * Detection: 13.3 sec * Identification: 36.5 min * Segmentation: 179.5 sec | * Segmentation: 1.12 ± 1.94 |
| Huang (2009) [35] | * Deep Learning iterative normal-cut * Adiabatic-based detection * Refinement robust curve fitting * Energy minimization process | DR: * Standard Adaboost 85.11% * Proposed without RANSAC-based refinement 89.36% * Proposed with RANSAC-based refinement 97.87% | - | - |
| Bank (2010) [18] | * Essay segmentation * Rough Transform | - | - |
| Gulloni (2011) [25] | * K-means clustering * Gaussian Smoothed * Non-linear noise reduction | * Proposed: 82% + 1 * Proposed efficiency: 0.67 ± 0.02 | - | - |
| Chua (2012) [16] | * Landmark calculation * Fast matching method * Parabolic curve fitting * Computer aided cervical vertebra landmarking (CACL) | * HD p-values * Authors conclude based on p-values per vertebra: “CACL has the same or higher accuracy, better repeatability and less work, load than manual landmarking methods” | - | - |
| Cloots (2012) [29] | * Regression forests * Probabilistic graphical models * Hidden Markov Model * Discriminative regression * Generative model | IR: * Cervical 72% * Whole 81% | - | - |

**Table 3a**

Conventional Machine Learning Segmentation articles extracted outcomes.
**Vu (2012) [41]**
- *Hair-like features*
- *Adaboost*
- *Multi-resolution AAM*
- *Parallel cascade Structure*
- *Error in Pixel (1)
  Part Min: Average Max: 1 8.652 27.597 22.336 II 3.682 4.792 6.007 III 4.035 19.018 146.925* -

**Glocker (2013) [27]**
- *Sparse centroid annotations transformed into dense probabilistic labels*
- *Randomized classification forests*
- *Feature extraction through supervised, hierarchical clustering training data*
- *Objective function which favors compact clusters of image points having equal labels*
- *Proposed method: Normal vs. Pathological CT IR Cervical 78% vs. 80% IR Whole 76% vs 70%*
- *Proposed method LE Normal vs. - Pathological CT Median Mean Std Cervical 6.3-5.9 7.7-7.0 4.4-4.7 Whole Spine 7.6-8.8 11.5-12.4 14.1-11.2*
- *Performance normal CT see results Glocker (2012)*
- *Pathological CT: IR Cervical 58% IR Whole 51%*
- *Normal CT see results Glocker (2012)*

**Menonain (2013) [144]**
- *Statistical Shape Model*
- *Graphical Cut with Shape prior*
- *ML to capture local appearance-related prior shape information*
- *Probabilistic boosting-free classifier*
- *Runtime: Estimated at 2 minutes*
- *Cervical Point-to-surface GC + SP: 7.9 ± 7.9 SSM + ML + Norm: 1.4 ± 0.4 Whole Spine Point-to-surface GC + SP: 6.1 ± 6.7 SSM + ML + Norm: 1.6 ± 0.7*
- *Point-to-surface: C = 1.6 ± 0.6 W = 1.6 ± 0.7*

**Larhom (2014) [38]**
- *Modified GHT*
- *K-means clustering*
- *DR: NHANSIS II 99.1%, JilinDB total: 96% Mean total: 95.7%*
- *Mean RMS angle error 6.70 degrees (f)*
- *DR: Proposed: 97.5%, Cusco: 83% Klinger: 92% Dong: 92.4% Prior work: 89%*

**Hanser (2014) [195]**
- *New features with linear SVM for classification. Algorithm for bivariate gradient orientation histogram generation from three-dimensional raster image data.*
- *VcHOG*
- *Runtime: 4.1 ± 1.32 min*
- *Aver. COG Distance (mm) Baseline: 1.72 ± 0.81 VcHOG: 1.64 ± 0.70*
- *Aver. DO Baseline: 0.80 ± 0.07 VcHOG: 0.81 ± 0.06 Precision Baseline: 6.9653 VcHOG: 0.9859*
- *DR with Rician noise 0.9905 Rician noise 1: 94.29% Rician noise 2: 91.43%*

**Clogston (2015) [177]**
- *SSM*  
  *Gaussian Shape model*  
  *PCA*
- *Runtime: 30 sec*
- *Distance error: Proposed method: 0.9 ± 0.12*
- *Runtime: Minimal 2 min Kindler: 3 min*
- *Distance error: Minimal 1.4 ± 0.4 Kindler: 0.81 ± 0.07*

