



OPEN

# Evaluation of U-Net models in automated cervical spine and cranial bone segmentation using X-ray images for traumatic atlanto-occipital dislocation diagnosis

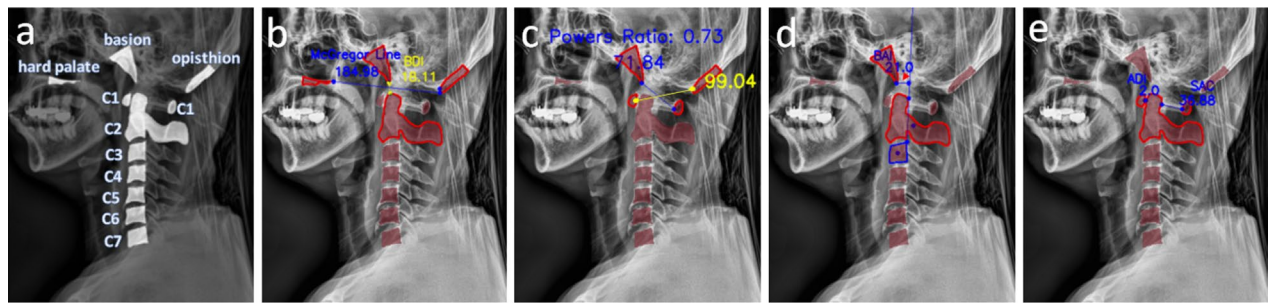
Jae-Hyuk Shim<sup>1,4</sup>, Woo Seok Kim<sup>2,4</sup>, Kwang Gi Kim<sup>1✉</sup>, Gi Taek Yee<sup>3✉</sup>, Young Jae Kim<sup>1</sup> & Tae Seok Jeong<sup>2</sup>

Segmentation of the cervical spine in tandem with three cranial bones, hard palate, basion, and opisthion using X-ray images is crucial for measuring metrics used to diagnose traumatic atlanto-occipital dislocation (TAOD). Previous studies utilizing automated segmentation methods have been limited to segmenting parts of the cervical spine (C3 ~ C7), due to difficulties in defining the boundaries of C1 and C2 bones. Additionally, there has yet to be a study that includes cranial bone segmentations necessary for determining TAOD diagnosing metrics, which are usually defined by measuring the distance between certain cervical (C1 ~ C7) and cranial (hard palate, basion, opisthion) bones. For this study, we trained a U-Net model on 513 sagittal X-ray images with segmentations of both cervical and cranial bones for an automated solution to segmenting important features for diagnosing TAOD. Additionally, we tested U-Net derivatives, recurrent residual U-Net, attention U-Net, and attention recurrent residual U-Net to observe any notable differences in segmentation behavior. The accuracy of U-Net models ranged from 99.07 to 99.12%, and dice coefficient values ranged from 88.55 to 89.41%. Results showed that all 4 tested U-Net models were capable of segmenting bones used in measuring TAOD metrics with high accuracy.

Traumatic atlanto-occipital dislocation (TAOD) is a traumatic injury stemming from damage to the spinal area resulting in dislocation of the upper cervical spine and the skull base<sup>1,2</sup>. Due to the high-energy nature of the trauma often associated with the injury such as motor vehicle accidents, most patients who experience atlanto-occipital dislocation have a high likelihood of mortality<sup>3</sup>. In cases of patients that survive the injury, it is vital that the dislocation is diagnosed and treated rapidly to improve the success rate of stabilization<sup>4,5</sup>. Patients who experience TAOD often present debilitating symptoms such as unconsciousness, respiratory arrest, and in severe cases, sensory, motor, and neurological deficits<sup>2</sup>. However, there lies some difficulty in consistently diagnosing TAOD in patients without neurological deficits due to the nature of certain symptoms, which can be as vague as severe neck pain, or injury to other areas of the body masking pain in the cervical area, to being completely asymptomatic<sup>2</sup>. To appropriately diagnose TAOD before complications from delayed treatment occur, radiological methods using MRI, CT, and X-rays are used to screen patients suspected of TAOD<sup>1,6</sup>.

There are several guidelines used to diagnose TAOD through radiological methods<sup>7,8</sup>. Images of the cervical spine are taken either through MRI, CT, or X-rays, then studied for particular dislocations of specific cervical and cranial areas. One metric used to diagnose TAOD is the 10 mm increase in the basion-dens interval (BDI), which refers to the distance between the basion and the top of the C1 cervical bone by 10 mm<sup>9</sup>. Another metric

<sup>1</sup>Department of Biomedical Engineering, Gil Medical Center, Gachon University College of Medicine, Incheon, Korea. <sup>2</sup>Department of Traumatology, Gil Medical Center, Gachon University College of Medicine, Incheon, Korea. <sup>3</sup>Department of Neurosurgery, Gil Medical Center, Gachon University College of Medicine, Incheon, Korea. <sup>4</sup>These authors contributed equally: Jae-Hyuk Shim and Woo Seok Kim. ✉email: kimkg@gachon.ac.kr; gtyee@gilhospital.com



