Check for updates

Received: 3 May 2024 | Revised: 6 July 2024 | Accepted: 6 September 2024

IET Image Processing

DOI: 10.1049/ipr2.13232

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Advancing precision in medical image segmentation: A performance analysis of loss functions for COVID-19 lung infection segmentation in computed tomography images

Emilio Delgado[1] | **Roberto Rodriguez-Echeverria**[1] | **Antonio Jesús Fernández-García**[2] | **Juan D. Gutiérrez**[3] | **Miguel Ángel Suero-Rodrigo**[4]

[1]Department of Computer Systems Engineering and Telematics, University of Extremadura, Cáceres, Spain

[2]Applied Computing Group (TIC-211), University of Almería, Almería, Spain

[3]Department of Electronics and Computer Science, University of Santiago de Compostela, Santiago de Compostela, Spain

[4]University Hospital of Cáceres, Extremadura Health Service (Servicio Extremeño de Salud), Cáceres, Spain

Correspondence
Emilio Delgado, Department of Computer Systems Engineering and Telematics, University of Extremadura, Cáceres 10003, Spain.
Email: edm@unex.es

Funding information
Ministerio de Ciencia e Innovación, Grant/Award Number: CPP2021-008491

Abstract

This study evaluates the effectiveness of three loss functions Asymmetric Unified Focal Loss (AUFL), Dice Similarity Coefficient Loss (DSCL), and Cross-Entropy (CE) for segmenting COVID-19 lung infections in computed tomography images. Detailed analyses using the intersection over union metric assessed each function's accuracy. AUFL achieved an average Dice Similarity Coefficient (DSC) of 85.18% ± 8.86%, outperforming DSCL 85.18% ± 8.86%, which had the same average DSC but less precise segmentation, and CE, which had an average DSC of 78.31% ± 11.93%. Segmentations using AUFL demonstrated more defined contours and better alignment with actual anatomical structures than those obtained with DSCL and CE. Observations revealed that AUFL-generated segmentations had more precise boundaries and were more consistent with the expected anatomical regions of lung infections. This study is the first to quantitatively and qualitatively compare the effectiveness of AUFL, DSCL, and CE in segmenting COVID-19 lung infections, providing concrete evidence of AUFL's superiority in segmentation performance and reliability for clinical applications. The findings underscore the importance of selecting appropriate loss functions to enhance segmentation in medical imaging, highlighting their crucial role in improving image-based diagnostics and treatment. The study emphasizes the need for ongoing research to optimize these segmentation techniques further.

1 | INTRODUCTION

In deep learning (DL) applied to medical image segmentation, the choice of loss function is a crucial factor significantly influencing model performance. This study segments lung infections in computed tomography (CT) images, particularly those caused by COVID-19. The accuracy of such segmentation is vital due to the clinical implications of early and accurate detection of severe lung pathologies.

A notable challenge in this domain is the class imbalance problem in CT images. This issue arises when the number of pixels representing the infection is significantly lower than the background or non-infected tissue. Such an imbalance can lead to biased model training and potentially suboptimal segmentation results. Addressing this imbalance is crucial, and choosing an appropriate loss function becomes even more pertinent. The selected loss function must handle the class imbalance effectively to ensure accurate differentiation between infected and non-infected regions.

The primary objective of this research is to assess how different loss functions affect the ability of various DL models to segment lung infections caused by COVID-19 in CT images effectively. To this end, a comparative study of three DL architectures, TransUNet, vision transformer (ViT), and Vanilla U-Net, employs three distinct loss function types: cross entropy (CE) as a distribution-based loss function, dice similarity coefficient loss (DSCL) as a region-based loss function, and asymmetric unified focal loss (AUFL) as a compound loss

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Author(s). *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

function. This comparison aims to analyse the capability of these combinations in segmenting lung infections, emphasizing the importance of selecting the appropriate loss function based on the specific requirements of the segmentation task.

The core research question addressed in this work is: How does the choice of loss function influence the efficacy of DL architectures in segmenting COVID-19 caused lung infections in CT images? The answer to this question is crucial for optimizing the performance of DL models in critical medical applications.

This study contributes significantly to medical image segmentation using DL architectures, focusing mainly on lung infections caused by COVID-19 in CT images. The main contributions include:

- **Selection and evaluation of loss functions**: This aspect combines the selection of loss function and investigation of specific loss functions. Emphasis is placed on choosing appropriate loss functions to address challenges such as class imbalance and small infection regions, including ground-glass opacities (GGO) and other types. The investigation focuses on applying specific loss functions like AUFL, DSCL, and CE to enhance accuracy in segmenting various sizes and types of lung lesions.
- **Comparative performance analysis**: A comparative analysis is conducted on the performance of different DL architectures—TransUNet, ViT, and Vanilla U-Net—in segmenting lung infections. This analysis provides a detailed evaluation of these techniques in clinical contexts, showing how different loss functions impact the segmentation of lesions of various types and sizes.
- **Illustrative examples of segmentation performance**: Concrete examples of segmentation performance are presented, providing a clear view of the effectiveness of different architectures and loss functions in identifying and delineating lung lesions of diverse types and sizes.

The methodology adopted to evaluate the impact of various loss functions on the segmentation of lung infections in CT images is in the following sections. The selection and configuration of specific loss functions used will be emphasized. Additionally, the experiments conducted and the comparative analyses performed will be described, focusing on the influence of each loss function on the accuracy and efficiency of segmentation, particularly in lesions of different types and sizes. Finally, the results will be discussed, conclusions drawn, and critical recommendations for future research in medical image segmentation using DL techniques will be provided.

2 | RELATED WORK

Biomedical image segmentation is a field of interest where the Vanilla U-Net architecture, initially proposed by Ronneberger et al. [1], has gained prominence. With its particular encoding and decoding design, this architecture has shown to be effective in segmenting COVID-19-associated lung infections.

Vanilla U-Net, despite its broad applications, faces limitations in accurately segmenting small objects or complex shapes, often resulting in imprecise edges. Despite these limitations, the architecture was applied in the Koudia et al. [2] study, which focused on automatically segmenting lung infections in CT images. Although limitations in segmentation performance were identified, the study highlighted the usefulness of Vanilla U-Net in identifying areas of infection.

Vanilla U-Net was compared with other architectures in a study of Saood and Hatem [3], where its superiority in automatic segmentation of medical images was evidenced. With proper preprocessing of the CT images, Vanilla U-Net demonstrated the ability to segment lung infections effectively.

While the Vanilla U-Net has been extensively applied, alternative approaches also show promise. For instance, the study by Enshaei et al. [4] presents a DL framework for automatically segmenting COVID-19 lesions in CT scans, focusing on GGO. Unlike the traditional Vanilla U-Net, this method employs an ensemble of four convolutional neural networks with encoder-decoder architectures based on pre-trained convolutional neural networks, integrating their outputs through smooth majority voting at the pixel level to enhance segmentation performance.

In his study, Qiu et al. [5] proposes using MiniSeg, an efficient encoder–decoder model for COVID-19 segmentation from CT images. MiniSeg integrates an attentional hierarchical spatial pyramid (AHSP) module for lightweight and effective multiscale learning and a dual-path encoder that combines deep and shallow convolutional paths for contextual feature extraction and fine details. This approach improves segmentation performance and reduces the propensity for overfitting.

To improve the accuracy of Vanilla U-Net, Raj et al. [6] proposed ADID-UNet, a variant that incorporates additional mechanisms such as attention gate and dense blocks. According to this study, such mechanisms can optimize the performance of Vanilla U-Net in segmenting more minor details like small lung infections.

Vanilla U-Net and its variants still retain their importance in medical image segmentation, especially for detecting and segmenting COVID-19 infections in CT. However, this architecture faces significant challenges, such as segmenting with precision irregular edges in the CT.

To address these challenges, Chi et al. [7] proposed a technique incorporating multidimensional inputs into a modified variant of Vanilla U-Net. According to this study, such an approach can more effectively capture visual features of COVID-19 infections in CT images, including complex textures and diffuse shadows.

Hu et al. [8] proposed DECOR-Net, a segmentation network based on Vanilla U-Net and designed to capture decor-related low-level features crucial for COVID-19 CT imaging. By applying a channel re-weighting strategy and a decorrelation loss, DECOR-Net reduces dependencies between channels and improves segmentation performance. This approach enhances the network's ability to accurately delineate infection regions, even in high heterogeneity and unclear boundaries.

Another approach is FV-Seg-Net, proposed by Abdel-Basset et al. [9]. This novel fully volumetric segmentation network

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Metrics associated with the studies used in the section. The images modalities used are CT for all studies.

Study	Architecture	Loss function	DSC	Recall	Specificity
Enshaei et al. [4]	Ensembled Convolutional Neural Networks (CNNs)	Categorical CE	62.70%	67.90%	98.00%
Qiu et al. [5]	MiniSeg (Vanilla U-Net based)	Custom loss function (CE Based)	75.91%	84.95%	97.72%
Koudia et al. [2]	Vanilla U-Net	Weighted categorical CE	72.60%	90.00%	96.20%
Saood and Hatem [3]	Vanilla U-Net y Seg-net	Weighted categorical CE	74.90%	95.60%	95.42%
Raj et al. [6]	ADID-UNET	Dice coefficient based loss	80.31%	–	99.66%
Chi et al. [7]	MID-UNet	Joint loss function	95.98%	96.27%	93.90%
Zhang et al. [10]	U-Net + ViT + own mechanisms	Compound loss	69.50%	72.10%	93.90%
Peng et al. [11]	U-Net + ViT + own mechanisms	Tversky loss	80.30%	77.20%	99.50%
Fan and Feng [12]	U-Net + ViT	Compound loss	79.10%	73.60%	96.70%
Hu et al. [8]	Vanilla U-Net	Compound loss	63.78%	67.99%	–
Abdel-Basset et al. [9]	Vanilla U-Net	Compound loss	85.69%	–	–

is based on Vanilla U-Net. It effectively leverages both local and global spatial information, enabling the processing of the entire CT volume at once. FV-Seg-Net improves segmentation performance by utilizing a multi-scale feature extraction strategy, simultaneously capturing fine details and broader contextual cues.

Recent developments in the field of automatic segmentation of lung infections in CT have focused on the combination of two remarkable DL architectures: ViT and Vanilla U-Net. This combination is perceived as an innovative strategy for automatic interpretation, analysis, and segmentation of TC.

Zhang et al. [10] made a first attempt at combination by proposing a system that integrates Vanilla U-Net with a compound loss function and additional proprietary mechanisms. This system, incorporating elements of ViT along with U-Net, overcomes the limitations of Vanilla U-Net in segmenting small edges and fine details using a compound loss function, specifically a compound loss function.

Similarly, Peng et al. [11] presented an approach that significantly improves the accuracy of automatic segmentations using the Tversky loss function, a region-based loss function. This approach, also combining U-Net with elements of ViT, emphasizes the adaptability of loss functions to specific segmentation contexts, leveraging the advanced image processing and understanding capabilities of ViT.

Finally, Fan and Feng [12] explored the combination of Vanilla U-Net and ViT, proposing an architecture that uses the structure of Vanilla U-Net and the power of ViT for automatic segmentation of lung infections in CT images. This study, like the study of Zhang et al. [10], uses a compound loss function, which highlights the relevance of such functions in the accuracy and efficiency of segmentation models.

