Volume 14, Issue 2, Page 26-36, 2025 https://doi.org/10.46810/tdfd.1594252 Research Article

BİNGÖL UNIVERSITY

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science

www.dergipark.gov.tr/tdfd

Detection of Cervical Vertebrae Using Object Detection and Semantic Segmentation Methods in Lateral Cephalometric Radiographs

Mazhar KAYAOĞLU[1*], Abdulkadir ŞENGÜR[2], Saadet ÇINARSOY CİĞERİM[3], Sabahattin BOR[4]

[1] Bingöl University, Department of Informatics, Bingöl, Türkiye
[2] Fırat University, Faculty of Technology, Department of Electrical and Electronics Engineering, Elazığ, Türkiye
[3] Van Yüzüncü Yıl University, Faculty of Dentistry, Department of Clinical Sciences, Van, Türkiye
[4] İnönü University, Faculty of Dentistry, Department of Clinical Sciences, Malatya, Türkiye
Mazhar KAYAOĞU ORCID No: 0000-0002-5807-9781
Abdulkadir ŞENGÜR ORCID No: 0000-0003-1614-2639
Saadet ÇINARSOY CİĞERİM ORCID No: 0000-0002-4384-0929
Sabahattin BOR ORCID No: 0000-0001-5463-0057

**Corresponding author: mkayaoglu@bingol.edu.tr*

(Received: 30.11.2024, Accepted: 04.04.2025, Online Publication: 27.06.2025)

Keywords
Object detection, Semantic segmentation, Classification

Abstract: This study proposes an artificial intelligence-based method for the detection and semantic segmentation of C2, C3, and C4 cervical vertebrae in lateral cephalometric radiographs. The dataset used in the research consists of 3,085 lateral cephalometric radiographs provided by the Department of Orthodontics, Faculty of Dentistry, Van Yüzüncü Yıl University. Following evaluations by expert clinicians, 2,520 radiographs that met the criteria for diagnostic accuracy and clinical suitability were included in the study. In the initial stage, vertebral regions were identified using YOLOv8 and YOLOv11 object detection models, and these areas were meticulously annotated using QuPath software. The labelled data were then subjected to segmentation using advanced deep learning models such as Attention-UNet, Attention-ResUNet, SEEA-UNet, and ResAt-UNet.The study revealed that the object detection models achieved a high performance with an accuracy of 99.8%. Among the segmentation models, Attention-ResUNet demonstrated the best performance with an accuracy of 99.25%, while the ResAt-UNet model stood out with its balanced generalization capacity. The generated binary masks provided a reliable dataset for bone age estimation and skeletal maturity analysis. This study aims to reduce radiation exposure and streamline clinical workflows by eliminating the need for additional imaging. The findings indicate that AI-supported methods minimize errors caused by manual assessments and ensure standardization in skeletal analysis. It is anticipated that these methods could be widely utilized in orthodontic and pediatric medical applications in the future.

Nesne Algılama ve Semantik Bölütleme Yontemleri Kullanılarak Lateral Sefalometrik Radyografilerde Servikal Vertebra Analizi

Anahtar Kelimeler
Nesne tespiti, Semantik segmentasyon, Sınıflandırma

Öz: Bu çalışmada, lateral sefalometrik radyografilerde C2, C3 ve C4 servikal vertebralarının tespiti ve semantik segmentasyonu için yapay zeka tabanlı bir yöntem önerilmektedir. Araştırmada kullanılan veri seti, Van Yüzüncü Yıl Üniversitesi Diş Hekimliği Fakültesi Ortodonti Anabilim Dalı tarafından sağlanan 3085 lateral sefalometrik radyografiden oluşmaktadır. Uzman hekimler tarafından yapılan değerlendirme sonucunda, tanısal doğruluk ve klinik uygunluk kriterlerini karşılayan 2520 radyografi seçilerek çalışmaya dahil edilmiştir. İlk aşamada YOLOv8 ve YOLOv11 nesne algılama modelleri kullanılarak vertebra bölgeleri tespit edilmiş ve ardından bu alanlar QuPath yazılımı ile detaylı şekilde anotasyonlanmıştır. Etiketlenen veriler, Attention-UNet, Attention-ResUNet, SEEA-UNet ve ResAt-UNet gibi ileri seviye derin öğrenme modelleri kullanılarak segmentasyon işlemlerine tabi tutulmuştur. Çalışma, nesne algılama modellerinin %99,8 doğruluk oranıyla yüksek performans sergilediğini ortaya koymuştur. Segmentasyon modelleri arasında en iyi performansı %99,25 doğruluk oranı ile Attention-ResUNet gösterirken, ResAt-UNet modeli genelleme kapasitesindeki dengesiyle

dikkat çekmiştir. Elde edilen ikili maskeler, kemik yaşı tahmini ve iskeletsel olgunluk analizi için güvenilir bir veri seti oluşturmuştur. Bu çalışma, ek görüntüleme ihtiyacını ortadan kaldırarak radyasyon maruziyetini azaltmayı ve klinik süreçleri hızlandırmayı amaçlamaktadır. Sonuçlar, yapay zeka destekli yöntemlerin manuel değerlendirme kaynaklı hataları en aza indirdiğini ve iskeletsel analizde standardizasyon sağladığını göstermektedir. Gelecekte, bu yöntemlerin ortodonti ve pediatrik tıbbi uygulamalarda yaygın olarak kullanılabileceği öngörülmektedir.

1. INTRODUCTION

Bone age estimation is critically important for evaluating growth and development in children and adolescents. These estimations play a vital role in orthodontic treatment planning, diagnosing skeletal anomalies, and monitoring individual growth processes. Traditionally, hand-wrist radiographs have been one of the standard methods for this evaluation. However, hand-wrist radiographs require an additional imaging procedure, which may increase radiation exposure and prolong clinical workflows. As an alternative, the cervical vertebrae (C2, C3, C4) in lateral cephalometric radiographs offer an effective method for simultaneous bone age estimation and skeletal maturity assessment without the need for extra imaging.

To establish a foundation for bone age estimation, the detection and segmentation of the C2, C3, and C4 cervical vertebrae were addressed in this study. The morphological characteristics of cervical vertebrae serve as strong biological markers for determining skeletal maturity stages, and analyzing this region provides reliable results for bone age estimation. However, manual evaluations can be time-consuming and prone to subjective errors. Therefore, developing artificial intelligence-based automated methods enables a faster, more consistent, and highly accurate analysis process.

