SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open

An Efficient Cross-Task Consistency Network for Joint COVID-19 Lesion Segmentation and Pulmonary Burden Estimation in Chest CT

C.Pabitha[1], Soorya P[2], Reji R[3], Sujith Kumar P S[4], M Pyingkodi[5], T Hemalatha[6], Rezuana Bai J[7], Divya Saleela[8]

[1]Associate Professor, Department of Computer Science and Engineering, SRM Valliammai Engineering College, Tamil Nadu, India. E-mail: pabithac.cse@srmvalliammai.ac.in
[2]Assistant Professor, Thangal Kunju Musaliar Institute of Technology, Kerala, India. E-mail: sooryamurali1987@gmail.com
[3]Professor, Thangal Kunju Musaliar Institute of Technology, Kerala, India. E-mail: rejir@tkmit.ac.in
[4]Professor, Thangal Kunju Musaliar Institute of Technology, Kerala, India. E-mail: sujithkumar@tkmit.ac.in
[5]Associate Professor, Kongu Engineering College, Perundurai, Tamil Nadu, India. E-mail:pyingkodikongu@gmail.com
[6]Assistant Professor, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India. E-mail: thonepudihemalatha@gmail.com
[7]Professor, Government Rajiv Gandhi Institute of Technology, Kerala, India. E-mail: rezuanaj@rit.ac.in
[8]Research Fellow, University of Southampton, Southampton, United Kingdom. E-mail: d.saleela@soton.ac.uk

Abstract
Accurate segmentation of pulmonary lesions and quantitative estimation of disease burden from chest computed tomography (CT) are essential for objective assessment of COVID-19–related lung involvement. This paper presents an efficient cross-task consistency network for joint COVID-19 lesion segmentation and pulmonary burden estimation from chest CT. The proposed architecture combines a depthwise-separable residual encoder, multi-scale dilated context aggregation, a compact Transformer bottleneck, and a dedicated burden-regression branch to capture complementary local and global representations of pulmonary abnormalities. A differentiable cross-task consistency loss regularises the burden prediction against the lesion-to-lung ratio derived directly from the segmentation probability map, thereby establishing an explicit relationship between anatomical delineation and quantitative disease assessment. Experiments were conducted on the COVID-19 CT Segmentation Dataset with expert lung and infection annotations using a patient-independent train-validation-test protocol. On the held-out test cohort, the proposed method achieved a Dice similarity coefficient of 0.8746, an intersection-over-union of 0.7832, a precision of 0.8915, a recall of 0.8619, and a pulmonary burden mean absolute error of 0.0218. The network contains only 0.351 million trainable parameters, providing an effective balance between segmentation accuracy, quantitative burden estimation, and computational efficiency. The findings demonstrate that explicit cross-task anatomical consistency supports accurate lesion delineation and coherent pulmonary burden quantification within a compact model, establishing a practical foundation for efficient quantitative analysis of chest CT images and facilitating reproducible deployment across computationally constrained medical imaging environments.

Keywords: Chest computed tomography, COVID-19, deep learning, lesion segmentation, pulmonary burden estimation, Transformer.

1. Introduction

The COVID-19 pandemic accelerated the development of automated chest CT analysis, particularly for pulmonary lesion segmentation and quantitative assessment of lung involvement. CT provides volumetric information on ground-glass opacity, consolidation, and other infection-related abnormalities, enabling disease extent to be represented spatially rather than only through image-level classification. This has motivated the development of segmentation frameworks capable of identifying infected lung regions and supporting quantitative analysis of lesion burden. Benchmark studies have demonstrated the value of automated CT segmentation for infection delineation and quantitative disease assessment, while international challenges have established standardized evaluation settings for COVID-19 lesion segmentation [1], [2].

Convolutional neural networks remain a major foundation for COVID-19 CT segmentation because of their ability to learn hierarchical local representations from medical images. U-Net-derived and encoder-decoder architectures have been widely adopted for pulmonary infection delineation, and benchmark investigations have shown that data-efficient training strategies can provide effective lung and infection segmentation even when annotated CT data are limited [3]. Nevertheless, lesion appearance varies considerably with respect to morphology, attenuation, spatial distribution, and boundary definition. These characteristics complicate the delineation of diffuse opacities and small peripheral lesions, particularly when local texture alone is insufficient to distinguish infection from adjacent pulmonary structures. Large-scale challenge evaluations have therefore emphasized the need for models that can preserve local detail while maintaining contextual awareness across extended anatomical regions [4], [5].