Table 1 summarizes the approaches used in the studies analysed, using a ✓ to indicate the metrics used and a "-" for those not used.

The diversity of segmentation approaches in lung infection detection in CT images underscores the complexity of challenges in this field. While the combination of DL architectures like Vanilla U-Net, ViT, and TransUNet is crucial, the incorporation of loss functions genuinely stands out. These loss functions improve accuracy and are fundamental in optimizing the efficiency of automated segmentation, highlighting their vital role in training DL models to the varied segmentation needs in medical imaging.

However, the current state of the art also highlights certain limitations, such as accuracy in segmenting small details. Although significant progress has been made, challenges still require more innovative and customized solutions. The state-of-the-art review presented here provides a comprehensive and detailed overview of current approaches in the segmentation of lung infections in CT images, laying the groundwork for the next section of this study, where a possible solution to these challenges will be discussed and proposed.

3 | BACKGROUND

3.1 | Computational features and pixel analysis of ground-glass opacities in COVID-19

GGO represent a visual pattern frequently observed in CT images of infected lungs. This pattern is especially prevalent in patients with COVID-19 and is our study's main focus of segmentation. GGO can be identified in lung CT images of patients with COVID-19 through computational analysis of pixel values. These values are positioned on the Hounsfield scale (HU), which measures the radiodensity of tissues. In this scale, air, with very low radiodensity, is assigned a value of −1000 HU, while dense bone tissue, with high radiodensity, can reach up to +1000 HU. GGOs are characterized by having pixel values in an intermediate range, typically between −700 and −300 HU, indicating an increase in lung tissue density.

From a computational standpoint, GGOs can vary in pixel intensity, size, and morphology. Studies like Yu et al. [13] and Oh et al. [14] have documented that GGOs' shape can vary from round and oval to more irregular configurations, resulting in variable pixel distribution patterns.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

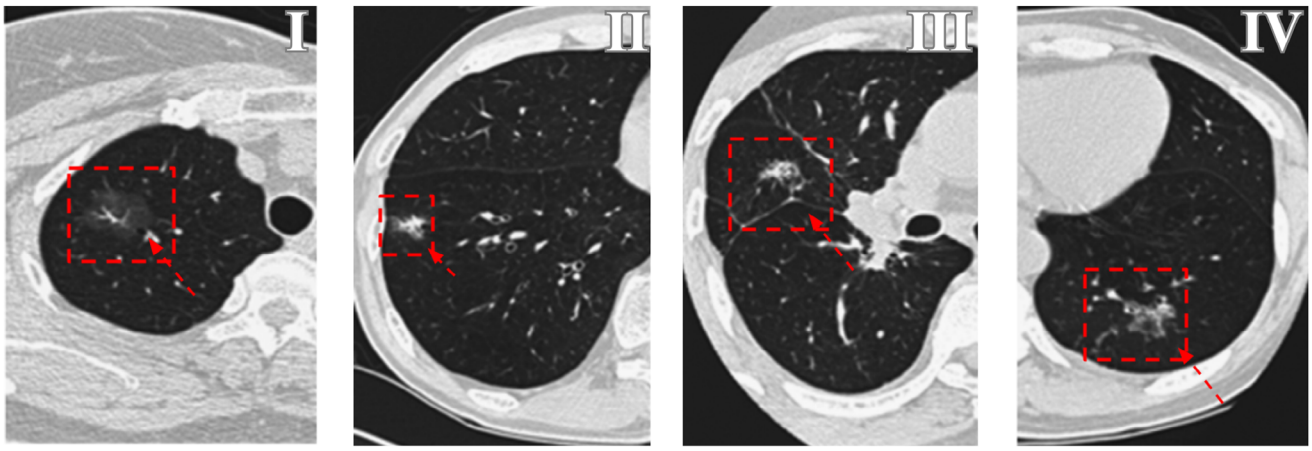

FIGURE 1 Types of ground-glass opacities. The images titles correspond to the types previously mentioned. The images have been extracted from the study of Oh et al. [14] and highlighted for better understanding.

Computational classification of GGOs can be described as follows:

- Type I: Present homogeneity in pixel values with well-defined edges, reflecting minimal alteration of the alveolar architecture. These pixel values indicate a slightly increased density compared to normal lung tissue.
- Type II: Characterized by a heterogeneous presentation in the CT image, with a mix of pixel values suggesting more severe changes in lung tissue, such as alveolar collapse and proliferation of elastic fibres.
- Type III: Distinctive for presenting higher density in the center, evidenced by higher pixel values, surrounded by peripheral GGOs with lower values. This may indicate more pronounced alveolar collapse and changes in the elastic fibre network.
- Type IV: Identified by a more homogeneous appearance in CT, reflected in a uniform distribution of pixel values throughout the lesion.

The different types of GGOs are graphically identified in the Figure 1.

The heterogeneous distribution of these lesions in the different lung lobes can be quantitatively assessed by analysing the distribution and variability of pixel values. Accuracy in detecting and characterizing GGOs, supported by image processing algorithms, is crucial for differential diagnosis and clinical assessment in patients with COVID-19. DL techniques have been proposed as a promising solution in this context (Section 2). This is because DL can learn complex and subtle features of medical images. This ability allows for more consistent and accurate detection and characterization of GGOs. Some of the advantages offered by DL in the detection of GGOs include:

- Automation and consistency: DL reduces interpretative variability and improves consistency in diagnosis by identifying patterns in pixels.
- Improved accuracy: These techniques can detect subtle changes in pixel values and morphologies, increasing accuracy in GGO detection.
- Potential for quantitative analysis: They allow the identification of GGOs and the quantification of their characteristics, facilitating more detailed and objective evaluations.
- Support in clinical decision making: Integrating DL into clinical practice supports radiologists, improving diagnostic and therapeutic accuracy for patients with COVID-19.

The inherent variability in the presentation of GGO, both terms of pixel intensity and morphology, poses significant challenges in segmenting lung infections caused by COVID-19. These challenges demand a sophisticated and adaptable analytical approach, where choosing an appropriate loss function becomes crucial. Advanced DL techniques, endowed with the ability to learn and discern complex patterns in medical image data, offer considerable promise in this field. The precise choice of the loss function is fundamental to improving accuracy and consistency in detecting and characterizing GGOs. This accuracy facilitates a more accurate diagnosis and strengthens clinical and therapeutic decision-making. This study focuses on exploring and validating the effectiveness of various loss functions in the field of DL, aimed at improving accuracy in segmenting GGOs in CT images. This improvement is crucial to contribute to treating and managing patients with COVID-19 significantly. Thus, the critical relevance of appropriately selecting the loss function, a key factor for precisely segmenting lung infections in such images, is highlighted.

3.2 | Loss functions in the field of medical imaging

Loss functions are essential in DL since they quantify the discrepancy between the output predicted by the model and the actual or desired output. Different types of loss functions suit different problems and objectives.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

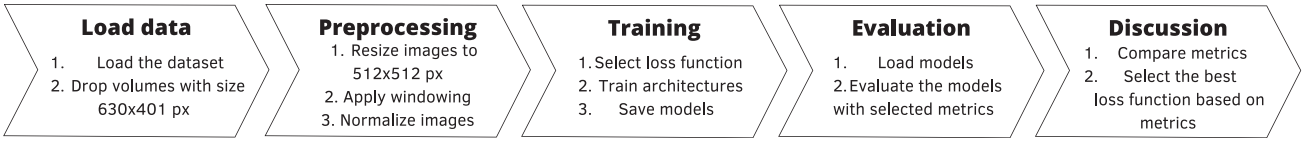

FIGURE 2 Work pipeline.

- Region-based loss functions define an area of the output space where the error is zero, and outside that region, the error grows according to a specific function.
- Distribution-based loss functions assume that the model output and the desired output follow certain probability distributions and measure the error as the divergence or distance between these distributions.
- Compound-based loss functions combine several simple functions, allowing a trade-off between accuracy and robustness.

Each function has its place in medical imaging and may be the most appropriate in different contexts.

4 | METHODOLOGY

This experiment is developed in two phases: a first phase of training nine different DL models with the selected combinations, and a second phase of evaluation of these models, comparing their results to select the combination offering the best performance. The whole work methodology used is graphically represented in Figure 2.

4.1 | Datasets

The "COVID-19 CT Lung and Infection Segmentation Dataset" is being considered meticulously curated by Jun et al. [15]. This dataset comprises 3520 carefully selected CT slices from 20 CT scans intricately associated with COVID-19-affected patients. Each dataset volume, on average, comprises a substantial collection of 176 axial slices, providing a comprehensive view of the patient's conditions.

Importantly, this dataset draws its richness from two distinct and invaluable data sources:

1. **Coronacases**: Within the Coronacases data source, we find CT images obtained explicitly from patients diagnosed with COVID-19. These CT images are meticulously analysed. These CT images are meticulously encoded in the HU, a quantitative rendering method highly valued for its usefulness in measuring the attenuation of X-rays by body tissues. These images have dimensions of 512 × 512 pixels, further accentuating the data's granularity.
2. **Radiopaedia**: These images, also associated with patients affected by COVID-19, presented in greyscale, indicate the variations of luminous intensity in the visual domain. It should be noted that most of the Radiopaedia data source volumes conform to dimensions of 630 × 630 pixels.

Furthermore, to underline the reliability and accuracy of the data set, it is paramount to acknowledge that two experienced radiologists meticulously prepared the segmentation masks accompanying these images [16]. Subsequently, these masks were subjected to thorough scrutiny and validation by a radiologist recognized for his extensive experience in the field, ensuring the highest standards of quality and accuracy in medical image segmentation.

4.2 | Preprocessing

The preprocessing step that all the CT follow consists of several steps:

1. **Resizing**: All images are resized to 512 × 512 pixels. Images with an original size of 630 × 401 pixels, which are 110 images, are excluded and correspond to only one volume of all volumes used. This information is summarized in Table 2. This decision is based on the consideration that these images may contain limited information content that, when resized, could generate a distortion in the data and negatively affect the effectiveness of the segmentation models.
2. **Windowing**: In the processing of the 'coronacases' files, the technique known as 'windowing' is applied [17]. Windowing involves adjusting the grayscale range in CT images to visually enhance certain structures. By selecting specific ranges of intensity values in the image data, windowing allows for improved differentiation and visualization of relevant tissues, such as lungs and lung infections. This is crucial for accurate identification of affected areas in medical imaging studies, particularly in diagnosing and analysing COVID-19. A window level (L) of -650 and a window width (W) of 1500 are used. The formula for calculating the gray level window can be expressed as follows:

$$I_{\text{windowed}} = \begin{cases} L + \frac{W}{2} & \text{if } I > L + \frac{W}{2} \\ L - \frac{W}{2} & \text{if } I < L - \frac{W}{2} \\ I & \text{in other cases} \end{cases} \quad (1)$$

where L is the centre of the window in HU, W is the width of the window, I is the intensity value of the processed pixel, and I_{windowed} is the final intensity value of the processed pixel.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Dataset summary. Images and volumes show totals versus selected after preprocessing.