Lateral cephalometric radiographs obtained from the Orthodontics Department of Van Yüzüncü Yıl University Faculty of Dentistry were utilized for the detection and segmentation of C2, C3, and C4 cervical vertebrae. Specifically, YOLOv8 and YOLOv11 models were employed for object detection tasks, while segmentation was carried out using advanced deep learning models such as Attention-UNet, Attention-ResUNet, SEEA-UNet, and ResAt-UNet. Initial labeling of the cervical vertebral regions in the images was performed on the Roboflow platform, followed by detailed annotations using QuPath software. As a result of these processes, a high-quality dataset was created for use in bone age estimation and other clinical analyses.

The primary aim of this study is to propose a method for the detection and segmentation of C2, C3, and C4 cervical vertebrae in lateral cephalometric radiographs and to demonstrate the potential of artificial intelligence-based approaches in this field. The study seeks to eliminate the need for additional diagnostic methods, such as hand-wrist radiographs, in clinical workflows, providing a solution that saves time for orthodontists and minimizes human error. In the future, the segmentation outputs obtained through this study could serve as a reference for bone age estimation processes, establishing a new standard for skeletal maturity analysis. Cervical vertebra analysis is not only useful for growth and development predictions but also has broader clinical applications, such as diagnosing skeletal anomalies and planning surgeries. The integration of artificial intelligence-supported methods into this domain minimizes subjective errors in manual evaluations, enabling faster and more precise analyses. This enhances patient care quality while reducing the workload of clinicians. Therefore, cervical vertebra analysis is becoming an increasingly common method for skeletal evaluation in modern medical practices.

2. RELATED STUDIES

Makaremi et al. (2019) developed a deep learning-based method to classify cervical vertebral maturation (CVM) stages in lateral cephalometric images. The study proposed a custom-designed deep convolutional neural network (CNN) model for classifying CVM stages across six categories extracted from X-ray images. Tests conducted with various image preprocessing techniques and datasets demonstrated accuracy rates exceeding 95%. The results highlighted the potential of the proposed method to classify CVM stages accurately and efficiently, emphasizing its utility as a significant tool in orthodontic treatment planning. The study also underscored the importance of tailored network architectures in small and balanced datasets [1].

Masuzawa et al. (2020) introduced a multi-stage deep learning model for vertebral segmentation, localization, and identification in 3D CT images. In the first stage, vertebral classes (cervical, thoracic, lumbar) were segmented using 3D Fully Convolutional Networks, followed by individual vertebra identification through an iterative network in the second stage. The proposed method achieved a 96% segmentation accuracy with a Dice score, 8.3 mm localization error, and an 84% identification rate, surpassing existing methods. This study presented an integrated framework for automated vertebral analysis in 3D CT images [2].

Demirel and Sonuç (2021) developed a semi-automatic method for bone age estimation to monitor children's growth and for forensic applications. The method combined the areas of carpal bones and the distal epiphyseal region of the radius with an artificial neural network model. Applied to a dataset of radiographs from children aged 1–7 years, the model achieved 87% training accuracy and 85% test accuracy, offering effective results. This method aimed to support physicians by reducing observational errors and enhancing prediction accuracy in age determination [3].

Chen et al. (2022) proposed a method combining U-Net and Mask R-CNN models for the segmentation and identification of cervical and lumbar vertebrae. The method achieved accuracy rates above 90% in lateral X-ray images of patients with ankylosing spondylitis (AS). The study aimed to enhance automation and precision in clinical assessments, even under pathological conditions [4].

Khazaei et al. (2023) developed a deep convolutional neural network (CNN) model to classify adolescent growth spurts based on CVM stages. The study utilized 1,846 lateral cephalograms from an Iranian subpopulation, focusing on the C2, C3, and C4 vertebrae. The model achieved an accuracy of 82% in three-class scenarios and 93% in two-class scenarios. The ConvNeXtBase-296-based CNN model was optimized through transfer learning, achieving high accuracy with limited data. This work highlighted the potential of deep learning-based tools for automated growth stage assessment in orthodontic treatment planning [5].

Li et al. (2023) developed a fully automated deep learning-based system called psc-CVM for evaluating cervical vertebral maturation (CVM) stages. Trained on a dataset of 10,200 lateral cephalograms, the system operated in three stages: detection of C2, C3, and C4 vertebrae positions, shape extraction, and CVM assessment based on the extracted shapes. Tests showed an average AUC value of 0.94, an accuracy rate of 70.42%, a Cohen's Kappa value of 0.645, a weighted Kappa value of 0.844, and an intraclass correlation coefficient of 0.946, indicating high consistency with expert evaluations. The study demonstrated the system's potential as an accurate, reliable, and efficient tool for clinicians in assessing growth and developmental stages [6].

Kresnadhi et al. (2023) compared ResNet-101, InceptionV3, and InceptionResNetV2 architectures for classifying CVM stages using deep learning methods. Images from the CVM-900 dataset were processed with a focus on C2–C4 and C2–C6 regions, supported by data augmentation techniques. InceptionResNetV2 performed best with 54.1% accuracy in the C2–C6 region. However, issues such as overfitting and insufficient multi-scale feature extraction limited performance gains, emphasizing the need for more advanced approaches to address these challenges [7].

Akay et al. (2023) aimed to automatically determine CVM stages in lateral cephalometric radiographs using a deep learning-based CNN model. Data from 588 radiographs were divided into six stages, and the model achieved a 58.66% accuracy rate after 40 epochs of training. While high F1 scores were obtained in CVM Stage 1, classification errors occurred in transitional stages due to similarities. The results suggested the model could serve as a fast and suitable tool for clinical use, with potential for improved performance through larger datasets [8].

Atıcı et al. (2023) developed a two-stage deep learning-based model for evaluating CVM stages using a continuous classification system. Trained on 1,398 lateral cephalometric radiographs, the model combined images with chronological age, achieving an accuracy of 81.17%. The continuous classification method represented growth and development processes more accurately than traditional discrete classifications, with a Pearson correlation coefficient exceeding 0.9, demonstrating high reliability. This study provided a more precise and clinically applicable method for skeletal maturity assessment [9].

Motie et al. (2024) introduced a three-stage deep learning model to classify CVM stages. Using 2,325 lateral cephalograms, the method involved region detection with Faster R-CNN and classification through two ResNet101 models. The first model categorized images into two main groups (C1–C3 and C4–C6), while the second model further classified these into subcategories. The proposed method achieved an overall accuracy of 82.96%, outperforming previous single-stage models and offering high accuracy in classification processes. This study provided an effective approach for automated CVM evaluation in clinical applications [10].

