OPEN ACCESS

PRZEGLĄD ELEKTROTECHNICZNY

Wojciech ŚWIĄTKOWSKI[1], Zuzanna KRAWCZYK-BORYSIAK[2], Wojciech NALEPA[3], Radosław KOZIARA[4]

ORCID: [1]0000-0002-4416-7289, [2]0000-0002-3897-1078, [3]0009-0000-2760-6422

DOI: 10.15199/48.2025.11.40

Evaluating the efficacy of the Segment Anything Model (SAM) in streamlining dental cyst segmentation within CBCT volumes

Ocena skuteczności modelu Segment Anything Model (SAM) w usprawnianiu segmentacji torbieli w danych obrazowych CBCT

[1] dr Wojciech Świątkowski, Clinic of Maxillofacial Surgery, Medical University of Lublin, ul. Staszica 11, 20-081 Lublin, e-mail: wojciech.swiatkowski@umlub.pl

[2] dr inż. Zuzanna Krawczyk-Borysiak, Faculty of Electrical Engineering, Warsaw University of Technology, IETISIP PW, ul. Koszykowa 75, 00-662 Warszawa, e-mail: zuzanna.krawczyk@pw.edu.pl

[3] Wojciech Nalepa, Chair and Department of Dental Surgery, Medical University of Lublin, e-mail: adwa834@gmail.com

[4] Radosław Koziara, AGH University of Krakow, E-mail: rkoziara@agh.edu.pl

Correspondence address: wojciech.swiatkowski@umlub.pl

Abstract: The paper aims to assess the usefulness of the recent foundation model called Segment Anything Model (SAM) as semi-automatic help during the cyst annotation step from cone-beam computed tomography (CBCT) Volumes, which can speed-up creation of extensive training dataset. The authors compare the accuracy of the SAM model and its version fine-tuned on medical data, MedSAM, against segmentations of 15 use cases performed by an expert. Test cases include both cysts in the jaw and sinus area. The influence of the model size on the accuracy of the results is also investigated.

Keywords: Segment Anything Model (SAM), segmentation, cone-beam computed tomography (CBCT), cysts

Streszczenie: Artykuł ma na celu ocenę przydatności modelu o nazwie Segment Anything Model (SAM) jako półautomatycznej pomocy podczas etapu adnotacji torbieli z tomografii stożkowej CBCT, co może przyspieszyć tworzenie obszernego zbioru danych szkoleniowych. Autorzy porównują dokładność modelu SAM i jego wersji wytrenowanej na danych medycznych, MedSAM, z segmentacją 15 przypadków użycia przeprowadzoną przez eksperta. Przypadki testowe obejmują zarówno torbiele w obszarze szczęki, jak i zatok. Zbadano również wpływ rozmiaru modelu na dokładność wyników.

Słowa kluczowe: SAM, segmentacja, tomografia stożkowa, torbiel

Introduction

In Poland, as in many other countries, the number of cone-beam computed tomography (CBCT) devices and the number of examinations performed using this modality continue to increase. [1]. According to current national regulations, dentists are authorized to interpret CBCT scans, provided that the field of view (FOV) does not exceed 8×8 cm. In clinical practice, this regulation frequently results in dentists analyzing CBCT images acquired in their own practices without a formal radiological report. In this context, artificial intelligence (AI)–based algorithms, particularly those utilizing deep learning techniques, may offer significant support in identifying potential pathological findings. One example includes cystic lesions of the head and neck region, which can remain asymptomatic and undetected in routine examinations. The integration of such AI tools into dental workflows may enhance diagnostic accuracy and facilitate early detection of clinically relevant anomalies.

A cyst is defined as a pathological cavity enclosed by a distinct epithelial lining, containing various substances such as air, fluid, mucus, keratin, or lipid material [2]. Odontogenic cysts originate from remnants of odontogenic epithelium that become entrapped within bone or gingival tissues, whereas non-odontogenic cysts arise from epithelial remnants unrelated to the tooth-forming apparatus. These lesions typically exhibit slow, expansive growth patterns, and in some instances, are associated with significant bone destruction and a risk of recurrence [3-5]. Odontogenic cysts are commonly classified based on their etiology as either inflammatory, such as radicular cysts (RC), or developmental, such as dentigerous cysts [6].