Data source	# Images	# Volumes	Scale	Range	Dimensions
Coronacases	2681/2681	10/10	Hounsfield scale	$-1023. - 10,000.$	512×512 *px*
Radiopaedia	729/839	9/10	Greyscale	$0. - 255.$	630×630 *px*
Processed dataset	3410/3520	19/20	Normalized scale	$0. - 1.$	512×512 *px*

3. **Normalization**: Normalizes the pixel values of the images to be in the range of 0 to 1. The Normalization is described as:

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

Where:
- $X_{\text{normalized}}$ is the normalized value of X.
- X is the original value to be normalized.
- $X_{\min}$ is the minimum value in the original range of X.
- $X_{\max}$ is the maximum value in the original range of X.

These steps are implemented to ensure uniformity in the dimensions of the CT, facilitate image processing by the DL models, and operate with a standard range of values. Additionally, this fact increases the computational speed during the training phase and facilitates the calculations performed during model training.

4.3 | Evaluation

In order to objectively evaluate the different models, this work uses different performance metrics that are widely used in the automatic segmentation of organs and infections. These, in particular, are: DSC and intersection over union (IoU). These metrics had been chosen based on the literature [18] and their relevance in evaluating the quality of the segmentation performed by the models.

4.3.1 | DSC

The DSC is a metric that compares the similarity between two given sets of elements. That is, it quantifies the degree to which they are similar regarding the number of elements they share. This coefficient is commonly used in many fields. However, it is especially relevant in artificial intelligence (AI), more specifically, in evaluating the automatic medical image segmentation offered by a DL model.

As seen in Equation (3), it is calculated as twice the number of common elements between sets A and B divided by the sum of the total elements in both sets. A high value of DSC indicates a good match between automatic and manual segmentation, while a low value indicates a poor match. One of the primary uses of this metric is to evaluate the quality of the automatic segmentation performed by a DL model by comparing it with the manual segmentation provided by an expert.

In Equation (3) A and B are two sets of pixels, and |A| and |B| are their sizes. The symbol ∩ represents the intersection of the two sets, i.e. the set of elements belonging to both A and B.

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

4.3.2 | IoU

IoU is a performance metric that measures the proportion of overlap between two areas, the areas being the predictions made by the model and its corresponding mask (ground truth) respectively. The metric is calculated by dividing the intersection area (the region where the two areas overlap) by the joint area (the region that covers the entirety of both areas without duplicating the intersection). IoU outputs values between 0 and 1, where 1 indicates a perfect segmentation, and 0 indicates an incorrect segmentation.

This metric is defined in Equation (4).

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

where A is the set of segmented pixels, and B is the set of pixels associated to the "ground truth".

4.4 | Training

An experiment is proposed that evaluates the performance of three DL architectures, Vanilla U-Net, ViT and an improved version of TransUNet, in the segmentation of lung infections from CT. This allows us to analyse the impact of three loss functions, AUFL, DSCL, and CE, on segmentation quality, keeping all other parameters and hyperparameters constant and using a fixed dataset for training and validation.

The dataset, detailed in Section 4.1, presents CT cases featuring lung infections caused by COVID-19. The distribution of images for training, validation, and testing, along with their corresponding steps, is presented in Table 3.

The hyperparameters configuring the different generated models are shown in Table 4. These are kept invariant during their respective training to evaluate only and exclusively the impact of the loss functions.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Distribution of images in the training, validation, and test sets.

Dataset	Step	Number of images used	Percentage represented
Training	Training	2387	70%
Validation	Training	682	20%
Test	Evaluation	341	10%
Total	–	3410	100%

TABLE 4 Hyperparameters used in the experiment configuration.

Hyperparameter	Value
Learning rate	1×10^{-4}
Optimizer	Adam
Batch size	16
Epochs	200

In addition to the hyperparameters presented in Table 4, it is important to highlight our general approach regarding hyperparameter configuration:

- We maintained consistency across all experiments to ensure a fair comparison and isolate the impact of different loss functions.
- We used the default hyperparameters for the loss functions, as defined by the original authors or implementations, to reflect common usage in the field and avoid inadvertently favouring one function over another.
- The general training parameters (learning rate, optimizer, batch size, and number of epochs) were kept constant across all experiments, as detailed in Table 4.

To further ensure the efficiency and practicality of the training process, we used an EarlyStopping callback. This approach allowed us to terminate the training of a model when no significant improvement in the validation loss was observed for a predefined number of epochs. Consequently, while the maximum number of epochs was set to 200, the actual number varied between models. This variation reflects a more realistic training scenario where the number of epochs is dynamically adjusted based on the model's performance, ensuring that each model is trained optimally without unnecessary computation. This method enhances the robustness of our comparison by preventing overfitting and ensuring that the models are evaluated at their best performance.

This approach allows us to directly compare the inherent properties and effectiveness of the loss functions. Future work could explore the impact of hyperparameter optimization on each loss function's performance in this specific context.

After training the models, the performance of each architecture is compared with each loss function on an independent test data set using standard metrics for image segmentation: DSC and IoU.

TABLE 5 Table of experiments performed.

Experiment number	Architecture	Loss function
1	TransUNet	AUFL
2	TransUNet	DSCL
3	TransUNet	CE
4	*ViT*	AUFL
5	ViT	DSCL
6	ViT	CE
7	Vanilla U-Net	AUFL
8	Vanilla U-Net	DSCL
9	Vanilla U-Net	CE

Experimentation provides information, in terms of DSC and IoU, on the optimal combination of DL architecture and loss function for lung infection segmentation in CT. All experiments performed during the experimentation are collected in Table 5. All the different models generated were trained with a 12GB Nvidia RTX-3060 GPU. No significant differences were observed in the training and evaluation times for each loss function.

4.5 | Results

The current section presents the results obtained from evaluating the loss functions selected for training DL models, focused on segmenting specific lung infections caused by COVID-19, such as GGO. Before detailing these results, it is crucial to understand the context and rationale for choosing these loss functions framed within the class imbalance challenge. This phenomenon is characterized by the disproportionate presence of specific data categories, is especially problematic in segmentation tasks since it can lead to a model bias towards more frequent classes, impairing the detection of minority but equally essential classes.

In response to this issue, three categories of loss functions were selected and evaluated. The DSC loss function was chosen for its ability to accurately measure the spatial similarity between model predictions and reference annotations, a fundamental aspect in medical image segmentation. On the other hand, the CE loss function was selected for its ability to calibrate the prediction probability in classification problems efficiently, thus providing a robust metric for model performance evaluation. Finally, the AUFL loss function was included for its effectiveness in balancing the importance between unbalanced classes and improving the focus on complex examples, essential in class-imbalanced contexts.

These functions were selected based on a thorough review of the specialized literature, orienting their evaluation towards their effectiveness in training accurate and robust models for the task at hand. The results of this evaluation, presented below, provide detailed insight into the advantages, limitations, and practical applications of each loss function in the specific context of COVID-19 lung infection segmentation.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.5.1 | Test data selection

Two types of CT scans were integrated for the model's training: high resolution CT (HRCT) and cone beam CT (CBCT). HRCT is an imaging technique that provides detailed, high-definition images specialized in capturing thin tissue slices, allowing precise visualization of internal structures. CBCT, on the other hand, is a tomography technique that uses a cone-shaped X-ray beam to generate three-dimensional images, notable for its ability to provide a detailed volumetric view of anatomical structures. While CBCT offers advantages regarding reduced radiation dose and rapid acquisition times, it typically produces images of lower spatial resolution and contrast compared to HRCT.

However, it is crucial to note that in the evaluation phase, these two data sets were carefully separated. Such a division has allowed a specific and detailed comparison of the performance of the different loss functions in each image type and ensured an accurate assessment of clinically relevant performance metrics.

Only CT slices exhibiting signs of lung infections caused by COVID-19 were employed in the test data set for the calculation of the evaluation results. This inclusion criterion was applied to preserve the clinical relevance and integrity of the evaluative results. This methodological approach ensures that the results accurately reflect the model's performance in pathology detection and are not affected by the inclusion of standard data that could distort the interpretation of the model's effectiveness.

4.5.2 | Loss functions

In the quest to study the impact caused by adequately selecting a loss function for segmentation of COVID-19 lung infections in medical imaging, such as CT, this study has implemented and evaluated several loss functions. Each loss function has been carefully chosen for its theoretical merits and potential to improve the stability and segmentation performance of the segmentation models, as measured by IoU and, and for its ability to integrate harmoniously into a joint solution.

In this context, we define stability as the consistency of segmentation results across different model inferences, with minimal variability between segmentations performed by the trained models. Quantitatively, stability is measured through the standard deviation of DSC and IoU metrics: a lower standard deviation indicates higher stability, while a higher standard deviation reflects lower stability. Segmentation performance, on the other hand, refers to the accuracy with which the trained models delineate the contours and areas of the regions of interest. Specifically, a precise segmentation in our study is one where the boundaries and extent of the COVID-19 lung infections are adequately delimited by the models trained with these loss functions.

Several vital considerations guided our selection of loss functions for the segmentation task:

- **Relevance to segmentation tasks**: We prioritized loss functions that have demonstrated effectiveness in medical image segmentation, particularly for tasks involving complex structures such as lung infections.
- **Complementary strengths**: We opted for a combination of loss functions to address different aspects of segmentation quality:
  - DL: To optimize overall overlap between predicted and ground truth segmentations.
  - CE: To enhance pixel-wise classification accuracy.
  - AUFL: To address class imbalance, a common issue in medical imaging where lesions may occupy a small portion of the image.
- **Current state-of-the-art**: Our choices reflect the most frequently used and successful loss functions in recent literature on COVID-19 lung segmentation [19].

It is worth noting that we considered distance-based loss functions, such as Hausdorff distance. However, we ultimately excluded them for several reasons:

- They are computationally more expensive, which could potentially slow down the training process.
- They can be sensitive to outliers, which might be problematic given the variability in COVID-19 lung lesions.
- Recent literature suggests that combinations of area-based and distribution-based loss functions often provide better results than distance-based loss functions [19].

This approach allowed us to leverage the strengths of multiple loss functions while avoiding potential drawbacks associated with distance-based metrics in this context.

This selection, based on the public availability of these functions and the previous criteria, allows us to build a segmentation model that not only benefits from the individual strengths of each function but also takes advantage of their combined potential to provide a more robust and effective solution to the challenges of the problem under study. A detailed description of each loss function is provided below, highlighting how these contribute specifically to the context of our research.

- **DSC loss**

  The DSCL function is a region-based loss function focused on evaluating the spatial similarity of segmented regions compared to the proper areas of interest (ground truth). This makes it ideal for applications where the shape and size of the target regions are critical factors, such as in the segmentation of lung infections in medical imaging.

  DSCL is often used to take advantage of its ability to accurately segment structures with defined borders, such as lung infections visible on HRCT and CBCT images.

  The DSCL loss function can be expressed as:

$$\mathcal{L}^{\mathrm{DSCL}} = 1 - \frac{2 \cdot \sum_i^N p_i \cdot y_i + \epsilon}{\sum_i^N p_i^2 + \sum_i^N y_i^2 + \epsilon} \tag{5}$$

  where p_i is the predicted probability that pixel i belongs to the region of interest, y_i is the corresponding field truth label, and

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ϵ is a smoothing term to avoid division by zero and maintain numerical stability of the loss function.

- **CE**

The CE loss function is classified within distribution-based loss functions, which measure the divergence between the probability distributions predicted by a model and the actual probability distributions of the data labels. CE is recognized for its simplicity and effectiveness, especially in classification tasks.