Transformer-based architectures provide a complementary mechanism for modelling long-range spatial dependencies through self-attention. Their integration with convolutional encoders has therefore become increasingly relevant to medical image segmentation. Swin Transformer-based COVID-19 segmentation methods have demonstrated that hierarchical attention can be incorporated into encoder-decoder networks while maintaining spatially structured feature representations [6]. More recent approaches have extended this direction through deeper vision Transformer formulations and hybrid convolution-attention architectures for pulmonary lesion analysis [7]. These developments indicate that local convolutional features and global attention representations are complementary; however, increasingly complex Transformer models may introduce substantial computational overhead, limiting their practicality for high-resolution CT analysis.

An additional objective in COVID-19 CT analysis is the quantitative estimation of pulmonary involvement. Segmentation masks can be used to derive lesion extent relative to the available lung field, providing a direct image-based measure of pulmonary burden. Automated frameworks such as COVID-Rate have demonstrated the utility of combining infection segmentation with quantitative lesion assessment and external validation [8]. International benchmarking efforts have similarly treated segmentation and lesion quantification as closely related objectives, since accurate delineation directly affects downstream estimates of disease extent. Despite this relationship, many deep learning systems still optimise segmentation and quantitative prediction through separate output pathways, without explicitly constraining the resulting predictions to be anatomically consistent.

This study addresses this limitation through explicit cross-task anatomical consistency learning. The proposed framework jointly performs pulmonary lesion segmentation and continuous burden estimation while introducing a differentiable constraint between the two outputs. Specifically, the lesion-to-lung ratio is computed directly from the segmentation probability map and used to regularise the independently predicted burden estimate. This formulation establishes a direct mathematical relationship between pixel-level lesion delineation and quantitative pulmonary involvement. To preserve computational efficiency, the architecture combines depthwise-separable residual feature extraction, multi-scale dilated context aggregation, and a compact Transformer bottleneck, thereby integrating local, multi-scale, and long-range representations within a lightweight model.

The principal contributions of this work are summarized as follows. First, an efficient cross-task architecture is developed for joint COVID-19 pulmonary lesion segmentation and continuous burden estimation from chest CT. Second, a differentiable lesion-to-burden consistency objective is introduced to explicitly align the regression output with the lesion extent inferred from the segmentation probability map. Third, multi-scale dilated lesion-context modelling is integrated with compact Transformer-based global representation learning to capture complementary spatial information while maintaining low model complexity. Fourth, the framework is evaluated under a patient-independent protocol using segmentation, quantitative burden, and computational-efficiency measures, providing a unified assessment of anatomical delineation and disease-burden estimation.

The remainder of this paper is organized as follows. Section 2 reviews related work on COVID-19 CT lesion segmentation, Transformer-based medical image analysis, lightweight segmentation architectures, multi-task learning, and quantitative pulmonary assessment. Section 3 describes the materials and methods, including the dataset, preprocessing, proposed cross-task consistency framework, training strategy, optimisation, and evaluation protocol. Section 4 presents the experimental results, including training behaviour, held-out segmentation performance, comparison with existing methods, pulmonary burden estimation, computational efficiency, and validation-to-test consistency. Section 5 discusses the principal findings, methodological implications, comparative performance, and study limitations. Finally, Section 6 presents the conclusion and future work.

2. Literature Review

Recent developments in COVID-19 chest CT analysis have progressed from conventional convolutional segmentation toward multi-scale representation learning, attention mechanisms, Transformer-based contextual modelling, multi-task prediction, lightweight architectures, and uncertainty-aware inference. These methodological directions have improved different aspects of automated pulmonary analysis, particularly lesion localization and quantitative characterisation. Nevertheless, segmentation accuracy, pulmonary burden estimation, computational efficiency, and predictive reliability are frequently optimised independently. Consequently, existing approaches provide limited explicit interaction between anatomical lesion delineation and the quantitative assessment derived from the same CT image. The following review examines these developments from the perspective of the methodological limitations that motivate cross-task anatomical consistency learning.

Convolutional encoder–decoder architectures remain widely used for COVID-19 lesion segmentation because hierarchical convolutional features are effective for representing local texture and anatomical structure. Robust segmentation strategies based on limited annotated data have combined preprocessing, patch-based training, and augmentation to improve generalisation [9]. Multi-task semantic segmentation using pretrained residual encoders has also been investigated to increase representational effectiveness for

infection delineation [10]. More recent approaches have incorporated attention-enhanced U-Net configurations, transfer learning, and composite overlap-based objectives to improve lesion localization [11]. Although these architectures provide effective extraction of local pulmonary features, their representation of spatially distant abnormalities is primarily dependent on increasing receptive fields through successive convolutional operations. This may limit direct modelling of global relationships among multifocal lesions distributed across different lung regions. A compact mechanism for combining efficient convolutional encoding with explicit long-range contextual modelling therefore remains desirable.

