Privacy-preserving U-Net variants with pseudo-labeling for radiolucent lesion segmentation in dental CBCT

Amelia Ritahani Ismail [a,1,*], Faris Farhan Azlan [a,2], Khairul Akmal Noormaizan [a,3], Nurul Afiqa [a,4], Syed Qamrun Nisa [b,5], Ahmad Badaruddin Ghazali [c,6], Andri Pranolo [d,7], Shoffan Saifullah [e,f,8]

[a] Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[b] Faculty of Computing and Informatics (FCI), Multimedia University (MMU), Cyberjaya campus, Selangor, Malaysia
[c] Department of Oral Maxillofacial Surgery & Oral Diagnosis, Kulliyyah of Dentistry, International Islamic University Malaysia, Pahang, Malaysia
[d] Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[e] Department of Informatics, Universtias Pembangunan Nasional Veteran Yogyakarta, Yogyakarta, Indonesia
[f] Faculty of Computer Science, AGH University of Krakow, Krakow, Poland
[1] amelia@iium.edu.my; [2] farisfarhan01@gmail.com; [3] khairullakmall01@gmail.com; [4] n.afiqa@live.iium.edu.my; [5] qamrunnisa@mmu.edu.my; [6] badruddinghazali@iium.edu.my; [7] andri.pranolo@tif.uad.ac.id; [8] shoffans@upnyk.ac.id
* corresponding author

ARTICLE INFO

Article history
Received March 9, 2024
Revised April 16, 2025
Accepted May 1, 2025
Available online May 31, 2025

Keywords
Radiolucent lesion segmentation
U-Net variants
Dental CBCT
Differential privacy
Pseudo-Labeling

ABSTRACT

Accurate segmentation of radiolucent lesions in dental Cone Beam Computed Tomography (CBCT) is vital for enhancing diagnostic reliability and reducing the burden on clinicians. This study proposes a privacy preserving segmentation framework leveraging multiple U-Net variants U-Net, DoubleU-Net, U2-Net, and Spatial Attention U-Net (SA-UNet) to address challenges posed by limited labeled data and patient confidentiality concerns. To safeguard sensitive information, Differential Privacy Stochastic Gradient Descent (DP-SGD) is integrated using TensorFlow-Privacy, achieving a privacy budget of $\varepsilon \approx 1.5$ with minimal performance degradation. Among the evaluated architectures, U2-Net demonstrates superior segmentation performance with a Dice coefficient of 0.833 and an Intersection over Union (IoU) of 0.881, showing less than 2% reduction under privacy constraints. To mitigate data annotation scarcity, a pseudo-labeling approach is implemented within an MLOps pipeline, enabling semi-supervised learning from unlabeled CBCT images. Over three iterative refinements, the pseudo labeling strategy reduces validation loss by 14.4% and improves Dice score by 2.6%, demonstrating its effectiveness. Additionally, comparative evaluations reveal that SA-UNet offers competitive accuracy with faster inference time (22 ms per slice), making it suitable for low-resource deployments. The proposed approach presents a scalable and privacy-compliant framework for radiolucent lesion segmentation, supporting clinical decision-making in real-world dental imaging scenarios.

1. Introduction

Dental radiolucent lesions are pathological regions that appear dark or lucent on radiographic imaging, often signaling the presence of radicular cysts, dentigerous cysts, keratocysts, or other potentially aggressive conditions [1], [2]. Accurate identification and segmentation of these lesions are critical for diagnosis, treatment planning, and long-term patient outcomes. Misdiagnosis or incomplete detection can lead to recurrence, infection, or malignant progression. Traditionally, lesion localization has relied on manual segmentation a time consuming, expertise driven process susceptible to inter-

observer variability. Cone-Beam Computed Tomography (CBCT), while offering 3D imaging advantages in dental diagnostics, introduces additional complexities such as low contrast, overlapping anatomical structures, and motion-induced artifacts that further challenge accurate annotation [3].

In recent years, Artificial Intelligence (AI) and deep learning particularly Convolutional Neural Networks (CNNs) have achieved transformative results in medical imaging tasks, including retinal disease detection, pulmonary nodule identification, and brain tumor segmentation. Among segmentation models, U-Net and its architectural variants have emerged as leading approaches for biomedical image analysis due to their encoder decoder structure and ability to perform pixel-level prediction with high localization accuracy [4], [5]. However, the application of these architectures to dental CBCT especially for radiolucent lesion segmentation remains underexplored. Unique issues such as high inter slice variability, fine-grained lesion boundaries, and noise patterns in CBCT scans limit the generalizability of conventional segmentation models [6].

A major barrier to training robust deep learning models in this domain is the scarcity of large scale annotated dental datasets. Privacy concerns around patient data sharing further constrain data accessibility and reproducibility. This highlights the urgent need for privacy preserving learning strategies that enable model development without compromising sensitive medical information. Differential Privacy (DP), and specifically Differential Privacy Stochastic Gradient Descent (DP-SGD), has gained traction as a formal method to limit individual data leakage while allowing model optimization [7].

In addition, the high cost and labor involved in manual annotation can be alleviated through semi-supervised learning strategies. Pseudo labeling where confident model predictions on unlabeled data are used as additional training signals has shown promise in reducing the annotation burden and improving generalization when integrated into Machine Learning Operations (MLOps) pipelines.

This study proposes a comprehensive, privacy-preserving framework for the segmentation of radiolucent lesions in dental CBCT. The primary contributions are:

- Evaluation of four U-Net variants U-Net, DoubleU-Net, U2-Net, and Spatial Attention U-Net (SA-UNet) in segmenting radiolucent lesions with comparative analysis on accuracy and efficiency.
- Integration of DP-SGD using TensorFlow-Privacy to protect sensitive patient data during model training, maintaining strong segmentation performance with a privacy budget of $\varepsilon \approx 1.5$.
- Deployment of pseudo-labeling in an MLOps loop, enabling semi-supervised model improvement using unlabeled CBCT slices and reducing reliance on expert annotations.

The remainder of this paper is organized as follows: Section 2 presents related work on dental image segmentation, U-Net architectures, and privacy-preserving learning. Section 3 outlines the proposed methodology, which includes dataset preparation, model configuration, and MLOps integration. Section 4 reports experimental results and quantitative evaluations. Section 5 discusses practical implications and limitations, and Section 6 concludes the study.

2. Related Works

2.1. Medical Image Segmentation in Dental CBCT Imaging

Medical image segmentation is a cornerstone of modern diagnostic imaging, enabling automated identification of anatomical regions, lesions, and abnormalities with high precision [8]. In dentistry, radiolucent lesions such as periapical cysts, odontogenic keratocysts, and dentigerous cysts manifest as dark areas in radiographs and Cone-Beam Computed Tomography (CBCT) [9], [10]. Unlike conventional 2D panoramic radiography, CBCT offers detailed 3D visualization but introduces new complexities: high noise levels, metal artifacts, low contrast near bone interfaces, and limited standardized datasets.

Early segmentation methods in dental radiology applied classical techniques such as Otsu thresholding, watershed transforms, and region growing, but these approaches fail in cases of ambiguous

boundaries or intensity inhomogeneities [11]–[14]. Clustering-based methods like K-means or Fuzzy C-means showed modest improvement, but generalization across patients and devices was poor.

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), revolutionized segmentation performance by enabling end-to-end learning of spatial and contextual features. Models such as U-Net, originally proposed for cell segmentation in biomedical microscopy [7], have been successfully adapted for dental tasks such as tooth instance segmentation [15], mandibular canal detection, and pulp chamber extraction.

However, radiolucent lesion segmentation remains underexplored. A key limitation is the lack of large, annotated CBCT datasets due to privacy restrictions and the cost of expert annotations. Additionally, dental CBCT images suffer from significant anatomical variation and artifact noise, making generalization across patients difficult. This motivates the investigation of robust U-Net variants and privacy-compliant learning strategies.