The CE loss function is mathematically defined as:

$$\mathcal{L}^{\mathrm{CE}} = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{6}$$

where M is the number of classes, y is a binary vector of indicators (if the class is correct for the observation, $y_{o,c} = 1$, and $y_{o,c} = 0$ otherwise), and $p_{o,c}$ is the predicted probability that the observation o belongs to the class c.

- **AUFL**

The AUFL loss function is an example of a composite loss proposed by Yeung et al. [20], that integrates the elements of several loss functions to take advantage of their strengths and compensate for their weaknesses. In this case, it combines the CE and DSC, adapting them to address the challenge of class imbalance in segmenting tiny elements, such as lung infections caused by COVID-19, in medical imaging. This adaptation allows the model to focus on underrepresented classes, thus improving its ability to discriminate without compromising accuracy.

The modified asymmetric focal loss function $\mathcal{L}^{\mathrm{maF}}$ is defined as:

$$\begin{aligned}\mathcal{L}^{\mathrm{maF}} = &-\frac{1}{N}\sum_{c=r} \delta y_c \log(p_{c,r}) \\ &-\frac{1}{N}\sum_{c\neq r}(1-\delta)(1-p_{c,r})^{\gamma}\log(p_{c,r}).\end{aligned} \tag{7}$$

where N represents the total number of classes, δ is the parameter to focus more on the rare class r, y_c is the actual label for the class c, $p_{c,r}$ is the predicted probability for the class c given the rare element r, and γ is the focusing parameter in the modified asymmetric focal loss.

To complement this, a term based on Tversky's similarity $\mathcal{L}^{\mathrm{maFT}}$ is included:

$$\mathcal{L}^{\mathrm{maFT}} = \sum_{c\neq r}(1-mT) + \sum_{c=r}(1-mT)^{-1} \tag{8}$$

Here, mT is a measure derived from Tversky's similarity, which allows adjustment for the contribution of false positives and false negatives.

The final AUFL loss function is composed of the linear combination of these two terms:

$$\mathcal{L}^{\mathrm{AUFL}} = \lambda\mathcal{L}^{\mathrm{maF}} + (1-\lambda)\mathcal{L}^{\mathrm{maFT}} \tag{9}$$

The parameter λ provides a mechanism for balancing the relative importance between the modified CE and the modified Tversky similarity-based term.

5 | DISCUSSION OF RESULTS

This section aims to perform a comparative analysis focused on the segmentation of lung infections associated with COVID-19, paying particular attention to the impact that different loss functions have on the stability and segmentation performance when using various DL models under uniform configurations. This feature is crucial to ensure the reliability and accuracy of medical diagnoses that rely on image interpretation, preventing variability in the quality of the predicted masks from negatively affecting clinical decision-making.

The methodology implemented in this study allows us to determine with a high degree of confidence how the selected loss functions affect the segmentation performance of COVID-19-induced lung lesions. For this purpose, advanced imaging modalities such as HRCT and CBCT, essential tools in detecting and evaluating these lung infections, are used as a reference.

The following sections address the evaluation and comparison of different loss functions in the segmentation of lung infections by COVID-19 using the generated models. Specific functions such as DSCL, CE, and AUFL are analysed, focusing on their impact on segmentation performance and stability. In addition, the adaptability of these functions under different imaging modalities is explored, highlighting their relevance for accurate and reliable results in clinical practices.

5.1 | Comparison of performance associated with loss functions

The comparison of model loss curves using different loss functions (AUFL, DSCL, and CE) across TransUNet, ViT, and Vanilla U-Net models provides insights into the training dynamics of these approaches for medical image segmentation (see Figure 3).

A key observation from these curves is the overall similarity in training behaviour across the different loss functions. Despite minor variations, there are no dramatic differences in the loss values or convergence patterns between AUFL, DSCL, and CE for any of the three model architectures during the training process.

For TransUNet, the loss curves for all three functions follow remarkably similar trajectories. While slight differences exist in the loss values, particularly after the initial epochs, these variations are minimal during training.

In the case of ViT, we observe a comparable pattern. The overall behaviour of all three loss functions is mainly consistent throughout the training phase, with only marginal differences in loss values.

Vanilla U-Net exhibits the most noticeable variations among the three models. However, the differences between the loss functions during training are relatively minor, even in this case.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

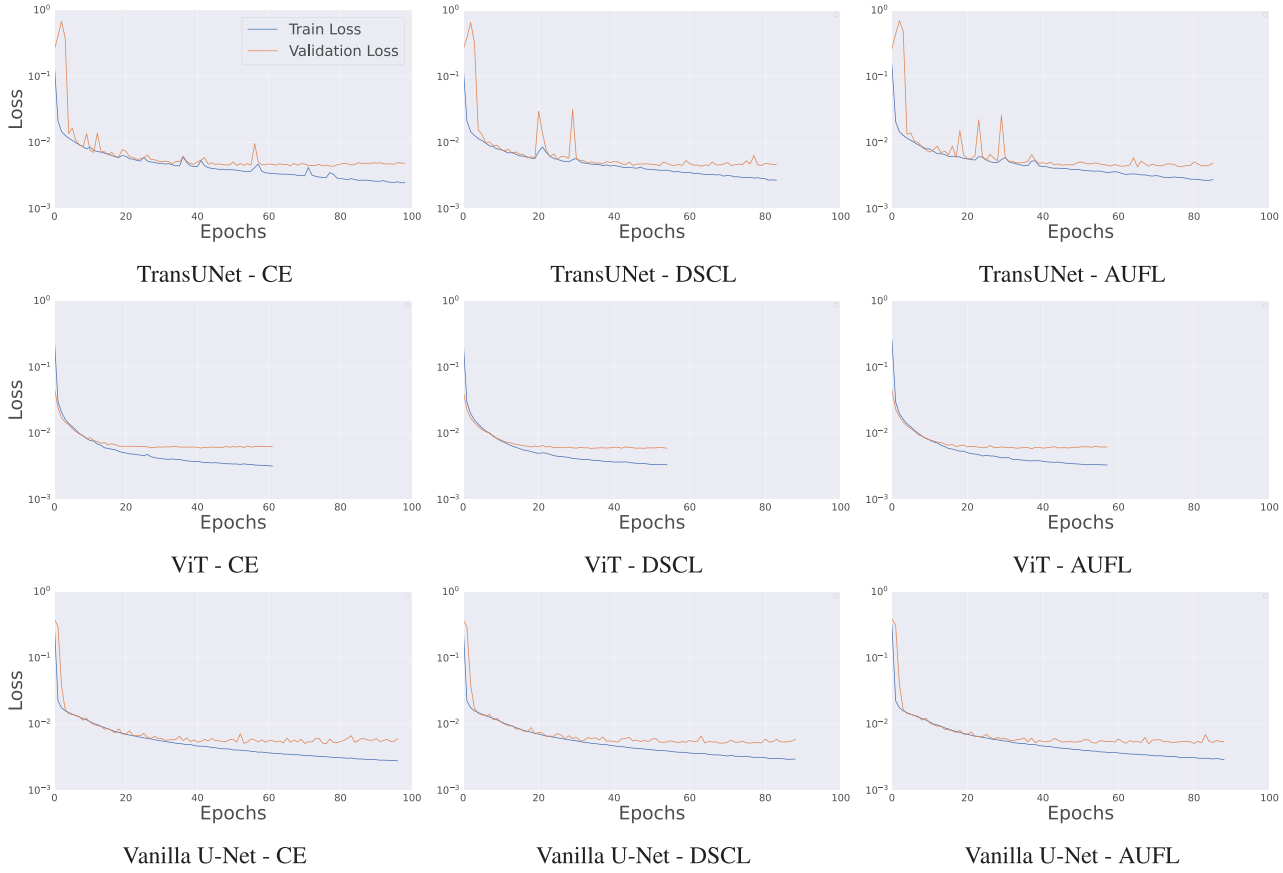

FIGURE 3 Comparison of model loss curves with different loss functions using different models.

Importantly, all three loss functions successfully guide the models to convergence, reaching low loss values by the end of the training process. This convergence occurs within a similar timeframe across all functions (±20 epochs), indicating that none of the loss functions require additional training time or computational resources.

The close alignment between training and validation curves across all scenarios suggests that overfitting is not a significant issue during training, regardless of the chosen loss function. This consistency between training and validation performance is maintained across all three model architectures and loss functions.

However, it is crucial to note that despite the similarities in training loss curves, AUFL demonstrates superior segmentation performance on the test data, as presented in the Table 6. This discrepancy between training behaviour and test performance underscores an important point: the effectiveness of a loss function cannot be judged solely by its training dynamics.

As introduced in the results section, the superior test results achieved by AUFL suggest that it guides the models to learn more relevant features for the segmentation task, even though this advantage is not apparent in the training loss curves. This improved performance on unseen data indicates that AUFL may be better at promoting generalization, a critical factor in medical image segmentation tasks.

The ability of AUFL to yield superior test results without requiring changes in the training process or additional computational resources makes it an attractive choice for medical image segmentation tasks. These findings emphasize the need for comprehensive evaluation methods in DL for medical imaging, where the actual value of a technique may only become apparent when applied to new, unseen data. The discrepancy between training curves and test performance highlights the complexity of model evaluation in this field. It underscores the importance of rigorous testing on independent datasets to assess a model's capabilities honestly.

The Table 6 illustrates performance differences, providing detailed insights into how each model responds to the varying quality of CBCT and HRCT images. The table shows that despite the lower quality of CBCT images, our models maintain a reasonable level of performance, highlighting their effectiveness and adaptability.

Table 6 establishes a meticulous comparison of AUFL, DSCL, and CE loss functions employed in different CT architectures and modalities through the analysis of DSC and IoU. These metrics constitute fundamental tools for the quantitative assessment of medical image segmentation performance, a critical aspect in the accurate delineation of lesions.

In the table breakdown, the architectures employed, such as TransUNet, ViT, and Vanilla U-Net, are listed and

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Comparison of loss functions for different models and modalities.

Architecture - Modality	Metrics	AUFL (Mean ± Std)	DSCL (Mean ± Std)	CE (Mean ± Std)
TransUNet - CBCT	DSC	86.69% ± 7.01%	**86.71% ± 7.91%**	81.76% ± 9.70%
	IoU	**90.06% ± 7.29%**	86.45% ± 11.70%	86.21% ± 12.23%
ViT - CBCT	DSC	**77.67% ± 22.69%**	72.83% ± 29.50%	74.97% ± 15.66%
	IoU	**79.90% ± 23.25%**	73.35% ± 30.70%	81.77% ± 19.83%
Vanilla U-Net - CBCT	DSC	**83.94% ± 9.35%**	82.12% ± 16.31%	79.15% ± 11.79%
	IoU	**86.99% ± 10.46%**	81.33% ± 19.86%	83.44% ± 12.64%
TransUNet - HRCT	DSC	**85.18% ± 8.86%**	84.13% ± 12.86%	78.31% ± 11.93%
	IoU	**87.57% ± 9.47%**	84.91% ± 13.98%	80.53% ± 13.93%
ViT - HRCT	DSC	**79.85% ± 16.97%**	65.32% ± 33.75%	74.89% ± 12.15%
	IoU	**83.50% ± 17.67%**	64.00% ± 35.13%	80.44% ± 12.60%
Vanilla U-Net - HRCT	DSC	82.75% ± 10.04%	**82.95% ± 13.42%**	77.01% ± 12.90%
	IoU	**87.70% ± 11.14%**	83.11% ± 15.57%	84.18% ± 15.22%

correlated with specific CT modalities, differentiated into HRCT and CBCT. The latter represent advanced imaging techniques in lung diagnostic evaluation, particularly in detecting COVID-19-related abnormalities.