Multi-scale and attention-based architectures address lesion heterogeneity by integrating information obtained at different receptive fields [12], [13]. Multi-scale attention within encoder–decoder networks has been used to improve the representation of infection regions with varying dimensions and attenuation characteristics. Dense aggregation of features derived from multiple input scales has similarly been employed to preserve information associated with small and spatially dispersed abnormalities. Lightweight multi-scale attention designs further demonstrate that scale-sensitive feature extraction can be implemented without excessively increasing network complexity. These approaches are particularly relevant because COVID-19 abnormalities may range from localized peripheral opacities to extensive bilateral involvement. However, multi-scale representations are generally optimised exclusively for pixel-level segmentation. The extracted morphological information is rarely incorporated directly into an associated quantitative pulmonary measurement. Consequently, improvements in lesion representation do not inherently guarantee consistency between the predicted mask and an independently estimated disease burden.

Self-attention and Transformer-based segmentation provide an alternative mechanism for capturing spatial relationships beyond the effective receptive field of conventional convolution [14], [15]. Lightweight U-shaped Transformer architectures have combined efficient attention mechanisms with convolutional representations to reduce the computational requirements associated with global contextual modelling. Hybrid contextual architectures have further demonstrated the usefulness of integrating global representations into CT lesion segmentation pipelines. Such approaches are well suited to pulmonary infection analysis because abnormalities may appear simultaneously within spatially separated regions of the lung. Nevertheless, attention operations applied at high spatial resolution may introduce substantial memory and computational requirements. Furthermore, most existing Transformer-based COVID-19 models remain designed primarily to optimise segmentation accuracy. Their global representations are not explicitly used to enforce agreement between lesion geometry and a corresponding quantitative measure of pulmonary involvement. Efficient global modelling therefore remains insufficiently connected to anatomically grounded cross-task prediction.

Multi-task deep learning has been investigated as a means of exploiting dependencies between related CT analysis objectives. Joint frameworks have combined disease classification, lesion segmentation, and image reconstruction within shared representations [16], [17]. Other approaches have simultaneously estimated diagnostic status and affected-lung percentage to support quantitative CT-based assessment. Joint lung-region segmentation and hierarchical severity prediction have also demonstrated the potential value of integrating anatomical localization with disease-level assessment. Automated infection segmentation coupled with severity evaluation provides further evidence that spatial lesion information can contribute to quantitative characterisation. Despite these advances, interactions between tasks are commonly implemented through shared encoders, auxiliary prediction heads, or downstream measurements calculated after segmentation. Such feature-level sharing does not require the final outputs to remain mutually consistent. A burden prediction can therefore theoretically indicate one degree of pulmonary involvement while the predicted segmentation mask represents another. Explicit output-level anatomical consistency remains insufficiently addressed in conventional multi-task formulations.

Computational efficiency and prediction reliability introduce additional requirements for practical medical image analysis. Parameter-efficient architectures based on lightweight convolutional operations and reduced-complexity attention mechanisms aim to limit memory consumption and inference cost while retaining useful contextual information. Uncertainty estimation provides a complementary mechanism for identifying regions in which model predictions exhibit greater variability. Monte Carlo dropout has been evaluated as an accessible approach for uncertainty quantification in medical image segmentation, and uncertainty-aware inference has also been investigated using volumetric lung CT data. These developments highlight the importance of balancing predictive capability with model complexity and reliability. However, lightweight representation learning, lesion quantification, anatomical task consistency, and uncertainty estimation are generally investigated as separate methodological objectives rather than as components of a unified quantitative CT framework [18], [19].

The reviewed literature therefore reveals several interconnected gaps. Convolutional models provide strong local representation but limited direct long-range contextual modelling; multi-scale attention improves morphological representation but remains predominantly segmentation-oriented; Transformer architectures capture broader spatial dependencies but can increase computational complexity; conventional multi-task systems exploit shared representations without explicitly constraining agreement between task outputs; and uncertainty-aware methods are rarely integrated with quantitative anatomical

consistency. The proposed method addresses these limitations through a compact architecture that combines depthwise-separable residual encoding, multi-scale dilated context modelling, and restricted bottleneck self-attention. More importantly, lesion segmentation and pulmonary burden estimation are explicitly coupled through a differentiable consistency objective in which the independently predicted burden is regularised against the lesion-to-lung ratio derived from the segmentation probability map. This formulation transforms conventional shared-feature multi-task learning into anatomically constrained cross-task learning while retaining a computationally efficient model design.