Mohammed et al. (2024) developed a CNN-based method for estimating skeletal growth maturity based on CVM and lower second molar calcification levels. Using 1,200 lateral cephalograms and 1,200 panoramic images, the method achieved six-class classification accuracy rates of 98% for CVM prediction in males and 97% for second molar calcification prediction in females. The study demonstrated the efficacy of AI-based approaches in assessing growth and development stages with high accuracy, aligning with traditional orthodontic imaging methods [11].

In light of the reviewed studies, AI-based methods in cervical vertebra analysis and bone age estimation provide significant advantages in terms of accuracy and reliability. Accordingly, this study develops a method for the automatic detection, segmentation, and classification of cervical vertebrae (C2, C3, and C4). The goal is to establish an AI-supported framework for lateral cephalometric radiographs, reducing dependency on manual evaluations and automating the bone age estimation process. Details of the datasets, algorithms, and workflows used in this study are comprehensively explained in the materials and methods section.

3. MATERİALS AND METHODS

In this study, a method was developed for analyzing cervical vertebrae (C2, C3, and C4) from raw radiographic images. In the first stage, the regions containing the C2–C4 vertebrae were identified on lateral cephalometric radiographs and then annotated in detail to prepare them for analysis. The detected regions provide the necessary data for semantic segmentation and classification processes of the C2, C3, and C4 vertebrae.

3.1. Data Collection

The dataset used in this study consists of lateral cephalometric radiographs obtained from the Department of Orthodontics, Faculty of Dentistry, Van Yüzüncü Yıl University **(Decision No: 2023/09-12)**, and serves as the primary data source for the research. Initially, all images underwent a comprehensive evaluation by expert clinicians and were assessed for diagnostic adequacy, visibility of anatomical structures, and technical suitability. Following these evaluations, a total of 2,520 radiographs from patients aged 7 to 22 were selected. This dataset includes images from 1,302 female and 1,218 male patients and supports the reliability of the study and the accuracy of analytical processes due to its high-quality standards. The selection criteria were based on the technical characteristics of the images and their adequacy for clinical analyses.

3.2. Dataset Preparation

Figure 1 shows an example of a raw cephalometric image used in the study, obtained from the hospital. These images were annotated on the Roboflow platform for training YOLOv8 and YOLOv11 models, with regions containing C2, C3, and C4 vertebrae meticulously labeled. Adopting a top-down approach, the study first ensured the general localization of the vertebrae, followed by a more detailed detection of the specific vertebral regions. This method aims to increase the analytical accuracy of the images and ensure a more precise detection process in the selected regions.

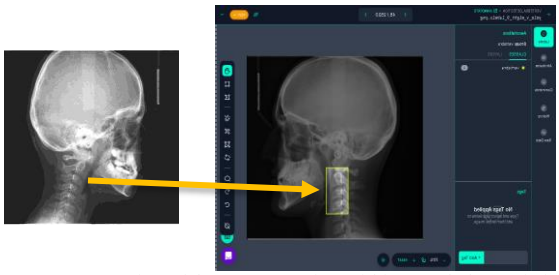


Figure 1. Annotation of the C2–C4 region within a cephalometric image

Figure 1 presents the labeling results of C2-C4 vertebra regions determined by YOLOv8 and YOLOv11 models. These areas underwent further analysis, and precise annotations for each vertebra (C2, C3, and C4) were made using the QuPath software. The flexibility offered by QuPath enabled accurate labeling of complex anatomical structures. During the annotation process, the boundaries of each vertebra were manually marked, taking into account their morphological features. This process contributed both to enriching the dataset for segmentation models and to obtaining a more detailed dataset for subsequent analysis phases. Figure 2 illustrates a visual example of this process performed using QuPath, detailing the annotation procedure for vertebra regions. This step is a critical part of model training, aiming to enhance segmentation and classification accuracy.

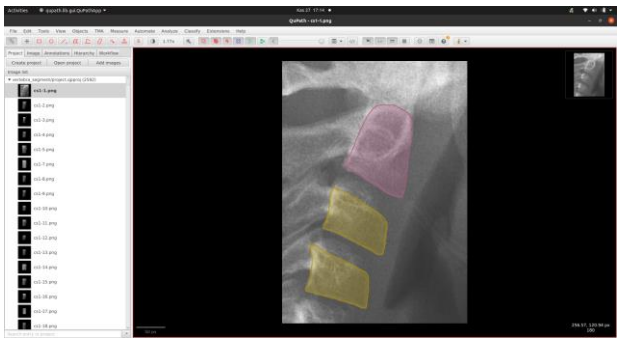
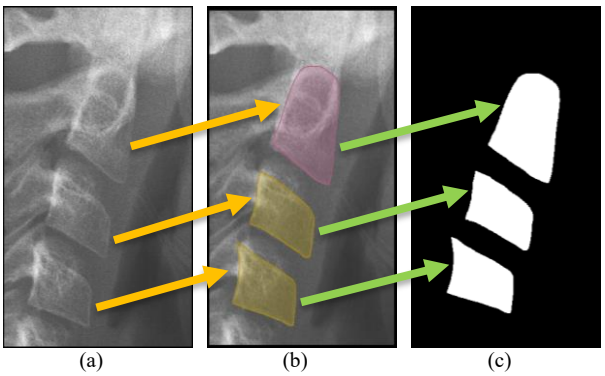


Figure 2. Annotation process for C2-C4 vertebrae

As a result of this process, a dataset containing input data for the deep learning-based U-Net models previously used in our study was created. Following the completion of the labeling and annotation processes, the obtained input images were prepared in RGB format and saved as PNG files, each with a resolution of 512x512 pixels. This process aimed to ensure the consistency and quality standards of the input data to optimize the model's segmentation capabilities.

Additionally, binary masks of the designated regions for each vertebra (C2, C3, C4) were generated as output images using QuPath. These masks were created in black-and-white format to emphasize only the relevant anatomical regions in the images and were also saved in PNG format with a resolution of 512x512 pixels. Defining the masks as black (background) and white (region of interest) facilitated the models' learning process by enabling a clearer distinction of target regions, thereby enhancing segmentation accuracy.



(a) (b) (c)

Figure 3. Segmentation process of C2-C4 vertebrae:
(a) Raw cephalometric image,
(b) Annotations of vertebra regions using QuPath,
(c) Binary masks created for the respective vertebrae.