Radiological features of cystic lesions in the oromaxillofacial region are not always pathognomonic and may vary depending on the cyst type, duration, anatomical location, extent of expansion, and presence of infection. Panoramic radiography and Cone Beam Computed Tomography (CBCT) are commonly used to assess these lesions, offering valuable information regarding size, margins, and spatial relationships.

Typically, cysts appear as well-defined unilocular or multilocular radiolucencies, with or without corticated borders. Septations may be present in some cases. The morphology varies by cyst type–radicular and dentigerous cysts tend to be round or oval; odontogenic keratocysts and traumatic bone cysts often show scalloped margins; nasopalatine duct cysts may exhibit pear-shaped or heart-shaped appearances. In advanced cases, cortical expansion, perforation, or even pathological fractures can occur. Additionally, adjacent structures such as teeth, the maxillary sinus, or neurovascular bundles may be displaced or compressed [7].

Early detection and appropriate management of cystic lesions in the head and neck region are of critical importance. If left undiagnosed and untreated, such lesions may progressively enlarge and exert a destructive effect on surrounding tissues. This can result in serious complications, including pathological fractures, inflammatory involvement of adjacent anatomical structures, and, in rare cases, malignant transformation [8─10].

Automated cyst segmentation remains challenging task due to the high variability of cyst shapes, sizes, locations, and tissue density, which is similar or the same as that of other soft tissues (Fig. 1). The lower resolution of CBCT compared to the traditional CT study and artefacts from metallic implants often present in the image make it even more challenging.

Convolutional neural networks (CNN) and alternative deep learning models became state of the art solutions for automated segmentation tasks, that can overcome at least partially forementioned issues, however they do require large, precisely annotated datasets for the training phase to make them reliable [11]. The manual annotation process is tedious, time-consuming and requires highly trained expert. In order to speed it up, semi-automatic methods like thresholding or active shape can be used and incorporated into annotation flow, under the expert supervision. Recently so-called foundation models – neural networks trained on massive datasets, which can be easily adapted to wide range of tasks and display significant level of generality – are reshaping the deep learning models landscape.

This paper aims to assess the usefulness of the recent foundation model called Segment Anything Model (SAM) as semi-automatic help during the annotation step. The authors compare the accuracy of the SAM model and its version fine-tuned on medical data, MedSAM, against segmentations of 15 use cases performed by an expert. Test cases include both cysts in the jaw and sinus area. The influence of the model size on the accuracy of the results will also be investigated.

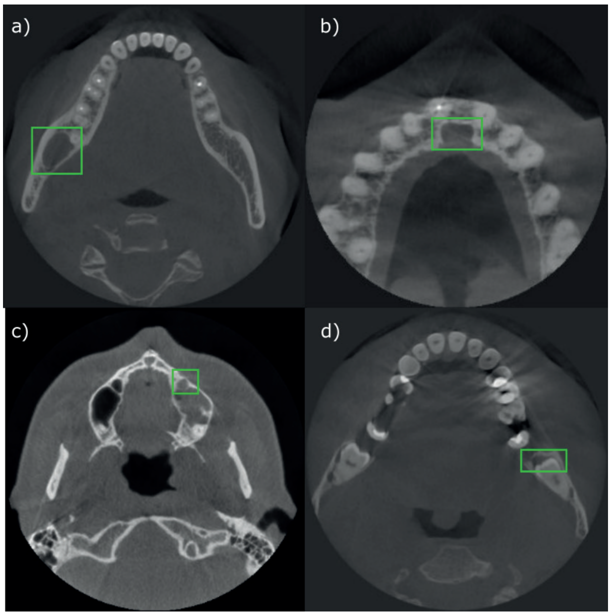

Fig. 1. Examples of cysts (marked with green box) variability and challenges across the dataset: a) the lesion demonstrates grayscale values similar to the surrounding soft tissues; b) no distinct boundary between the cystic capsule and the adjacent soft tissues of the palate; c) lesion hard to recognize without 3D context; d) atypical contour of the lesion on CBCT section, associated with its developmental origin as a dentigerous cyst

Literature review

While there have been several works on the detection of cysts [12─15] in panoramic radiographs, the CBCT modality has not yet been fully explored. However, several papers describe successful experiments in dental cyst segmentation tasks [16, 17]. In [16] authors proposed novel U-net based solution, broadly used in medical area [18-20], achieving Dice score of 88.84%, an IoU score of 81.23%, and an AUC score of 92.37%. 300 cases were used in model training process, which emphasizes the need to prepare a large set of training data in order to obtain satisfactory results. Authors of [17] achieved AUC equal to 91.4%, however they focused on detection task only and included in the dataset 2 to 4 slices from each CBCT set most representative to the lesion.