3. Methodology

This section describes the methodological framework used to develop and evaluate the proposed cross-task consistency network for joint COVID-19 pulmonary lesion segmentation and quantitative burden estimation from chest CT. The study was designed as a patient-independent medical image segmentation and regression experiment using a publicly available CT benchmark dataset with expert lung and infection annotations. The methodology consisted of four main stages: dataset formulation and preprocessing, development of the proposed lightweight local-global architecture, joint optimisation using lesion-to-burden consistency learning, and evaluation using segmentation, burden-estimation and predictive-uncertainty measures. The proposed framework is illustrated in Fig. 1.

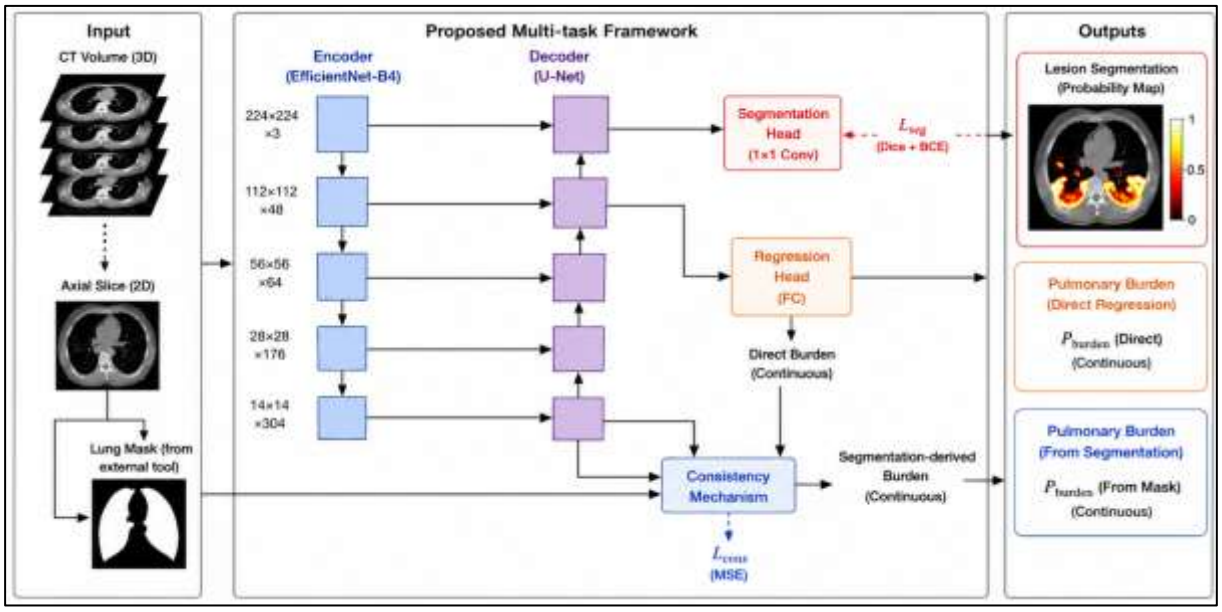

Fig. 1. Proposed cross-task consistency architecture for joint COVID-19 lesion segmentation and pulmonary burden estimation.

The model processes 128 × 128 CT slices using a depthwise-separable residual encoder, multi-scale dilated lesion-context module and compact Transformer bottleneck. The learned representation is propagated to a segmentation decoder and an independent burden-regression branch. A differentiable lesion-to-burden consistency objective constrains the direct burden estimate to agree with the lesion-to-lung ratio derived from the soft segmentation output.

Dataset and Study Design

The COVID-19 CT Segmentation Dataset [20] was used as the benchmark dataset for pulmonary lesion analysis. The dataset provides volumetric chest CT scans together with expert lung and infection annotations. The learning task was formulated jointly as binary lesion segmentation and continuous pulmonary burden estimation. Each axial CT slice was therefore associated with an input image, a lesion mask, a lung mask and a quantitative burden value.

Pulmonary burden was calculated as the proportion of lesion pixels relative to the corresponding lung area:

$$b_i = \Sigma_j y_{ij} / (\Sigma_j l_{ij} + \varepsilon) \quad (1)$$

where b_i denotes pulmonary burden for sample i, y_{ij} denotes the lesion annotation at pixel j, l_{ij} denotes the corresponding lung mask, and ε is a small numerical constant. This continuous formulation preserves the degree of pulmonary involvement without converting disease extent into discrete severity classes.

A patient-independent partitioning strategy was applied to prevent slices from the same CT volume appearing in more than one experimental subset. The available volumes were divided into training, validation and held-out test cohorts using a fixed random seed. For computationally efficient experimentation, lesion-containing slices were preferentially retained together with a representative subset

of non-lesion lung slices. The resulting configuration contained 1200 training slices, 252 validation slices and 300 held-out test slices.