Figure 3 presents a sample visual of the output images generated at the end of this process, providing a visual explanation of how the outputs are structured. These steps are crucial for ensuring the homogeneity of the dataset prepared for model training and enhancing the effectiveness of deep learning models. The dataset prepared in this manner offers a high standard to improve model performance and ensure suitability for future applications.

3.3. Object Detection

YOLOv8 and YOLOv11 models were employed for the automatic detection and segmentation of C2-C4 vertebra regions, forming a pivotal stage in the research methodology due to their rapid and accurate detection capabilities. YOLOv8 (You Only Look Once, Version 8) is a deep learning model that provides fast and effective results for object detection and classification tasks. As the latest version in the YOLO series, YOLOv8 introduces various improvements built upon previous models. The network architecture has been optimized to achieve more precise and faster object detection performance. YOLOv8 stands out with adaptive bounding box headers, better feature map extraction, and dynamic data augmentation techniques. The model is designed to be trained on large datasets and is preferred for real-time applications due to its low latency. Specifically, it provides an optimal balance between speed and accuracy [12].

YOLOv11 is one of the most up-to-date object detection models and offers significant improvements compared to earlier YOLO versions. This model utilizes deep learning techniques more efficiently, providing high accuracy in detecting both small and large objects. YOLOv11 uses multi-scale feature maps to improve performance across a wide range of scales and incorporates advanced regularization techniques that enhance the model's generalization ability. Additionally, the model is equipped with attention mechanisms, enabling more effective object detection, especially in complex scenes. YOLOv11 is a preferred model for both academic and industrial applications due to its real-time performance [13][14].

In this study, YOLOv8 was selected as the primary model for the automated, accurate, and rapid detection of C2, C3, and C4 vertebrae. YOLOv8 has demonstrated exceptional performance in object detection tasks, particularly excelling in mAP50 and mAP50-95 metrics. This capability ensures high accuracy, even in complex datasets such as lateral cephalometric radiographs with intricate anatomical structures. Additionally, the model's low latency facilitates efficient detection, enabling faster processing and minimizing potential diagnostic errors in clinical workflows. The advanced architecture of YOLOv8 effectively balances detection accuracy and computational efficiency. Its capacity to handle diverse and complex data structures provides a robust foundation for segmentation processes, where precision is critical. Moreover, the model simplifies the detection phase, reducing reliance on manual interventions and promoting a standardized, reproducible workflow. In contrast, YOLOv11, while being a lightweight model with lower hardware requirements, was not chosen as it could not match the high accuracy and rapid detection performance offered by YOLOv8. Based on these technical advantages, YOLOv8 was deemed the most suitable model for achieving the objectives of this study.

Figure 4 shows the workflow used for detecting the C2-C4 vertebra regions with the YOLOv8 architecture. In the input stage, cephalometric images were provided to the model. In the backbone section, image features were processed through convolutional layers, and key features were extracted. In the neck section, feature maps of different scales were merged and upscaled to enable multi-scale detection. In the prediction stage, the detected regions were identified, and outputs were generated. As a result, the C2-C4 vertebra regions were accurately detected, and the model's outputs were displayed.

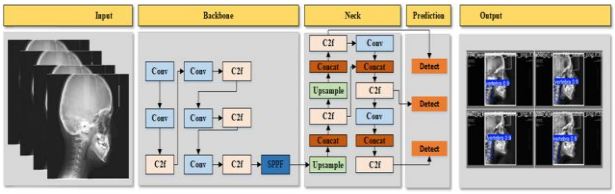


Figure 4. Workflow used for detecting the C2-C4 vertebra regions with the YOLOv8 architecture [15]

3.4. Semantic Segmentation

Semantic segmentation is a computer vision technique aimed at assigning a class to each pixel in an image. This method not only detects objects in an image but also defines the spatial boundaries of these objects in detail. Semantic segmentation is widely used in various fields, including medical image analysis, autonomous vehicles, satellite imaging, and augmented reality, where there is a need to distinguish objects or regions belonging to different classes, especially in complex scenes. It relies on deep learning models to classify pixels. The model extracts both low-level features (e.g., color and edges) and high-level features (e.g., shape and object meaning) from the input image. Each pixel is then assigned a class. Models such as U-Net, Fully Convolutional Networks (FCN), SegNet, and DeepLab are commonly used for this purpose. Semantic segmentation is frequently preferred in applications requiring high accuracy, such as medical image analysis. For instance, segmenting specific anatomical structures (e.g., organs or tissues) in X-ray or MRI images helps support diagnosis and treatment processes. Additionally, automatic and rapid segmentation reduces the time loss and human error associated with manual labeling. This method typically relies on an encoder-decoder architecture. The encoder extracts features from the input image, while the decoder uses these features to perform pixel-level classification. The model is often optimized using loss functions such as cross-entropy or dice loss in segmentation processes.

Semantic segmentation enables the precise identification of object or region boundaries by classifying pixels in detail, offering time savings and accuracy improvements compared to manual labeling. It is widely used in areas such as medical image analysis (organ and tumor segmentation), autonomous vehicles (road and pedestrian detection), satellite imaging (land classification), and industrial quality control [16] [17] [18].

Attention-UNet is a model that adds attention mechanisms to the classic U-Net architecture, highlighting important regions in the image. This model is particularly used to segment target regions more accurately in complex medical images. The attention mechanisms filter out irrelevant areas in the image while

improving classification and segmentation accuracy. Additionally, with attention gates, the model reduces unnecessary information overload by focusing only on the necessary features. This provides a significant advantage, especially in cases with limited datasets and low-contrast images [19].

Attention-ResUNet enhances the U-Net architecture by adding residual connections and attention mechanisms, which both simplify the learning process and improve segmentation performance. Residual connections speed up learning by reducing gradient loss encountered during training of deeper layers of the model. The attention modules ensure that critical regions in the image are emphasized more effectively. This model provides high accuracy in delicate tasks like tissue or organ segmentation in medical images, especially for low-contrast and complex structures [20].

SEEA-UNet integrates Squeeze-and-Excitation (SE) blocks into the U-Net architecture, enabling more effective emphasis on important features. SE blocks dynamically adjust the importance of each feature map, highlighting areas the model needs to focus on. This feature offers a significant advantage in medical image segmentation, especially when working with limited data. SEEA-UNet delivers high precision and enhances overall performance, particularly in areas like organ segmentation or distinguishing bone structures [21].