The structure of the columns reflects the evaluation metrics adopted, with the first column highlighting the uniform implementation of DSC and IoU across all instances evaluated. Subsequent columns detail the AUFL, DSCL, and CE loss functions used in the model training process. Each is presented along with its average and standard deviation, providing a comprehensive picture of each model's behaviour in the face of uniform configurations.

The significance of a reduced standard deviation is emphasized, as it denotes superior consistency and stability in model performance, a precious quality in the COVID-19 segmentation of lung lesions, where high accuracy is imperative for effective diagnosis and treatment. Therefore, the interpretation of these metrics is performed considering the direct impact they have on the reliability of clinical diagnoses.

The AUFL loss function, best overall, has demonstrated, regardless of the network architecture employed, remarkable efficiency and stability in the segmentation of lung infections by COVID-19. This is reflected in notably superior results, particularly in IoU, for most cases, with only two exceptions where it virtually ties with the DSCL loss function. The relevance of IoU, as pointed out by Müller et al. [21], lies in its ability to more effectively penalize both under- and over-segmentation compared to DSC. This feature makes it a robust and reliable metric, especially valuable for assessing the accuracy of segmentation models. Although DSC is still widely used in scientific publications for evaluation in medical imaging, the preference for IoU in this context is justified by its higher stringency in penalizing segmentation errors. This robustness of IoU is particularly significant in the use of CBCT images, where the integration of AUFL in the TransUNet model has demonstrated not only accuracy but also more excellent stability than that obtained by the other models using different loss functions, crucial aspects for clinical practice where reliability is as essential as the accuracy of the results.

It is observed that although the AUFL loss function generally exhibits high general performance, it is not uniformly the most effective. In the Vanilla U-Net model with the HRCT modality, DSCL achieves a slightly higher average DSC of 82.95%, compared to AUFL's 82.75%, but with a higher standard deviation (13.42% for DSCL versus 10.04% for AUFL), indicating more significant variability in DSCL's results. In the TransUNet model using the CBCT modality, DSCL shows a slightly higher DSC of 86.71% compared to AUFL's 86.69%, suggesting the need for careful selection of loss functions for optimal performance and consistency in different imaging modalities. The variability in the performance of the models trained with the DSCL loss function could derive from their sensitivity to fluctuations in image quality since observing Table 6 evidences a significant difference between the different standard deviations obtained for each modality by the different models. Thus, both AUFL and DSCL can be considered equally effective for use in different imaging modalities, with AUFL possibly offering slightly better consistency across varied scenarios.

As for the CE loss function, the results reflect a generally lower performance and lower stability. This could indicate a lower ability of CE to handle the particularities of COVID-19-specific lung lesion segmentation, perhaps due to lower robustness to variations in lesion characteristics or insufficient penalization of incorrect predictions at infection edges. A detailed understanding of these limitations is crucial and highlights the importance of selecting loss functions that promote accuracy and robustness to variations in disease manifestations.

The analysis highlights the importance of segmentation performance and stability in segmentation, emphasizing the need to employ composite loss functions in the clinical context of COVID-19, specifically in the segmentation of lung infections. These functions combine the individual advantages of each, with careful selection crucial to ensure accuracy, reliability, and stability.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 Tukey's HSD Test Results for the segmentation performance of the loss functions considered.

(a) DSC Tukey's HSD test results

Group1	Group2	meandiff	p-adj	Reject H_0
AUFL	CE	−0.0544	0.0	Yes
AUFL	DSCL	−0.0442	0.0002	Yes
CE	DSCL	0.0102	0.6173	No

(b) IoU Tukey's HSD test results

Group1	Group2	meandiff	p-adj	Reject H_0
AUFL	CE	−0.0389	0.0028	Yes
AUFL	DSCL	−0.0804	0.0	Yes
CE	DSCL	−0.0415	0.0013	Yes

It is established, after a meticulous analysis presented in Table 6, that the AUFL loss function proves to be the most effective in CT medical image segmentation, regardless of the neural network architecture used, such as TransUNet, ViT, and Vanilla U-Net. This conclusion is derived from the exhaustive comparison with DSCL and CE loss functions, using DSC and IoU indicators. The AUFL stands out for its ability to achieve greater segmentation performance and stability in delineating lung lesions related to COVID-19, evidenced by superior results, particularly in the IoU. It is observed that, although the DSCL function achieves slightly superior performance in some instances, this does not alter the overall dominance of AUFL, considering the more significant variability and lower stability of DSCL in different modalities. On the other hand, the CE loss function shows inferior segmentation performance and stability, indicating its limited efficacy in COVID-19 lung lesion-specific segmentation. Therefore, it is concluded that the choice of AUFL as a loss function is validated by its accuracy and robustness to variations in disease manifestations. It is crucial in clinical practice to ensure reliable and consistent results.

To assess the performance of the segmentation methods across different loss functions, we employed a one-way analysis of variance (ANOVA) followed by Tukey's honest significant difference (HSD) test for post-hoc comparisons. The analysis was conducted for each loss function, grouping the DSC coefficients and IoU of the three models generated for each loss function. This approach allowed us to evaluate the consistency of performance across multiple training instances. ANOVA was used to determine if there were any statistically significant differences among the means of the groups, while Tukey's HSD test allowed for pairwise comparisons between the methods.

Our null hypothesis (H_0) was that there is no significant difference in performance (measured by DSC coefficients and IoU) among the different segmentation methods (AUFL, CE, and DSCL). The alternative hypothesis (H_1) was that at least one of the methods performs significantly differently from the others.

ANOVA was used to determine if there were any statistically significant differences among the groups' means, while Tukey's HSD test allowed for pairwise comparisons between the methods.

The results of Tukey's HSD tests are summarized in Table 7. The table provides a detailed comparison between each pair of segmentation methods, indicating adjusted p-values (p-adj) and whether the null hypothesis (no difference) was rejected. The results show statistically significant differences between several pairs of segmentation methods, as evidenced by the adjusted p-values (p-adj) being less than 0.005 in most cases.

For instance, when comparing AUFL and CE methods using the DSC metric, we observed a mean difference of −0.0544 with an adjusted p-value of 0.0. Given that this difference is statistically significant (p-adj = 0.0), we can conclude with confidence that this observed difference is not due to chance.

In this context:

- A difference of −0.0544 can be moderate, as a difference of around 5% could represent a significant improvement or deterioration in practice, depending on the specific application.
- It is essential to contextualize this difference within its practical impact. In segmentation tasks, for example, a 5% difference can be considerable if it affects the quality of predictions.

Similarly, when comparing AUFL and DSCL methods using the IoU metric, we found a more substantial mean difference of −0.0804 (p-adj = 0.0), indicating an even more pronounced and statistically significant difference in performance. These findings suggest that the choice of segmentation method substantially impacts the outcomes, with certain methods demonstrating superior performance compared to others. The specific pairwise comparisons and their respective significance levels provide valuable insights into each segmentation approach's relative effectiveness within our study's context.

It is worth noting that while some differences, such as between CE and DSCL in the DSC metric (meandiff = 0.0102, p-adj = 0.6173), are not statistically significant, they may still have practical implications depending on the specific requirements of the segmentation task at hand.

These results underscore the importance of carefully selecting segmentation methods based on statistical significance and practical relevance to the specific application domain.

In order to deepen the analysis presented, it should be emphasized that the results discussed in the following subsections are derived exclusively from the application of the TransUNet architecture. This decision is justified by the effectiveness and relevance of such architecture in the context of our study. When considering the various loss functions applied, it is observed that TransUNet maintains a consistent performance, which underlines its robustness and adaptability against different evaluation criteria. This approach allows for a more detailed and specific comparison of the results, thus ensuring

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a more accurate and focused interpretation of TransUNet's unique characteristics.

5.2 | Implications and effectiveness of compound loss function in segmentation of lung lesions by COVID-19

Having addressed a general comparison of loss functions in lung lesion segmentation, the investigation now turns to a more detailed analysis of the performance and application of these functions, focusing on different sizes of infections. While the study includes an evaluation of loss functions across a full range of infection sizes, special attention is given to smaller lesion sizes due to their high prevalence and significant clinical importance in the context of COVID-19, as reflected in previous studies, exemplified by Yu et al. [13]. This thorough approach seeks to elucidate the influence of different loss functions on the accuracy and efficacy of lung lesion segmentation, especially in smaller lesions, which is crucial for the continuous improvement of disease diagnosis and treatment.

The choice of AUFL as a loss function is reinforced by findings such as those of Yu et al. [13] and Oh et al. [14], which indicate that more than 68% of COVID-19-associated lung lesions are GGO of less than 2 cm in diameter, equivalent to a total of 1887 pixels in the infection masks on the CT scans used. This approach aligns with the detailed study of loss functions in the context of the specific characteristics of COVID-19 lung lesions, underlining the need for balanced and reliable performance in clinical application for effective disease management.

A computational approach was adopted to address the challenge of classifying different masks into different sizes based on the metadata associated with CT scans collected in NIFTI format. This method, focused on estimating the number of infection-associated pixels within a circumference of diameter X cm, assuming a circular area, provides a valuable and adaptable estimate for classifying the size of lesions or infections on CT scans.

Applying the formula for the area of a circle $A = \pi r^2$, where r is half the diameter X (i.e. $r = \frac{X}{2}$, the area of the circumference turns out to be $A = \pi \left(\frac{X}{2}\right)^2$. With the area of an individual pixel being $0.8107 \times 0.8107\,\text{mm}^2$, the total number of pixels on the circumference is calculated by dividing the area of the circumference by the area of a pixel. This calculation is summarized in the Equation (10).

$$\text{Number of infection pixels} = \left\lceil \frac{\pi\left(\frac{X}{2}\right)^2}{0,8107^2} \right\rceil \tag{10}$$

Although an idealized circular shape was used for simplicity, it is recognized that in practice, many lesions may have oval or irregular shapes [13], underscoring the need to consider the simplifications made in assuming a circular shape.

To complement this analysis, we use a histogram in Figure 4 to categorize the infection points detected by the masks into four defined intervals: 44 to 2043 (<4.3 cm of diameter), 2044 to 4043 (<5.9 cm of diameter), 4044 to 8043 (<8.3 cm of diameter), and 8044 to 23,280 (<14 cm of diameter). The choice of these intervals is based on achieving a balance between representativeness and statistical significance within the data set. Each interval contains enough masks to allow for a robust and reliable assessment of the distribution of lesion sizes. The delineation of these intervals is intended to accurately capture the variability in lung lesion sizes caused by COVID-19, as seen in the images. This stratification is essential for a detailed and meaningful comparison between size categories, thus facilitating the identification of trends and patterns in the incidence and characteristics of lung lesions.

Subsequently, we aggregate the results obtained from the different CT modalities. This unification was performed to obtain a more detailed perspective of the lesions or infections.

Figure 5 presents a comprehensive analysis of the different loss functions used in the segmentation of lung infections by COVID-19 in CT images, using the metrics DSC and IoU. This figure shows how each loss function affects segmentation performance and consistency regarding these critical metrics.