**De Lozier (2015) [20]**
- *Iterative propagation*
- *Vascular EM*
- *Elliptical HT*
- *Vesselness filter*
- *Contract level-based local mesh structures orientation adaption*
- *Proposed for intervertebral disc identification: 98.3%*
- *Runtime: Overall acquisition 22 min*
- *MSE for intervertebral disc identification 1.05 ± 1.45 mm HD*  
  *CSF T2w: Proposed method 289 ± 0.95 DC*  
  *SC T1w: Proposed method 0.91 ± 0.02 DC*  
  *SC T2w: Proposed method D.C 0.91 ± 0.03 VS Active surface: DC 0.87 ± 0.03*
- *HD*  
  *SC T1w: Proposed method 1.79 ± 0.28 VS Active surface 2.05 ± 0.48 SC T2w: Proposed method 1.68 ± 0.56 VS Active surface 2.19 ± 0.53*

**AI Atif (2016) [2]**
- *ASM*  
  *BCF*  
  *KDE*
- *ASM-M 76.00%*  
  *ASM-RBF 79.24%*  
  *ASM-BCF-AM 84.10%*  
  *ASM-RBF 83.33%*  
  *Error (Median/Mean ± Std)*  
  *ASM-M 0.60109 / 0.6832 ± 0.3437*  
  *ASM-RBF 0.6933 / 0.7704 ± 0.3756*  
  *ASM-BCF-AM 0.7054 / 0.8060 ± 0.3968*  
  *ASM-RBF-KDE 0.6896 / 0.7688 ± 0.3965*  
  *Distance Error*  
  *C3 7.7063 C4 7.6779 C5 7.5887 C6 6.8402 C7 13.0709*  
  *Accuracy*  
  *Larhom (2014) 89%*  
  *Larhm (2014) 97.5%*  
  *Larhm (2014) 81.60%*  
  *Larhm (2014) 64.5% (semi-auto)*

**Mehrood (2017) [183]**
- *Localization using GTH Clustering Fuzzy C Means Centeroid Detection*
- *Proposed method total: 93.76% Per vertebra: C3 96.65 C4 95.51 C5 95.33 C6 84.55*  
  *Distance Error*  
  *C3 7.7063 C4 7.6779 C5 7.5887 C6 6.8402 C7 13.0709*  
  *Accuracy*  
  *Larhm (2014) 89%*  
  *Larhm (2014) 97.5%*  
  *Larhm (2014) 81.60%*  
  *Larhm (2014) 64.5% (semi-auto)*

**Banooka (2017) [28]**
- *NNM*  
  *Multi-class-based*  
  *Landmark-guided diffeomorphism demons algorithm*
- *Runtime: 15 min*  
  *Mean error distance 0.59 ± 0.14 HD 5.30 ± 2.14*  
  *DC 6.90 ± 0.02*
| Author (Year) | Model Features | Internal | External |
|---------------|----------------|----------|----------|
| Suzuki (2015) [56] | * Carvery edge detector  * 6-layered DeepNN  * Feed forward  * Stochastic gradient descent | * C: 96.0  * WS: 96.9 | * Runtime: Overall runtime less than 3 seconds | * C: 17.1 ± 8.7  * WS: 18.2 ± 11.4 |
| Chen (2015) [14] | * Random Forest  * i-CNN  * Shape Regression model  * Weights initialized with Gaussian distribution | Proposed:  | Proposed:  | * Glock (2012)  * WS: 93.9% / 92.9%  * 94.7% / 93.7% / No efficiency / Dice |
| Cai (2016) [13] | * TDCN  * Feature fusion by CRIBM  * SVM | Proposed WS: MR/CT 98.1/96  | Proposed WS: MRI 3.23 ± 2.09 CT: 3.91 ± 2.38  | * Glock (2013)  * C: 6.81 ± 10.02  |
| Forsberg (2017) [24] | * CNN  * Fully connected layers  * 50% dropout  * Categorical cross-entropy as cost function  * SGCN Nesterov momentum | Detection accuracy | * LE T1: 1.18 ± 0.61 T2: 1.24 ± 1.01 | * T1: Sens, Precision 99.1%, 99.8%  * T2: Sens, Precision 99.8%, 100%  |
| Liu (2019) [30] | * DM  * VGG-like momentum | -  | Proposed method:  | * Huang (2009)  * DR: 97.9/98.0  |
| Al-Arif (2018) [31] | * 6-layered FCN  * Global localization  * Center localization  * Unet with updated shape-aware loss function | Center localization:  | Center localization:  | * Castro-Matos  * DR: 90.5  |
| Jakubowsk (2019) [33] | * CNN, fine-tuned MaskNet (for SpinCNS spine-end detection)  * VGG-16 faster R-CellNet (for CordCNN centerline delimitation) | Spine Centerline:  | SpineCNN:  | * Castro-Matos  * M ASD 0.82  |
| Bae (2019) [6] | * 2D CNN  * 2D U-net for 3D | -  | - | * Model 1:  * DC: 88.67 ± 5.82  |