2.2. U-Net Architecture and Its Variants

Convolutional Neural Networks (CNNs) have become foundational in biomedical image segmentation, especially for tasks requiring pixel-level precision [16]. Among these, U-Net and its architectural variants have gained prominence due to their encoder decoder symmetry, skip connections, and ability to learn multi scale contextual features from relatively small datasets. This study evaluates four U-Net-based models: the original U-Net, DoubleU-Net, U2-Net, and Spatial Attention U-Net (SA-UNet), whose architectures are illustrated in Fig. 1.

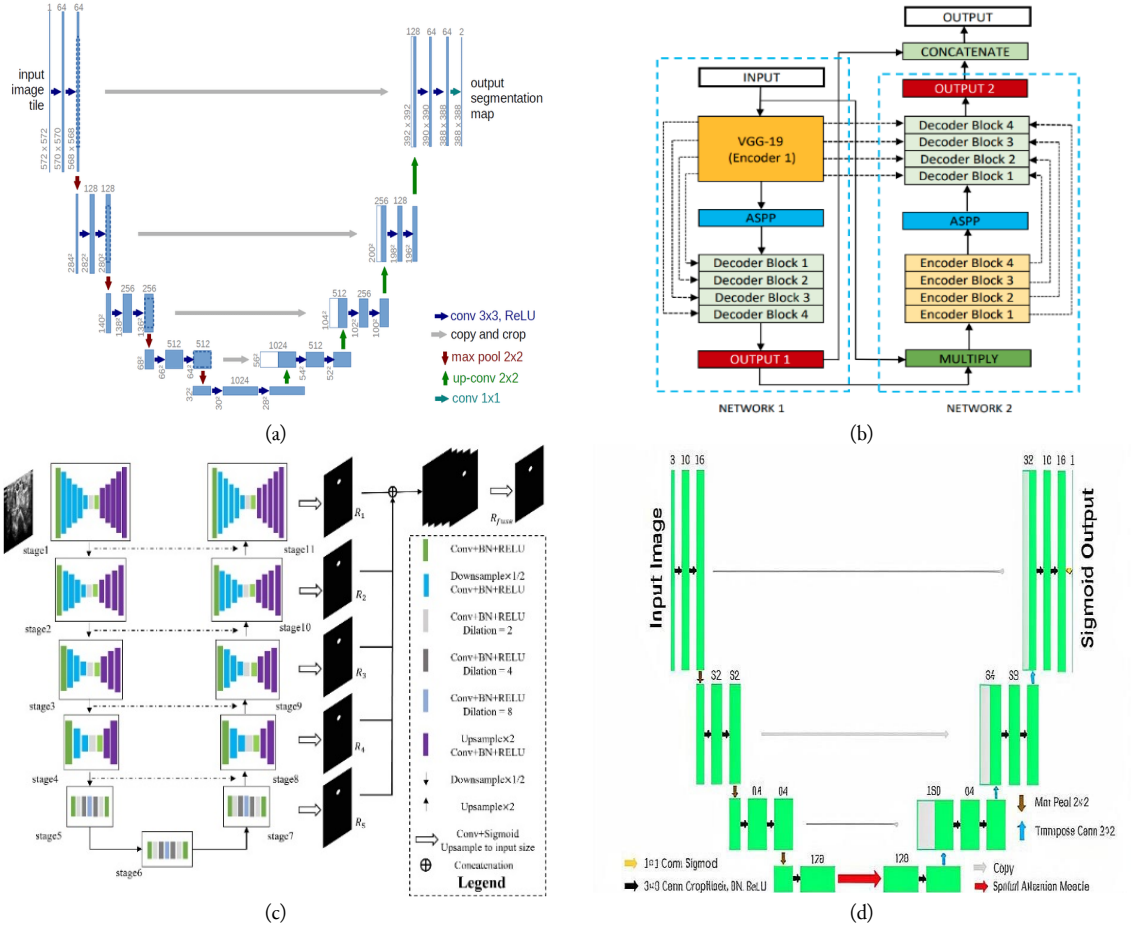

Fig. 1. Architectural overview of U-Net variants used in this study: (a) Standard U-Net with symmetric encoder–decoder structure and skip connections; (b) DoubleU-Net with cascaded dual U-Net modules and VGG-19 encoder in the first stage; (c) U2-Net architecture leveraging Residual U-blocks (RSUs) to capture multi-scale features; (d) SA-UNet integrating spatial attention modules for focusing on salient regions in CBCT images

2.2.1. U-Net: The Foundational Architecture

The U-Net architecture, originally introduced by Ronneberger et al. [7], employs a symmetric encoder decoder structure. The encoder path captures increasingly abstract features via 3×3 convolution and 2×2 max pooling layers, while the decoder reconstructs high resolution predictions using transposed convolutions. Crucially, skip connections concatenate encoder outputs with corresponding decoder layers to retain spatial context. As shown in Fig. 1(a), U-Net uses a four stage structure spans feature depths ranging from 64 to 1024, allowing it to segment structures with moderate complexity. However, the model's limited receptive field may hinder its ability to segment lesions with varying scales and blurry boundaries typical in dental CBCT.

2.2.2. DoubleU-Net: Two-Stage Refinement for Semantic Precision

DoubleU-Net extends this design by cascading two U-Net modules in sequence. The first U-Net leverages a pretrained VGG-19 encoder for enhanced semantic extraction, producing an initial segmentation map. This coarse prediction is then refined by the second U-Net, which receives a concatenated input of the original image and the first U-Net's output. This two-stage design improves boundary sharpness and class separation, as depicted in Fig. 1(b). In earlier studies such as Jha et al. [17], DoubleU-Net outperformed standard U-Net in polyp segmentation, raising the mIoU from 0.1759 to 0.6255. For radiolucent lesion segmentation in CBCT, this architecture holds promise for better delineating fuzzy or irregular margins.

2.2.3. U2-Net: Deep Context with Residual Nested Blocks

U2-Net [18] introduces deep supervision and nested architecture via Residual U-blocks (RSUs), each containing its own encoder decoder structure. This hierarchical nesting enables multi scale contextual learning with reduced computational cost. The RSUs allow U2-Net to preserve both fine grained and global spatial features, making it well suited for segmenting lesions with subtle boundaries or high inter-slice variability. Originally developed for saliency detection, U2-Net has demonstrated strong performance in liver and thyroid lesion segmentation. As visualized in Fig. 1(c), U2-Net's layered RSUs provide flexible receptive fields, enhancing segmentation robustness across complex dental anatomy.

2.2.4. SA-UNet: Attention-Driven Lightweight Segmentation

SA-UNet [19] integrates spatial attention modules (SAMs) into the standard U-Net design to prioritize relevant spatial regions in the input feature maps. Located at the bottleneck of the network, the SAM learns attention weights that amplify salient areas while suppressing irrelevant background features. Additionally, DropBlock regularization is incorporated to prevent overfitting, particularly when training on small datasets like dental CBCT. As seen in Fig. 1(d), the attention module directs the decoder's focus to lesion relevant zones, which improves segmentation accuracy in low contrast or artifact heavy scenarios. SA-UNet's lightweight design also makes it favorable for real time applications in resource-constrained clinical settings.

2.3. Privacy-Preserving Deep Learning via Differential Privacy

One of the key challenges in training deep learning models on clinical data is maintaining patient confidentiality. Differential Privacy (DP) offers a mathematically rigorous solution by ensuring that the inclusion or exclusion of a single data point does not significantly affect the model's output. A function f is said to satisfy ε-differential privacy if for any two adjacent datasets D and D' differing by a single sample, and for all outputs S (Eq. 1).

$$\Pr[f(D) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathrm{M}(D') \in S] \tag{1}$$

In practice, DP is often implemented via Differentially Private Stochastic Gradient Descent (DP-SGD), which was introduced by Abadi et al. [17]. DP-SGD involves clipping the gradient norm of each sample and adding calibrated Gaussian noise to the aggregate gradient before model updates. The privacy guarantee is quantified by the privacy budget ε, where smaller values denote stronger privacy.