The predominant concentration of masks in the first interval underscores the high incidence of smaller lesions in the dataset, highlighting the importance of accurate segmentation for these smaller lesions. This information complements the findings presented in Figure 5, where the boxplots corresponding to each size interval provide detailed insight into the performance of the loss functions in segmenting lesions of different sizes.

The box plots in Figure 5 display the DSC and IoU metrics distribution for the AUFL, DSC, and CE loss functions across the different bins. These plots show the individual performance of each loss function and reveal a clear trend: as lesion size decreases, all loss functions tend to show reduced efficiency. However, it is notable that AUFL exhibits a less pronounced decrease in performance with smaller lesions, maintaining compact and high median boxes reflecting superior segmentation performance and consistency compared to and CE. In contrast, DSCL and CE exhibit more significant variability and a steeper decline in their effectiveness as lesion size decreases. In particular, the DSCL function shows many outliers, especially in the IoU metric, indicating less optimal segmentation performance in small lesions. Although DSCL may be effective in specific contexts, its general applicability for smaller lesions is compromised. It may require additional optimizations to ensure consistent and reliable results across all lesion size ranges.

The CE function also shows a noticeable scatter in segmentation results. Wider IQRs and lower medians compared to AUFL indicate lower stability, which is critical when accuracy is critical for clinical applications, as is the case for detecting small lung lesions caused by COVID-19.

Detailed analysis in Figure 5 reflects distinct trends in the effectiveness of AUFL, DSCL, and CE loss functions in segmenting lung infections by COVID-19 in CT images across DSC and IoU metrics. Box plots for AUFL show lower scatter and consistently higher medians across all infection size

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

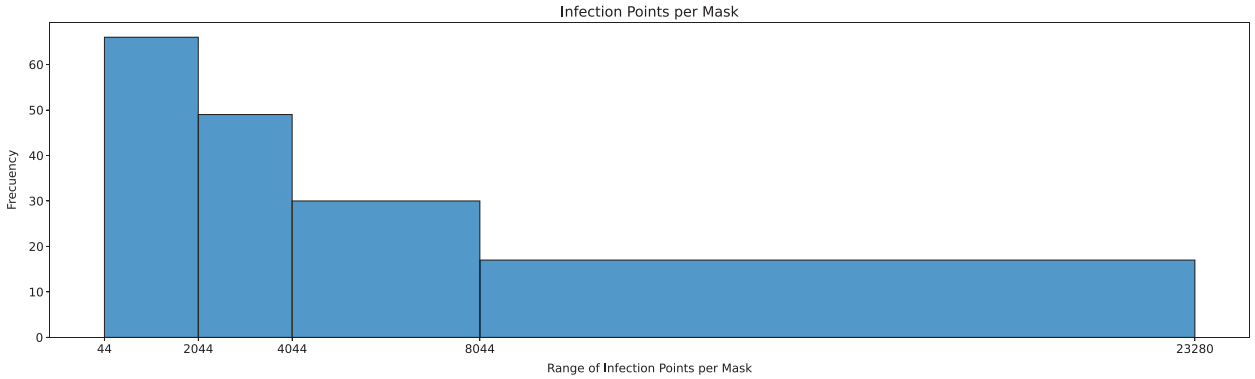

FIGURE 4 Histogram of the distribution of lung lesion sizes, categorized by mask infection sites.

TABLE 8 Comparison of performance metrics of segmentation with TransUNet model with AUFL, DSCL, and CE loss functions grouped by lesion size.

Loss function	Bin	DSC (Mean ± Std)	IoU (Mean ± Std)
AUFL	44–2043	**82.97% ± 10.01%**	**87.68% ± 10.55%**
DSCL		82.96% ± 15.86%	86.07% ± 18.22%
CE		76.53% ± 13.66%	82.11% ± 17.79%
AUFL	2044–4043	90.10% ± 6.20%	**93.77% ± 7.91%**
DSCL		**90.27% ± 5.85%**	91.68% ± 7.26%
CE		84.83% ± 7.13%	89.24% ± 9.99%
AUFL	4044–8043	**88.81% ± 4.61%**	**89.94% ± 6.38%**
DSCL		88.71% ± 4.00%	87.47% ± 6.14%
CE		83.69% ± 4.26%	83.93% ± 6.25%
AUFL	8044–23,280	90.43% ± 4.94%	**90.51% ± 5.31%**
DSCL		**90.59% ± 4.30%**	90.21% ± 6.52%
CE		86.48% ± 5.06%	88.89% ± 5.30%

intervals, indicating superior and stable segmentation performance regardless of lesion size. In contrast, DSCL and CE exhibit greater data scatter, with noticeable variability increasing at smaller bins, suggesting a decrease in segmentation performance as lesion size decreases. Specifically, the DSCL function shows substantial outliers, particularly in IoU metrics, implying more frequent incidences of inaccurate segmentations. These observations are crucial, as they highlight the importance of selecting an appropriate loss function that can maintain accuracy across various lesion sizes, with AUFL demonstrating the slightest marked decrease in performance with minor lesions.

While the results obtained with the AUFL loss function indicate superior performance in CT segmentation, particularly for small-sized lung lesions associated with COVID-19, it is imperative to take a critical stance and meticulously analyse its potential limitations. Table 8 contains the means and standard deviations associated with the DSC and IoU metrics for each bin loss function exposed in Figure 5. It is observed that, for all cases, AUFL obtains a higher IoU than that offered by the rest of the loss functions regardless of lesion size.

5.3 | Comparative evaluation of AUFL in the segmentation of lung infections by COVID-19

Figure 6 compares false negative rates for the AUFL, DSCL, and CE loss functions. This visualization is critical for analysing the results obtained by applying the AUFL loss function, highlighting its impact on the accuracy of segmentation of lung infections caused by COVID-19.

In the analysis of AUFL, CE, and DSCL loss functions applied to different bins and synthesized in Table 9, a distinctive pattern is observed regarding false negative rate (FNR) and false positive rate (FPR). In terms of FNR, AUFL demonstrates consistent efficacy in reducing false negatives (infection points considered as benign by the models) across all bins. This feature is especially relevant in contexts where the omission of positive instances could have critical consequences. The superiority of AUFL in minimizing false negatives is maintained across the different mask sizes represented by the bins, although the magnitude of this superiority varies.

In contrast, FPR analysis reveals a more balanced scenario among the three loss functions. CE is slightly more effective at smaller bins in controlling false positives (points with no infection considered as infection points by the models), which is crucial as a higher FPR indicates over-segmentation. This control is an essential advantage in applications where false positives may generate unnecessary costs or actions. However, as we move into larger bins, the difference between CE, DSCL, and AUFL narrows, showing closer competition. AUFL performs slightly better in the larger bins, although the difference is minimal. It is worth noting that while minimizing FPR is important, achieving a lower false negative rate (FNR) is generally more critical in infectious disease contexts. A lower FNR means fewer infected points are missed, which is crucial for effective disease control and prevention of outbreaks, even if it comes at the cost of a slightly higher FPR.

The lower section of Figure 6 presents violin plots corresponding to the AUFL, DSCL, and CE loss functions. There is a noticeable variability in the density of false negative rates for each loss function, as manifested in the diversity of the violin plots' shapes and widths. This variability indicates differences in the capacity of the loss functions to identify infected pixels in lung segmentation images effectively. It is crucial to note that all

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

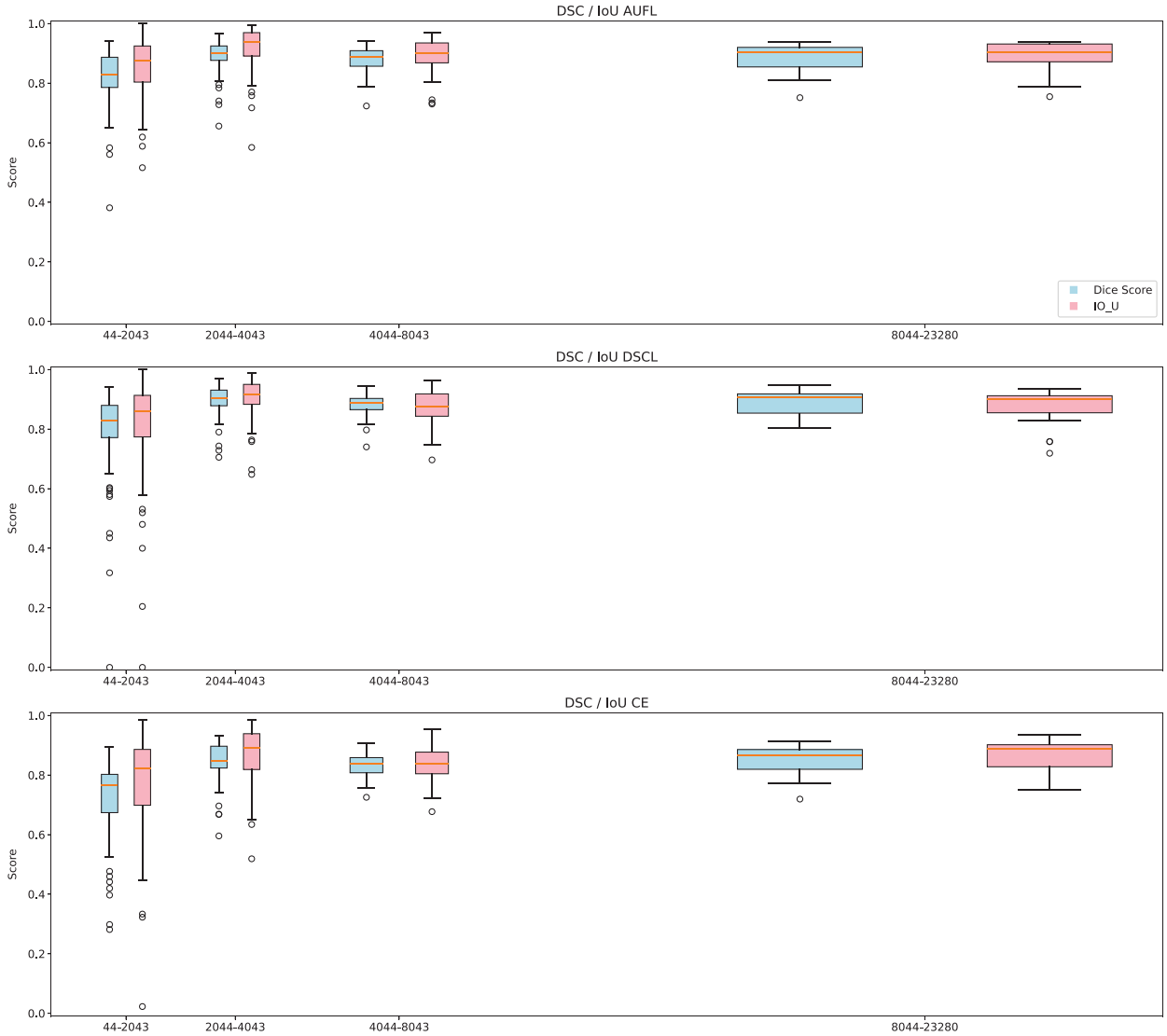

FIGURE 5 Performance evaluation of the different loss functions in test set segmentation.