ResAt-UNet is a deep learning model that stands out in the field of medical image segmentation. This model combines the ResNet-based encoder structure with the U-Net architecture to provide effective segmentation performance. The model aims to improve segmentation accuracy with attention mechanisms and residual connections. Attention mechanisms allow the model to focus on important areas during segmentation, while residual connections prevent information loss and enable effective learning in deeper layers of the model. The main advantage of ResAt-UNet is that it provides more precise segmentation, especially when working with low-contrast and complex structures in medical images. This model is typically evaluated using metrics like Dice Similarity Coefficient (DSC) and Intersection over Union (IoU), where it yields higher results compared to the traditional U-Net architecture. ResAt-UNet is applied in various medical tasks such as brain tumor segmentation, lung lesion detection, and organ delineation. By combining ResNet and attention mechanisms, the model ensures both efficient feature extraction and the ability to focus on critical regions. These features make ResAt-UNet a strong choice for medical segmentation problems [22].

4. EXPERIMENTAL STUDIES

In our study, a powerful hardware infrastructure was preferred for processing large volumes of data and training deep learning models. In this context, a system with 100 GB of RAM capacity was used, and a dual GPU configuration from the NVIDIA RTX A4000 model, which is known for providing high performance in artificial intelligence applications, was specifically chosen. These GPUs played a critical role in accelerating deep learning algorithms and processing large datasets. The stages of dataset preparation, labeling, and image processing were carried out using the Python programming language. Thanks to Python's flexible structure and extensive library support, image processing and the implementation of deep learning models were efficiently completed. Popular libraries such as TensorFlow and NumPy were used during model training and validation stages. Additionally, tools like Pandas and Matplotlib were preferred to enhance the accuracy of data preprocessing and analytical processes.

With this robust hardware and software infrastructure, model training and testing processes were completed quickly, and it was also possible to optimize complex models and work with large datasets. The provided infrastructure contributed to the efficiency of the study and enhanced the reliability of the results.

As shown in Figure 5, the main workflow of our study is systematically visualized. This process encompasses all steps, starting from raw data processing and leading to the final output.

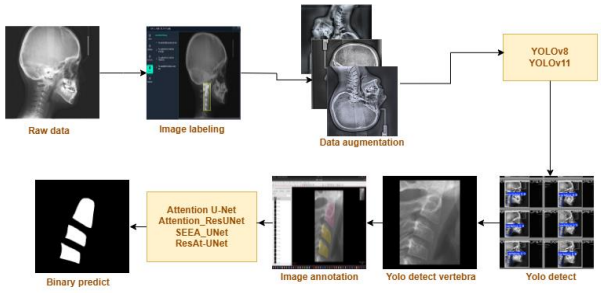


Figure 5. Flow diagram of the study

The study begins with the collection of raw lateral cephalometric X-ray images. On these images, the vertebra regions were highlighted and labeled. During the labeling process, particular focus was placed on the C2, C3, and C4 vertebrae, ensuring that these regions were clearly defined. Following the labeling process, the dataset was augmented with techniques such as rotation and scaling, as well as methods like Gaussian noise and contrast adjustments. These steps aimed to improve the models' generalizability under varying imaging conditions and to evaluate their robustness to noise. This comprehensive data augmentation strategy significantly contributed to assessing model performance in a broader context and adapting them for clinical applications.

Before segmentation, vertebra regions were automatically detected using YOLOv8 and YOLOv11 models. The speed and accuracy provided by these models effectively isolated critical vertebra regions, after which the necessary annotations for segmentation were performed. In the segmentation phase, various deep learning models with attention mechanisms were used. These included Attention-UNet, Attention-ResUNet, SEEA-UNet, and ResAt-UNet models, each aimed at more accurately differentiating the vertebra regions. The binary prediction

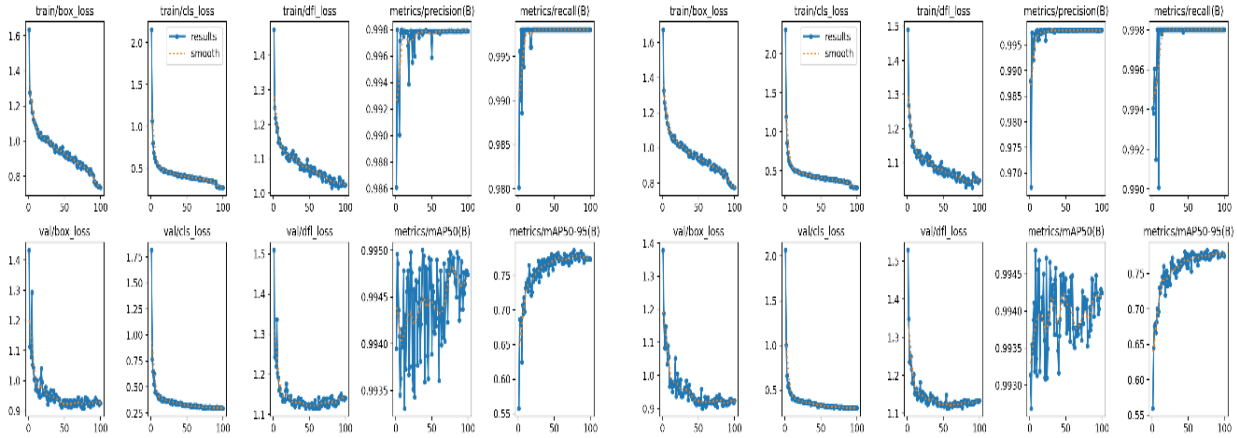
masks obtained from the application of these models clearly delineated the vertebral structures.

In this study, the cervical vertebra regions were detected using the YOLOv8 and YOLOv11 models. Both models were trained on a total of 2520 lateral cephalometric X-ray images. The data was divided into 70% for training, 20% for validation, and 10% for testing. During the training process, the batch size was set to 16 and the image size to 640 pixels. The training continued for a total of 100 epochs. The results for the models are listed in Table 1, and the outcome metrics are visually presented in Figure 6.

