Article

Artificial Intelligence Assisted Detection and Localization of Spinal Metastases

Edgars Edelmers [*] , Artūrs Ņikuļins , Klinta Luīze Sprūdža , Patrīcija Stapulone , Niks Saimons Pūce , Elizabete Skrebele , Everita Elīna Siņicina , Viktorija Cīrule , Ance Kazuša , Katrina Boločko

Article

Artificial Intelligence Assisted Detection and Localization of Spinal Metastases

Edgars Edelmers [1],*, Artūrs Ņikuļins [2], Klinta Luīze Sprūdža [3], Patrīcija Stapulone [4], Niks Saimons Pūce [5], Elizabete Skrebele [6], Everita Elīna Siņicina [7], Viktorija Cīrule [8], Ance Kazuša [9] and Katrina Boločko [10]

[1] Faculty of Medicine, Rīga Stradiņš University, LV-1010, Riga, Latvia

[2] Faculty of Computer Science, Information Technology and Energy, Riga Technical University, LV-1048, Riga, Latvia

[3] Faculty of Medicine, Rīga Stradiņš University, LV-1010, Riga, Latvia

[4] Faculty of Medicine, Rīga Stradiņš University, LV-1010, Riga, Latvia

[5] Faculty of Computer Science, Information Technology and Energy, Riga Technical University, LV-1048 Riga, Latvia

[6] Faculty of Civil and Mechanical Engineering, Riga Technical University, LV-1048, Riga, Latvia

[7] Faculty of Biology, University of Latvia, LV-1004, Riga, Latvia

[8] Faculty of Medicine, Department of Radiology, Rīga Stradiņš University, LV-1010, Riga, Latvia

[9] Faculty of Medicine, Rīga Stradiņš University, LV-1010, Riga, Latvia

[10] Department of Computer Graphics and Computer Vision, Riga Technical University, LV-1048, Riga, Latvia

* Correspondence: edgars.edelmers@rsu.lv

Abstract: The objective of this study was to develop and evaluate artificial intelligence (AI) models for the detection and instance segmentation of vertebrae and spinal metastases in computer tomography (CT) scans. The models were trained on datasets consisting of patients with polytrauma and relatively undamaged spines, as well as patients diagnosed with spinal metastases. Our results indicate that the models achieved high performance in vertebra segmentation, with F-beta scores ranging from 0.88 to 0.96 across all vertebra classes. For spinal metastases, the model attained F-beta scores of 0.68 for lytic type and 0.57 for sclerotic type metastases. Additionally, the models demonstrated the capability to detect isolated metastatic nodes in other bones, highlighting their robustness and potential for broader clinical application. Despite these promising results, the study faced limitations including a relatively small and homogeneous dataset, variability in segmentation mask quality, and the need for real-world clinical validation. Future clinical trials are necessary to evaluate the practical utility and effectiveness of these AI models in improving patient outcomes. Our study emphasizes the potential of AI-assisted detection and segmentation models in enhancing diagnostic accuracy and efficiency in clinical practice. The data used in this research are published on Zenodo platform.

Keywords: artificial intelligence; spinal metastases; vertebras segmentation; computer tomography; medical imaging; instance segmentation; radiomics

1. Introduction

1.1. Machine Learning in Healthcare

Radiomics presents itself as a novel discipline of artificial intelligence. It refers to the analysis of medical images extracting malignancy-associated information – pathophysiology significant data integrated into mathematical parameters. [1] There are clinical portfolios for the integration of radiomics. First, the initial step relates to a clearly defined target population and the use of radiomics, which leads to improvements compared with standard-of-care examination diagnostics. In a second aspect, the technical approach of statistical elements must be considered. Throughout image stratification, minimizing variability is suggested. Machine learning aspects such as possible bias and other computational parameters could be specified, for example, within Bayesian updating. In a third

aspect, the experiment is expected to have reproducible test performance. [2] Another important aspect reflects the European Union's General Data Protection Regulation to enhance transparency in clinical applications. [1]

A deep residual convolutional neural network is used for the detection of metastatic bone lesions with automatically segmented regions. Deep-learning segmentation can be adapted into Computed Tomography scan (CT scan) and Magnetic Resonance Imaging (MRI). [3] Furthermore, a CT scan can reach up to 74% sensitivity and 56% specificity. CT scan provides whole skeletal system assessment and systemic staging with the benefit minimizing patient`s radioactivity performance. Moreover, MRI shows the finest resolution for bone and soft tissue. The proposed sensitivity is 95% and specificity is 90%. [1]

Bone metastases can be diagnosed using a variety of imaging modalities, each offering specific advantages and limitations. Plain radiographs (X-rays) are typically the first imaging method for patients presenting with bone pain. However, X-rays have limited utility in asymptomatic patients and for evaluating bones with a high cortex-to-marrow ratio, such as ribs, due to their low sensitivity in detecting subtle or early-stage metastases. [4] Computed tomography (CT) is preferred for these cases, offering superior resolution of cortical and trabecular bone structures, and allowing for adjustments in window width and level. CT scans provide a detailed view across multiple planes, improving diagnostic sensitivity and specificity. According to recent studies, CT has a pooled sensitivity of 72.9% and a specificity of 94.8% in detecting bone metastases, particularly in areas like the ribs. [5]

In oncology, CT is commonly used for staging and follow-up in cancers affecting the thorax and abdomen, as it covers large portions of the axial skeleton. This enables clinicians to not only assess the spread of metastatic disease but also distinguish between metastatic and degenerative changes. CT also assists in evaluating structural abnormalities that may have been detected with other modalities, such as magnetic resonance imaging (MRI) and scintigraphy. [6]

Although other imaging modalities, such as MRI, positron emission tomography/computed tomography (PET/CT), single-photon emission computed tomography (SPECT), and bone scintigraphy, offer higher sensitivity (91% for MRI, 90% for PET/CT, and 86% for scintigraphy), these methods are often not used for routine screening or follow-up due to their higher cost and limited availability, particularly in countries with constrained healthcare resources. [7] For instance, MRI is typically reserved for cases requiring detailed soft-tissue contrast or assessing bone marrow involvement, while PET/CT and scintigraphy are often used in specific cases but are less frequently employed for general screening due to expense and logistical challenges. [8]

Selecting the appropriate imaging modality based on clinical needs is crucial for optimizing diagnostic accuracy. CT remains an accessible and reliable choice for detecting bone metastases, especially in complex areas like the ribs, where other imaging techniques may offer limited utility.