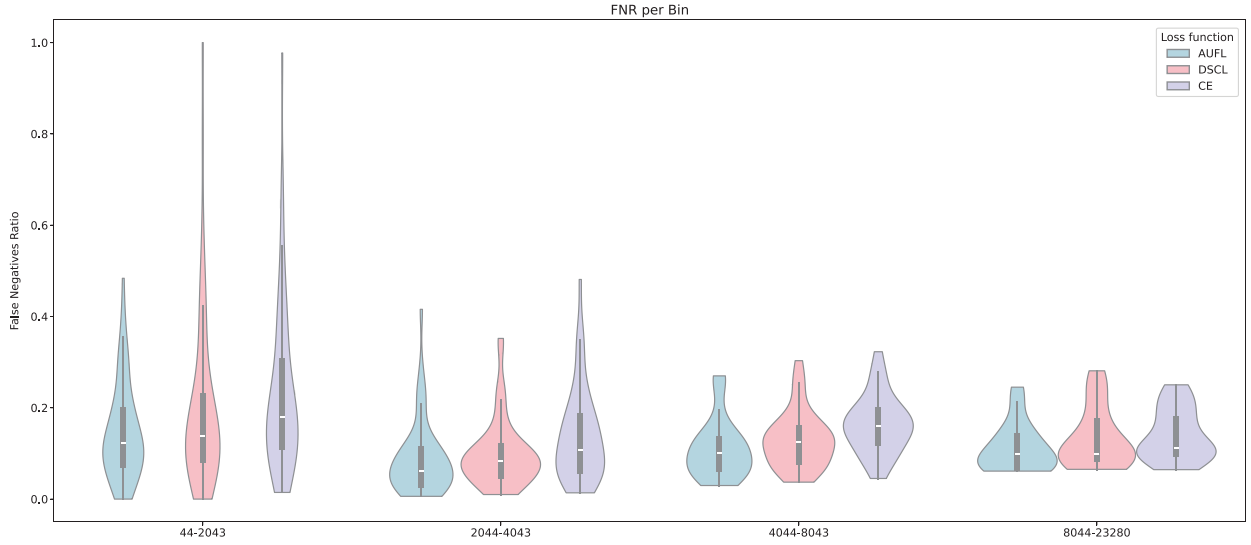

FIGURE 6 Comparative evaluation of false negative rates in lesion size classification using the TransUNet model with AUFL, DSCL, and CE loss functions.

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 9 Comparison of false negative (FNR) and false positive (FPR) rates with the TransUNet model with AUFL, DSCL and CE loss functions grouped by lesion size.

Bin	Metric	AUFL	CE	DSCL
44–2043	FNR	**14.65% ± 10.64%**	23.05% ± 17.92%	19.50% ± 18.36%
2044–4043	FNR	**8.75% ± 8.00%**	13.91% ± 10.10%	9.85% ± 7.34%
4044–8043	FNR	**11.05% ± 6.50%**	16.33% ± 6.35%	12.97% ± 6.24%
8044–23280	FNR	**11.17% ± 5.36%**	14.06% ± 5.65%	13.29% ± 6.62%
44–2043	FPR	0.08% ± 0.05%	**0.06% ± 0.07%**	0.07% ± 0.06%
2044–4043	FPR	0.15% ± 0.08%	**0.10% ± 0.07%**	0.13% ± 0.08%
4044–8043	FPR	0.30% ± 0.14%	**0.19% ± 0.12%**	0.25% ± 0.12%
8044–23,280	FPR	0.66% ± 0.40%	0.54% ± 0.34%	**0.52% ± 0.33%**

loss functions demonstrate a more consistent false negative rate in more extensive lesions. The AUFL loss function is depicted with a violin plot that shows a denser distribution towards the lower end of FNR, suggesting higher efficiency in detecting infected areas. The density concentration at the lower ranges for AUFL indicates a more precise performance in identifying pathological areas, potentially resulting from more effective handling of the class imbalance in lung segmentation tasks. Conversely, the DSCL function exhibits a broader spread of false negative rates, indicative of variability in its performance. This could stem from the class imbalance problem posed by different types of infections and their sizes. Meanwhile, the CE function is characterized by a distribution that suggests a capacity intermediate between DSCL and AUFL. However, it is less optimized than AUFL in minimizing false negatives, particularly in larger lesions with more uniform performance.

This analysis suggests that the choice of the optimal loss function may vary depending on the specific bin and, hence, mask size. While AUFL excels at minimizing FNR across all bins, the choice between CE and DSCL to optimize FPR may depend on the size of the mask in question. These observations emphasize the importance of considering error type (false positives versus false negatives) and mask size when selecting the most appropriate loss function for a specific application.

5.4 | Illustrative examples of segmentation performance

The images presented in Figures 7 and 8 illustrate the segmentation results using the AUFL, DSCL, and CE loss functions for the CBCT and HRCT imaging modalities, respectively. For each modality, the same image slice is used across all three loss functions to compare their performance on an identical section. It is important to note that these representations have been chosen primarily for illustrative purposes. While these sections are distinguished by their atypical IoU values, the selection of these specific images should not be considered as a generalization of the results. Therefore, they should be considered as representative examples. Additionally, the images have been magnified to allow for a more detailed examination of the segmentation results.

For the CBCT modality shown in Figure 7, the AUFL loss function achieves an IoU of 87.301%, which reflects a high agreement between the prediction and the actual mask, although a slight over-segmentation is observed. This is manifested in the inclusion of additional areas beyond the mask contour, mainly observed in the lower central area. In contrast, the DSCL function presents an IoU of 64.285%, and a segmentation is identified that, although adequate, shows areas that seem not to be delineated entirely, which could be interpreted as under-segmentation. On the other hand, the CE function evidences an IoU of 32.275%, with a marked over-segmentation, where the prediction significantly exceeds the mask limits in several regions, indicating a reduced precision in delineating the areas of interest.

Concerning the HRCT modality shown in Figure 8, the AUFL function reaches an IoU of 69.083%, showing moderate segmentation performance although over-segmentation continues to be noted to a lesser extent than in CBCT. The DSCL function, with an IoU of 58.778%, shows similar segmentation fidelity, with a slightly higher tendency towards under-segmentation, suggesting the possible omission of relevant areas. The CE function, with an IoU of 44.656%, confirms intense over-segmentation, aligning with the observations made in the CBCT modality, highlighting this function's difficulty in accurately delineating areas of interest.

It highlights, therefore, the relevance of adequately selecting the loss function to optimize the performance of the segmentation algorithms. The AUFL function, despite its slight over-segmentation in the CBCT modality, demonstrates superior performance compared to the other loss functions evaluated, evidencing its applicability in clinical contexts where accuracy is paramount. The need to balance over- and under-segmentation to maximize segmentation performance in different medical imaging modalities is emphasized.

6 | LIMITATIONS AND THREADS TO VALIDITY

This section is expanded to provide a deeper understanding of the restrictions and assumptions underlying the analysis of

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

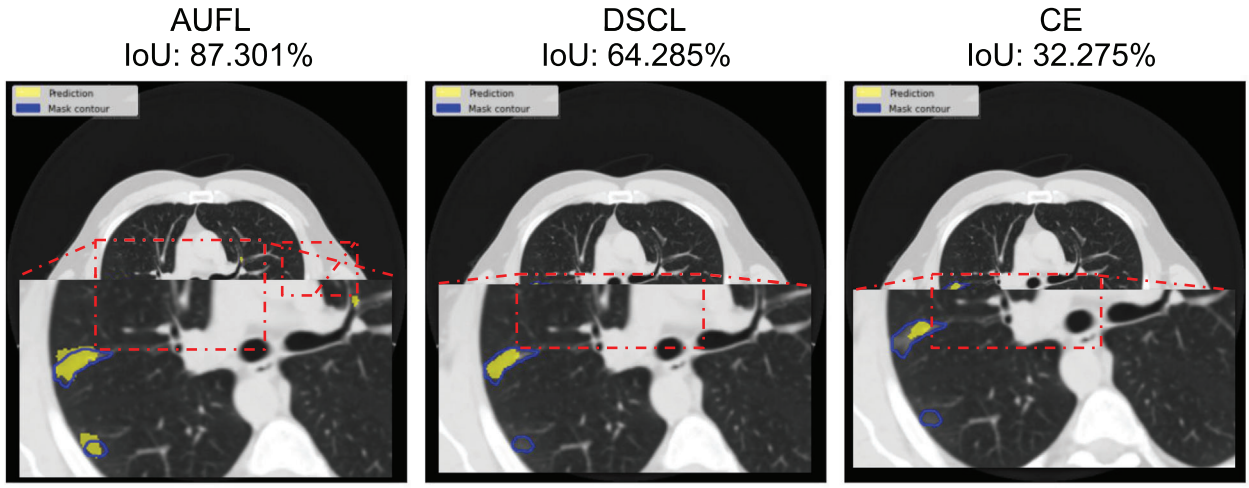

FIGURE 7 Segmentation results for CBCT modality.

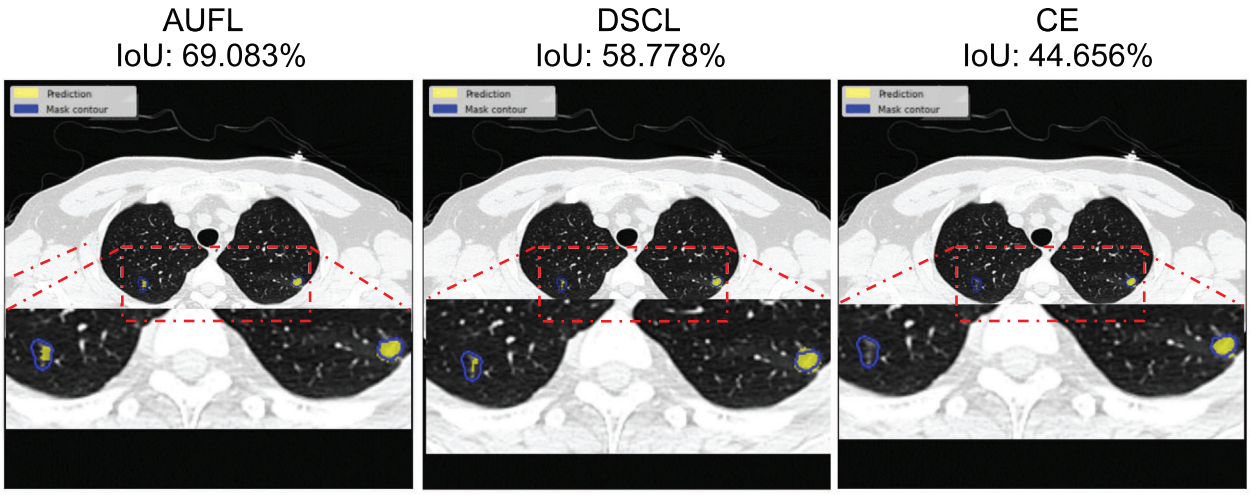

FIGURE 8 Segmentation results for HRCT modality.

loss functions in image segmentation. The following additional limitations are detailed:

- **Generalization to other imaging modalities**: Although the study focuses on the segmentation of lung infections in CT images, it should be noted that the results may not apply to other medical imaging modalities, such as magnetic resonance or ultrasound. Each modality has unique characteristics regarding resolution, contrast, and types of artefacts, which could require adaptations in the loss functions used.
- **Instrumentation dependence**: It should be considered that the CT images used come from different equipment and manufacturers. Variations in the quality and characteristics of these images can influence the effectiveness of the loss functions and the replicability of the results in other CT data sets.
- **Diversity of pathologies**: The current research is limited to lung infections, and the representativeness of these in the dataset may not cover the wide range of clinical presentations. Loss functions might require specific adjustments to accommodate the heterogeneity of other lung conditions or pathologies.
- **Clinical interpretation and applicability**: This study has not addressed the integration of automatic segmentation into clinical practice and its interpretation by specialists. Clinical validation and acceptance of these tools are crucial for their adoption into diagnostic routines.
- **Variability in segment delineation**: The inter- and intra-observer variability in the definition of the true segmentation masks can introduce biases in evaluating the loss functions. This variability underscores the need to standardize segmentation procedures to improve the objectivity of the results.