The SAM model has a wide range of applications in various fields [21─22]. Authors of [23] performed detailed analysis of Segment Anything Model (SAM) on 19 publicly available medical datasets of various modalities including MR, X-ray, ultrasound, CT and PET imaging, concluding that model outperforms other interactive segmentation approaches, however its performance varied significantly across datasets and images within the dataset itself. Nevertheless, authors consider SAM model as perspective one in medical annotation pipeline. Paper [24] also indicates multiple possible applications of the model in medical segmentation area.

The natural consequence: training the model with medical data was achieved in [25], which presented MedSAM model finetuned on 1,570,263 medical image-mask pairs from 10 modalities, improving its capabilities in this area. First version of SAM model operates on 2D images only, which can reduce its usefulness for 3D medical imaging modalities such as CBCT. Second version of the model SAM2 allows also to process video images and by that can be adapted to 3D medical data. MedSAM2 [26] version of the model allow to process 3D medical modalities.

In this paper, the authors focus primarily on evaluation of the first version of SAM model planning to extend it for other aforementioned alterations.

Segment Anything Model

The Segment Anything Model (SAM), developed by Meta AI Research and first introduced in 2023 [27], is a foundational model for image segmentation. It was trained on over 1 billion masks across 11 million images, forming the largest segmentation dataset to date at the time. SAM supports zero-shot segmentation by accepting various types of input prompts, such as points, bounding boxes, or a combination of both, enabling it to generalize to a wide range of tasks and previously unseen images.

SAM's architecture is composed of three main components:

1. an image encoder, specifically a Vision Transformer (ViT), which generates dense image embeddings;
2. a prompt encoder, which processes user prompts (points, boxes, masks, or text)
3. a mask decoder, which combines the image and prompt embeddings to produce the final segmentation mask.

Data preparation

The study protocol was reviewed and approved by the Bioethics Committee at the Medical University of Lublin (approval no. BKB/827/10/2024). CBCT scans depicting cystic lesions in the maxillofacial region were selected from the database of the Clinic of Maxillofacial Surgery at the Medical University of Lublin. All data were anonymized using OnDemand software. Digital processing of the CBCT images was performed in ITK-SNAP [28]. First, the region of interest (ROI) was marked in three planes, followed by semi-automatic segmentation using the active contour method (also known as the "snake" algorithm). Thresholding was applied as a pre-segmentation step, with both lower and upper thresholds defined, along with a smoothness parameter. Finally, segmentation results were verified and refined manually. The segmentations were saved in NIfTI format.

Following the recommendation of [23], which states that box prompts gave better outcome then the point-based ones the authors focused on bounding-box prompting. In order to create prompts the script was prepared which created bounding-box over the segmentation for each slice where it was present. Only slices where there was cyst segmentation present were passed further to the model. The input slices were transformed to 2D grayscale images. No prior cropping was applied.

Experiment

Three versions of SAM model were investigated with different size of weights: base one (93.7 million of parameters), large (312 million parameters) and huge (641 million). Models were prompted with bounding-box of the lesion, imitating area which could be chosen by medical expert during annotation task. Out of 3 segmentation results generated by the model, the output with the highest prediction score was chosen and saved. Then the outputs were compared with reference segmentations with the use of Dice Similarity Coefficient (DSC)(1), Intersection over Union (IoU)(2), Precision (positive predictive value)(3) and Recall (sensitivity / true positive rate)(4) metrices, which can be defined as follows:

$$(1) \qquad DSC = \frac{2TP}{2TP + FP + FN}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

where:
TP (true positive) – pixels of tumor area correctly segmented by the model,
FP (false positive) – pixels misclassified by the model as tumor,
FN (false negative) – pixels in the tumor area not recognized by the model.

After careful investigation authors noticed that cases 36, 31 and 34 present significantly worse results than other cases, which is reflected in Dice coefficient analysis performed for each case, depicted at Fig 2: median of case 36 has significantly lower value than other cases, cases 31 and 34 have the widest interquartile ranges. The visual analysis showed that reference mask was wrongly generated in case of 36. For cases 31 and 34 – there were several lesions visible on single slices which caused generation of one bounding-box for all of them. Such prompting was not precise enough for the model and indicating segmentations of other structures. Because lower performance was in that case the fault of preprocessing steps not the model itself aforementioned cases were excluded from further analysis.