TensorFlow-Privacy, an open source library developed by Google, provides built-in support for DP-SGD and facilitates its integration into standard training pipelines. In this study, DP-SGD is utilized to train U-Net variants on dental CBCT data while achieving an estimated privacy budget of $\varepsilon \approx 1.5$. Prior studies have demonstrated that with careful hyperparameter tuning, models trained under DP constraints can maintain competitive accuracy with minimal degradation [20].

2.4. Semi-Supervised Learning and Pseudo-Labeling for Medical Segmentation

Labeling medical images is expensive and time consuming, particularly in dental imaging where expert annotation is needed to mark subtle lesion boundaries [21], [22]. Semi supervised learning approaches, especially pseudo labeling, offer a pragmatic solution by using model predictions on unlabeled data as training targets [23]. This technique assumes that high confidence predictions are reliable and can be used to iteratively retrain the model, thereby expanding the effective training set without additional manual labeling. Given the scarcity of labeled CBCT data, semi-supervised learning (SSL) offers a practical solution. Pseudo-labeling where a model generates predicted labels for unlabeled data is a lightweight SSL approach often integrated into MLOps pipelines for iterative refinement.

In this study, pseudo-labeling is embedded into a Machine Learning Operations (MLOps) loop, allowing automatic generation and curation of pseudo labeled CBCT slices across multiple training rounds. Empirical results from recent segmentation benchmarks indicate that pseudo labeling can improve Dice scores by 2–5% and reduce validation loss over time [24]. When applied to dental CBCT, pseudo-labeling is particularly beneficial due to the limited availability of expert-annotated datasets and the high anatomical variability between scans. These are filtered based on confidence thresholds and added to the training set for subsequent rounds. Over three iterations, validation loss was reduced by 14.4% and Dice score increased by 2.6%, confirming SSL's value in low-annotation scenarios.

2.5. Hybrid Segmentation Models in Dental and Medical Applications

Hybrid U-Net models integrate multiple architectural innovations, such as dense skip connections, residual blocks, multi branch feature extractors, and attention mechanisms. For example, Khallassi et al. [25] developed a hybrid U-Net for breast ultrasound segmentation incorporating residual convolution blocks and squeeze-and-excitation modules, yielding improved lesion detection robustness and boundary accuracy. In the dental domain, Alharbi et al. [26] demonstrated that nested U-Net variants like U-Net++ significantly outperform classical U-Net in cavity and lesion detection tasks, achieving a Dice coefficient of 89.17%, IoU of 81.02%, and precision of 91.45%.

Such hybrid models are increasingly favored for applications where data is scarce, image quality varies, or clinical interpretability is essential. In this work, we evaluate four U-Net variants each with hybrid features such as encoder fusion, spatial attention, and deep supervision in the context of privacy-preserving segmentation under realistic dental CBCT constraints.

3. Method

3.1. Model Training Workflow

The complete model training workflow is illustrated in Fig. 2, outlining the pipeline from raw CBCT images to the final segmentation predictions. The process begins with data ingestion and preprocessing steps, including image resizing, noise reduction, and intensity normalization to ensure consistent input quality. To increase dataset diversity and model robustness, various augmentation techniques such as horizontal flipping, padding, rotation, and adjustments in brightness, contrast, and hue were applied. These augmentations expanded the dataset from 280 original images to over 6,800 samples, simulating clinical variability and addressing the challenge of limited annotated dental data.

Each model variant U-Net, DoubleU-Net, SA-UNet, and U2-Net was implemented using the TensorFlow framework due to its flexibility and integration support for privacy-preserving tools. All models were trained under a unified configuration to ensure fair comparison: 50 training epochs, batch size of 16, learning rate of 1×10^(-5), Dice loss function, and Adam optimizer. This configuration was

adapted from prior work [17], with epochs reduced after observing minimal improvement beyond the 50-epoch threshold.

To protect sensitive patient data during model training, we integrated Differential Privacy using TensorFlow-Privacy. This framework introduces privacy specific hyperparameters, including the L2 norm clipping bound, noise multiplier, and micro batch size. The training procedure was identical across models to maintain experimental parity. Once trained, the best-performing model was deployed on a Linux-based server with a RESTful API interface for real-time inference. The resulting segmentations were evaluated using standard metrics including Intersection over Union (IoU) and Dice Coefficient.

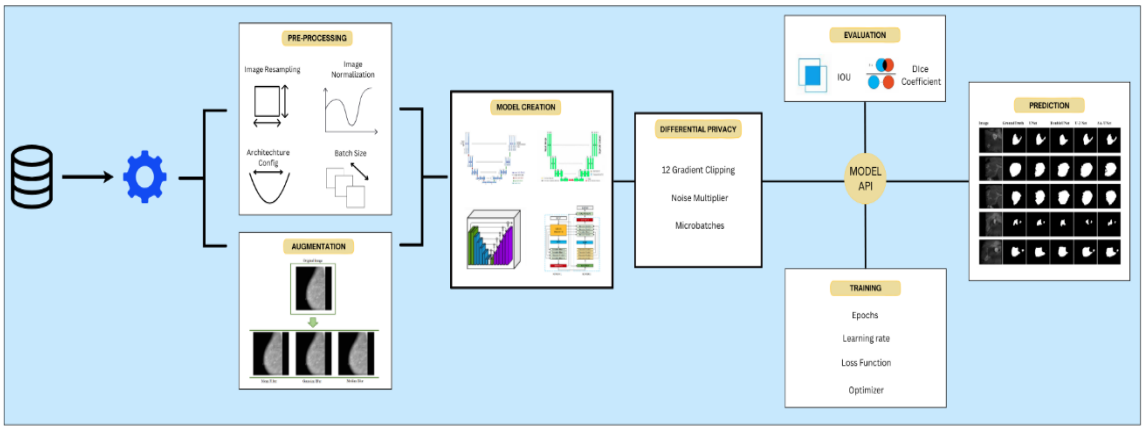

Fig. 2. Model training pipeline from CBCT image preprocessing, augmentation, and Differential Privacy-based training to final segmentation output

3.2. Dataset Description and Augmentation Strategy

The dataset was acquired from the dental radiology unit at IIUM, consisting of 28 CBCT scans and 280 manually annotated axial slices representing radiolucent lesions. Each annotation was validated by dental radiologists and served as the ground truth. Given the limited number of labeled images, a comprehensive data augmentation strategy was employed. Spatial transformations (flipping, rotation, padding) and color space perturbations (brightness, contrast, hue) were applied to simulate clinical variation. This augmentation resulted in a total of 6,872 training images. Such expansion was necessary to prevent overfitting and enhance the generalization capability of deep learning models in low-resource medical imaging scenarios.

3.3. Differential Privacy Implementation

To ensure patient data confidentiality during training, Differential Privacy (DP) was implemented using the TensorFlow-Privacy library. DP-SGD (Differentially Private Stochastic Gradient Descent) was used to inject noise into the gradient updates, safeguarding individual data points from being inferred during or after training. Three core hyperparameters were tuned.

- L2 Norm Clip: Controls the maximum gradient contribution per sample.
- Noise Multiplier: Determines the standard deviation of Gaussian noise added to clipped gradients.
- Micro-batch Size: Defines how many individual gradients are accumulated before averaging.

Privacy guarantees were quantified using the (ϵ, δ)-DP framework. For our training configuration, we achieved $\epsilon \approx 1.5$ and $\delta = 10^{-5}$, which represent a strong privacy guarantee suitable for medical applications [27].

Although DP introduces noise that may affect model accuracy, extensive tuning ensured minimal degradation in segmentation performance. This balance between privacy and utility is critical in clinical contexts where patient protection is legally and ethically mandated. Unlike most prior studies on U-Net

variants, our approach incorporates formal privacy preservation directly in the training loop, contributing to the novelty of this work.