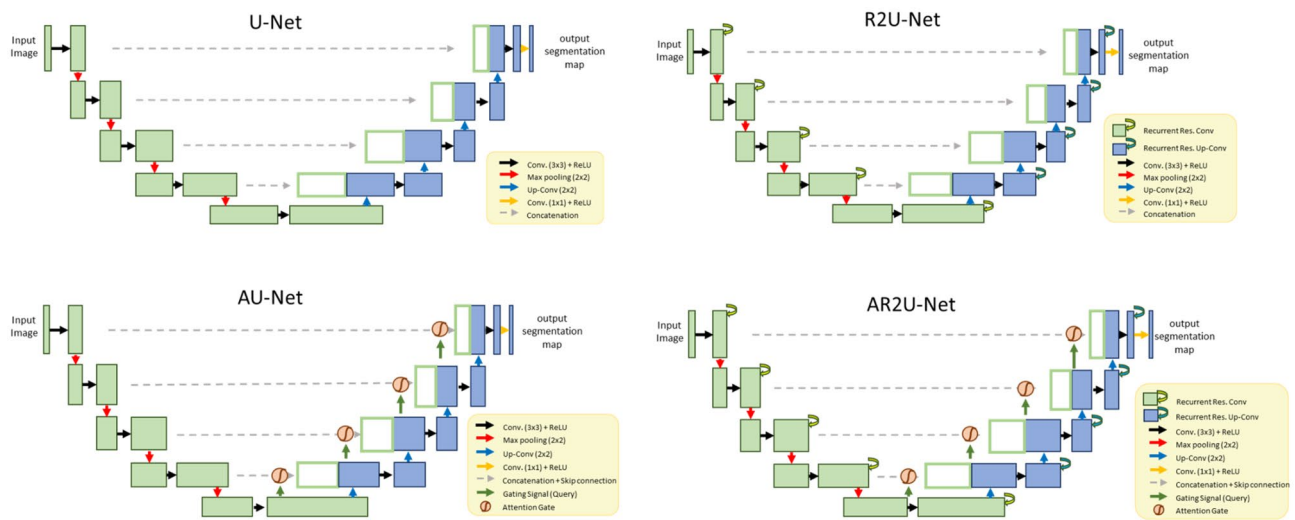
**Figure 1.** Example of metrics measured for TAOD diagnosis. (a) Bone structures of the cervical spine used for measuring TAOD diagnosing metrics. (b) McGregor line is a line that connects the posterior edge of the hard palate and the most caudal point of the opisthion. BDI is defined by the distance from the most inferior edge of the basion to the closest point of the superior edge of C2. (c) Power's ratio, calculated by dividing the distance of the most inferior edge of the basion to the center edge of the right C1 by the distance of the leftmost point of the opisthion to the center point of the left C1. (d) BAI (line designated by the red arrow) is measured by the closest distance between the most inferior edge of the basion to the line that extends cranially along the posterior cortex of C2. (e) ADI refers to the closest distance between the center point of left C1 to the superior region of C2, while SAC refers to the closest distance between the center point of right C1 to the superior region of C2. All distances shown in this figure are in mm.

is the Power's ratio, which refers to the ratio of the distance between the opisthion to the anterior arch of C1, and the distance between the basion to the posterior arch of C1<sup>10</sup>. The basion-axial interval (BAI), which measures the distance between the basion and the C2 line is checked for a displacement of more than 12 mm or less than 4 mm for TAOD diagnosis<sup>10</sup>. The atlanto-dens interval (ADI), which refers to the distance between the anterior arch of the atlas and the dens of the axis, as well as space-available-cord (SAC), which refers to the distance between the posterior surface of the dens to the anterior surface of the posterior arch of the dens, are used to radiologically examine for TAOD<sup>7</sup>. McGregor line, which is a line that connects the hard palate to the most caudal point of the midline occipital curve, is also used to diagnose TAOD<sup>11</sup>. Examples of metrics measured for TAOD diagnosis are shown in Fig. 1. Depending on the injury sustained by the patient, each diagnostic method may show results with inconsistent conclusions, often requiring the clinician to use multiple techniques to assess for TAOD with sufficient confidence<sup>6</sup>.

With recent advances in CT and MRI imaging, high resolution, high contrast, 3D images of the cervical spine can be acquired with low scanning time. Compared to X-ray scans, CT and MRI scans can provide a much higher level of detail for accurately identifying key structures used to measure TAOD metrics such as BDI, BAI, and the power's ratio<sup>12</sup>. However, there are certain constraints or situations when X-rays are recommended before CT or MRI scans are utilized. In the case of patients that report very mild symptoms of TAOD such as neck pain, it is likely that clinicians would recommend X-ray scans for cheap, fast, and accessible imaging of the cervical spine, especially if the clinic does not carry the necessary equipment for CT or MRI scans<sup>12</sup>. Additionally, X-rays may show sufficient resolution for identifying necessary metrics for TAOD diagnosis, making CT or MRI scans unnecessary<sup>13</sup>. As such, there is a benefit to developing methods for improving X-ray diagnosis of TAOD to ensure that the patients are rapidly diagnosed before complications of undiagnosed TAOD occur<sup>6</sup>.

Previous studies have explored methods to segment the cervical spine using X-rays. Early studies in segmenting the cervical spine drew references from spinal segmentation, which involved using techniques to obtain basic landmarks for a generalized location of the spine, then deforming a 2D contour model to match each individual vertebra based on boundary detection<sup>14,15</sup>. Various vertebrae corner and center detection algorithms were first used to establish the locations of each cervical vertebra<sup>16,17</sup> then contours of each cervical vertebra were deformed to match the intensity of its reference using edge detection<sup>18–20</sup>. Machine learning approaches, such as Hough forest-based architectures and U-Net were adopted for a significantly stronger performing segmentation, being able to better discriminate between cervical vertebra and its surroundings<sup>21–23</sup>. However, previous studies on segmenting the cervical spine have been limited to bones up to C3, where outlines are clearly defined and unobstructed by the stacking effect of bones common in X-rays<sup>22,23</sup>. Certain aspects of the cervical spine make it difficult to delineate, some being the circular shapes of C1 and C2 bones causing stacking and the lack of clear boundaries for cranial bones like the opisthion or hard palate<sup>22</sup>.