Artificial intelligence (AI) is a rapidly advancing technology that has demonstrated significant potential across various domains, including medicine. AI is poised to transform numerous aspects of the medical field, such as patient care, administrative processes, diagnostics, treatment planning, and scientific research. In radiology, AI is often used in conjunction with radiomics, a technique that extracts quantitative features from medical images to uncover patterns not visible to the human eye. These radiomic features, when combined with AI methodologies such as machine learning and deep learning, can enhance diagnostic accuracy, prognostic predictions, and personalized treatment strategies. AI technologies, including machine learning, natural language processing, and robotics, can be applied independently or synergistically to analyse clinical data, generate reports, assist in diagnosing conditions, and predict treatment outcomes based on patient-specific variables. The integration of radiomics and AI has the potential to refine medical imaging analysis, offering deeper insights into disease characterization and treatment efficacy. [9]

Artificial intelligence makes sense of massive amounts of clinical, genomic, and imaging data. It can improve physician efficiency, increase diagnostic accuracy, and personalize treatment. Perioperative management, rehabilitation assistance, drug production, and new specialist education are only a few of the fields where artificial intelligence is showing its endless possibilities.

Higher diagnostic accuracy prevents unnecessary tests that take time and finances and can be a psychological burden to patients, as well as potentially dangerous due to use of ionising radiation and toxic contrast media. Based on research, artificial intelligence is showing promise with high accuracy in diagnostics of the following specializations:

- Radiology: recognition of tuberculosis in chest X-ray images, benign and malignant nodules in lungs based on CT data, breast cancer lesions in mammography and detection and classification of other tumours.
- Pathology: differentiation of melanocytic lesions, gastric cancer types, prediction of gene mutations associated with cancer, determination of kidney function from biopsy results.
- Ophthalmology: diagnostic of retinal diseases, glaucoma, keratoconus, grading cataracts.
- Cardiology: improvement in cardiovascular risk and pulmonary hypertension patient's outcome prediction accuracy.
- Gastroenterology: endoscopic detection of colorectal polyps, gastric and esophageal cancer, Barret's oesophagus, squamous carcinoma cell, and other lesions.
  Artificial intelligence can be also used in the therapy processes:
- Three-dimensional printing: it is possible to print individual models from patient's data that help doctors get more visualization, make more detailed preoperative plans, and practice simulated surgery in advance. This technology can be also used with bioactive materials to make body implants.
- Virtual reality: an opportunity for surgeons to practice, improve skills, get assistance during surgeries.
- Da Vinci surgical artificial intelligence system: it has proven to be minimally invasive, provide a clearer image, make operation more accurate and convenient, provide the possibility to do the operation remotely, lower complication rate, and be beneficial in terms of postoperative recovery. [10]

The need for health care services and the advancement of artificial intelligence have resulted in the creation of conversational agents – chatbots, speech recognition screening systems – that can assist with various health-related tasks such as change of behaviour, treatment support, health monitoring, training, triage, and screening. Most studies have shown that these conversational agents are generally effective and satisfactory. [11]

Jobs in healthcare specialties involving digital information, like radiology and pathology, are more likely to get automated than those requiring direct doctor-patient contact. However, artificial intelligence is not expected to replace healthcare specialists but to support their skills and help them put in more effort in patient care and jobs that require unique human skills like empathy, persuasion, and big-picture integration. Artificial intelligence is expected to have a significant impact on the future of medicine, but its integration in healthcare also presents ethical, legal, and practical challenges that need to be carefully addressed and limited. However, further research is needed to fully understand the long-term effects and ensure the safe and effective integration of artificial intelligence-based technologies into healthcare. [12]

1.2. Vertebral Metastases

Vertebral metastases represent the secondary involvement of the vertebral spine by hematogenous disseminated metastatic cells. [13] It is the 3rd most common site of metastasis, after lungs and liver and is a major cause for morbidity, characterized by severe pain, impaired mobility, pathologic fractures etc. [14] In 90% of cases they are asymptomatic and are present in 60-70% of patients with systemic cancer.

80% of primary tumours give off bone metastases. [14] They are classified as osteolytic, osteoblastic or mixed, according to the primary mechanism of interference with normal bone remodelling. Osteolytic characterized by destruction of normal bone and sclerotic by deposition of new bone. [15] Primary tumours with predominantly osteolytic metastases are breast cancer (65-75%), thyroid cancer (65-75%), urothelial cancer (20-25%), renal cell carcinoma (20-25%), melanoma (14-45%), non-Hodgkin lymphoma and multiple myeloma. Primary tumours with predominantly sclerotic metastases are prostate (60-80%), small cell lung cancer (30-40%), Hodgkin lymphoma. [13–

15] Mixed type of lesions is present in breast cancer (15-20%), gastrointestinal cancers and squamous cancers. [15]

Nowadays incidence of spinal metastatic disease increases due to improved patient survival and advanced diagnostic techniques. [16] The median-survival from diagnosis of bone metastasis is: 6 months in melanoma; 6-7 months in lung; 6-9 months in bladder; 12 months in renal cells carcinoma; 12-53 months in prostate; 19-25 months in BC and 48 months in thyroid. [15]

Unfortunately, no treatment has been proven to increase the life expectancy of patients with spinal metastases. The goals of therapy are pain control and functional preservation. [17] Thus it is important not only to diagnose spinal metastases, but also to follow progression of disease, evaluate stability of the vertebral column and identify patients who will benefit from surgical consultation or intervention.