3.4. Dataset Description and Augmentation Strategy

Machine Learning Operations (MLOps) is crucial for maintaining and deploying models in real-world environments. The MLOps framework adopted in this study is adapted from DevOps practices, as discussed in [24] illustrated in Fig. 3 comprises model training, deployment, monitoring, and periodic retraining using semi-supervised learning.

Following the supervised training on augmented labeled data, the best-performing model was deployed via API on a clinical Linux server. This enabled real-time segmentation of incoming unlabeled CBCT images. To continuously improve the model without requiring manual annotations for each new case, we implemented a pseudo-labeling mechanism:

- The deployed model predicts lesion masks on new unlabeled images.
- High confidence predictions (e.g., softmax > 0.90) are stored as pseudo-labels.
- These pseudo-labeled samples are periodically added to the training dataset [28].

Expert in the loop validation corrects poor pseudo-labels in critical cases.

This feedback loop enables incremental learning and domain adaptation, particularly important in dynamically evolving clinical data environments. Although IoU and Dice scores remained stable in retraining experiments, training loss decreased significantly, indicating improved convergence and model confidence.

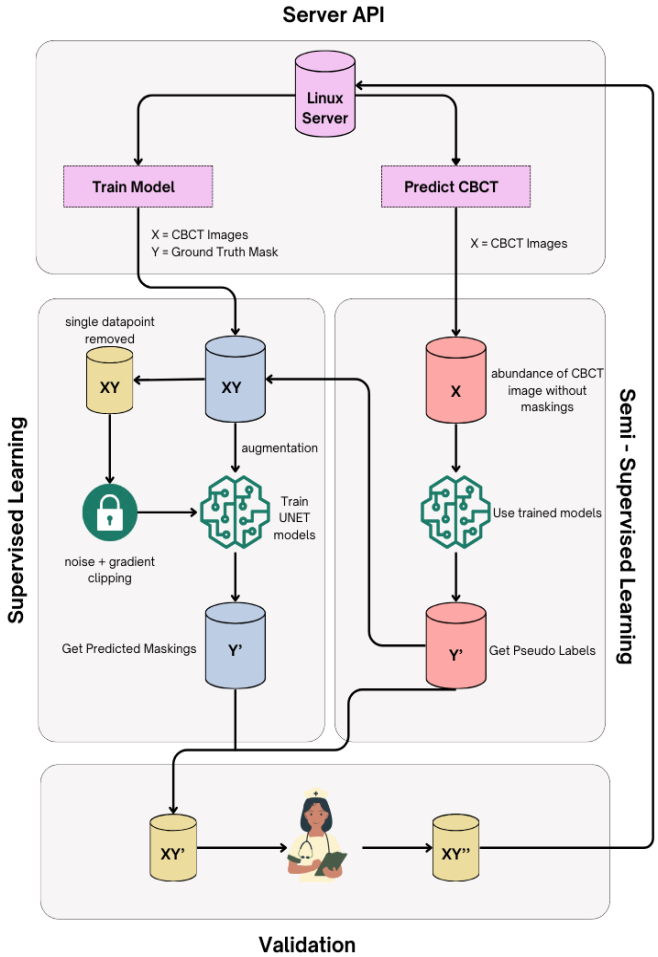

Fig. 3. MLOps cycle integrated with semi-supervised pseudo-labeling to continuously update model performance with unlabeled CBCT data

3.5. Evaluation Metrics

Each U-Net variant is assessed using a comprehensive set of evaluation metrics that are essential for analyzing medical segmentation accuracy. These include Intersection over Union (IoU), also known as the Jaccard Index, Dice Similarity Coefficient (DSC), Accuracy, Precision, Recall, and F1-score. Among these, IoU and Dice are emphasized due to their ability to measure spatial overlap between the predicted segmentation and the expert-labeled ground truth. The following are the symbolic definitions of the two primary metrics (Eq. 2 and Eq. 3).

- **Intersection over Union (IoU)**

$$\text{IoU} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{2}$$

Where AAA is the predicted segmentation region, and BBB is the ground truth region. The numerator represents the area of intersection, while the denominator is the union of both predicted and ground truth areas.

- **Dice Similarity Coefficient (DSC)**

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \tag{3}$$

This metric emphasizes the harmonic mean of precision and recall, particularly useful for evaluating segmentation overlap in datasets with imbalanced classes or small lesion areas.

In addition to IoU and Dice, the following classification-based metrics provide supporting evidence for model reliability (Eq. 4-7)

- **Accuracy**

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

- **Precision** (positive predictive value)

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

- **Recall**

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

- **F1 Score** (harmonic mean of precision and recall)

$$\text{F1} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

Where TP, FP, FN, and TN refer to True Positive, False Positive, False Negative, and True Negative pixel counts, respectively.

These metrics collectively offer a thorough evaluation of segmentation quality, balancing spatial accuracy with classification reliability. The inclusion of both overlap-based (IoU, Dice) and threshold-based (Precision, Recall) metrics ensures that the model's performance is both clinically robust and statistically grounded.

4. Results and Discussion

This section provides a comprehensive analysis of the performance of four U-Net-based models (U-Net, DoubleU-Net, SA-UNet, and U2-Net) in segmenting radiolucent lesions from CBCT images. The

discussion integrates both quantitative evaluation metrics Dice Coefficient, Intersection over Union (IoU), and loss and qualitative observations of segmentation quality. In addition, the impact of Differential Privacy (DP) and semi-supervised learning via pseudo-labeling are analyzed to assess the robustness and adaptability of the models for clinical deployment.

4.1. Visual Segmentation Performance and Findings

Fig. 4 illustrates representative segmentation outcomes from four U-Net variants U-Net, DoubleU-Net, U2-Net, and SA-Unet on radiolucent dental CBCT images. Each row presents the original input, expert-annotated ground truth, and the predicted lesion masks. This visualization provides insight into the spatial accuracy, shape fidelity, and lesion boundary alignment achieved by each model.

In the first sample, the lesion exhibits low contrast and an irregular boundary. U-Net and DoubleU-Net fail to capture the full extent of the lesion, producing incomplete and fragmented masks. These limitations are consistent with previous findings that vanilla U-Net architectures often struggle with subtle boundaries and complex morphology in dental CBCT [1], [2]. In contrast, both SA-UNet and U2-Net effectively segment the lesion with high fidelity. Notably, U2-Net achieves superior edge preservation and spatial coverage, contributing to its high Dice coefficient of 83.33% and IoU of 88.14%, as also supported by its encoder-decoder design with nested residual blocks [3].