Table 1. Comparative results of YOLOv8 and YOLOv11 models

Metrik	Yolov8	Yolov11
Precision (P)	0.998	0.998
Recall (R)	0.998	0.998
Map50	0.995	0.994
Map50-95	0.783	0.782
Inference Time	1.8 Ms	1.9 Ms
Preprocess Time	0.3 Ms	0.3 Ms
Postprocess Time	0.3 Ms	0.3 Ms
Total Number of Layers	168	238
Number of Parameters	3,005,843	2,582,347
Gflops	8.1	6.3

When the results of both models were examined, it was observed that both YOLOv8 and YOLOv11 performed exceptionally well in object detection tasks. Both models detected the vertebra regions with high accuracy, achieving precision and recall values of 99.8%. YOLOv8 exhibited a slight advantage over YOLOv11 in mAP50 (99.5%) and mAP50-95 (78.3%) metrics. Additionally, YOLOv8 was observed to provide faster inference with a time of 1.8 ms. On the other hand, the YOLOv11 model operates with fewer parameters (2,582,347) and lower computational power requirements (6.3 GFLOPs). This indicates that YOLOv11 is a lighter model and offers a suitable alternative for applications that need to run on lower hardware resources. In terms of preprocessing and postprocessing times, both models provided similar results (0.3 ms).

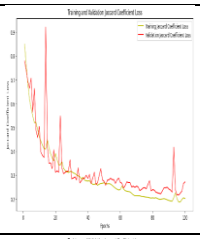
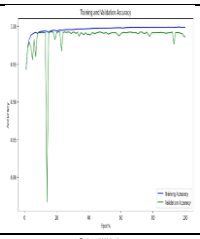
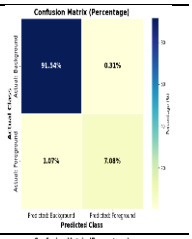
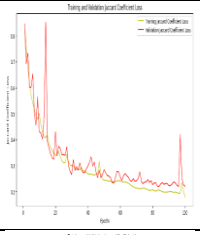
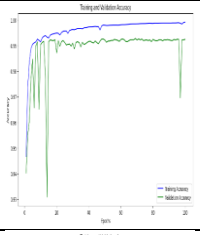
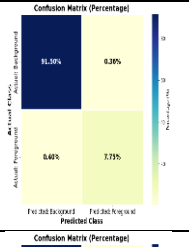
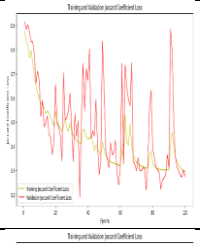
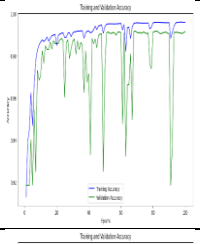
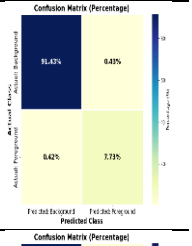
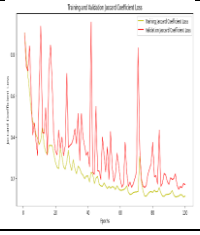
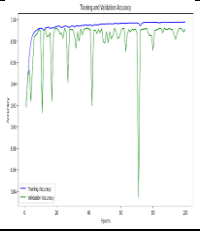
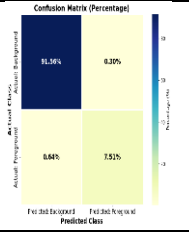


(a) (b)

Figure 6. Graphical representation of the results for the YOLOv8 (a) and YOLOv11 (b) models.

As a result, the graphs show a consistent decrease in the loss values during the training process for both models. In both YOLOv8 and YOLOv11 models, the metrics of train/box_loss, train/cls_loss, and train/dfl_loss steadily decreased after 100 epochs. Similarly, a significant reduction was observed in validation losses, and the models showed an increased generalization capacity throughout the training process. Looking at the Precision and Recall graphs, both models reached levels of 99.8%, with a rapid improvement early in the training process. For the mAP50 metric, YOLOv8 reached 99.5%, while YOLOv11 showed a similar result of 99.4%. In the mAP50-95 values, YOLOv8 slightly outperformed with 78.3%, while YOLOv11 showed 78.2%. YOLOv8 demonstrated superior performance in terms of accuracy and speed, while YOLOv11 stands out with its lower computational load. Both models offer effective and reliable options for cervical vertebra detection. These detections were used as input data for segmentation models, enabling a detailed analysis of the vertebra regions. During the segmentation phase, the aim was to reveal finer details and classify the detected regions more accurately. This approach provided a solid foundation for subsequent analysis and model accuracy by clearly differentiating the vertebral structures. The results for the models used in this context are shown in Table 2.

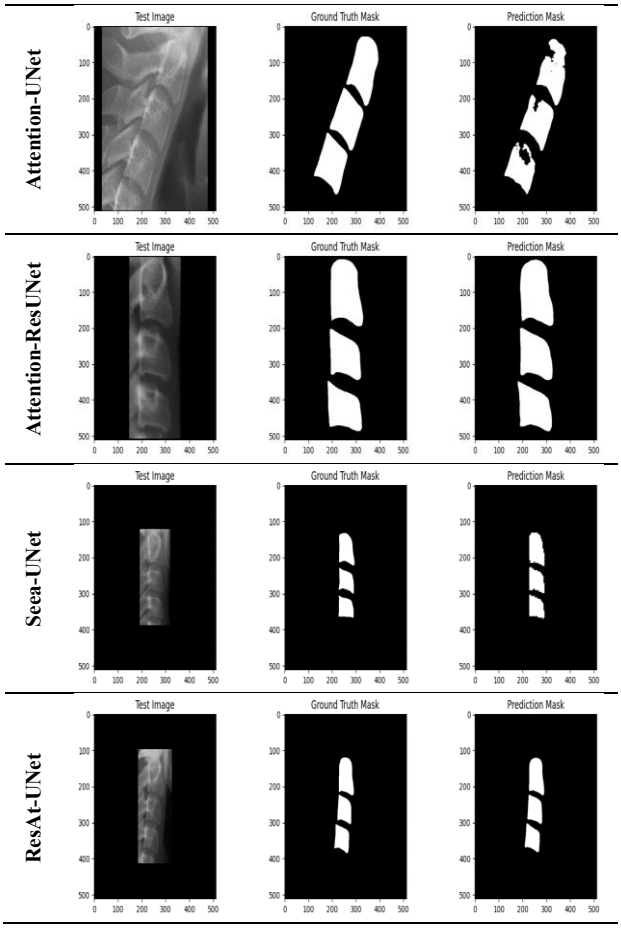
Table 2. Accuracy/loss graphs and confusion matrix of segmentation models' performance results