There are multiple scoring systems available for evaluation of different aspects of well-being of a patient with metastatic spine disease. One of them is The Spinal Instability Neoplastic Score (SINS), which is used to evaluate spinal instability. It also acts as a prognostic tool for surgical decision making. [18]

The SINS is a scoring system based on one clinical factor (pain) and five radiographic parameters (location, bone lesion quality, spinal alignment, vertebral body collapse and involvement of posterolateral spinal elements). Each component is assigned a score reflecting its contribution to overall instability of the spinal segment. The six individual scores of each component are added for a cumulative score ranging from 0 to 18. The higher the total score signals more severe instability. [19]

In the evaluation of 131 surgically stabilized spine metastasis patients the SINS [18] demonstrated near-perfect inter- and intra-observer reliability in determining three clinically relevant categories of stability. With SINS ≥7 surgical stabilization significantly improved patients' quality of life. [16]

The presented research serves as a proof of concept for the upcoming project in terms of which we plan to create a representative cohort group with even age, sex and oncology stage distribution among the dataset data.

Our main objective is to locate metastases in patient computer tomography if there are such. On daily bases small, vaguely visible occurrences of metastases in computer tomography could be easily missed by healthcare professionals. A well-tuned system which could indicate the region of possible disease could be crucial for a patient's life. If there is any deformation of the bone morphology structure, the artificial intelligence can record it. What has been done manually by professionals could be done faster and with better quality by AI. After deformation region indication healthcare professionals can decide is it a metastasis or not.

U-net segmentation architecture initially was built for medical imaging data analyses. [20] Architecture as an output provide segmentation masks with the same size as input, which is ideal indication of possible metastases. For this project, a 3D version of the U-net architecture is utilized to work with the 3-dimensional nature of computer tomography data along with the 2D version. [20,21]

One study showed that a deep-learning algorithm (DLA) could help radiologists detect possible spinal cancers on CT scans. The system, which uses a U-Net-like architecture, had a sensitivity of 75% in identifying possibly malignant spinal bone lesions, considerably boosting radiologists' ability to detect incidental lesions that would otherwise go unnoticed due to scan focus or diagnostic bias. In this situation, AI serves as a second reader, significantly increasing detection sensitivity without leading. [22]

Another important component of AI in spinal metastatic imaging is its involvement in early detection and therapy, which is key for avoiding problems and enhancing patient quality of life. Recent research has looked into the use of AI approaches in image processing, diagnosis, decision support, and therapy aid, summarizing the current evidence of AI applications in spinal metastasis care. These technologies have demonstrated encouraging outcomes in boosting work productivity and lowering unpleasant events, but further study is needed to evaluate clinical performance and allow adoption into ordinary practice. [23]

There is a similar study that introduced a deep learning (DL) algorithm designed for diagnosing lumbar spondylolisthesis using lateral radiographs. This research aimed at improving the accuracy of medical diagnostics by assisting doctors in reducing errors in disease detection and treatment. The study was retrospective, involving multiple institutions, and focused on patients with lumbar spondylolisthesis. The DL model utilized included Faster R-CNN and RetinaNet for lumbar spondylolisthesis detection, showing the potential of AI to significantly enhance diagnostic accuracy in spinal condition. [24]

2. Materials and Methods

Our research is structured into two distinct stages, employing a previously validated methodology for the preprocessing of radiological data. [25] The initial stage focuses on the localization of the patient's spine, a critical step for highlighting regions potentially affected by spinal metastases. In this stage, each vertebra is isolated and segmented starting from the cervical region down to the lower spine, including the combination of sacrum and coccyx. The second stage is dedicated to the identification of metastases through the application of segmentation masks, identifying two types of metastases in the spinal region: lytic and sclerotic. Our methodology employs two U-Net based neural networks: one network is trained to locate the spine, while the second network is trained for the instance segmentation of metastases, including type prediction. This structured approach ensures precise localization and identification of spinal metastases, facilitating targeted clinical interventions.

Before training, the data were converted into a "nearly raw raster data" format to optimize input/output operations during the training process. The first dataset for vertebra segmentation comprises 115 patients diagnosed with polytrauma but with relatively undamaged spines, all of whom underwent full-body CT scanning at the local hospital RAKUS (Rīgas Austrumu klīniskā universitātes slimnīca).

The second dataset, used for metastasis detection, consists of patients diagnosed with spinal metastases and is detailed in Table 1. [26]

Table 1. Descriptive statistics of the dataset utilized for training the vertebra segmentation model.

Sex	Age	Metastasis Type	Primary Metastatic Site
Female	57	Sclerotic	Melanoma
Female	72	Sclerotic	Lungs
Female	72	Sclerotic	Lungs
Female	68	Lytic	Ovary
Female	39	Sclerotic	Breast
Male	74	Lytic	Prostate
Female	82	Sclerotic	Breast
Female	82	Sclerotic	Breast
Female	64	Sclerotic	Breast
Female	65	Sclerotic	Breast
Female	65	Sclerotic	Breast
Female	61	Sclerotic	Breast
Female	45	Sclerotic	Breast
Female	45	Sclerotic	Breast
Female	70	Sclerotic	Breast
Male	66	Sclerotic	Lungs
Female	52	Sclerotic	Breast

Male	53	Lytic	Kidney
Female	60	Sclerotic	Breast
Male	74	Sclerotic	Blader
Female	79	Lytic	Kidney
Female	48	Lytic	Ovary
Male	66	Sclerotic	Large intestine
Female	73	Lytic	Multiple myeloma
Male	66	Lytic	Multiple myeloma
Female	61	Sclerotic	Breast
Female	73	Lytic	Breast
Female	79	Lytic	Kidney
Female	48	Lytic	Ovary
Male	75	Lytic	Stomach
Male	75	Lytic	Stomach
Male	64	Lytic	Kidney
Female	39	Lytic	Ovary
Male	55	Lytic	Multiple myeloma
Female	60	Lytic	Multiple myeloma
Female	70	Lytic	Breast
Female	32	Lytic	Multiple myeloma
Female	61	Lytic	Kidney