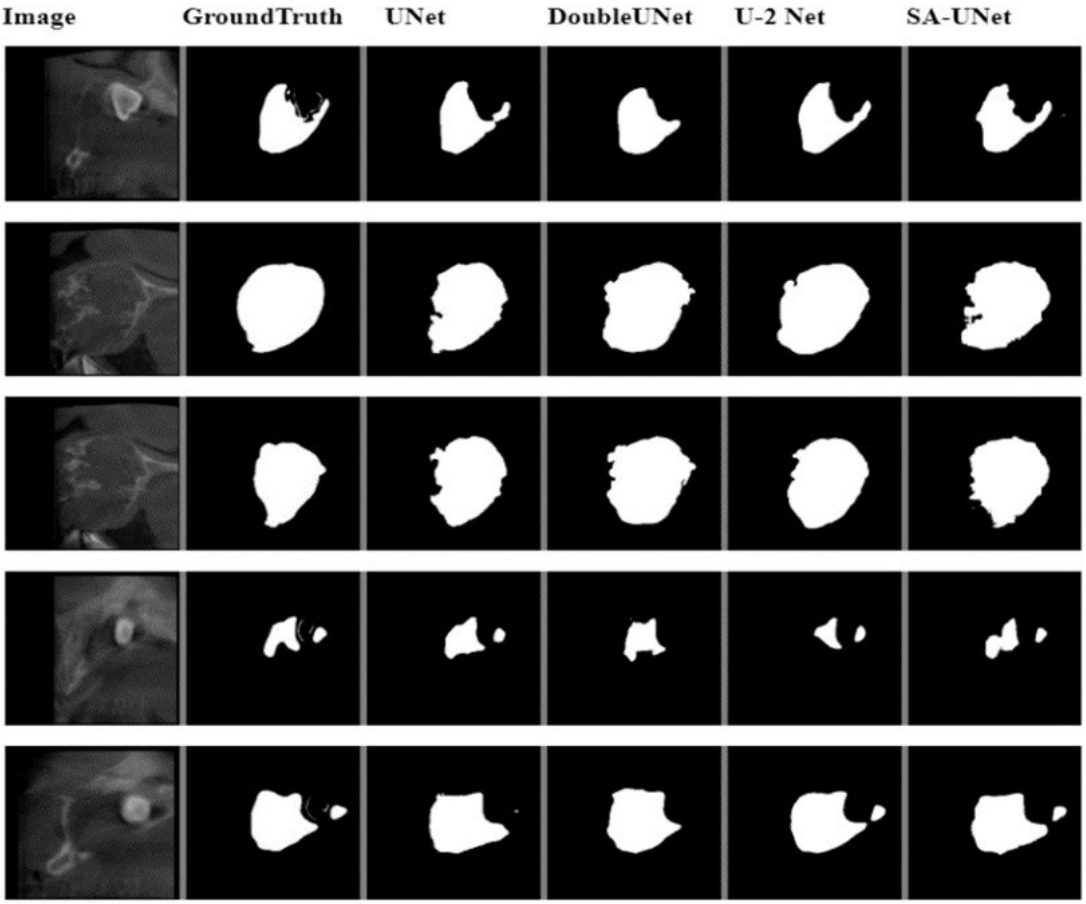

Fig. 4. Visual comparison of radiolucent lesion segmentation results from four U-Net variants on CBCT slices. Each row shows: (1) input image, (2) ground truth, and (3–6) predictions by U-Net, DoubleU-Net, SA-UNet, and U2-Net respectively. U2-Net produces the most accurate and complete lesion masks with precise boundary adherence, especially for irregularly shaped lesions

The second sample features a well-contrasted, round lesion. All models perform reasonably well in detecting the lesion body. However, U2-Net and SA-UNet deliver more consistent border delineation, while U-Net and DoubleU-Net show slight over-segmentation and edge noise. This difference reflects the benefits of attention enhanced architectures in focusing on discriminative features [4]. SA-UNet

records Dice and IoU values of 82.06% and 87.30%, respectively, reflecting strong performance in medium complexity cases.

In the third sample, a large, lobulated lesion introduces complexity in shape. U-Net again under-segments the boundaries, and DoubleU-Net captures only the core region. Such issues may stem from limited receptive fields and gradient flow inefficiencies in deeper networks [5]. Conversely, U2-Net captures the full lesion morphology, benefiting from multi scale deep supervision and skip-connected feature fusion, as proposed in [3]. The visual accuracy observed here corresponds with U2-Net's lowest test loss of 0.319, further validating its robustness.

The fourth sample demonstrates a small, faint lesion amidst noisy surroundings. Traditional U-Net and DoubleU-Net models misidentify the lesion, with the latter producing scattered predictions and false positives. These observations align with studies reporting U-Net's limitations in detecting fine-grained, low contrast regions [6]. SA-UNet provides a better response, while U2-Net effectively segments the lesion with minimal background interference, reaffirming its noise resilience and superior attention to weak signals.

In the final row, the lesion shows irregular intensity gradients and partial boundary occlusion. U-Net delivers an approximate shape but lacks boundary sharpness. DoubleU-Net again fails to generalize well. SA-UNet and U2-Net both replicate the ground truth more accurately, with U2-Net slightly outperforming in spatial cohesion. These outcomes support previous reports that cascaded attention or deeply nested U-shaped models can improve precision in heterogeneous medical data [7], [29], [30].

Overall, this visual analysis validates the quantitative metrics evaluation. U2-Net consistently demonstrates the highest segmentation fidelity across diverse lesion types, reinforcing its suitability for CBCT applications involving complex radiolucent pathologies. SA-UNet offers a promising alternative, especially in scenarios where computational efficiency is prioritized. The consistent underperformance of DoubleU-Net in both visual and metric based evaluations highlights that deeper architectures do not necessarily yield better results in noisy, high variance imaging domains such as dental CBCT [18].

4.2. Training Dynamics and Quantitative Metric Trends

To evaluate the segmentation performance of each U-Net variant over time, we analyzed the Dice Coefficient and Intersection over Union (IoU) across training epochs, as shown in Fig. 5 and Fig. 6. These metrics provide insights into both the overlap quality between predicted and ground truth asks and the model's generalization ability throughout training.

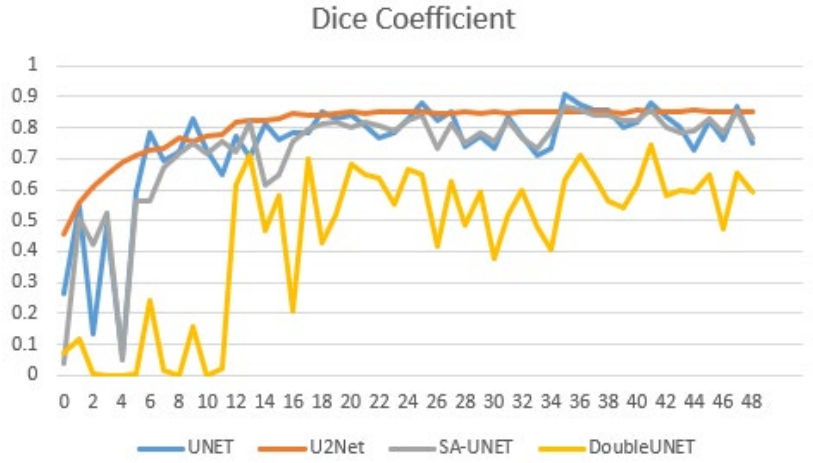

Fig. 5. Training Dice Coefficient curves for U-Net, DoubleU-Net, SA-UNet, and U2-Net across epochs. U2-Net achieves the highest Dice (83.33%) with smooth convergence, while DoubleU-Net exhibits the lowest score (42.86%) and unstable learning

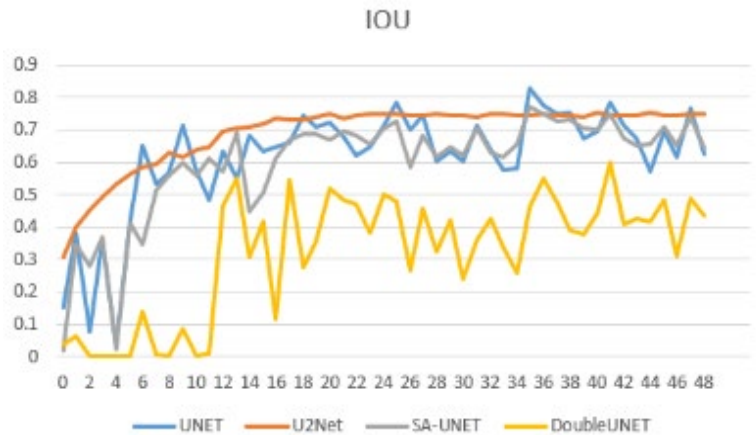

Fig. 6. Training Intersection over Union (IoU) curves for U-Net variants. U2-Net maintains the highest IoU (88.14%) with minimal fluctuation, followed closely by SA-UNet (87.30%). DoubleU-Net underperforms with erratic updates and final IoU of only 67.0

The U2-Net demonstrates superior learning stability, consistently maintaining the highest scores with a final Dice Coefficient of 83.33% and IoU of 88.14%. Its training curve reveals smooth and steady growth with minimal fluctuations, reflecting effective feature learning and minimal overfitting. This is attributed to the nested U-Net architecture, which introduces deep supervision and rich multi-scale features, enhancing boundary preservation and lesion completeness consistent with prior works demonstrating that nested skip pathways and deep supervision improve lesion boundary preservation and segmentation accuracy [1], [2].