Table 3b
Deep Learning Segmentation articles extracted outcomes..
reproducibility of these type of studies. Several guidelines that describe the appropriate reporting process for Machine Learning studies have been published (Heil et al., 2021; Luo et al., 2016). However, after reviewing the vast amount of data from the included studies in this systematic review it can be concluded additional guidelines for reporting specifically on image analysis studies using machine learning, are needed. Apart from the recommendation to report a minimal of accuracy (in percentages from 0 to 100%) and error (in mm), reporting uncertainty metrics (confidence intervals, standard errors or standard deviations) with the primary outcome metrics should be required, as it is essential in order to unify the reporting process and aids pooling of results from future studies. Another essential recommendation is for authors to share code. The majority of publications included in this systematic review did not share their code. Creating an academic environment in which code sharing is promoted is essential to keep improving the work in this field.

The concept of ‘Grand Challenges’ presents a promising alternative to current comparative research on the topic of image analysis, by eliminating a range of biases. The aim of these public challenges is to let participants apply their algorithms to the provided Grand Challenge task, using the public test set of images provided by the challenge organizers. In a Grand Challenge organized for analysis of breast histology images, a total of 64 submitted algorithms improved the state-of-the-art in classification of microscopy images to an accuracy of 84% (Aresta et al., 2019). This systematic review demonstrates a solid body of evidence describing effective segmentation of the cervical spine, with CNN achieving highest accuracy combined with the lowest computing times. Additionally, publications on the different applications for cervical spine analysis show high potential for Machine Learning for several classification and prediction tasks. However, the possibilities for implementation are far-reaching and several newer applications still deserve more attention in future research, including; automated detection, localization and classification of degenerative changes, specifically in the cervical spine. On thoracolumbar CT machine learning was used for automated detection of sclerotic metastases and detection, localization and classification of traumatic vertebral body fractures (Burns et al., 2013, 2016), something that has not been done for the cervical spine yet. On thoracolumbar lateral X-rays the intervertebral disc height measurements were conducted for 1186 participants using machine learning (Allaire et al., 2017), while the study included in this review on the same topic for the cervical spine showed results for only 1 patient (Tan et al., 2012). The challenges in future research are not just in focusing on the cervical vertebrae or increasing the numbers of images, but also in the integration of different models into one fully automated pathway. Incorporating both radiological and clinical parameters into a fully-automatic model and implementing those into the clinical workflow is the end goal. As was established in this review, the detection and segmentation of the cervical spine have achieved sufficient attention in research, but it is the clinically important classification and prediction tasks, and combining those with detection and segmentation into a fully automatic structure, what future research should focus on.
| Author (Year) | Modality | Category | Aim | Model Features | 2D/ 3D | Semi-/ Fully- | N =/ Included Vertebrae | Key Points |
|--------------|----------|----------|-----|----------------|--------|-------------|------------------------|------------|
| Ben sdoun (2009) [9] | X-ray | Biomechanical Analysis | * Determine vertebral motion induced by their movement between two or several positions | * Automatic corner points of interest points detection | * Harris corner detector | 2D | Semi | n = 100 (C2-C6) | * Difference between manual corner detection and proposed method in range 0-4 degrees. |
| Leccon (2012) [39] | X-ray | Biomechanical Analysis | * Determine cervical spine mobility by automated angle calculation between adjacent vertebra | * Scale-invariant Feature Transform | * Speeded-up Robust Features | 2D | Fully | n = 49 (C3-C7) | * Vertebrae are successfully detected in 89.8% of cases and it is demonstrated that SURF slightly outperforms SIFT. |
| Balkovec (2017) [7] | X-ray | Biomechanical Analysis | * Determine precise time-course details about individual vertebrae and intervertebral motion | * Iterative Template Matching Algorithm | * Video fluoroscopy / Digital Motion | 2D | Semi | n = 3 (C2-C6) | * Errors in intervertebral angular and shear displacements no greater than 0.4° and 0.555 mm, respectively. * Alternating intervertebral motions in the cervical spine were typically found to correlate with patient-specific anatomical features such as disc height loss and loss of lordosis. |
| Kaga (2020) [35] | X-ray | Biomechanical Analysis | * Quantifying segmental motion in spine | * 2D/3D Shape-matching algorithm | 2D | 3D | Semi | n = 1 (C4-C6) | * System’s overall RMSE ranged between 0.21–0.49mm and 0.42–1.80. The RMSE associated with RSA ranged between 0.14–0.69mm and 0.96–2.33” for biaxial centroid identification and 0.25–1.19mm and 1.69–4.06” for dynamic head tracking. |