We have utilized the built-in validation mechanism of the nnUnet library in the form of 5 cross-validation. In its turn, all the segmentation masks for the training data were created by medical professionals from Rigas Stradiņš University using the 3D Slicer software platform for medical image informatics, image processing, and three-dimensional visualization. [27] Following data preparation, four U-Net architecture subtypes were trained: 2D images in the form of single slices from CT scans, 3D low resolution with downsampled input image data, 3D full resolution utilizing the original resolution of the CT scans, and 3D cascade full resolution which uses downsampled images to understand the overall structure on a large scale before learning details from the full resolution image data as is stated in the utilized library. [28] The best performing subtype was selected for testing purposes.

3. Results

The process begins with identifying the Region of Interest (ROI), with the spine being the ROI for metastases in CT scans. In the initial stage, the spine is segmented to facilitate metastases analysis. The U-Net architecture, trained using cross-validation, employs a total loss function combining Dice loss and cross-entropy loss. The model is trained on patches extracted from the original image, with the Dice metric calculated based on these patches. Post-training, the sliding window method is used for inference, which may result in patches different from those used during training, potentially decreasing the Dice score. Validation patches are sampled similarly to training patches, and the Dice coefficient is computed over all sampled patches collectively. **Figures 1** and **2** depict the pseudo-Dice metric, which is used to monitor training progress and detect overfitting. This metric is updated continuously and is distinct from the actual Dice coefficient.
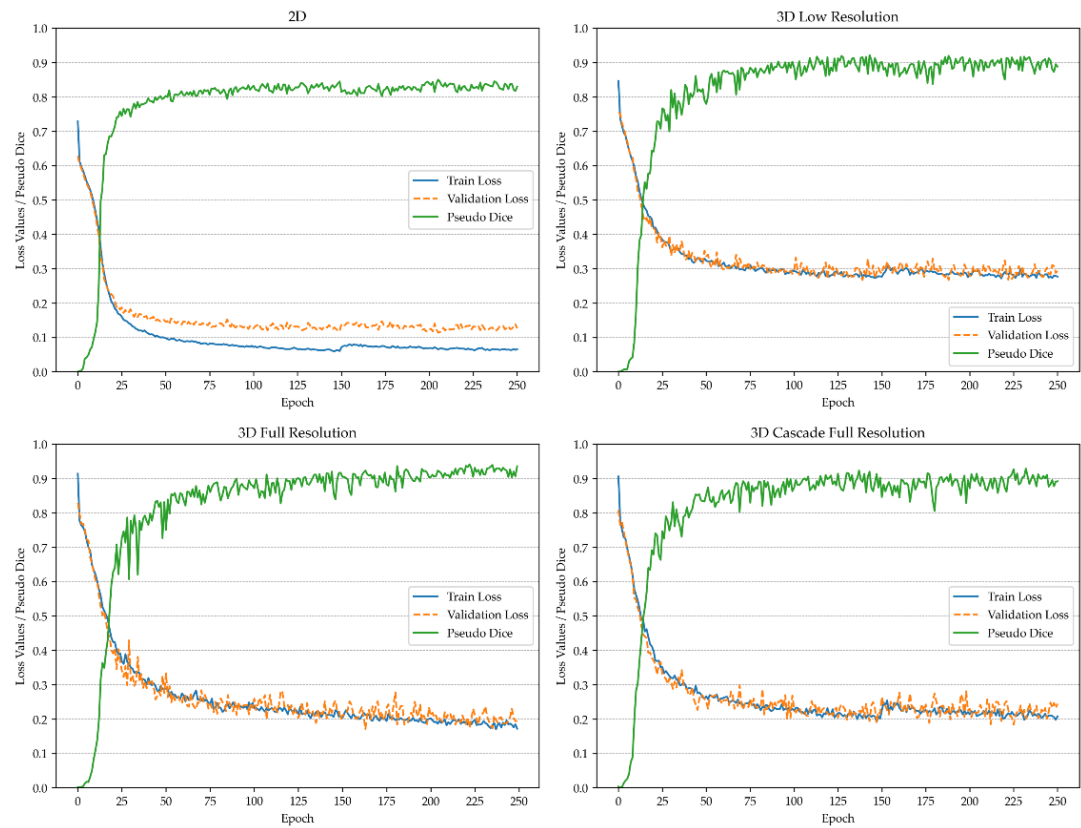
Figure 1. Training process of the model for vertebra detection and instance segmentation.
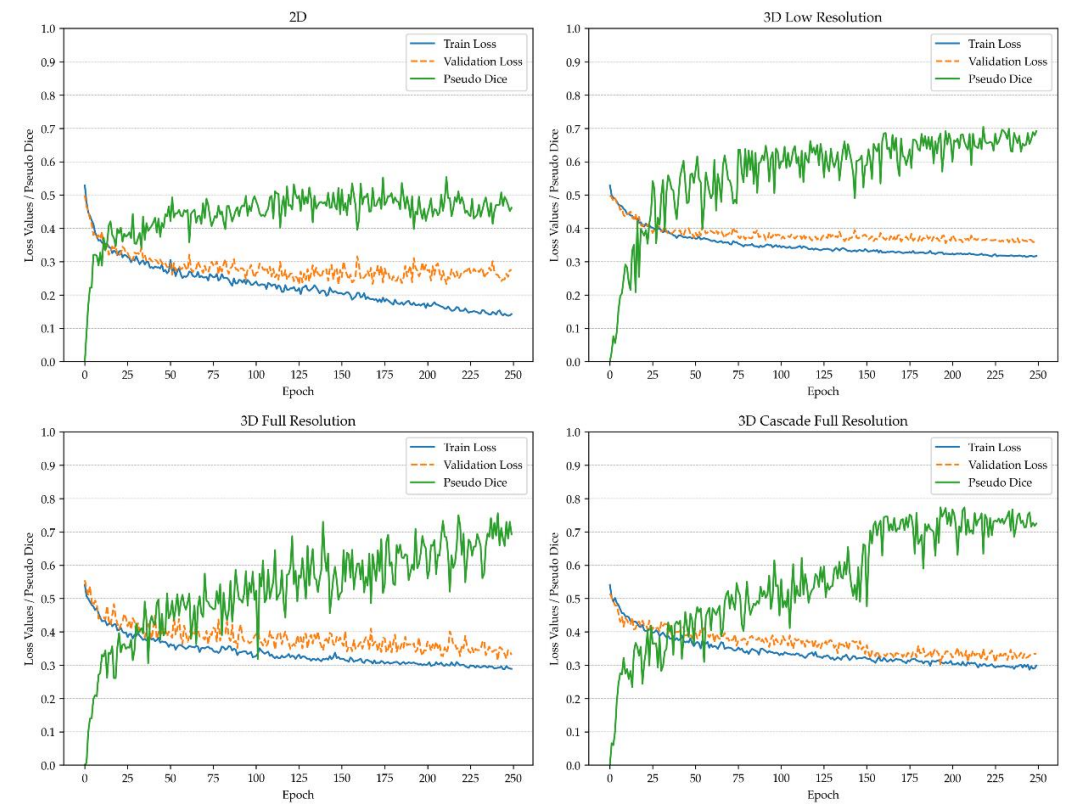