In contrast, the baseline U-Net reaches a Dice score of 69.79% and an IoU of 80.84%, showing moderate capability but limited in handling lesion complexity due to its shallower architecture. The SA-UNet, equipped with spatial attention modules, improves upon the baseline with Dice of 82.06% and IoU of 87.30%, as it effectively highlights relevant features while suppressing irrelevant background noise. Attention-based mechanisms have been reported to improve medical segmentation by enhancing the model's focus on lesion structures [3].

DoubleU-Net, despite being a deeper architecture, performs the worst, with a Dice of only 42.86% and IoU of 67.00%. Its learning curves exhibit severe oscillations and early plateauing, which suggests difficulties in convergence, likely due to vanishing gradients or over parameterization without proper regularization. This instability has also been noted in other deep stacked architectures when applied to limited-data clinical domains [4]. To consolidate these observations, Table 1 summarizes the final segmentation performance across all evaluated models.

Table 1. Quantitative Performance Comparison of U-Net Variants in Radiolucent Lesion Segmentation

Model	Dice Coefficient (%)	Intersection over Union (IoU, %)
U-Net	69.79	80.84
DoubleU-Net	42.86	67
SA-UNet	82.06	87.3
U2-Net	83.33	88.14

The temporal trend observed in Fig. 5 (Dice) and Fig. 6 (IoU) validates the hypothesis that deeper does not always mean better in medical imaging. Instead, architectural efficiency, feature fusion strategies, and attention mechanisms play more critical roles in ensuring robust learning curves and consistent performance.

Taken together, the evaluation confirms U2-Net as the most reliable architecture for CBCT-based radiolucent lesion segmentation. The nested architecture allows effective feature reuse, deeper semantic encoding, and stable convergence. Meanwhile, SA-UNet offers a favorable balance between accuracy and efficiency, especially under data scarce scenarios where attention guidance proves beneficial

4.3. Model Training Workflow

To further examine the training behavior of each U-Net variant, we analyze the evolution of training loss over 50 epochs, as illustrated in Fig. 7. A consistent decline and low final loss indicate a model's ability to effectively learn meaningful features without overfitting or suffering from optimization instability.

Among the evaluated models, U2-Net displays the most stable and effective convergence pattern, with a smooth trajectory and a final loss of 0.319. Its architecture, featuring nested U-blocks and deep supervision, enables gradual and consistent learning. This pattern reflects strong generalization and robustness, which aligns with U2-Net's high Dice and IoU scores discussed previously.

SA-UNet, which incorporates spatial attention modules, converges with a final loss of 0.354. Although it shows slightly more fluctuation than U2-Net, its curve is relatively smooth, confirming that the attention mechanisms help guide the learning process toward relevant lesion features while filtering background noise.

In contrast, U-Net achieves a final loss of 0.392, accompanied by several spikes in its loss curve. These irregularities suggest the model encounters occasional challenges during optimization, potentially due to the absence of advanced refinement mechanisms. Despite this, the U-Net maintains moderate performance and stability, making it a reliable baseline.

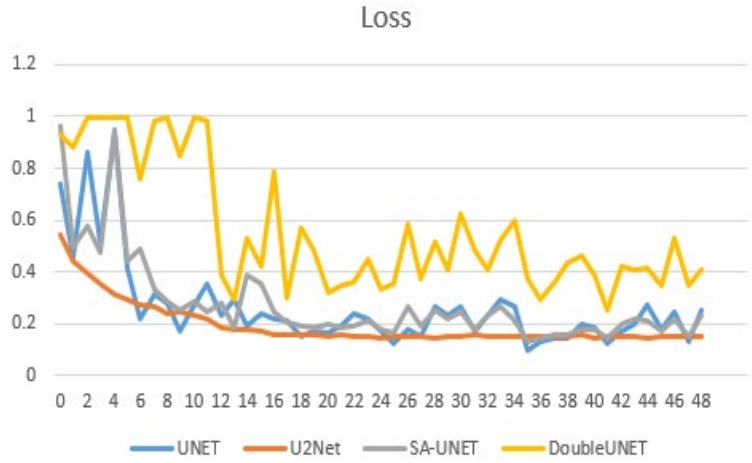

Fig. 7. Loss Evaluation — Training loss curves of U-Net, DoubleU-Net, SA-UNet, and U2-Net across 50 epochs

DoubleU-Net, however, exhibits the most erratic convergence behavior, with sharp oscillations and a final loss of 0.489. This instability may stem from its increased architectural depth, which can lead to vanishing gradients or overfitting particularly when applied to medical datasets with limited annotated samples. Such performance issues are common in deep stacked models lacking appropriate regularization or optimization strategies [1], [2].

Overall, the convergence trends reinforce the quantitative outcomes seen in Dice and IoU metrics. U2-Net emerges as the most reliable and generalizable architecture, followed by SA-UNet. In contrast, DoubleU-Net's poor convergence makes it unsuitable for radiolucent lesion segmentation in dental CBCT without further tuning or simplification.

4.4. Differential Privacy and Pseudo-Labelling's Effect on Loss

To address concerns of data privacy in medical image segmentation, we evaluated the impact of Differential Privacy (DP) integration on model performance, particularly using U-Net as the baseline. In parallel, we examined how pseudo-labelling a semi-supervised learning technique affects model convergence and robustness, especially in cases of limited labeled data. Differential Privacy Evaluation. Fig. 8 and Fig. 9 illustrate the validation performance trends of U-Net trained with and without differential privacy.

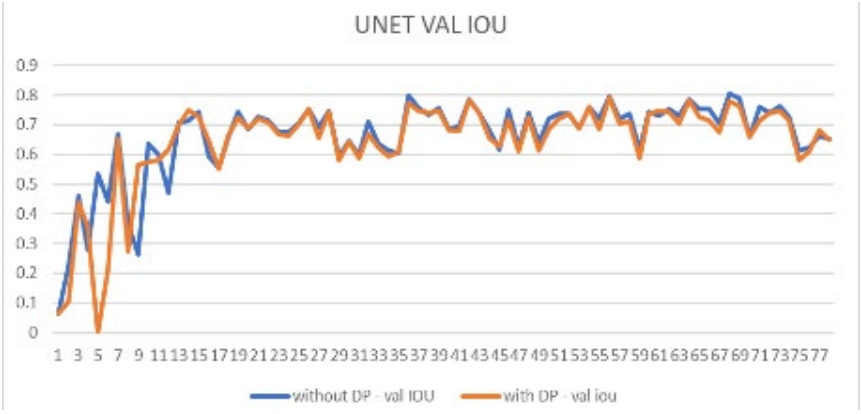

Fig. 8. Validation IoU Curves for U-Net with and without Differential Privacy (DP)

As seen in Fig. 8, the validation Intersection over Union (IoU) curves for both settings show similar progression. Although the model with DP presents a slight performance reduction (final IoU ≈ 66.32%) compared to its non-DP counterpart (final IoU ≈ 68.04%), the stability and consistency of the curve indicate that the added noise does not drastically impair the segmentation capability. This observation confirms that privacy-preserving training can be incorporated with minimal trade offs in accuracy, an important finding for sensitive domains such as healthcare, where data confidentiality is paramount.