Table 1 presents general overview of metrics for three versions of the model, with the large model achieving the highest IoU (77.73%) and Dice score (87%) values. Surprisingly, huge model, with the highest number of parameters, achieved the lowest value of those scores out of three models. This is because the Recall of the model was dropping with the model size, which may indicate that less pixels belonging to the cyst class were detected. On the other hand model increased its precision with the size – larger model generates less false positive pixels.

We evaluated also basic version of MedSAM model which achieved Dice score results by over 6% lower than standard large SAM model and was the weakest from all considered versions. MedSAM model however, excels in precision measure value. For further analysis we focused on large model version as the one achieving the highest score on general Dice measure, ranging from 75.32% to 94.83% for the best case.

Out of 12 cases 5 achieved global Dice coefficient value higher than 90% (Case 35, 11, 2, 16 and 1).

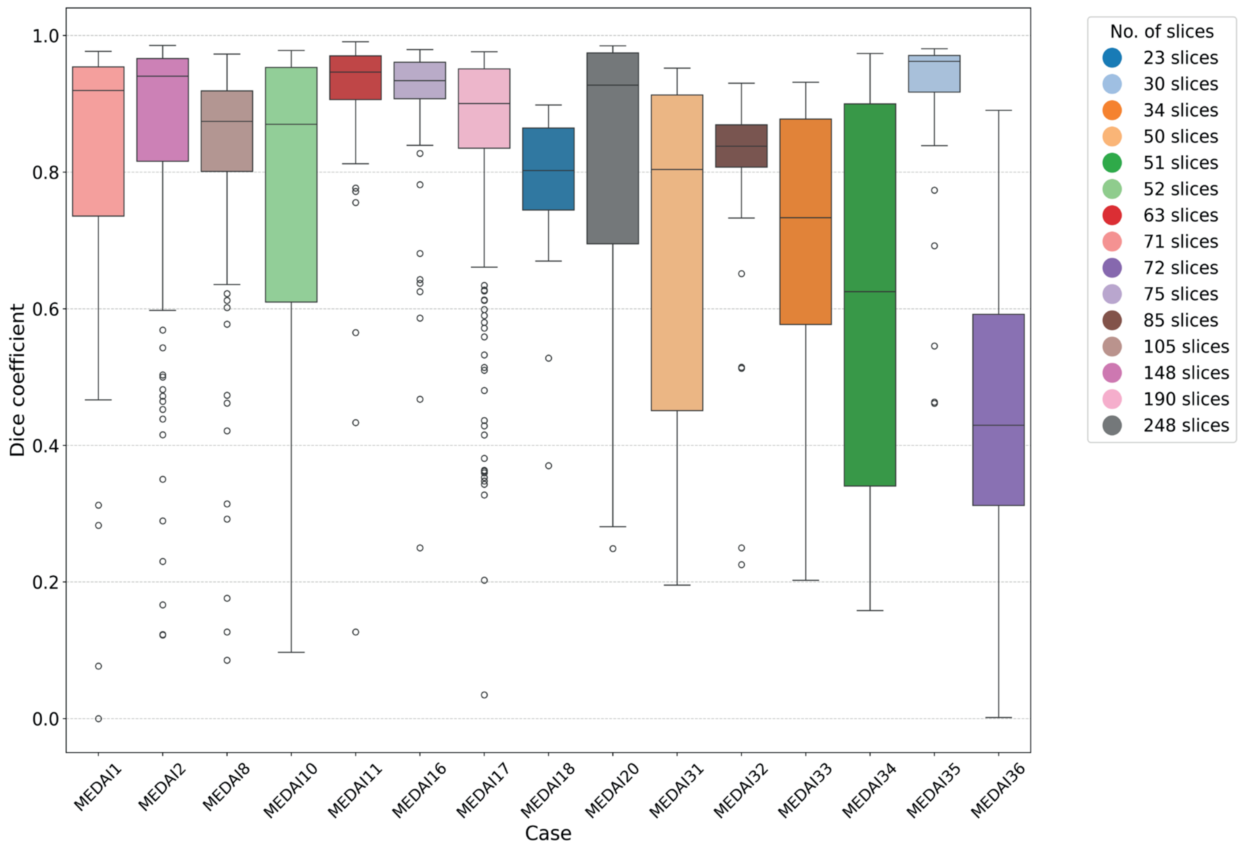

Fig. 2. Box plot representing Dice coefficient for each evaluated case (SAM large model). Color of boxes refers to the number of slices containing reference segmentation