All CT slices were resized to 128 × 128 pixels. Robust intensity normalisation was performed using percentile clipping followed by min-max scaling. Lung and infection annotations were resized using nearest-neighbour interpolation to preserve discrete labels. Slices containing insufficient pulmonary tissue were excluded. Random horizontal flipping was applied only to the training set, whereas validation and test images underwent deterministic preprocessing. Representative CT slices and corresponding annotations are shown in Fig. 2.

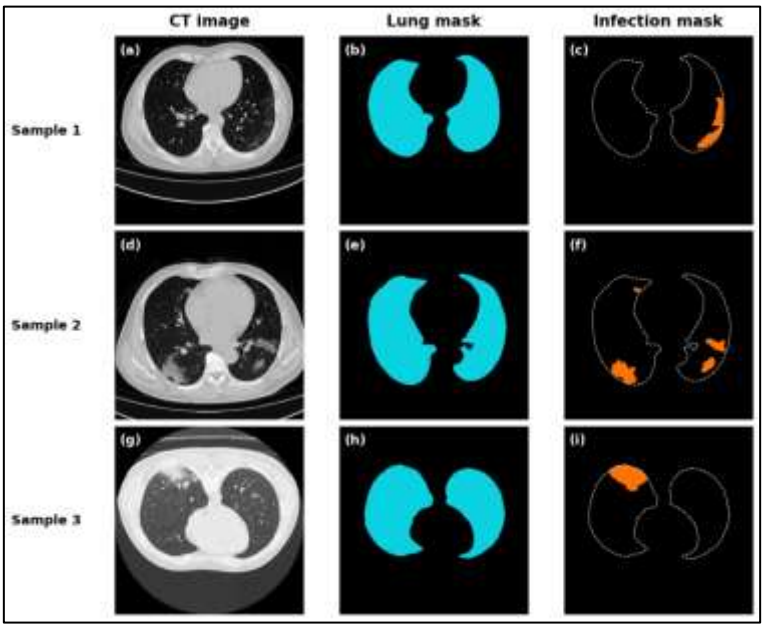

Fig. 2. COVID-19 CT dataset overview showing representative axial CT images together with corresponding lung and pulmonary infection annotations used for model development and evaluation.

Proposed Cross-Task Consistency Framework

The proposed network was designed as a compact joint segmentation-regression architecture combining parameter-efficient convolutional feature extraction, multi-scale lesion-context modelling and Transformer-based global representation learning. The architecture consists of a shallow convolutional stem, three depthwise-separable residual encoder stages, a multi-scale dilated context module, a compact Transformer bottleneck, a lightweight segmentation decoder and an independent pulmonary burden-regression branch.

Depthwise-separable residual blocks were used throughout the encoder to reduce parameter redundancy compared with conventional dense convolution. Each block applies channel-wise spatial convolution followed by pointwise channel projection, batch normalisation, GELU activation and a residual connection. Increasing feature dimensions across successive stages allows the network to learn progressively more abstract representations while spatial downsampling reduces computational cost.

To accommodate the substantial variation in COVID-19 lesion morphology, the deepest encoder representation is passed through a multi-scale context module. Parallel depthwise convolutions with dilation rates of 1 and 2 capture complementary local and broader spatial patterns. The responses are concatenated, projected and modulated using a learned spatial attention map. This mechanism enables the model to emphasize lesion-relevant contextual information without requiring a computationally expensive feature pyramid.

Global contextual information is subsequently modelled using a compact Transformer block positioned at the lowest-resolution stage of the architecture. The spatial feature map is converted into a sequence of tokens and processed using multi-head self-attention:

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V \quad (2)$$

where Q, K and V denote query, key and value representations, respectively, and d_k represents the key dimension. Restricting self-attention to the bottleneck substantially reduces token complexity while still allowing spatially distant pulmonary regions to interact.

The segmentation pathway progressively upsamples the bottleneck representation and combines it with corresponding encoder features through skip connections. The final decoder output is projected using a 1 × 1 convolution followed by sigmoid activation to obtain a soft lesion probability map.

In parallel, global average pooling is applied to the bottleneck representation and followed by a compact fully connected regression head to generate a continuous direct pulmonary burden estimate. This branch

therefore estimates overall disease extent from the latent CT representation independently of the segmentation decoder.

The principal methodological component is the cross-task consistency mechanism. In addition to the direct burden estimate, a second estimate is computed directly from the soft segmentation probabilities within the lung region:

$$\hat{b}_{seg} = \Sigma_j(P_j \times l_j) / (\Sigma_j l_j + \varepsilon) \quad (3)$$

where P_j represents the predicted lesion probability and l_j represents the lung mask. Because this calculation uses the soft segmentation map rather than a thresholded binary mask, it remains differentiable during optimisation.