For this study, we prepared a dataset of segmentations involving the cervical spine (C1 ~ C7) and three bones: opisthion, hard palate, and the basion. We then applied four different machine learning U-Net models: standard U-Net<sup>24</sup>, Attention U-Net (AU-Net)<sup>25</sup>, recurrent residual U-Net (R2U-Net)<sup>26</sup>, and Attention R2U-Net (AR2U-Net)<sup>27</sup>, then observed their performance metrics in the form of dice coefficient, sensitivity, specificity, and accuracy as well as model complexity through parameter count and floating-point operations per second (FLOPs). We quantitatively evaluated the segmentation metrics of each bone class between the 4 U-Net models in the form of dice coefficient, Hausdorff distance (HD) and mean square distance (MSD). Additionally, we qualitatively evaluated the common patterns of each U-Net model displayed in segmentations. Through our research, we aim to show that key bone structures involving in evaluating TAOD metrics can be automatically segmented using U-Net. Moreover, we aim to show unique patterns in segmentations that could arise utilizing each U-Net model based on the quantitative evaluation of segmentation metrics and qualitative evaluation of segmentation masks.



**Figure 2.** Diagrams of U-Net, R2U-Net, AU-Net and AR2U-Net architecture used for cervical spine segmentation.

## Materials and methods

**Data acquisition.** Sagittal X-rays of 707 subjects for delineation were collected at Gachon University Gil Medical Center. The average age of subjects was  $53.9 (\pm 16.7)$ , with 400 of them being male and 307 being female. Images had resolutions ranging from 0.139 mm per pixel to 0.194 and had an average of 1688 pixels as width and 2102 as height. All images were adjusted to face left, horizontally flipping X-rays of right-facing subjects. This study protocol was approved by the institutional review board at Gachon University Gil Medical Center (GDIRB2022-190). Methods used in this study were all in accordance with the relevant guidelines and regulations of the declaration of Helsinki, and informed consent was obtained from all participants involved in this study.

Contours of C1 ~ C7 bones along with the three bones, opisthion, hard palate, and basion, were delineated by two experts that received training from qualified medical doctors. An example of delineation is shown in Fig. 1a.

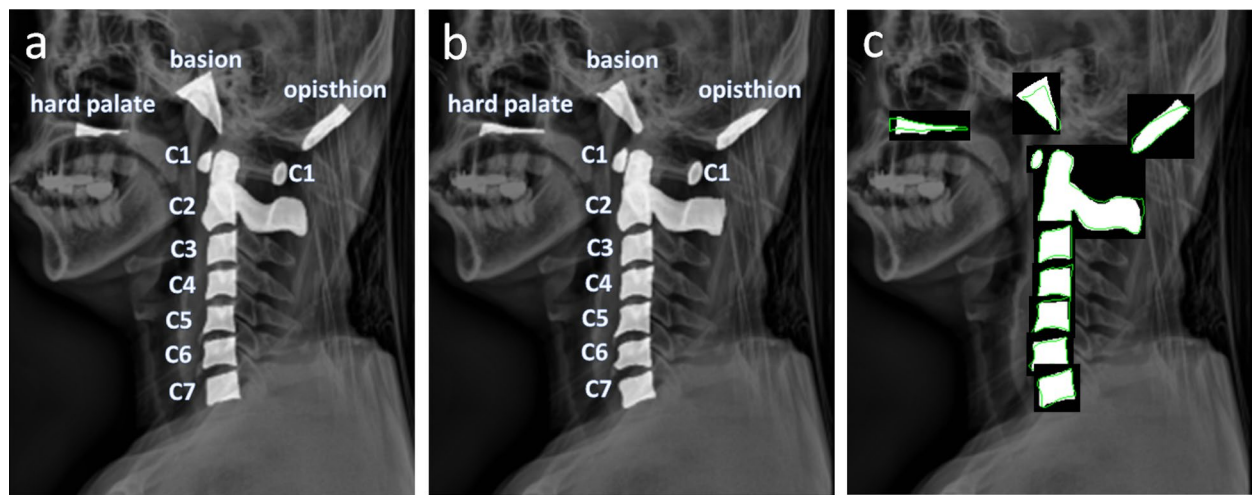
X-rays were resized to  $512 \times 512$ , and histogram equalized using contrast limited adaptive histogram equalization (CLAHE). 80% of the images (513 images) were randomly designated as the training and validation dataset. The remaining 20% (194 images) were used for testing the trained models.

**Network.** The main framework for the three models used is U-Net, a commonly used convolutional neural network for the purpose of biomedical image semantic segmentation. U-Net consists of two pathways responsible for classifying and localizing each object in an image. The first pathway is referred to as the encoder, which involves multiple convolution, rectified linear unit (ReLU), and max-pooling layers to downsample data for feature extraction. The second pathway is referred to as the decoder, which involves up-sampling the extracted feature maps with convolution, concatenation, and ReLU layers for localization information. Through the encoder and decoder pathways, the U-Net architecture can efficiently extract both segmentation and location data, which enables low-cost training of highly accurate biomedical segmentation models.

For better segmentation performance, previous studies have attempted to incorporate various techniques and adjustments to the original U-Net architecture. One method used in this study with modified U-Net modules is R2U-Net, which utilizes residual layers with skip connections that forward propagate low-level and high-level information while outputs are recurred through feedback connections to store information over time, increasing performance with the same input. Another method used in this study with modified U-Net modules is AU-Net, which incorporates attention gates to suppress feature information in irrelevant regions. AR2U-Net integrates R2U-Net convolution neural networks with attention gates for suppressing irrelevant background features while enhancing segmentation performance with R2U-Net benefits. Diagrams of the four U-Net models are shown in Fig. 2.

The three U-Net networks were trained on a desktop with an Intel i9-10900 CPU, 32 GBs of RAM, and NVIDIA RTX 3060 with 12 GBs of dedicated RAM for 200 epochs, with a batch size of 3, Adam optimizer, a binary cross-entropy loss function, and a learning rate of 0.001.