Table 1 Average metrics values of SAM segmentation

Model version	IoU [%]	Dice [%]	Precision [%]	Recall [%]
Base	76.76	86.45	79.07	96.24
Large	77.73	87.11	81.26	94.38
Huge	74.74	84.65	83.35	86.31
MedSAM Base	68.92	80.80	88.22	75.25

In all aforementioned cases the cyst is easily distinguishable from the surrounding due to contrast between tissues or having well-defined sclerotic border in most slices, which results with nearly perfect segmentation as shown in Fig 3. However, cases 2 and 1 have higher variability of Dice coefficient value across slices than other best cases. Significantly lower results in some slices are caused by lack of bone barrier between cyst and rest of soft tissues. In case no 2 model sometimes segments tooth as the part of the lesion.

Significant errors in segmentation mostly include cases where bone structure was detected instead of the cyst, Such examples are shown in Fig. 4. Some of the mis-segmentations might be mitigated introducing more granular bounding-box prompting (like in Fig. 4c) and 4d), where one bounding-box encompassed two parts of the segmentation). Whereas, for cases a) and b) possible way of improvement would be usage of the model able to detect 3D relationships, or model fine-tuning.

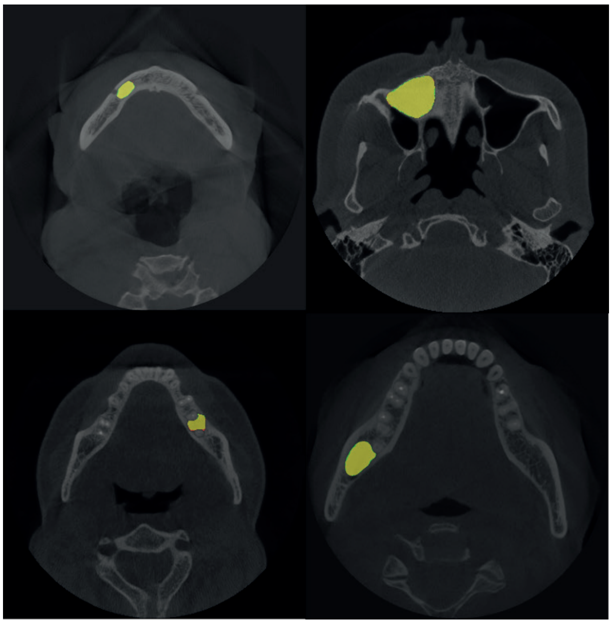

Fig. 3. Examples of nearly perfect segmentations of SAM large model (yellow color – overlap between ground truth and segmentation (TP), green – false positive (FP), red – false negative (FN))

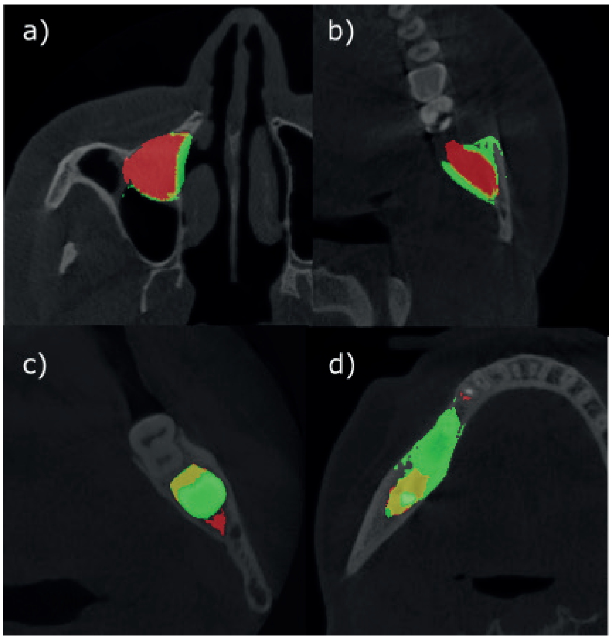

Fig. 4. Examples of bad segmentations of SAM large model (yellow color – overlap between ground truth and segmentation (TP), green – FP (model only segmentation), red – FN (reference only))