Figure 2. Training process of the model for spinal metastasis detection and instance segmentation.

The actual Dice similarity coefficient, computed at the end of the training, is not calculated on patches but on the whole images using a sliding window approach. This provides the actual performance metrics, whereas the pseudo-Dice metrics are used solely to sanity-check the training process. All metrics are calculated based on validation, utilizing the network architecture that demonstrated the best performance during the training process. For vertebra segmentation, these metrics are depicted in Table 2, while for metastasis instance segmentation, they are shown in Table 3.

Table 2. Metrics for the evaluation of the "3D Full Resolution" model for vertebra detection and instance segmentation.

Vertebra	Dice Similarity Coefficient	F-beta Score	Panoptic Quality
C1	0.94	0.94	0.75
C2	0.95	0.95	0.82
C3	0.93	0.93	0.75
C4	0.93	0.93	0.75
C5	0.93	0.94	0.75
C6	0.93	0.93	0.75
C7	0.94	0.93	0.79
T1	0.94	0.94	0.81
T2	0.95	0.95	0.83
T3	0.95	0.95	0.82
T4	0.95	0.95	0.83
T5	0.94	0.94	0.82
T6	0.88	0.87	0.69
T7	0.87	0.88	0.70
T8	0.91	0.92	0.75
T9	0.93	0.93	0.77
T10	0.94	0.94	0.81
T11	0.95	0.95	0.85
T12	0.95	0.94	0.84
L1	0.95	0.94	0.83
L2	0.94	0.94	0.83
L3	0.93	0.92	0.81
L4	0.94	0.89	0.84
L5	0.95	0.94	0.86
Sacrum	0.96	0.96	0.89

Table 3. Metrics for the evaluation of the "3D Cascade Full Resolution" model for metastasis instance segmentation.

Metastasis type	Dice Similarity Coefficient	F-beta Score	Panoptic Quality
Lytic	0.71	0.68	0.45
Sclerotic	0.61	0.57	0.30

4. Discussion

Mathematical metric properties are often neglected, for example, when using the Dice similarity coefficient (DSC) in the presence of particularly small structures, making it a poor metric for metastasis segmentation. Recent research suggests that more preferable metrics for evaluating model performance in instance segmentation tasks are the F-beta score and panoptic quality. [29] However, the panoptic quality metric (presented in **Table 3**) has been severely criticized for use in cases rich in small segmentations with variable shapes and where the background is treated as a separate class, as is the case in metastasis segmentation. [30] Consequently, the F-beta score has been adjusted to suit the needs of oncological diagnostics, where an F-beta score with greater emphasis on recall is preferable, such as when minimizing false negatives is critical but false positives remain significant. This adjustment resulted in a beta value of 2.

The highest overall performance for the vertebra segmentation model was achieved using the 3D full-resolution architecture. The predicted mask utilizing this model is demonstrated in Figure 3.
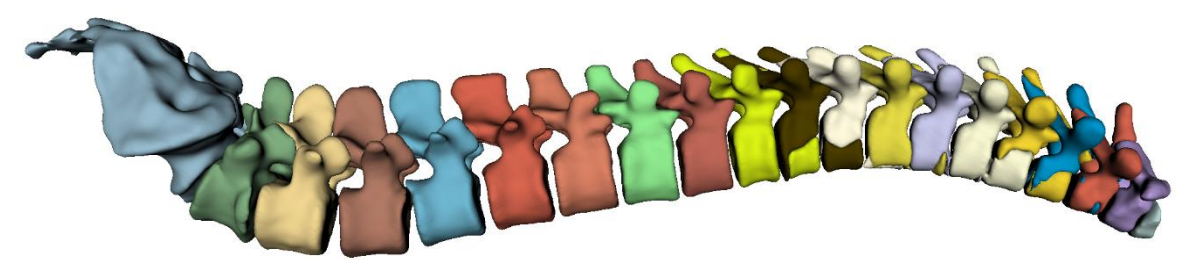

Figure 3. Predicted mask utilizing vertebras segmentation model.