	Accuracy & Loss Graphs		Confusion Matrix
Attention-UNet			
Attention-ResUNet			
Seea-UNet			
ResAt-UNet			

In Table 2, where the performance results of the segmentation models are examined, various findings have been obtained based on accuracy and loss graphs, as well as confusion matrices. The four different models used in the study Attention-UNet, Attention-ResUNet, SEEA-UNet, and ResAt-UNet demonstrated different performances during the training and validation phases. The Attention-UNet model achieved 99.88% training accuracy and 98.62% validation accuracy. However, fluctuations in the validation loss indicate that the model showed signs of overfitting in some cases and that its generalization ability might be limited. The Attention-ResUNet model, on the other hand, performed the best, with 99.92% training accuracy and 99.25% validation accuracy. The consistency between the training and validation losses suggests that the model has a strong generalization capacity. In the SEEA-UNet model, training accuracy was 99.59% and validation accuracy was 99.16%. The increase in validation loss suggests that the generalization ability of this model might be slightly more limited compared to the other models. The ResAt-UNet model achieved 99.78% training accuracy and 99.87% validation accuracy, obtaining a balanced result between training and validation performance. Low validation loss and Jaccard index loss indicate that the model has a high generalization capacity. When performance results based on the confusion matrix are evaluated, it is observed that the models show varying success levels in the "background" and "foreground" classes. The Attention-UNet model achieved high accuracy in the background class (99.25% F1-score), but drew attention to a lower recall value in the foreground class (86.86%), indicating difficulty in detecting some true positive examples. The Attention-ResUNet model reached 95.35% F1-score in the foreground class, demonstrating a balanced performance in both precision and recall. This shows that the model has a high detection rate and accurately identifies most of the true positive examples. The SEEA-UNet model also achieved high precision and recall values, but the recall value in the foreground class was slightly lower. The ResAt-UNet model stands out with a balanced result between training and validation accuracy and low loss values. It achieved 94.14% F1-score in the foreground class, but the recall value was slightly lower compared to the other models (92.16%), indicating that the model has strong generalization ability but may miss some true positives. Overall, the highest accuracy rate of 99.24% was achieved by the Attention-ResUNet model. The ResAt-UNet model, however, demonstrated strong generalization capacity with a balanced performance between training and validation and low loss values. The SEEA-UNet model stands out with its fast learning ability, while the Attention-UNet model, despite high training accuracy, lags behind other models in terms of generalization due to fluctuations in validation loss.

The Attention-ResUNet model demonstrated the highest inter-class agreement with a Cohen's Kappa value of 94.9%, exhibiting superior performance in segmentation accuracy. Statistical power analysis conducted for this model indicated that the dataset provided 100% power, confirming that the results are highly generalizable. SEEA-UNet achieved a Cohen's Kappa value of 94.4%, displaying high agreement and particularly excelling in small and detailed regions. Similarly, the statistical power for this model was calculated as 100%, establishing the reliability of its outcomes. The ResAt-UNet model, with a Cohen's Kappa value of 93.6%, stood out for its balanced generalization capacity, and power analysis demonstrated that this model also operated with 100% power, reinforcing the statistical significance of its results. The Attention-UNet model achieved a strong agreement with a Cohen's Kappa value of 90.4%, and power analysis revealed that the dataset used for this model also had 100% power. The power analyses conducted for each model demonstrated that the dataset, consisting of 2,520 images, is more than sufficient for ensuring the generalizability and reliability of the results. These findings confirm that the models provide effective and standardized methods suitable for clinical applications.

Table 3. Contains visuals of the predictions made by the segmentation models.



When the visuals of the segmentation model predictions in Table 3 are evaluated alongside the performance results of the models in Table 2, it can be observed that the Attention-UNet model provided very accurate predictions for the segmentation masks, especially in large and distinct vertebra regions. However, some deficiencies in the predicted masks were observed in smaller and boundary areas. This is consistent with the recall value of 86.86% in the foreground class. The model's precision value is at 95.74%, indicating a high true positive prediction rate, but some true positives have been missed. The Attention-ResUNet model demonstrated a more balanced performance in the segmentation masks, both in the background and foreground classes. The precision value for the foreground class was recorded at 95.61%, and the recall value at 95.10%. The segmentation masks showed that the vertebra boundaries were accurately segmented. Especially in complex boundary structures, the predicted masks closely matched the ground truth masks. The SEEA-UNet model stood out for its fast learning ability and low loss values. The segmentation masks in the visuals produced accurate results in large vertebra areas. However, the model's recall value is at 94.81%, and some deficiencies were observed in smaller vertebra regions. The precision value of 94.76% indicates a high accuracy rate. The ResAt-UNet model, especially with its segmentation accuracy in boundary areas, stood out. The segmentation masks showed that the vertebra boundaries were largely accurately segmented in both the background and foreground classes. The precision value for the foreground class was 96.20%, and the recall value was 92.16%. This shows that the model's accuracy is high, but some small vertebra areas were missed. Overall, the visual analysis of the segmentation masks indicates consistency with the models' classification metrics. The Attention-ResUNet model stands out in terms of overall accuracy and balance of the segmentation masks. The SEEA-UNet and ResAt-UNet models provided effective predictions in prominent vertebra areas but showed slight deficiencies in smaller areas. The Attention-UNet model, despite its high accuracy, made more errors in smaller and complex areas compared to the other models.

Table 4. IoU results of segmentation models

Model	IoU
Attention-ResUNet	0.9506
SEEA-UNet	0.9419
ResAt-UNet	0.9405
Attention-UNet	0.9219

Table 4 presents the Intersection over Union (IoU) values of the four different models used in this study. IoU is a critical metric for evaluating the performance of semantic segmentation models, quantitatively expressing the overlap ratio between the predicted and actual regions. According to Table 4, the highest IoU value, 95.06%, was achieved by the Attention-ResUNet model, which stands out with its superior segmentation accuracy. SEEA-UNet and ResAt-UNet demonstrated balanced and high accuracy, with IoU values of 94.19% and 94.05%, respectively. The Attention-UNet model, on the other hand, showed a comparatively lower performance with an IoU value of 92.19%. These differences in IoU values clearly highlight the sensitivity and generalization capacities of the models on the dataset. Particularly, Attention-ResUNet is considered a strong candidate for clinical applications in terms of segmentation accuracy. These findings demonstrate that AI-based approaches can achieve high levels of accuracy in semantic segmentation processes, significantly contributing to clinical decision-making workflows.