**Segmentation evaluation.** Each individual bone segment (hard palate, opisthion, basion, C1 ~ C7) was extracted and isolated from each test subject's predicted segmentation masks and ground truth masks. The dice coefficient, HD, and MSD values of each predicted segmentation bone mask and its respective ground truth mask were calculated using the segmentation metrics python package<sup>28</sup>. The process was repeated for results obtained with each U-Net model. An example of predicted segmentation masks and ground truths overlapping to calculate dice coefficient is shown in Fig. 3.



**Figure 3.** Example of overlapping predicted masks and ground truth masks for calculating dice coefficient. (a) Ground truth masks manually annotated. (b) Predicted segmentation masks (c) overlap of ground truth masks (white contours) and predicted segmentation masks (green outline) to calculate dice coefficient.

**Segmentation metrics.** The similarities between model predicted segmentations and manually annotated ground truths were defined through the Dice coefficient, a commonly used metric for validating segmentation accuracy and reproducibility. The Dice coefficient formula is described as 2 times the intersection of predicted masks and ground truth masks, divided by the combined number of pixels of both predicted masks and ground truth masks. The formula of the Dice coefficient is as follows:

$$DICE = \frac{2|P \cap G|}{|P| + |G|},$$

where P and G are the number of pixels in predicted segmentation and ground truth masks, respectively, and  $P \cap G$  is the intersection of the predicted segmentation and ground truth masks.

The similarities of each boundary between predicted masks and reference masks were evaluated through HD and MSD. HD measures the largest difference between the surface distances of predicted mask and reference mask, while MSD describes the average difference between the boundaries of each surface.

## Results

The performance metrics of images tested as well as the parameters and FLOPs of the four trained U-Net models are shown in Table 1. All four models showed segmentation performance metrics of the following: Standard U-Net showed an overall average sensitivity of 89.86%, specificity of 99.52%, dice-coefficient of 89.21%, and accuracy of 99.12%. Residual U-Net showed an average sensitivity of 89.00%, specificity of 99.54%, dice-coefficient of 88.85%, and accuracy of 99.10%. Attention U-Net showed the highest average sensitivity (90.44%), dice-coefficient (89.41%), and accuracy (99.13%), but lower specificity (99.51%). Attention residual U-Net showed the highest specificity of 99.58%, but the lowest average sensitivity (87.67%), dice coefficient (88.55%), and accuracy (99.07%). AR2U-Net showed the highest computational cost of 2.454 M parameters and 0.324 FLOPs, while standard U-Net showed the lowest with 1.968 M parameters and 0.279 FLOPs.

Table 2 shows the average dice coefficient, HD, and MSD values of each individual bone part, obtained from segmenting each test subject's (n = 193) X-ray image with the trained U-Net models. AU-Net showed the highest dice coefficient in most bone parts (C3, 95.30%; C4, 95.29%; C5, 95.17%; C6, 95.21%; C7, 94.66%). R2U-Net had the highest dice coefficient in segmenting the hard palate (87.76%) and the opisthion (85.71%). AR2U-Net had the highest dice coefficient in segmenting C1 (95.18%). U-Net had the lowest HD and MSD values in basion and C1 segmentations. AU-Net had the lowest HD and MSD values in C3, C4, C5, C6 and C7. AR2U-Net had the lowest HD and MSD values in hard palate, opisthion, and C2.

## Discussion

For this study, four different U-Net models, standard U-Net, attention U-Net, residual U-Net, and attention residual U-Net, were trained using X-ray images of the cervical spine to segment key bone structures used to diagnose TAOD. Previous research in using machine learning models to segment the cervical spine has been limited to bones C3 and below due to clear, defined shapes and unobstructed clarity from a lack of overlap in X-rays<sup>22,23</sup>. As such, no methods have been yet tested for segmenting the entirety of the cervical spine in tandem with the basion, opisthion, and hard palate. All four U-Net segmentation models provided adequate segmentation performance, with average specificity and dice coefficient approaching 90%. As such, the results showed that automated segmentation of bone structures used to diagnose TAOD is possible with benefits such as rapid speed of the segmentation and less reliance on manual intervention.

The high performance of the U-Net semantic segmentation model has been a strong motivation for its utilization in many studies<sup>24</sup>. The encoder and decoder pathways of U-Net enable the retention of rich feature



	U-Net	AU-Net	R2U-Net	AR2U-Net
Sensitivity	89.86%	<b>90.44%</b>	89.00%	87.67%
Specificity	99.52%	99.51%	99.54%	<b>99.58%</b>
Dice coefficient	89.21%	<b>89.41%</b>	88.85%	88.55%
Accuracy	99.12%	<b>99.13%</b>	99.10%	99.07%
Parameters	1.968 M	2.342 M	2.081 M	<b>2.454 M</b>
FLOPs	0.279	0.306	0.298	<b>0.324</b>

**Table 1.** Average sensitivity, specificity, dice coefficient, accuracy, parameter count, FLOP, of U-Net, attention U-Net and residual U-Net models. Bolded numbers represent the largest value in that category. M in parameters represents million.