Summary

The following paper investigated the capabilities of foundation SAM model towards cyst segmentation in CBCT images. The best results were achieved by large version of the model with general Dice score value of 87.11%. The authors believe the model can be of substantial help in speeding up the segmentation process when used in combination with other tools allowing to easily perform its modification. The advantages of the model include:

- high general value of segmentation score and ability to adapt the segmentation to variety of slices with different anatomical features,
- nearly perfect segmentation results for areas with clear barrier between bone and cyst,
- easiness of implementation with no need of prior model training.

The weakness of the model is lack of 3D context which sometimes is causing lack of proper segmentation as well as over-segmentation on other structures. The model requires also fairly exact bounding-box prompts to be accurate.

Despite the previously described obstacles, the results are promising. The authors remain committed to utilizing all CBCT cross-sections encompassing the lesion, rather than relying solely on selected slices or panoramic radiographs [17]. The primary objective of the ongoing project is to substantially increase the sample size, in which we believe SAM will help us.

Received: 10.11.2025, **Accepted:** 12.11.2025, **Published:** 24.11.2025

REFERENCES

[1] Lam M, Critchley S, Zhang A, Monsour P., Current trends in the adoption and education of cone beam computed tomography and panoramic radiography machines across Australia. Dentomaxillofac Radiol, 50 (2021), No. 5, doi: 10.1259/dmfr.20200380

[2] Qiu W, Gu PR, Chuong CM, Lei M., Skin Cyst: A Pathological Dead-End with a New Twist of Morphogenetic Potentials in Organoid Cultures., Front Cell Dev Biol., 8 (2021), doi: 10.3389/fcell.2020.628114

[3] Da Silva LP, Gonzaga AK., Et al. Epidemiologic study of odontogenic and non-odontogenic cysts in children and adolescents of a Brazilian population., Med Oral Patol Oral Cir Bucal., 23 (2018), No. 1, doi: 10.4317/medoral.22138

[4] De Souza LB, Gordón-Núñez MA, Nonaka CF, de Medeiros MC, Torres TF, Emiliano GB. Odontogenic cysts: demographic profile in a Brazilian population over a 38-year period. Med Oral Patol Oral Cir Bucal., (15) 2010, No. 4

[5] Nonaka CF, Henriques AC, de Matos FR, de Souza LB, Pinto LP. Nonodontogenic cysts of the oral and maxillofacial region: demographic profile in a Brazilian population over a 40-year period, Eur Arch Otorhinolaryngol., (268) 2011, No. 6, doi: 10.1007/s00405-010-1458-x.

[6] Weber M, Ries J, Büttner-Herold M, Geppert CI, Kesting M, Wehrhan F., Differences in Inflammation and Bone Resorption between Apical Granulomas, Radicular Cysts, and Dentigerous Cysts., J Endod. (45) 2019, No. 10, 1200–1208, doi: 10.1016/j.joen.2019.06.014

[7] Malik, N., Cysts of the “Oro-Maxillofacial Region”. In: Bonanthaya, K., Panneerselvam, E., Manuel, S., Kumar, V.V., Rai, A. (eds) Oral and Maxillofacial Surgery for the Clinician. Springer, Singapore, 2021, doi: 10.1007/978-981-15-1346-6_27

[8] Kouhsoltani M, Mesgarzadeh AH, Moradzadeh Khiavi M., Mandibular Fracture Associated with a Dentigerous Cyst: Report of a Case and Literature Review. J Dent Res Dent Clin Dent Prospects. 9 (2015), No. 3, PMID: 26697153; PMCID: PMC4682017

[9] Kim GM, Oh JS, You JS, Moon SY, Choi HI., Deep Neck Infection Caused by Infected Dentigerous Cyst: A Case Report. J Oral Med Pain, 46 (2021), No. 4, 150–154

[10] Frank E, Macias D, Hondorp B, Kerstetter J; Inman JC. Incidental Squamous Cell Carcinoma in an Epidermal Inclusion Cyst: A Case Report and Review of the Literature, Case Rep Dermatol., 10 (2018), vol.1, 61–68, doi: 10.1159/000487794