The highest overall performance for the metastasis segmentation model was achieved using the 3D full-resolution architecture. The predicted mask utilizing this model is demonstrated in Figure 4. It worth to mention that the model was able to correctly detect and segment metastasis not only in spine, but also in others bone structures, which can be seen in aforementioned figure (sternum).
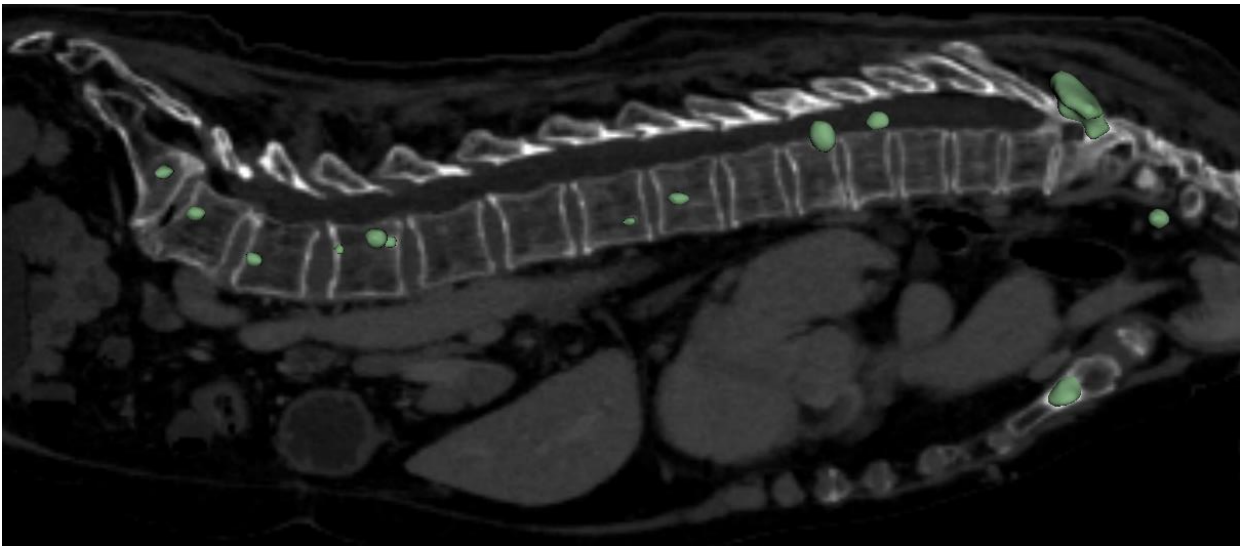
Figure 4. Predicted mask utilizing metastases segmentation model.

Recent studies on spinal metastasis segmentation using deep learning have shown notable advancements, particularly in MRI applications. Our research aligns with these trends, offering valuable metrics that demonstrate the model's performance across lytic and sclerotic metastases.

For lytic metastases, our model achieved a Dice Similarity Coefficient (DSC) of 0.71, an F-beta score of 0.68, and a Panoptic Quality of 0.45. These results are comparable to Kim et al. (2024), who reported a mean per-lesion sensitivity of 0.746 and a positive predictive value of 0.701 using a U-Net-based model. [31] Similarly, Liu et al. (2021) achieved a precision of 0.76 and a recall of 0.67 for pelvic bone metastasis segmentation. [10] Our F-beta score and DSC values indicate a strong ability to detect and segment lytic metastases, consistent with these findings.

However, the performance for sclerotic metastases was lower, with a DSC of 0.61, an F-beta score of 0.57, and a Panoptic Quality of 0.30. Ong et al. (2022) reported a DSC of up to 0.78 for spinal metastasis segmentation with sensitivity rates of 78.9%, suggesting that sclerotic metastases pose greater challenges due to their subtle imaging characteristics. [32] Our results reflect this difficulty, with lower scores across all metrics for sclerotic lesions.

It is worth highlighting the VerSe: Large Scale Vertebrae Segmentation Challenge, where state-of-the-art performance achieved a mean vertebrae identification rate of 96.6% and a Dice coefficient of 91.7%. The challenge made a significant contribution to the field by providing a dataset of 374 multi-detector CT scans, which has been instrumental in advancing vertebrae segmentation research. [33] In comparison, our dataset is considerably smaller, and our training process was shorter (250 epochs). Despite these limitations in dataset size and training duration, our model demonstrated competitive performance, particularly for lytic metastasis segmentation, though there remains room for improvement in handling more complex lesion types like sclerotic metastases. This comparison underscores the potential of our approach even with more constrained computational and data resources.

Despite the promising results of our study, several limitations need to be acknowledged: The dataset used for training the model primarily consisted of patients from a single medical centre, which may limit the generalizability of the results. A more diverse and larger dataset would be beneficial to improve the robustness of the model across different populations and imaging conditions. Additionally, the quality of the segmentation masks created by medical professionals could introduce variability, as inter-observer variability in manual segmentations can affect the training quality and performance evaluation of the model. Furthermore, while the model shows relatively high accuracy in detecting spinal metastases in the provided dataset, it has not yet been validated in a real-world clinical setting. Further clinical trials and validations are required to assess its practical utility and effectiveness in improving patient outcomes.

Another critical aspect to consider is the technical limitations related to the scanning protocols, which can vary significantly between cases. For optimal performance, inference should be conducted on radiological data that closely match the technical properties of the data used to train the model. The resolution of the radiological image data is particularly crucial, as variations in resolution can significantly impact the model's accuracy and effectiveness. The relatively limited dataset for metastasis detection negatively impacted the overall performance of the model, highlighting an area for enhancement. Expanding the dataset to include a broader range of cases with diverse patient demographics and imaging conditions would likely improve the model's robustness and generalizability. A larger dataset would also allow for better training and validation of the model, enabling it to accurately detect and segment metastases across various clinical scenarios. This enhancement is crucial for increasing the model's diagnostic accuracy and effectiveness in real-world applications. [34,35]

5. Conclusions

In this study, we have developed and trained two models for the detection and instance segmentation of vertebrae and spinal metastases using advanced AI techniques. The results

demonstrate that our models are capable of accurately identifying and segmenting these structures, providing a valuable tool for clinical applications. Specifically, the F-beta score for vertebra segmentation ranged from 0.88 to 0.96 across all vertebra classes, indicating high performance in this task. For spinal metastases, the model achieved an F-beta score of 0.68 for lytic type and 0.57 for the sclerotic type, showcasing the potential of our approach in identifying different types of metastases.