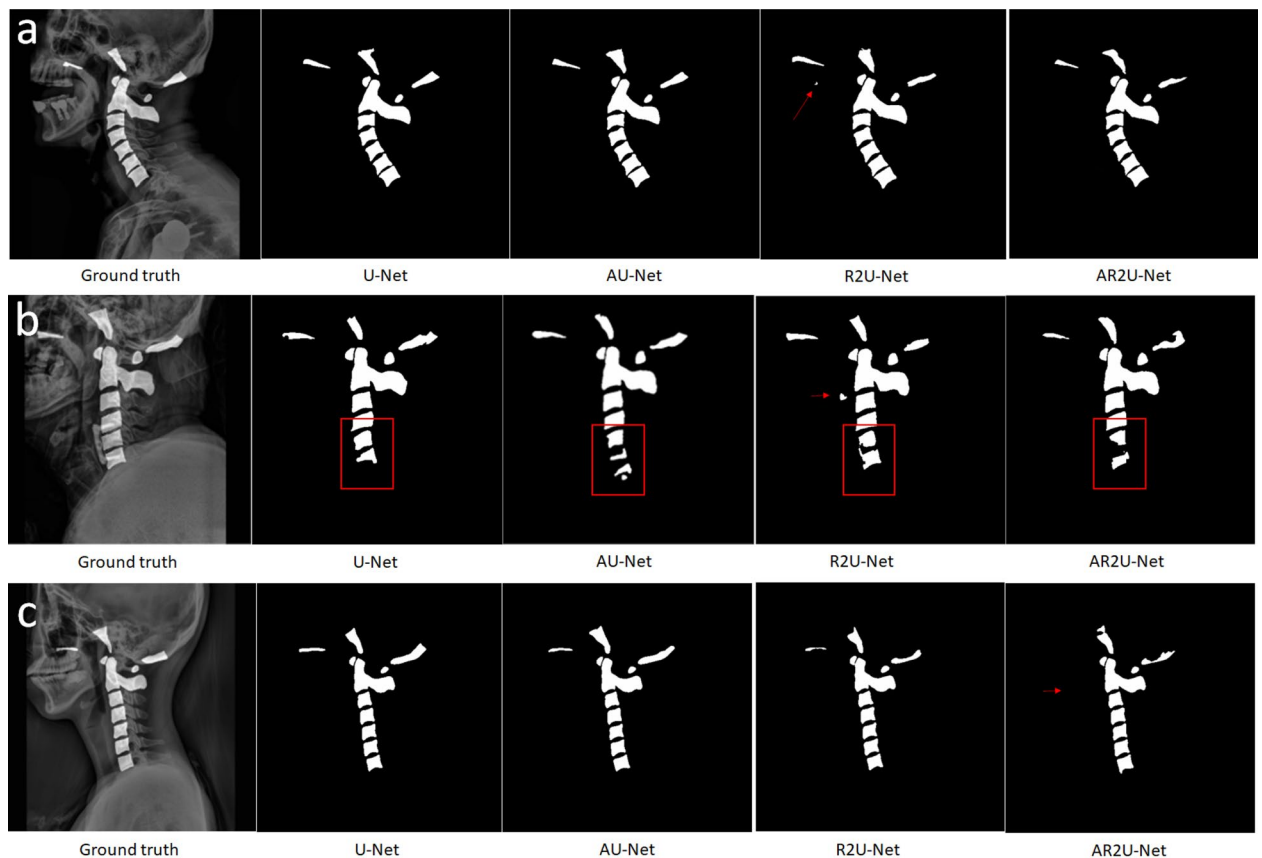
	Metric	Hard palate	Opisthion	Basion	C1	C2	C3	C4	C5	C6	C7
U-Net	DICE	87.60%	85.07%	89.29%	94.82%	<b>96.79%</b>	95.24%	95.26%	94.62%	94.06%	93.78%
	HD	16.03	13.22	<b>8.63</b>	<b>2.16</b>	4.62	2.13	2.26	2.27	2.30	2.76
	MSD	2.70	2.71	<b>2.03</b>	<b>0.49</b>	0.80	0.56	0.57	0.53	0.59	0.64
AU-Net	DICE	87.47%	85.67%	<b>89.69%</b>	95.06%	96.78%	<b>95.30%</b>	<b>95.29%</b>	<b>95.17%</b>	<b>95.21%</b>	<b>94.66%</b>
	HD	13.56	11.45	9.56	2.29	4.55	<b>2.07</b>	<b>2.10</b>	<b>2.24</b>	<b>2.28</b>	<b>2.67</b>
	MSD	2.38	2.62	2.16	0.49	0.80	<b>0.54</b>	<b>0.53</b>	<b>0.53</b>	<b>0.59</b>	<b>0.64</b>
R2U-Net	DICE	<b>87.76%</b>	<b>85.71%</b>	88.94%	94.81%	96.58%	95.00%	95.07%	95.13%	94.89%	94.60%
	HD	15.68	12.55	8.80	2.25	4.60	2.20	2.20	2.25	2.43	2.92
	MSD	2.64	2.67	2.03	0.51	0.82	0.57	0.55	0.56	0.61	0.65
AR2U-Net	DICE	86.65%	85.70%	88.49%	<b>95.18%</b>	96.48%	94.34%	94.55%	94.32%	94.47%	94.14%
	HD	<b>13.43</b>	<b>11.34</b>	9.71	2.52	<b>4.48</b>	2.21	2.34	2.44	2.37	2.91
	MSD	<b>2.37</b>	<b>2.60</b>	2.19	0.54	<b>0.80</b>	0.56	0.58	0.58	0.60	0.65

**Table 2.** Average dice coefficient values of each bone mask segmented by different U-Net models. Bolded numbers represent the largest DICE value, lowest HD and MSD value.

and location information, reducing the resources and time necessary to train a model with good semantic segmentation performance. The standard U-Net model, trained on 513 X-ray images with no modifications showed a sensitivity of 89.86% and an accuracy of 99.12%. In order to further optimize segmentation performance, the same dataset was trained on three different U-Net models with modifications to the standard U-Net encoder and decoder pathway. The first U-Net model trained on our dataset was AU-Net, a model that incorporates attention gates to modify feature maps and suppress features in irrelevant areas<sup>27</sup>. AU-Net performed the best out of the 4 models tested in average sensitivity (90.44%), dice coefficient (89.41%), and accuracy (99.13%). The second model trained on our dataset was R2U-Net, a modified U-Net neural network that incorporated recurring residual blocks that forward propagate and recur information to reduce computation resources while improving generalization. In this particular case, training R2U-Net on 513 X-ray images showed that segmentation performance suffered in all categories compared to the standard U-Net, except in specificity (99.54%). Additionally, R2U-Net introduced various false-positive errors in predicted images, segmenting areas located away from the ground truth segmentations (Fig. 4a,b). The last U-Net model trained on our dataset was AR2U-Net, a model that incorporated both attention gates and recurring residual blocks to the standard U-Net model. AR2U-Net trained with our dataset showed the lowest average sensitivity (87.67%), dice coefficient (88.55%), and accuracy (99.07%) but the highest specificity (99.58%). Segmentations with trained AR2U-Net also showed various instances with severely fragmented masks and even some cases where masks were omitted, as shown in Fig. 4b,c.