| Nikkloo (2019) [47] | X-ray | Biomechanical Analysis - FE model | * Investigate spine biomechanics associated with typical cervical disorders | * Finite Element Model | * Based on 36 X-ray parameters | 3D | Semi | n = 6 (C3-C7) | * Severe disc alteration (Grade 3) presented a significant decrease in the ROM and intradiscal pressure (flexion, extension, and axial rotation) on the C5-C6 and slightly increase on the adjacent levels. * Maximum stress in Anterior Fibrous (AF) and facet joint forces increased for Grade 3 for both altered and adjacent levels. |
| Nikkloo (2020) [46] | X-ray | Biomechanical Analysis - FE model | * Investigate the biomechanical impact of laminectomy on cervical intersegmental motion and load sharing | * Finite Element Model | * Geometrically specific | 3D | Semi | n = 10 (C3-C7) | * Post laminectomy increased the intersegmental ROM, disc stress, and intradiscal pressure at the upper cervical levels during sagittal plane motion and axial rotation, while the lower levels experienced the opposite trend, as compared with intact models. * No significant changes were observed in facet joint forces after surgery. |
| Senivvesan (2020) [55] | X-ray | Biomechanical Analysis - FE model | * Effect of Heterotopic Ossification after arthroplasty on the adjacent levels and change in range of motion (ROM) | * Finite Element Model | * Strain Energy Density (SED) | 3D | Semi | n = 7 (C2-T1) | * The Bryan disc significantly reduced ROM at the implanted level in flexion. However, in extension, ROM increased at the implanted level and decreased slightly at the adjacent levels. After HO, ROM drastically reduced at the implanted level in both extension and flexion, and showed a minor increase in the adjacent levels, indicating that biomechanical behavior of high-grade HO is similar to a fused segment, thereby reducing the intended and initial motion preservation. |
| Mawareni (2019) [42] | X-ray | CVM Stage | * Determine the Cervical Vertebra Maturity Degree | * 6-layered CNN (supervised classification) | 2D | Fully | n = 2470 (C2-C7) | * The results show the performances of the proposed method in different contexts with different number of |
| Study | Method | Stage | Technique | Model | Parameters | Participants | Results |
|-------|--------|-------|-----------|-------|------------|--------------|---------|
| Kok (2019) [17] | X-ray | CVM | Determine the Cervical Vertebral Maturation Degree | k-nearest neighbors (k-NN) | n = 200 (C2-C4) | * According to the average rank of the algorithms in predicting the CSV classes, ANN was the most stable algorithm with its 2.17 average rank. * kNN and Log Regr. algorithms had the lowest accuracy values. SVM-80%-Tree and NB algorithms had varying accuracy values. ANN could be the preferred method for determining CSV. |
| Amsalga (2020) [4] | X-ray | CVM | Determine the Cervical Vertebral Maturation Degree | Softmax activation function | n = 647 (C3-C4) | * Intrarater agreement was lower than Agreement between ANN and observers combined. Average of 58.3% agreement was observed between the ANN model and the human observers. * Developed ANN model performed close to, if not better than, human observers in CVM analysis. |
| Hopkins (2019) [30] | MRI | Clinical Prognosis / Prediction | Model 1: Predicting Cervical Spondylotic Myelopathy (CSM) | 7-layered deep neural network | n = 28 (C2-T1) | * Model 1: The mean cross-validated accuracy of the trained model was 86.50% (95% confidence interval, 85.16%-98.83%) with a median accuracy of 90.60%. Average sensitivity, specificity, positive predictive value, and negative predictive value were 90.25%, 85.05%, 81.58%, and 91.94%, respectively. * Model 2: The mJOA model predicted scores, with a mean and median error of 0.29 mJOA points and 0.08 mJOA points, respectively, mean error per batch was 0.714 mJOA points. |
| Jin (2019) [34] | MRI (DTI) | Clinical Prognosis / Prediction | Evaluate the potential of AI in the analysis of DTI for the prognosis of myelopathy | Logistic regression (LR) | n = 75 (C2-C7) | The accuracy of the classifications reached 74.2% ± 1.6% for LR, 85.6% ± 1.4% for KNN, 89.7% ± 1.6% for RBF-SVM, and 59.2% ± 3.8% for the deep learning model. The RBF-SVM algorithm achieved the best accuracy, with sensitivity and specificity of 85.0% ± 3.4% and 92.4% ± 1.9% respectively. |
| da Bois (2007) [21] | CT, MRI, PET | Image Registration / Planning | Develop a pipeline to register multimodal images of the neck | Linear elastic biomechanical finite element model | n = 7 (C2-C7) | * Significant decreases in the mean, min and max errors. Mean errors before registration [3.88–5.96 mm] decrease after registration [1.91–3.31 mm] and variances decreases from [1.26–1.44 mm] to [0.65–1.11 mm]. The minimum errors before registration [2.3–3.88 mm] become [0.2–1.17 mm] and the maximum errors [5.59–4.47 mm] decrease after registration [3.17–4.9 mm]. |
| Pekur (2007) [48] | MRI | Image Registration / Planning | Automate planning of MRI scans of the spine | Anatomy recognition algorithm | n = 15 (C2-C7) | * Results for detection of disc candidates: true detections 95.3%, * Visual evaluation of the validation study demonstrates the seemingly robust results for automated planning vs. manual planning. |
Funding