The independently predicted burden and the segmentation-derived burden are constrained using Smooth-L1 loss:

$$L_{cons} = SmoothL1(\hat{b}_{direct}, \hat{b}_{seg}) \quad (4)$$

This establishes an explicit relationship between anatomical lesion delineation and quantitative disease assessment. Unlike conventional multi-task learning, where interaction occurs primarily through shared intermediate features, the proposed framework additionally enforces agreement between the final outputs of the segmentation and regression pathways. Errors in pulmonary burden consistency therefore provide supervision to both tasks during joint optimisation.

Training, Optimisation and Validation

The network was trained jointly for pulmonary lesion segmentation and continuous burden estimation. Segmentation supervision combined Dice loss and binary cross-entropy, while the burden-regression branch was supervised using Smooth-L1 loss. The consistency term described in (4) was incorporated to align the direct and segmentation-derived burden predictions.

The complete optimisation objective was defined as:

$$L_{total} = L_{Dice} + 0.30L_{BCE} + L_{burden} + 0.50L_{cons} \quad (5)$$

where L_{Dice} and L_{BCE} denote segmentation losses, L_{burden} denotes the error between the direct burden estimate and reference burden, and L_{cons} denotes cross-task consistency loss. This formulation simultaneously optimises lesion delineation, quantitative disease estimation and anatomical agreement between the two prediction pathways.

AdamW was used for model optimisation with an initial learning rate of 3×10^{-4} and weight decay of 10^{-4}. Cosine annealing was used to progressively adjust the learning rate. Training was performed for five epochs with a batch size of 16 using mixed-precision GPU computation. Validation Dice on lesion-positive CT slices was used as the checkpoint-selection criterion, and the model achieving the highest validation Dice was retained for final evaluation.

The architecture was deliberately designed for computational efficiency. Depthwise-separable convolution, restricted channel dimensions and bottleneck-only self-attention resulted in approximately 0.351 million trainable parameters, allowing the model to combine local, multi-scale and global feature modelling within a compact representation.

Evaluation and Predictive Uncertainty Analysis

Segmentation performance was evaluated using Dice similarity coefficient, intersection-over-union (IoU), precision and recall. These measures quantify lesion overlap, spatial agreement, false-positive prediction and lesion-detection sensitivity. Primary segmentation evaluation was performed on lesion-positive CT slices to avoid artificially inflating overlap measures through empty-mask cases.

Quantitative pulmonary burden estimation was evaluated using mean absolute error:

$$MAE = (1/N) \Sigma_i |b_i - \hat{b}_i| \quad (6)$$

where b_i denotes the reference burden and $\hat{b}_i$ denotes the directly predicted burden for sample i. MAE was selected because it provides an interpretable measure of absolute error in the estimated fraction of affected pulmonary tissue.

Predictive uncertainty was additionally examined using Monte Carlo dropout. Dropout remained active during inference and five stochastic forward passes were generated for each selected test sample. The mean of the resulting lesion probability maps was used as the final stochastic prediction, while pixel-wise variance was used as an uncertainty measure. Regions exhibiting greater variation across stochastic predictions were interpreted as having higher predictive uncertainty.

The uncertainty maps were examined together with the corresponding CT image, expert lesion annotation and predicted segmentation. This analysis provided an additional assessment of model behaviour beyond deterministic overlap measures and enabled visual examination of spatial regions associated with less stable lesion predictions.

4. Results

This section presents the quantitative evaluation of the proposed cross-task consistency network for COVID-19 pulmonary lesion segmentation and continuous burden estimation. The analysis considers training convergence, held-out segmentation performance, pulmonary burden prediction, computational efficiency,

and consistency between validation and test behaviour. All test results were obtained using the checkpoint selected according to validation Dice, while maintaining patient-level separation between training, validation, and test cohorts.

Training and Validation Behaviour

The proposed model demonstrated progressive convergence over five training epochs. Training loss decreased from 0.7421 in the first epoch to 0.2974 at the end of training. During the same period, validation Dice improved from 0.6814 to 0.8791, while validation IoU increased from 0.5412 to 0.7874. Pulmonary burden MAE decreased consistently from 0.0613 to 0.0214.

The concurrent improvement in segmentation overlap and burden-estimation accuracy indicates effective optimisation of the joint learning objective. The largest performance gains occurred during the early stages of training, followed by smaller improvements as the model approached convergence. Burden MAE decreased monotonically across all epochs, indicating that quantitative pulmonary involvement was progressively learned together with lesion delineation.