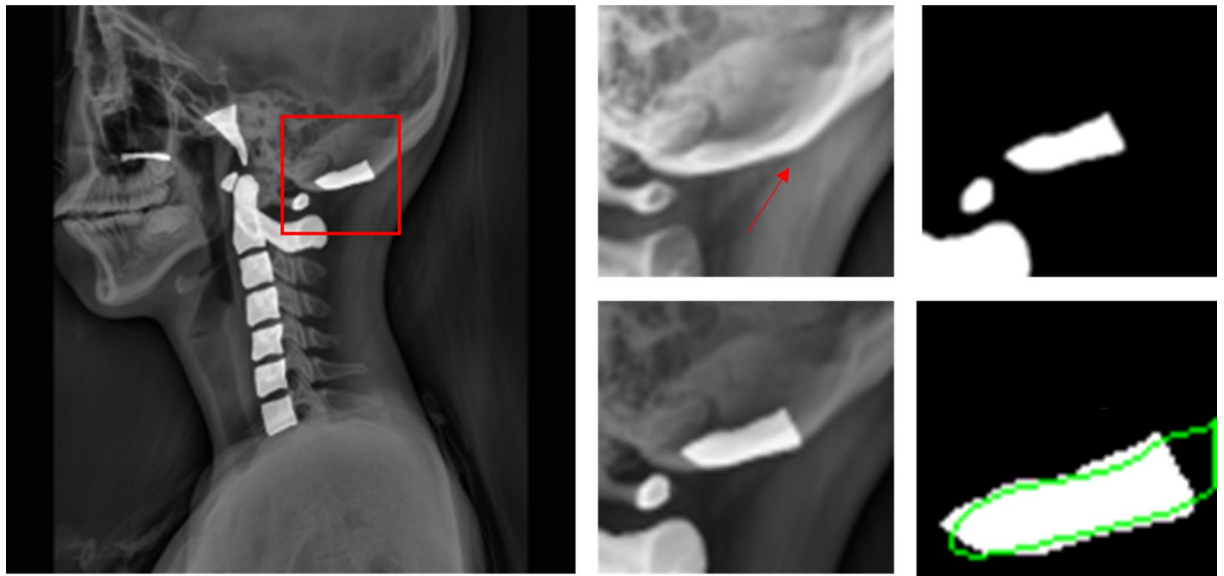
As shown in Table 1 describing computational costs, standard U-Net with no attention gates or residual blocks had the lowest number of parameters (1.968 M) and FLOPs (0.279), requiring the least number of resources and time to run out of the 4 U-Net models. R2U-Net uses slightly more parameters (2.081 M, 5.74% higher) and FLOPs (0.298, 6.8% higher) than U-Net for included residual blocks supplementing standard U-Net. However, the increased computational costs of R2U-Net results in lower sensitivity, dice coefficient and accuracy compared to U-Net results. AU-Net with attention gates uses more parameters (2.342 M) than both U-Net and R2U-Net but similar FLOPs (0.306) with R2U-Net. AR2U-Net requires the greatest number of parameters (2.454 M) and FLOPs (0.324), while providing a minor improvement in specificity and performing the worst in sensitivity, dice coefficient and accuracy. Purely based on model metrics, AR2U-Net were the least efficient in terms of computational power, requiring 16.13% FLOPs over standard U-Net for worse sensitivity, dice coefficient and accuracy. On the contrary, AU-Net showed minor improvements over standard U-Net in sensitivity, dice coefficient and accuracy at the cost of 9.68% increased FLOPs.

The average dice coefficient, HD, and MSD values of each bone mask segmented with the 4 different U-Net models were compared for evaluation. Dice coefficient described how well the surfaces of each segmentation