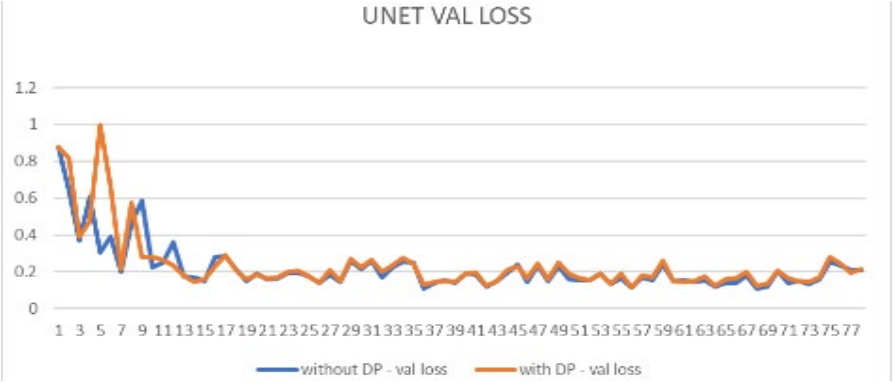

Fig. 9. Validation Loss Curves for U-Net with and without Differential Privacy (DP)

Similarly, Fig. 9 demonstrates the validation loss trends. Initially, the DP-enabled model experiences slightly elevated loss due to noise injection. However, both curves eventually converge toward a similar value range (-0.18–0.22). This convergence implies that the model maintains learning capability even under privacy constraints, validating the robustness and practicality of DP in real world deployments. The alignment of these results with prior DP literature confirms that privacy aware training mechanisms can retain acceptable segmentation quality in sensitive imaging contexts [1], [2].

Effect of Pseudo-Labelling on Convergence. In addition to privacy mechanisms, we employed pseudo-labelling to explore its role in improving training stability and reducing annotation burden. Fig. 10 presents the training loss progression before and after pseudo-labelling. The baseline model (blue line) shows a slower convergence and higher fluctuation in early epochs, with final loss at approximately 0.2112. After applying pseudo-labelling (orange line), the model achieves smoother learning dynamics, with a final loss reduced to 0.1635.

This improvement can be attributed to the additional training signal provided by high confidence predictions on unlabeled data, which serves to guide the model in regions of uncertainty. These findings are consistent with prior work suggesting that pseudo-labelling enhances representation learning and boosts segmentation performance when labeled data is limited [3], [4]. Furthermore, the smoother convergence suggests enhanced generalization and lower risk of overfitting an essential attribute for clinical model reliability.

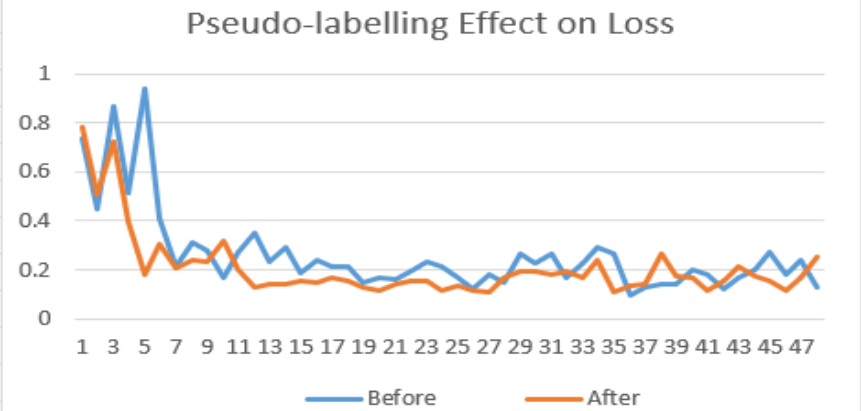

Fig. 10. Pseudo-Labelling's Effect on Training Loss for U-Net

4.5. Model Training Workflow

The complete model training workflow is illustrated in Fig. 2, outlining the pipeline from raw CBCT images to the final segmentation predictions. The process begins with data ingestion and preprocessing steps, including image resizing, noise reduction, and intensity normalization to ensure consistent input quality. To increase dataset diversity and model robustness, various augmentation techniques such as horizontal flipping, padding, rotation, and adjustments in brightness, contrast, and hue were applied. These augmentations expanded the dataset from 280 original images to over 6,800 samples, simulating clinical variability and addressing the challenge of limited annotated dental data.

5. Clinical Implications, Limitations, and Future Work

The outcomes of this study offer notable implications for the clinical adoption of privacy-preserving lesion segmentation in dental radiology. Specifically, the high Dice coefficient (83.33%) and Intersection over Union (IoU, 88.14%) achieved by U2-Net underscore its ability to generate consistent and precise lesion masks in radiolucent regions, which are often ambiguous to delineate in CBCT. Such segmentation accuracy holds promise for augmenting diagnostic workflows in dental informatics, enabling early intervention, better treatment planning, and reduced diagnostic variance among clinicians.

Visual inspection (Fig. 7) confirms that U2-Net and SA-UNet effectively capture lesion boundaries, aligning closely with expert annotations, especially for small and complex structures. This reinforces their clinical utility, as segmentations that preserve morphological accuracy are critical in surgical and endodontic contexts. Moreover, the minimal performance trade-off observed with differential privacy (DP) integration IoU reduction of only 0.7% demonstrates that patient confidentiality can be maintained without significantly sacrificing diagnostic utility. This is particularly relevant in the context of GDPR-compliant AI systems in healthcare, where balancing privacy and accuracy remains a pressing challenge.

The incorporation of pseudo-labeling further addresses one of the biggest bottlenecks in medical imaging limited annotated data. Our results reveal that pseudo-labeling reduced the validation loss by 14.9%, improving generalization and robustness. This indicates its potential to extend model scalability across multi-center datasets or rare conditions where manual annotation is scarce or costly.

Despite these strengths, certain limitations remain. The DoubleU-Net architecture performed poorly, with high training loss and unstable convergence, likely due to excessive parameterization and lack of sufficient regularization in a limited-data setting. This highlights the need to tailor model complexity to dataset size and quality. Another limitation is the lack of multi-view and multi-modal input integration, which could provide complementary anatomical cues for more robust segmentation, especially for ambiguous or overlapping lesions.

From a deployment perspective, although the proposed framework demonstrates promise in experimental settings, real-world implementation would require further validation using larger and more

heterogeneous datasets, potentially under varied imaging protocols and device manufacturers. Additionally, training with DP-SGD incurs longer training times and requires careful tuning of the noise multiplier and gradient clipping thresholds to prevent underfitting an aspect to be further explored in future iterations.

In terms of research impact and funding relevance, the outcomes of this study align with the goals of developing secure, AI-driven solutions for dental diagnostics. The demonstrated effectiveness of hybrid U-Net models with integrated privacy mechanisms supports investment into translating these methods into deployable clinical tools. Future directions include exploring transformer-based hybrid architectures, federated learning frameworks for decentralized training, and extending evaluation to real-time segmentation scenarios in clinical workflows.

6. Conclusion

This research developed and evaluated privacy-preserving U-Net variants for automatic segmentation of radiolucent lesions in dental cone-beam computed tomography (CBCT). By integrating U2-Net, SA-UNet, U-Net, and DoubleU-Net architectures with differential privacy (DP-SGD) and semi-supervised pseudo-labeling, the framework addresses critical challenges in dental imaging such as limited labeled data and patient confidentiality. Among the models, U2-Net delivered the best segmentation performance (Dice = 83.33%, IoU = 88.14%) with smooth convergence and stable learning, while SA-UNet followed closely due to its effective spatial attention mechanism. In contrast, DoubleU-Net showed limited generalizability, likely due to its over-parameterization and unstable convergence on relatively small clinical datasets. The integration of differential privacy resulted in only a minor decrease in IoU (78.2% vs. 78.9%), confirming its suitability for clinical-grade deployment without compromising accuracy. Additionally, the use of pseudo-labeling reduced validation loss by 14.9%, demonstrating the benefit of leveraging unlabeled data to improve generalization. These findings highlight that a lightweight, explainable, and privacy-aware segmentation model like U2-Net is a promising candidate for real-world dental diagnostic systems. Future directions will involve expanding to multi-class lesion classification, adopting federated learning for decentralized model training, and validating the framework across diverse CBCT scanners and institutions.

Acknowledgment

The study was approved by the International Islamic University Malaysia Research Ethics Committee, approval no. (IIUM IREC 2022-152).