[11] Wang L, Xu Y, Wang W, Lu Y., Application of machine learning in dentistry: insights, prospects and challenges, Acta Odontologica Scandinavica, 85 (2025), 145–154, doi: 10.2340/aos.v84.43345

[12] Yang, H.; Jo, E.; Kim, H.J.; Cha, I.-h.; Jung, Y.-S.; Nam, W.; Kim, J.-Y.; Kim, J.-K.; Kim, Y.H.; Oh, T.G.; et al., Deep Learning for Automated Detection of Cyst and Tumors of the Jaw in Panoramic Radiographs., J. Clin. Med., 9 (2020), No. 6, doi:10.3390/jcm9061839

[13] Kwon, O., Yong, T. H., et.al., Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network, Dento maxillo facial radiology, 49 (2020), No. 8, doi: 10.1259/dmfr.20200185

[14] Tajima, S., Okamoto, Y.,Et al., Development of an automatic detection model using artificial intelligence for the detection of cyst-like radiolucent lesions of the jaws on panoramic radiographs with small training datasets., Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology, 34 (2022), No. 5, 553-560, doi: 10.1016/j.ajoms.2022.02.004.

[15] Yu, D., Hu, J., Feng, Z., Song, M., & Zhu, H., Deep learning based diagnosis for cysts and tumors of jaw with massive healthy samples., Scientific reports, 1855 (2022), No. 12, doi: 0.1038/s41598-022-05913-5

[16] Huang, Z., Li, B., Cheng, Y., Kim, J., Odontogenic cystic lesion segmentation on cone-beam CT using an auto-adapting multi-scaled UNet., Frontiers in oncology, 14 (2024), doi: 10.3389/fonc.2024.1379624

[17] Lee, J. H., Kim, D. H., Jeong, S. N., Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. Oral diseases, 26 (2020), No. 1, 152–158, doi: 10.1111/odi.13223

[18] Krishnamoorthy V., Logeswari S., Sivanantham S., Akshaya V., Ganagavalli K., Mohanraj V., Glaucoma detection with the segmentation of optic disc and cup region using improved u-net and classification algorithm, Electrical Review, (2025), No. 5, doi: 10.15199/48.2025.05.15

[19] Sathiyamurthy B. K., Madhaiyan V. K., Automated lung cancer T-Stage detection and classification using improved U-Net model, International Journal of Electrical and Computer Engineering (IJECE), 14 (2024), No. 6, 6817–6826, doi: 10.11591/ijece.v14i6.pp6817-6826

[20] Gautam P., Singh M., A multidimensional U-net architecture for Alzheimer's disease classification using T1 weighted brain structural MRI, Computers and Electrical Engineering, 128 (2025), doi: 10.1016/j.compeleceng.2025.110687

[21] Tran D.K., Nguyen V.N. Et al., WOODWORK: A deep-learning based framework for woodpecker damage detection in powerline inspection, International Journal of Electrical Power & Energy Systems, 171 (2025), doi: 10.1016/j.ijepes.2025.110900

[22] Y. -J. Ng, Y. -Y. Tee, D. Cheng and B. -H. Gwee, SAMIC: Segment Anything Model for Integrated Circuit Image Analysis, TENCON 2024, Singapore, 2024, pp. 418–421, doi: 10.1109/TENCON61640.2024.10902875

[23] Mazurowski M. A., Dong H., Gu H., Yang J., Konz N., Zhang Y., Segment anything model for medical image analysis: An experimental study, Medical Image Analysis, 89 (2023), doi: 10.1016/j.media.2023.102918.

[24] L. Zhang, X. Deng, Y. Lu, Segment Anything Model (SAM) for Medical Image Segmentation: A Preliminary Review, 2023 IEEE Int Conf Bioinformatics Biomed (BIBM), Istanbul, Turkiye, (2023), 4187–4194, doi: 10.1109/BIBM58861.2023.10386032.

[25] Ma, J., He, Y., Li, F. et al. Segment anything in medical images, Nat Commun, 654 (2024), No. 15, doi: 10.1038/s41467-024-44824-z

[26] Ma, J. et al., MedSAM2: Segment Anything in 3D Medical Images and Videos, arXiv preprint, arXiv:2504.03600 (2025)

[27] Kirillov A. et.al, Segment Anything, arXiv preprint, doi: 10.48550/arXiv.2304.02643 (2023)

[28] Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage, 31 (2006), No. 3, doi:10.1016/j.neuroimage.2006.01.015.