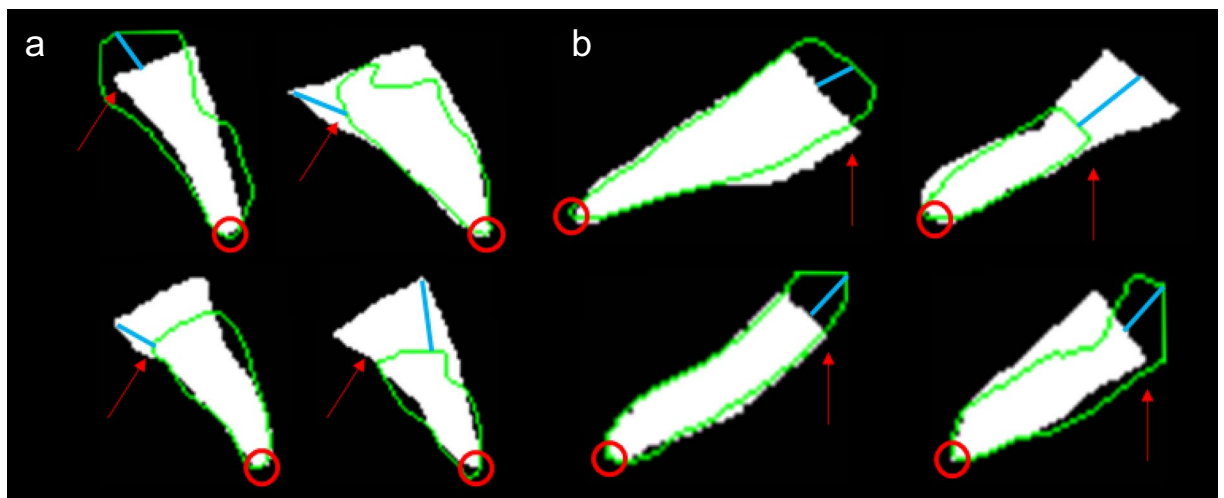


**Figure 4.** Example of errors found in U-Net predicted segmentation masks. Red arrows and red boxes refer to errors found in segmentations with certain U-Net models. (a) Example of a false positive in R2U-Net predicted segmentation masks. (b) Example of fragmented segmentations affected by spinal implants and false positives in R2U-Net predicted segmentation masks. (c) Example of omitted segmentation in AR2U-Net predicted segmentation masks.

overlapped with each other, while HD and MSD values described how well the predicted segmentation matched its boundaries with the reference mask. Results showed standard U-Net had the highest dice coefficient when segmenting C2 (96.79%), and the lowest HD and MSD value in basion and C1 segmentations. R2U-Net had the highest dice coefficient when segmenting the the hard palate (87.76%) and opisthion (85.71%). AU-Net showed the highest dice coefficient in segmenting the basion (89.69%), C3 (95.30%), C4 (95.29%), C5 (95.17%), C6 (95.21%), C7 (94.66%). AU-Net also showed the lowest HD and MSD values in C3, C4, C5, C6, and C7 segmentations. AR2U-Net showed the highest dice coefficient in C1 segmentations (95.18%) and the lowest HD and MSD in hard palate, opisthion and C2 segmentations. Compared to bones C1 through C7, the hard palate, opisthion, and basion showed noticeably lower dice coefficient values and higher HD and MSD in all U-Net models, partially due to vague boundaries from X-ray blurs making some segmentations dependent on human judgment as shown in Fig. 5. At a first glance, it would seem that a noticeably lower hard palate HD (13.43) and MSD (2.37) in AR2U-Net over standard U-Net hard palate HD (16.03) and MSD (2.70) shows the advantages of AR2U-Net despite the lower dice coefficient (86.65% compared to 87.60%) due to how TAOD metrics are mainly calculated using the boundaries of segmentations. However, the inconsistent boundary selection of the three cranial bones: the hard palate, opisthion, and the basion, are usually isolated to the posterior ends of each bone structure, areas that are not considered for calculating TAOD related metrics. For example, the TAOD metrics such as the BDI, which takes the distance between the inferior (lowest) point of the basion and the closest point of superior C2<sup>9</sup> as shown in Fig. 1b. As such, evaluating each U-Net model's performance in accurately determining TAOD metrics may not be consistent with dice coefficient, HD, or MSD values as visualized in Fig. 6. The dice coefficient, HD, and MSD values of bones C1 through C7 were mostly similar among the four U-Net models, due to unobscured boundaries making segmentation and prediction consistent. However, minor variations of metrics were noticeable in C5, C6, and C7 bones, in part due to a number of subjects with spinal implants commonly located in C5, C6, and C7 bones, as shown in Fig. 4b.



**Figure 5.** An example of obfuscated boundaries of an opisthion's posterior end. Red arrow points to the ground truth boundary of the opisthion's posterior end. The green outline shows an example of an AU-Net predicted opisthion mask, which extends further beyond the ground truth.



**Figure 6.** Example of masks that show poor dice coefficient, high HD, and MSD but with high accuracy in regions used for measuring metrics for diagnosing TAOD. Green outlines refer to AU-Net segmentation masks, white masks refer to ground truths. Blue lines refer to a visual depiction of how HD can be calculated. Red arrows refer to posterior ends of cranial bones with inaccurate posterior boundaries but are not used in determining TAOD metrics. Red circles refer to anterior ends that are used for determining TAOD metrics. (a) Basion masks. (b) Opisthion masks.

There are some limitations to consider for this study. First, as previously mentioned, the manual segmentation of ground truths is significantly influenced by human judgment. Regions with the highest likelihood for error are mainly located at the posterior ends of cranial bones where boundaries are heavily obfuscated by bone stacking effects in X-rays. While it may take significant resources to do so, making additional redundant ground truths segmented by a multitude of trained professionals can help improve consistency. Second, it is difficult to conclude which U-Net model is the most useful in determining TAOD metrics due to only measuring segmentation performance. As previously mentioned, segmentations of cranial bones with low dice coefficient values or high HD and MSD values can still show accurate measurements of TAOD metrics. Third, the patterns shown in our segmentation results may differ depending on the training or testing data used. Since our data used to train U-Net

models was moderately consistent, it is possible that each U-Net model trained or tested on data with heavily fractured bones or with spinal implants can show different results.

To the best of our knowledge, this study was the first to train U-Net and 3 other U-Net variant models for segmenting the cervical spine, as well as the 3 cranial bones, hard palate, opisthion, and the basion, areas in which are involved in calculating metrics used for diagnosing TAOD. Our methods shown in this study present a potential avenue for automated rapid diagnosis of TAOD using X-rays. Additionally, we show that modifications to the standard U-Net model can be pursued to potentially further improve cervical spine segmentation performance.

## Data availability

The U-Net models and code generated for this study are available from the corresponding author on responsible request.