5. RESULTS AND FUTURE WORK

In this study, an artificial intelligence-based method was developed for the detection and semantic segmentation of the C2, C3, and C4 cervical vertebrae on lateral cephalometric radiographs. Our goal was to create a reliable infrastructure that will lay the groundwork for bone age prediction and skeletal maturity analysis, while also evaluating the potential of artificial intelligence technologies in this field.

In the study, the YOLOv8 and YOLOv11 models were used for object detection tasks, with both models achieving a 99.8% accuracy rate, enabling the rapid and effective detection of cervical vertebra regions. For the segmentation tasks, detailed annotations were performed using QuPath software, and the data were subsequently analyzed using deep learning models such as Attention-UNet, Attention-ResUNet, SEEA-UNet, and ResAt-UNet. In the comparative evaluation of these models, the Attention-ResUNet model stood out with a segmentation performance of 99.25% accuracy, while ResAt-UNet

drew attention with its superior generalization capacity. The resulting segmentation masks were prepared with high precision and created a suitable dataset for bone age prediction and related analyses. This work contributes to accelerating clinical processes and obtaining more reliable results, particularly in fields such as orthodontics and pediatric medicine. The developed method eliminates the need for additional imaging, reduces radiation exposure, and prevents subjective errors in manual evaluations. With these features, it offers an alternative approach to traditional methods such as hand-wrist radiographs. Additionally, this method has the potential to create a standardized framework for skeletal maturity analysis and bone age prediction processes.

This study demonstrates that the methods developed for the automatic detection and segmentation of C2, C3, and C4 vertebrae provide an innovative contribution by minimizing the subjective errors encountered in manual analyses in the literature and reducing radiation exposure. While similar studies in the literature have made significant contributions to advancements in this field, the proposed methods stand out due to the use of YOLO models and detailed segmentation stages. In this study, the original improvements made in model configurations have enhanced segmentation accuracy, offering a more efficient application in clinical processes.

In future studies, it is planned to perform automatic bone age prediction using the binary datasets obtained in this study. To improve the accuracy of these predictions, the existing dataset will be enriched, and analyses for different age groups will be conducted. The goal in this process is to enhance model performance using advanced artificial intelligence algorithms. It is expected that the results will allow for a more precise assessment of individuals' growth and development processes. Additionally, it is anticipated that these methods will also be used in the diagnosis of skeletal anomalies and in surgical planning processes.

In conclusion, this study represents an important step in the integration of artificial intelligence-based methods in medical image analysis. The developed method not only accelerates clinical processes and reduces error rates but also contributes to the digital transformation of modern medicine. In future studies, it is hoped that these approaches will be integrated into different medical application areas, providing more comprehensive and widely applicable solutions.

Notice

This study is derived from the doctoral dissertation entitled “Bone Age Assessment through the Analysis of Cervical Vertebrae in Lateral Cephalometric Radiographs Using Semantic and Instance Segmentation Methods,” conducted by Mazhar Kayaoğlu under the academic supervision of Abdulkadir Şengür.

REFERENCES

[1] M. Makaremi, C. Lacaule, and A. Mohammad-Djafari, “Deep learning and artificial intelligence for the determination of the cervical vertebra maturation degree from lateral radiography,” Entropy, vol. 21, no. 12, p. 1222, 2019.

[2] N. Masuzawa et al., “Automatic segmentation, localization, and identification of vertebrae in 3D CT images using cascaded convolutional neural networks,” in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2020, LNCS 12261. Springer, 2020.

[3] O. Demirel and E. Sonuç, “Yapay Zeka Teknikleri Kullanılarak Kemik Yaşı Tespiti,” Türkiye Sağlık Enstitüleri Başkanlığı Dergisi, vol. 4, no. 3, pp. 17–30, 2021.

[4] Y. Chen et al., “Vertxnet: Automatic segmentation and identification of lumbar and cervical vertebrae from spinal x-ray images,” arXiv preprint, arXiv:2207.05476, 2022.

[5] M. Khazaei et al., “Automatic determination of pubertal growth spurts based on the cervical vertebral maturation staging using deep convolutional neural networks,” J. World Fed. Orthod., vol. 12, no. 2, pp. 56–63, 2023.

[6] H. Li et al., “The psc-CVM assessment system: A three-stage type system for CVM assessment based on deep learning,” BMC Oral Health, vol. 23, no. 1, p. 557, 2023.

[7] G. A. Kresnadhi et al., “Comparative Analysis of ResNet101, InceptionV3, and InceptionResnetV2 Architectures for Cervical Vertebrae Maturation Stage Classification,” in Proc. 2023 Int. Conf. on Electrical Engineering and Informatics (ICE), 2023.

[8] G. Akay et al., “Deep convolutional neural network—The evaluation of cervical vertebrae maturation,” Oral Radiol., vol. 39, no. 4, pp. 629–638, 2023.

[9] S. F. Atici et al., “A Novel Continuous Classification System for the Cervical Vertebrae Maturation (CVM) Stages Using Convolutional Neural Networks,” 2023.

[10] P. Motie et al., “Improving cervical maturation degree classification accuracy using a multi-stage deep learning approach,” 2024.

[11] M. H. Mohammed et al., “Convolutional Neural Network-Based Deep Learning Methods for Skeletal Growth Prediction in Dental Patients,” J. Imaging, vol. 10, no. 11, p. 278, 2024.

[12] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” arXiv preprint, arXiv:1804.02767, 2018.

[13] Z. Ge, “Yolox: Exceeding yolo series in 2021,” arXiv preprint, arXiv:2107.08430, 2021.

[14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” arXiv preprint, arXiv:2004.10934, 2020.

[15] H. Herfandi et al., “Real-Time Patient Indoor Health Monitoring and Location Tracking with Optical Camera Communications on the Internet of Medical Things,” Appl. Sci., vol. 14, no. 3, p. 1153, 2024.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, 2015, pp. 5–9.
[18] A. Garcia-Garcia et al., "A review on deep learning techniques applied to semantic segmentation," arXiv preprint, arXiv:1704.06857, 2017.
[19] O. Oktay et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint, arXiv:1804.03999, 2018.
[20] J. Schlemper et al., "Attention-gated networks for improving ultrasound scan plane detection," arXiv preprint, arXiv:1804.05338, 2018.
[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[22] Z. Fan et al., "ResAt-UNet: a U-shaped network using ResNet and attention module for image segmentation of urban buildings," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 16, pp. 2094–2111, 2023.