Declarations

Author contribution. The contribution or credit of the author must be stated in this section.
Funding statement. The unding agency should be written in full, followed by the grant number in square brackets and year.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

Data and Software Availability Statements

Data and Software availability statements provide a statement about where data and software supporting the results reported in a published article can be found, including hyperlinks to publicly archived datasets and software analyzed and generated during the study/experiments.

References

[1] A. W. K. Yeung, "Radiolucent Lesions of the Jaws: An Attempted Demonstration of the Use of Co-Word Analysis to List Main Similar Pathologies," *Int. J. Environ. Res. Public Health*, vol. 19, no. 4, p. 1933, Feb. 2022, doi: 10.3390/ijerph19041933.

[2] S. Essaket, L. Benjelloun, and S. Chbicheb, "Odontogenic Keratocyst Mimicking a Radicular Cyst," *Integr. J. Med. Sci.*, vol. 8, p. 256, Jan. 2021, doi: 10.15342/ijms.2021.356.

[3] P. Cimflova, J. M. Ospel, M. Marko, B. K. Menon, and W. Qiu, "Variability assessment of manual segmentations of ischemic lesion volume on 24-h non-contrast CT," *Neuroradiology*, vol. 64, no. 6, pp. 1165–1173, Jun. 2022, doi: 10.1007/s00234-021-02855-z.

[4] A. Deshmukh, "Artificial Intelligence in Medical Imaging: Applications of Deep Learning for Disease Detection and Diagnosis," *Univers. Res. Reports*, vol. 11, no. 3, pp. 31–36, Jun. 2024, doi: 10.36676/urr.v11.i3.1284.

[5] M. Al-Asali, A. Y. Alqutaibi, M. Al-Sarem, and F. Saeed, "Deep learning-based approach for 3D bone segmentation and prediction of missing tooth region for dental implant planning," *Sci. Rep.*, vol. 14, no. 1, p. 13888, Jun. 2024, doi: 10.1038/s41598-024-64609-0.

[6] Z. Huang, B. Li, Y. Cheng, and J. Kim, "Odontogenic cystic lesion segmentation on cone-beam CT using an auto-adapting multi-scaled UNet," *Front. Oncol.*, vol. 14, p. 1379624, Jun. 2024, doi: 10.3389/fonc.2024.1379624.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, Springer, Cham, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[8] M. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, "Deep learning for medical image segmentation: State-of-the-art advancements and challenges," *Informatics Med. Unlocked*, vol. 47, p. 101504, Jan. 2024, doi: 10.1016/j.imu.2024.101504.

[9] H. C. Karabas, I. Ozcan, M. S. Tekkesin, S. A. Tasyapan, B. Guray, and M. M. Atapek, "Evaluation of Radiolucent Lesions Associated with Impacted Teeth: A Retrospective Study," *Curr. Med. Imaging Former. Curr. Med. Imaging Rev.*, vol. 16, no. 10, pp. 1332–1339, Jan. 2021, doi: 10.2174/1573405616666200206115827.

[10] E. Cotti and E. Schirru, "Present status and future directions: Imaging techniques for the detection of periapical lesions," *Int. Endod. J.*, vol. 55, no. S4, pp. 1085–1099, Oct. 2022, doi: 10.1111/iej.13828.

[11] A. H. Abdel-Gawad, L. A. Said, and A. G. Radwan, "Optimized Edge Detection Technique for Brain Tumor Detection in MR Images," *IEEE Access*, vol. 8, pp. 136243–136259, 2020, doi: 10.1109/ACCESS.2020.3009898.

[12] M. T. Nyo, F. Mebarek-Oudina, S. S. Hlaing, and N. A. Khan, "Otsu's thresholding technique for MRI image brain tumor segmentation," *Multimed. Tools Appl.*, vol. 81, no. 30, pp. 43837–43849, Dec. 2022, doi: 10.1007/s11042-022-13215-1.

[13] H. Mittal, A. C. Pandey, M. Saraswat, S. Kumar, R. Pal, and G. Modwel, "A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets," *Multimed. Tools Appl.*, vol. 81, no. 24, pp. 35001–35026, Oct. 2022, doi: 10.1007/s11042-021-10594-9.

[14] M. Juneja, N. Aggarwal, S. K. Saini, S. Pathak, M. Kaur, and M. Jaiswal, "A comprehensive review on artificial intelligence-driven preprocessing, segmentation, and classification techniques for precision furcation analysis in radiographic images," *Multimed. Tools Appl.*, pp. 1–54, Jul. 2024, doi: 10.1007/s11042-024-19920-3.

[15] I. S. Bayrakdar *et al.*, "Deep-learning approach for caries detection and segmentation on dental bitewing radiographs," *Oral Radiol.*, vol. 38, no. 4, pp. 468–479, Oct. 2022, doi: 10.1007/s11282-021-00577-9.

[16] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, Jul. 2021, doi: 10.1016/j.neucom.2020.04.157.

[17] M. Abadi *et al.*, "Deep Learning with Differential Privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Oct. 2016, vol. 24-28-Octo, pp. 308–318, doi: 10.1145/2976749.2978318.

[18] X. Wei *et al.*, "Building Outline Extraction Directly Using the U2-Net Semantic Segmentation Model from High-Resolution Aerial Images and a Comparison Study," *Remote Sens.*, vol. 13, no. 16, p. 3187, Aug. 2021, doi: 10.3390/rs13163187.

[19] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Jan. 2021, pp. 1236–1242, doi: 10.1109/ICPR48806.2021.9413346.

[20] C. Dwork, "Differential Privacy," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4052 LNCS, Springer, Berlin, Heidelberg, 2006, pp. 1–12, doi: 10.1007/11787006_1.

[21] Z. Zheng, H. Yan, F. C. Setzer, K. J. Shi, M. Mupparapu, and J. Li, "Anatomically Constrained Deep Learning for Automating Dental CBCT Segmentation and Lesion Detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 603–614, Apr. 2021, doi: 10.1109/TASE.2020.3025871.

[22] Y. Chen, M. Mancini, X. Zhu, and Z. Akata, "Semi-Supervised and Unsupervised Deep Visual Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1327–1347, Mar. 2024, doi: 10.1109/TPAMI.2022.3201576.

[23] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, Feb. 2020, doi: 10.1007/s10994-019-05855-6.

[24] M. M. John, H. H. Olsson, and J. Bosch, "Towards MLOps: A Framework and Maturity Model," in *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Sep. 2021, pp. 1–8, doi: 10.1109/SEAA53835.2021.00050.

[25] L. Feng *et al.*, "MLU-Net: A Multi-Level Lightweight U-Net for Medical Image Segmentation Integrating Frequency Representation and MLP-Based Methods," *IEEE Access*, vol. 12, pp. 20734–20751, 2024, doi: 10.1109/ACCESS.2024.3360889.

[26] A. M. Dostovalova, A. K. Gorshenin, J. V. Starichkova, and K. M. Arzamasov, "Comparative analysis of modifications of U-Net neuronal network architectures in medical image segmentation," *Digit. Diagnostics*, vol. 5, no. 4, pp. 833–853, Nov. 2024, doi: 10.17816/DD629866.

[27] D. Yu *et al.*, "Individual Privacy Accounting for Differentially Private Stochastic Gradient Descent," *arXiv*, pp. 1–22, Jun. 2022. [Online]. Available at: https://arxiv.org/pdf/2206.02617.

[28] M. Barry *et al.*, "StreamMLOps: Operationalizing Online Learning for Big Data Streaming & Real-Time Applications," in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, Apr. 2023, vol. 2023-April, pp. 3508–3521, doi: 10.1109/ICDE55515.2023.00272.

[29] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Jul. 2020, vol. 2020-July, pp. 558–564, doi: 10.1109/CBMS49503.2020.00111.

[30] C. Zhang, X. Deng, and S. H. Ling, "Next-Gen Medical Imaging: U-Net Evolution and the Rise of Transformers," *Sensors*, vol. 24, no. 14, p. 4668, Jul. 2024, doi: 10.3390/s24144668.