Received: 16 June 2022; Accepted: 7 November 2022

Published online: 12 December 2022

## References

- Kim, Y. J. *et al.* Traumatic atlanto-occipital dislocation (AOD). *Korean J. Spine*. **9**(2), 85 (2012).
- Hall, G. C. *et al.* Atlanto-occipital dislocation. *World J. Orthop.* **6**(2), 236 (2015).
- Cooper, Z. *et al.* Identifying survivors with traumatic craniocervical dissociation: A retrospective study. *J. Surg. Res.* **160**(1), 3–8 (2010).
- Kenter, K., Worley, G., Griffin, T. & Fitch, R. D. Pediatric traumatic atlanto-occipital dislocation: Five cases and a review. *J. Pediatr. Orthop.* **21**(5), 585–589 (2001).
- Chang, D. G. *et al.* Traumatic atlanto-occipital dislocation: Analysis of 15 survival cases with emphasis on associated upper cervical spine injuries. *Spine* **45**(13), 884–894 (2020).
- Joaquim, A. F., Schroeder, G. D. & Vaccaro, A. R. Traumatic atlanto-occipital dislocation—A comprehensive analysis of all case series found in the spinal trauma literature. *Int. J. Spine Surg.* **15**(4), 724–739 (2021).
- Glaun, G. D. & Phillips, J. H. Occipitocervical dissociation in three siblings: A pediatric case report and review of the literature. *J. Am. Acad. Orthop. Surg. Glob. Res. Rev.* **2**(5), e067 (2018).
- Yang, S. Y. *et al.* A review of the diagnosis and treatment of atlantoaxial dislocations. *Glob. Spine J.* **4**(3), 197–210 (2014).
- Singh, A. K. *et al.* Basion-cartilaginous dens interval: An imaging parameter for craniocervical junction assessment in children. *Am. J. Neuroradiol.* **38**(12), 2380–2384 (2017).
- Rojas, C. A., Bertozzi, J. C., Martinez, C. R. & Whitlow, J. Reassessment of the craniocervical junction: normal values on CT. *Am. J. Neuroradiol.* **28**(9), 1819–1823 (2007).
- Benke, M., Yu, W. D., Peden, S. C. & O'Brien, J. R. Occipitocervical junction: Imaging, pathology, instrumentation. *Am. J. Orthop.* **40**(10), E205–E215 (2011).
- Yelamarty, P. K. *et al.* Radiological protocol in spinal trauma: Literature review and Spinal Cord Society position statement. *Eur. Spine J.* **29**(6), 1197–1211 (2020).
- Marques, C. *et al.* Accuracy and reliability of X-ray measurements in the cervical spine. *Asian Spine J.* **14**(2), 169 (2020).
- Long, L. R., Thoma, G. R. Use of shape models to search digitized spine X-rays. In *Proceedings 13th IEEE Symposium on Computer-Based Medical Systems*. CBMS 2000 255–260 (IEEE, 2000).
- Long, L. R., Thoma, G. R. Identification and classification of spine vertebrae by automated methods. In *Medical Imaging 2001: Image Processing*, vol. 4322, 1478–1489 (SPIE, 2001).
- Benjelloun, M. & Mahmoudi, S. Spine localization in X-ray images using interest point detection. *J. Digit. Imaging* **22**(3), 309–318 (2009).
- Larham, M. A., Mahmoudi, S., Benjelloun, M. Semi-automatic detection of cervical vertebrae in X-ray images using generalized Hough transform. In *2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)* 396–401 (IEEE, 2012).
- Benjelloun, M. & Mahmoudi, S. X-ray image segmentation for vertebral mobility analysis. *Int. J. Comput. Assist. Radiol. Surg.* **2**(6), 371–383 (2008).
- Xu, X., Hao, H. W., Yin, X. C., Liu, N., Shafin, S. H. Automatic segmentation of cervical vertebrae in X-ray images. In *The 2012 International Joint Conference on Neural Networks (IJCNN)* 1–8. (IEEE, 2012).
- Mahmoudi, S. A., Lecron, F., Manneback, P., Benjelloun, M., Mahmoudi, S. GPU-based segmentation of cervical vertebra in X-ray images. In *2010 IEEE International Conference on Cluster Computing Workshops and Posters (CLUSTER WORKSHOPS)* 1–8. (IEEE, 2010).
- Al Arif, S. M., Asad, M., Gundry, M., Knapp, K. & Slabaugh, G. Patch-based corner detection for cervical vertebrae in X-ray images. *Signal Process. Image Commun.* **1**(59), 27–36 (2017).
- Al Arif, S. M., Knapp, K. & Slabaugh, G. Fully automatic cervical vertebrae segmentation framework for X-ray images. *Comput. Methods Programs Biomed.* **1**(157), 95–111 (2018).
- Rehman, F., Shah, S. I., Gilani, S. O., Emad, D., Riaz, M. N., Faiza, R. A novel framework to segment out cervical vertebrae. In *2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE)* 190–194. (IEEE, 2019).
- Ronneberger, O., Fischer, P., Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241. (Springer, 2015).
- Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. (2018).
- Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., Asari, V. K. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*. (2018).
- Zuo, Q., Chen, S. & Wang, Z. R2AU-Net: Attention recurrent residual convolutional neural network for multimodal medical image segmentation. *Secur. Commun. Netw.* **10**, 2021 (2021).
- Jia, J. A package to compute segmentation metrics: seg-metrics. [https://github.com/Ordgod/segmentation\\_metrics](https://github.com/Ordgod/segmentation_metrics). (2020).

## Acknowledgements

This work was supported by the GRRC program of Gyeonggi province. [GRRC-Gachon2020(B01), AI-based Medical Image Analysis, and by the Gachon University (GCU-202205980001) and by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program



(IITP-2022-2017-0-01630) supervised by the IITP (Institute for Information & communications Technology Promotion).

### Author contributions

All authors reviewed and approved the manuscript. All authors made significant contributions to the manuscript. K.G.K., Y.J.K., G.T.Y. designed and supervised the research. W.S.K., T.S.J. provided data and guidance for research. J.H.S. performed the experiments and wrote the main manuscript text. J.H.S and W.S.K made equal contributions to the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to K.G.K. or G.T.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2023