Furthermore, our model exhibited the ability to detect isolated metastatic nodes in other bones, extending its utility beyond the primary objective of spinal metastasis detection. This capability underscores the model's robustness and versatility in different clinical scenarios, providing comprehensive diagnostic support.

Author Contributions: Conceptualization, E.E.; methodology, E.E.; software, E.E., A.N.; validation, E.E.; formal analysis, E.E.; investigation, E.E.; resources, E.E., K.B., V.C.; data curation, E.E., A.N., K.L.S., P.S., N.P., E.S., E.S., A.K.; writing—original draft preparation, E.E., A.N.; writing—review and editing, E.E.; visualization, E.E.; supervision, E.E.; project administration, E.E.; funding acquisition, E.E. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethics Committee of Rīga Stradiņš University (2-PĒK-4/584/2023 at 20.10.2023).

Data Availability Statement: https://doi.org/10.5281/zenodo.13645871

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dong, X.; Chen, G.; Zhu, Y.; Ma, B.; Ban, X.; Wu, N.; Ming, Y. Artificial Intelligence in Skeletal Metastasis Imaging. *Computational and Structural Biotechnology Journal* **2024**, *23*, 157–164, doi:10.1016/j.csbj.2023.11.007.
2. Huang, E.P.; O'Connor, J.P.B.; McShane, L.M.; Giger, M.L.; Lambin, P.; Kinahan, P.E.; Siegel, E.L.; Shankar, L.K. Criteria for the Translation of Radiomics into Clinically Useful Tests. *Nat Rev Clin Oncol* **2023**, *20*, 69–82, doi:10.1038/s41571-022-00707-0.
3. Liu, S.; Feng, M.; Qiao, T.; Cai, H.; Xu, K.; Yu, X.; Jiang, W.; Lv, Z.; Wang, Y.; Li, D. Deep Learning for the Automatic Diagnosis and Analysis of Bone Metastasis on Bone Scintigrams. *CMAR* **2022**, *Volume 14*, 51–65, doi:10.2147/CMAR.S340114.
4. Heindel, W.; Gübitz, R.; Vieth, V.; Weckesser, M.; Schober, O.; Schäfers, M. The Diagnostic Imaging of Bone Metastases. *Dtsch Arztebl Int* **2014**, *111*, 741–747, doi:10.3238/arztebl.2014.0741.
5. Woo, S.; Suh, C.H.; Kim, S.Y.; Cho, J.Y.; Kim, S.H. Diagnostic Performance of Magnetic Resonance Imaging for the Detection of Bone Metastasis in Prostate Cancer: A Systematic Review and Meta-Analysis. *European Urology* **2018**, *73*, 81–91, doi:10.1016/j.eururo.2017.03.042.
6. Ellmann, S.; Beck, M.; Kuwert, T.; Uder, M.; Bäuerle, T. Multimodal Imaging of Bone Metastases: From Preclinical to Clinical Applications. *Journal of Orthopaedic Translation* **2015**, *3*, 166–177, doi:10.1016/j.jot.2015.07.004.
7. Liu, F.; Dong, J.; Shen, Y.; Yun, C.; Wang, R.; Wang, G.; Tan, J.; Wang, T.; Yao, Q.; Wang, B.; et al. Comparison of PET/CT and MRI in the Diagnosis of Bone Metastasis in Prostate Cancer Patients: A Network Analysis of Diagnostic Studies. *Front Oncol* **2021**, *11*, 736654, doi:10.3389/fonc.2021.736654.
8. Orcajo-Rincon, J.; Muñoz-Langa, J.; Sepúlveda-Sánchez, J.M.; Fernández-Pérez, G.C.; Martínez, M.; Noriega-Álvarez, E.; Sanz-Viedma, S.; Vilanova, J.C.; Luna, A. Review of Imaging Techniques for Evaluating Morphological and Functional Responses to the Treatment of Bone Metastases in Prostate and Breast Cancer. *Clin Transl Oncol* **2022**, *24*, 1290–1310, doi:10.1007/s12094-022-02784-0.
9. Davenport, T.; Kalakota, R. The Potential for Artificial Intelligence in Healthcare. *Future Healthc J* **2019**, *6*, 94–98, doi:10.7861/futurehosp.6-2-94.
10. Liu, P.; Lu, L.; Zhang, J.; Huo, T.; Liu, S.; Ye, Z. Application of Artificial Intelligence in Medicine: An Overview. *CURR MED SCI* **2021**, *41*, 1105–1115, doi:10.1007/s11596-021-2474-3.

11. Milne-Ives, M.; De Cock, C.; Lim, E.; Shehadeh, M.H.; De Pennington, N.; Mole, G.; Normando, E.; Meinert, E. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. *J Med Internet Res* **2020**, *22*, e20346, doi:10.2196/20346.

12. Ahuja, A.S. The Impact of Artificial Intelligence in Medicine on the Future Role of the Physician. *PeerJ* **2019**, *7*, e7702, doi:10.7717/peerj.7702.

13. Rasuli, B.; Dawes, L. Vertebral Metastases. In *Radiopaedia.org*; Radiopaedia.org, 2008.

14. Shah, L.M.; Salzman, K.L. Imaging of Spinal Metastatic Disease. *International Journal of Surgical Oncology* **2011**, *2011*, 1–12, doi:10.1155/2011/769753.

15. Macedo, F.; Ladeira, K.; Pinho, F.; Saraiva, N.; Bonito, N.; Pinto, L.; Gonçalves, F. Bone Metastases: An Overview. *Oncol Rev* **2017**, doi:10.4081/oncol.2017.321.