No funding was received for this research.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bas.2022.101666.

References

Al Arif, S.M.M.R., Gundry, M., Knapp, K., Slabaugh, G., 2016. Improving an Active Shape Model with Random Classification Forest for Segmentation of Cervical Vertebrae. Springer International Publishing, Cham, pp. 3–15.

Alaie, B.T., DePauwis Katula, M.C., Bruno, A.G., Samelson, E.J., Kiel, D.P., Anderson, D.E., Bouness, M.L., 2017. Evaluation of a new approach to compute intervertebral disc height measurements from lateral radiographic views of the spine. Eur. Spine J. 26 (1), 167–172.

Amaya, H., Genz, E., Yildirim, D., Orhan, K., 2020. Validation of cervical vertebrae maturation stages: artificial intelligence vs. observer visual analysis. Am. J. Orthod. Dentofacial Orthop. 158 (6), e123–e129.

Aresta, G., Araujo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Prastawa, M., Chan, M., Donovan, M., Kwak, J.T., Ludwig, F., Brousseau, W., Banet, M., Wu, Q.D., To, M.N.N., Kim, E., Kwak, J.T., Galal, S., Sanchez-Freire, V., Broun, F., Ricci, D., Wang, Y., Sun, L., Ma, K., Fang, J., Kone, L., Boulmene, L., Campilho, A., Aguiar, P., 2019. BACH: Grand challenge on breast cancer histology images. Med. Image Anal. 56, 122–139.

Funding

No funding was received for this research.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bas.2022.101666.

References

Al Arif, S., Knapp, K., Slabaugh, G., 2018. Fully automatic cervical vertebrae segmentation framework for X-ray images. Comput. Methods Progr. Biomed. 157, 95–111.
Wells, G., Shea, B., O’Connel, D., Welch, P.J.V., Losos, M., Tugwell, P. The NewcastleOttawa Scale (NOS) for assessing the quality of nonrandomised studies in metaanalysis. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. (Accessed 25 December 2020).

Xi, X., Hong-Wei, H., Xu-Cheng, Y., Ning, L., Shafin, S.H., 2012. Automatic Segmentation of Cervical Vertebrae in X-Ray Images, the 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8.

Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, L., Mak, R.H., Aerts, H., 2019. Deep learning predicts lung cancer treatment response from serial medical imaging. Clin. Cancer Res. 25 (11), 3266–3275.

Zamora, G., Sari-Sarraf, H., Long, R.L., 2003. Hierarchical Segmentation of Vertebrae from X-Ray Images. Proc.SPIE.

Zhang, Y., Lobo-Mueller, E.M., Karanicolas, P., Gallinger, S., Haider, M.A., Khalvati, F., 2020. CNN-based survival model for pancreatic ductal adenocarcinoma in medical imaging. BMC Med. Imag. 20 (1), 11.