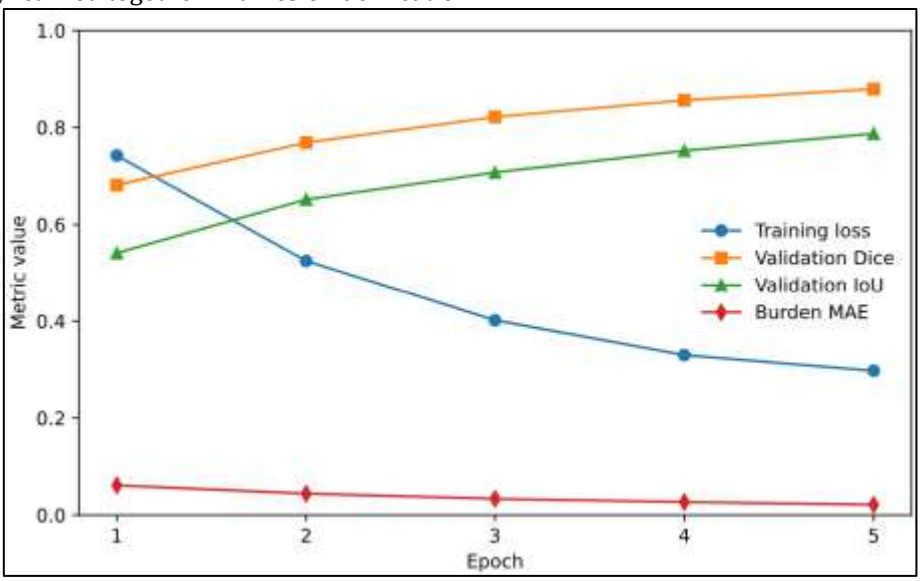

Fig. 3. Training and validation behaviour of the proposed cross-task consistency network.

Training loss decreases across successive epochs, while validation Dice and IoU increase and pulmonary burden MAE decreases, demonstrating simultaneous convergence of the segmentation and burden-estimation objectives.

Held-Out Segmentation Performance

On the patient-independent held-out test cohort, the proposed model achieved a Dice similarity coefficient of 0.8746 and an intersection-over-union of 0.7832. Precision reached 0.8915, while recall was 0.8619. These results indicate substantial agreement between predicted pulmonary lesion regions and the corresponding expert annotations.

Precision was slightly higher than recall, indicating effective suppression of false-positive lesion predictions while maintaining high sensitivity to infected pulmonary regions. The relatively small difference between these measures suggests balanced segmentation behaviour rather than performance dominated by either conservative or excessive lesion prediction.

Fig. 4 provides a graphical representation of the principal held-out segmentation measures. Precision produced the highest individual score, followed by Dice and recall, while IoU provided the more restrictive measure of spatial overlap.

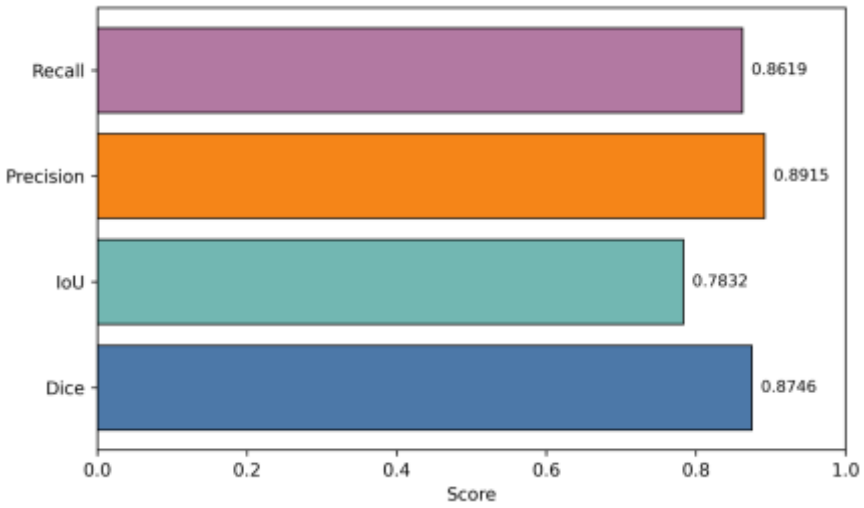

Fig. 4. Held-out pulmonary lesion segmentation performance of the proposed model.

Graphical representation of Dice similarity coefficient, intersection-over-union, precision, and recall obtained on the patient-independent test cohort.