16. Mossa-Basha, M.; Gerszten, P.C.; Myrehaug, S.; Mayr, N.A.; Yuh, W.T.; Jabehdar Maralani, P.; Sahgal, A.; Lo, S.S. Spinal Metastasis: Diagnosis, Management and Follow-Up. *BJR* **2019**, *92*, 20190211, doi:10.1259/bjr.20190211.

17. Wibmer, C.; Leithner, A.; Hofmann, G.; Clar, H.; Kapitan, M.; Berghold, A.; Windhager, R. Survival Analysis of 254 Patients After Manifestation of Spinal Metastases: Evaluation of Seven Preoperative Scoring Systems. *Spine* **2011**, *36*, 1977–1986, doi:10.1097/BRS.0b013e3182011f84.

18. Fox, S.; Spiess, M.; Hnenny, L.; Fourney, Daryl.R. Spinal Instability Neoplastic Score (SINS): Reliability Among Spine Fellows and Resident Physicians in Orthopedic Surgery and Neurosurgery. *Global Spine Journal* **2017**, *7*, 744–748, doi:10.1177/2192568217697691.

19. Murtaza, H.; Sullivan, C.W. Classifications in Brief: The Spinal Instability Neoplastic Score. *Clin Orthop Relat Res* **2019**, *477*, 2798–2803, doi:10.1097/CORR.0000000000000923.

20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. **2015**, doi:10.48550/ARXIV.1505.04597.

21. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation 2016.

22. Gilberg, L.; Teodorescu, B.; Maerkisch, L.; Baumgart, A.; Ramaesh, R.; Gomes Ataide, E.J.; Koç, A.M. Deep Learning Enhances Radiologists' Detection of Potential Spinal Malignancies in CT Scans. *Applied Sciences* **2023**, *13*, 8140, doi:10.3390/app13148140.

23. Ong, W.; Zhu, L.; Zhang, W.; Kuah, T.; Lim, D.S.W.; Low, X.Z.; Thian, Y.L.; Teo, E.C.; Tan, J.H.; Kumar, N.; et al. Application of Artificial Intelligence Methods for Imaging of Spinal Metastasis. *Cancers* **2022**, *14*, 4025, doi:10.3390/cancers14164025.

24. Zhang, J.; Lin, H.; Wang, H.; Xue, M.; Fang, Y.; Liu, S.; Huo, T.; Zhou, H.; Yang, J.; Xie, Y.; et al. Deep Learning System Assisted Detection and Localization of Lumbar Spondylolisthesis. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1194009, doi:10.3389/fbioe.2023.1194009.

25. Edelmers, E.; Kazoka, D.; Bolocko, K.; Sudars, K.; Pilmane, M. Automatization of CT Annotation: Combining AI Efficiency with Expert Precision. *Diagnostics* **2024**, *14*, 185, doi:10.3390/diagnostics14020185.

26. Edelmers, E. CT Scans of Spine with Metastases (Lytic, Sclerotic) 2024.

27. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.-C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging* **2012**, *30*, 1323–1341, doi:10.1016/j.mri.2012.05.001.

28. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat Methods* **2021**, *18*, 203–211, doi:10.1038/s41592-020-01008-z.

29. Maier-Hein, L.; Reinke, A.; Godau, P.; Tizabi, M.D.; Buettner, F.; Christodoulou, E.; Glocker, B.; Isensee, F.; Kleesiek, J.; Kozubek, M.; et al. Metrics Reloaded: Recommendations for Image Analysis Validation. *Nat Methods* **2024**, *21*, 195–212, doi:10.1038/s41592-023-02151-z.

30. Foucart, A.; Debeir, O.; Decaestecker, C. Panoptic Quality Should Be Avoided as a Metric for Assessing Cell Nuclei Segmentation and Classification in Digital Pathology. *Sci Rep* **2023**, *13*, 8614, doi:10.1038/s41598-023-35605-7.

31. Kim, D.H.; Seo, J.; Lee, J.H.; Jeon, E.-T.; Jeong, D.; Chae, H.D.; Lee, E.; Kang, J.H.; Choi, Y.-H.; Kim, H.J.; et al. Automated Detection and Segmentation of Bone Metastases on Spine MRI Using U-Net: A Multicenter Study. *Korean J Radiol* **2024**, *25*, 363, doi:10.3348/kjr.2023.0671.

32. Ong, W.; Zhu, L.; Zhang, W.; Kuah, T.; Lim, D.S.W.; Low, X.Z.; Thian, Y.L.; Teo, E.C.; Tan, J.H.; Kumar, N.; et al. Application of Artificial Intelligence Methods for Imaging of Spinal Metastasis. *Cancers* **2022**, *14*, 4025, doi:10.3390/cancers14164025.

33. Sekuboyina, A.; Husseini, M.E.; Bayat, A.; Löffler, M.; Liebl, H.; Li, H.; Tetteh, G.; Kukačka, J.; Payer, C.; Štern, D.; et al. VerSe: A Vertebrae Labelling and Segmentation Benchmark for Multi-Detector CT Images. *Medical Image Analysis* **2021**, *73*, 102166, doi:10.1016/j.media.2021.102166.

34. Papalia, G.F.; Brigato, P.; Sisca, L.; Maltese, G.; Faiella, E.; Santucci, D.; Pantano, F.; Vincenzi, B.; Tonini, G.; Papalia, R.; et al. Artificial Intelligence in Detection, Management, and Prognosis of Bone Metastasis: A Systematic Review. *Cancers* **2024**, *16*, 2700, doi:10.3390/cancers16152700.

35. Koike, Y.; Yui, M.; Nakamura, S.; Yoshida, A.; Takegawa, H.; Anetai, Y.; Hirota, K.; Tanigawa, N. Artificial Intelligence-Aided Lytic Spinal Bone Metastasis Classification on CT Scans. *Int J CARS* **2023**, *18*, 1867–1874, doi:10.1007/s11548-023-02880-8.