These limitations emphasize the need for future research to evaluate loss functions in broader and more diversified contexts,

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

considering collaboration with specialists for clinical validation and optimizing segmentation techniques for their practical application in medical diagnosis.

7 | CONCLUSIONS

This study has provided a comprehensive and comparative evaluation of different loss functions used in lung image segmentation, primarily focusing on COVID-19-related infections. It has been shown that the appropriate choice of loss function is crucial to achieving accurate segmentation, which has significant implications for clinical interpretation and patient management.

Among the main findings, it stands out that the AUFL loss function proved to be the most effective in most scenarios, particularly in the segmentation of small lesions, which is vital for accurate COVID-19 diagnosis. However, variations in performance were observed depending on the lesion size and image type (i.e. CBCT vs. HRCT), highlighting the importance of the careful selection of the loss function based on the specific application context.

Significant limitations were identified that should be considered for future research. These include the need to generalize findings to other imaging modalities, dependence on the type of instrumentation used, diversity in lung pathologies, computational considerations, and variability in segment delineation. Addressing these limitations is essential to expand the applicability of automatic segmentation tools, ensuring their validity and utility in a broader spectrum of clinical scenarios.

Our study employed a range of loss functions sufficiently representative of the different types typically used in this field. However, future lines of work will include:

1. **Incorporation of novel loss functions**: Implement area, distance, and entropy-based loss functions to advance clinical data analysis.
2. **3D medical image segmentation**: Apply advanced loss functions to improve 3D medical image segmentation precision.
3. **Expansion to various medical imaging modalities**: Extend research to diverse medical imaging modalities beyond CT scans, such as MRI or PET.
4. **Integration with the MONAI framework [22]**: Integrate solution with MONAI to enhance scalability and efficiency in clinical settings.
5. **Collaboration with clinicians**: Partner with clinicians to assess clinical impact of segmentation improvements.
6. **Architectural design exploration**: Investigate emerging architectural designs' interaction with loss functions in COVID-19 lung infection segmentation.

Through these future directions, we aim to further advance the field of medical image analysis, improve clinical outcomes, and contribute to the ongoing development of digital and personalized medicine.

AUTHOR CONTRIBUTIONS

Emilio Delgado: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; writing—original draft. **Roberto Rodriguez-Echeverria**: Writing—review and editing. **Antonio Jesús Fernández-García**: Writing—review and editing. **Juan D. Gutiérrez**: Writing—review and editing. **Miguel Ángel Suero-Rodrigo**: Writing—review and editing.

ACKNOWLEDGEMENTS

This work was supported by Grant CPP2021-008491 funded by MICIU/AEI/10.13039/50100011033 and by the European Union Next Generation EU/PRTR.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in COVID-19 CT Lung and Infection Segmentation Dataset at https://zenodo.org/records/3757476.

ORCID

Emilio Delgado https://orcid.org/0000-0002-9144-0698

REFERENCES

1. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, pp. 234–241. Springer, Cham (2015)
2. Koudia, A., Chouaba, S.E., Belkhiat, D.E.C.: Implementation and comparison of U-net networks for automatic COVID-19 lung infection segmentation. In: 2022 19th IEEE International Multi-Conference on Systems, Signals and Devices, SSD 2022, pp. 1067–1072. IEEE, Piscataway, NJ (2022)
3. Saood, A., Hatem, I.: Covid-19 lung CT image segmentation using deep learning methods: U-net versus SegNet. BMC Med. Imaging 21(1), 1–10 (2021)
4. Enshaei, N., Afshar, P., Heidarian, S., Mohammadi, A., Rafiee, M.J., Oikonomou, A., Fard, F.B., Plataniotis, K.N., Naderkhani, F.: An ensemble learning framework for multi-class COVID-19 lesion segmentation from chest CT images. In: 2021 IEEE International Conference on Autonomous Systems (ICAS), pp. 1–6. IEEE, Piscataway, NJ (2021)
5. Qiu, Y., Liu, Y., Li, S., Xu, J.: MiniSeg: an extremely minimum network based on lightweight multiscale learning for efficient COVID-19 segmentation. IEEE Trans. Neural Networks Learn. Syst. 35(6), 8570–8584 (2022). https://doi.org/10.1109/TNNLS.2022.3230821
6. Raj, A.N.J., Zhu, H., Khan, A., Zhuang, Z., Yang, Z., Mahesh, G.V., Karthik, G.: ADID-UNET - a segmentation model for COVID-19 infection from lung ct scans. PeerJ Comput. Sci. 7, e349 (2021). https://doi.org/10.7717/PEERJ-CS.349/SUPP-2
7. Chi, J., Zhang, S., Han, X., Wang, H., Wu, C., Yu, X.: MID-UNet: Multi-input directional UNet for COVID-19 lung infection segmentation from CT images. Signal Process. Image Commun. 108, 116835 (2022). https://doi.org/10.1016/J.IMAGE.2022.116835
8. Hu, J., Yang, Y., Guo, X., Ma, T.: DECOR-Net: a COVID-19 lung infection segmentation network improved by emphasizing low-level features and decorrelating features. arXiv:2302.14277 (2023)
9. Abdel-Basset, M., Hawash, H., Chang, V.I.: FV-Seg-Net: fully volumetric network for accurate segmentation of COVID-19 lesions from chest CT scans. IEEE Trans. Ind. Inf. 19, 3321–3330 (2023). https://doi.org/10.1109/TII.2022.3146175

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10. Zhang, J., Xiang, K., Wang, J., Liu, J., Kang, M., Pan, Z.: Trans-Inf-Net: COVID-19 lung infection segmentation based on transformer. In: International Conference on Virtual Rehabilitation, ICVR 2022, pp. 306–312. IEEE, Piscataway, NJ (2022)
11. Peng, Y., Zhang, T., Guo, Y.: Cov-TransNet: dual branch fusion network with transformer for COVID-19 infection segmentation. Biomed. Signal Process. Control 80, 104366 (2023). https://doi.org/10.1016/j.bspc.2022.104366
12. Fan, X., Feng, X.: SELDNet: sequenced encoder and lightweight decoder network for COVID-19 infection region segmentation. Displays 77, 102395 (2023). https://doi.org/10.1016/J.DISPLA.2023.102395
13. Yu, H., Liu, S., Zhang, C., Li, S., Ren, J., Zhang, J., Xu, W.: Computed tomography and pathology evaluation of lung ground-glass opacity. Exp. Ther. Med. 16, 5305–5309 (2018). https://doi.org/10.3892/etm.2018.6886
14. Oh, J.Y., Kwon, S.Y., Yoon, H.I., Lee, S.M., Yim, J.J., Lee, J.H., Yoo, C.G., Kim, Y.W., Han, S.K., Shim, Y.S., Kim, T.J., Lee, K.W., Chung, J.H., Jheon, S.H., Sung, S.W., Lee, C.T.: Clinical significance of a solitary ground-glass opacity (GGO) lesion of the lung detected by chest CT. Lung Cancer 55(1), 67–73 (2007). https://doi.org/10.1016/j.lungcan.2006.09.009
15. Jun, M., Cheng, G., Yixin, W., Xingle, A., Jiantao, G., Ziqi, Y., Minqing, Z., Xin, L., Xueyuan, D., Shucheng, C., et al.: COVID-19 ct lung and infection segmentation dataset, Zenodo. https://doi.org/10.5281/zenodo.3757476 (2020). Accessed 21 Sep 2022
16. Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., Cao, T., Zhu, Y., Nie, Z., Yang, X.: Towards data-efficient learning: a benchmark for COVID-19 CT lung and infection segmentation. Med. Phys. 48(3), 1197–1210 (2021). https://doi.org/10.1002/mp.14676
17. Kamalian, S., Lev, M.H., Gupta, R.: Computed tomography imaging and angiography - principles. Handbook Clin. Neurol. 135, 3–20 (2016). https://doi.org/10.1016/B978-0-444-53485-9.00001-5
18. Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M.D., Buettner, F., Christodoulou, E., Glocker, B., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M.A., Wiesenfarth, M., Emre Kavur, A., Sudre, C.H., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Rädsch, T., Acion, L., Antonelli, M., Arbel, T., Bakas, S., Benis, A., Blaschko, M.B., Jorge Cardoso, M., Cheplygina, V., Cimini, B.A., Collins, G.S., Farahani, K., Ferrer, L., Galdran, A., vanGinneken, B., Haase, R., Hashimoto, D.A., Hoffman, M.M., Huisman, M., Jannin, P., Kahn, C.E., Kainmueller, D., Kainz, B., Karargyris, A., Karthikesalingam, A., Kofler, F., Kopp-Schneider, A., Kreshuk, A., Kurc, T., Landman, B.A., Litjens, G., Madani, A., Maier-Hein, K., Martel, A.L., Mattson, P., Meijering, E., Menze, B., Moons, K.G.M., Müller, H., Nichyporuk, B., Nickel, F., Petersen, J., Rajpoot, N., Rieke, N., Saez-Rodriguez, J., Sá¡nchez, C.I., Shetty, S., van Smeden, M., Summers, R.M., Taha, A.A., Tiulpin, A., Tsaftaris, S.A., Van Calster, B., Varoquaux, G., Jäger, P.F.: Metrics reloaded: recommendations for image analysis validation. Nat. Methods. 21(2), 195–212 (2024)
19. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L.: Loss odyssey in medical image segmentation. Med. Image Anal. 71, 102035 (2021). https://doi.org/10.1016/J.MEDIA.2021.102035
20. Yeung, M., Sala, E., Schönlieb, C.B., Rundo, L.: Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Comput. Med. Imaging Graphics 95, 102026 (2022). https://doi.org/10.1016/j.compmedimag.2021.102026
21. Müller, D., Soto Rey, I., Kramer, F.: Towards a guideline for evaluation metrics in medical image segmentation. BMC Res. Notes 15, 210 (2022). https://doi.org/10.1186/S13104-022-06096-Y/FIGURES/2
22. Jorge Cardoso, D.P.: MONAI: an open-source framework for deep learning in healthcare. arXiv:2211.02701 (2022)

How to cite this article: Delgado, E., Rodriguez-Echeverria, R., Fernández-García, A.J., Gutiérrez, J.D., Suero-Rodrigo, M.Á.: Advancing precision in medical image segmentation: A performance analysis of loss functions for COVID-19 lung infection segmentation in computed tomography images. IET Image Process. 18, 4047–4065 (2024). https://doi.org/10.1049/ipr2.13232

17519667, 2024, 13, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/ipr2.13232 by Universidad De Extremadura, Wiley Online Library on [22/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License