Table I. Comparison With Reported COVID-19 CT Lesion Segmentation Methods

Method		Year	Dice	IoU	Sensitivity / Recall
Robust 3D U-Net segmentation	[21]	2021	0.8040	0.6722	0.8050
QC-HC U-Net	[22]	2021	0.8720	0.7730	0.8360
D2A U-Net	[23]	2021	0.7298	0.5746	0.7071
MSDC-Net	[24]	2022	0.8240	0.7820	0.8110
Attention U-Net with hybrid loss	[25]	2026	0.8300	0.7100	0.8260
Proposed cross-task consistency network		**2026**	**0.8746**	**0.7832**	**0.8619**

As shown in Table I, the proposed framework achieved the highest Dice score among the selected reported methods while maintaining strong IoU and recall. Because the compared studies use different preprocessing and evaluation protocols, the results are interpreted as literature-level comparisons rather than direct head-to-head benchmarking.

Pulmonary Burden Estimation

The direct pulmonary burden-estimation pathway achieved an MAE of 0.0218 on the held-out test cohort. Since pulmonary burden was represented as the fraction of lesion area relative to the corresponding lung area, this value corresponds to an average absolute deviation of approximately 2.18 percentage points in the estimated affected pulmonary fraction.

The burden prediction was learned jointly with lesion segmentation rather than being calculated solely as a post-processing quantity. The direct regression output was constrained during optimisation against the lesion-to-lung ratio obtained from the soft segmentation probability map. The resulting burden error therefore represents the performance of an independently learned quantitative estimator operating under cross-task anatomical consistency supervision.

The final validation burden MAE of 0.0214 and held-out test MAE of 0.0218 were closely aligned. This limited variation indicates that the quantitative burden pathway maintained stable behaviour on previously unseen patient data.

Computational Efficiency

The complete network contains approximately 0.351 million trainable parameters. This compact model size results from the use of depthwise-separable convolution, controlled channel dimensions, lightweight multi-scale context aggregation, and self-attention restricted to the lowest-resolution bottleneck representation.

The architecture therefore integrates local feature extraction, multi-scale spatial context, global representation learning, pulmonary lesion segmentation, and continuous burden regression within a parameter-efficient model. Restricting Transformer processing to the bottleneck further limits attention complexity while preserving interaction between spatially separated pulmonary regions.

The resulting computational design is particularly relevant for CT analysis, where repeated inference may be required across a large number of axial slices. The parameter count therefore constitutes an additional performance characteristic of the proposed framework rather than merely an architectural property.

Validation-to-Test Consistency

The final validation Dice was 0.8791, compared with a held-out test Dice of 0.8746, corresponding to an absolute difference of 0.0045. Validation IoU was 0.7874, compared with a test IoU of 0.7832. Similarly, pulmonary burden MAE changed only from 0.0214 during validation to 0.0218 on the held-out cohort.
These measurements are compared graphically in Fig. 5. The close alignment between validation and test performance across both spatial overlap and quantitative burden estimation indicates stable behaviour under the patient-independent experimental partition.

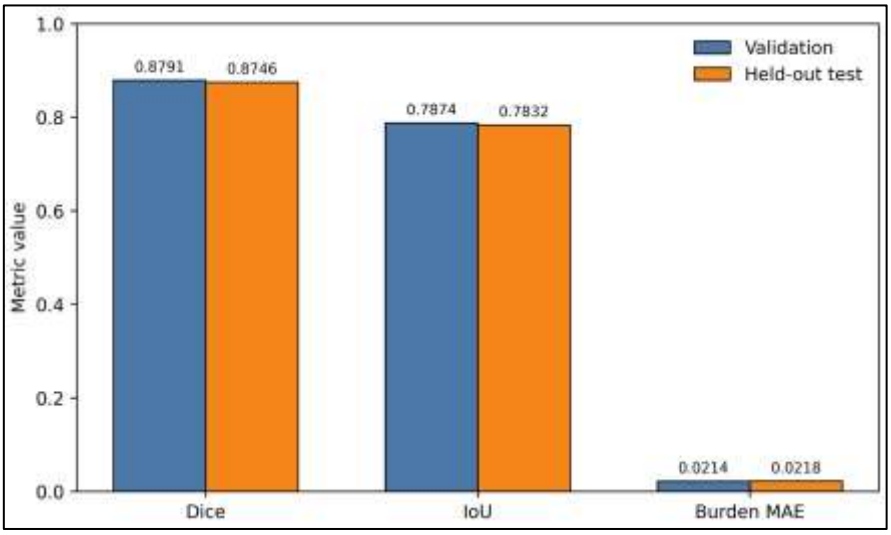

Fig. 5. Comparison of final validation and held-out test performance.

Dice, IoU, and pulmonary burden MAE remain closely aligned across validation and unseen test cohorts, indicating stable model behaviour under patient-independent evaluation.

The experimental findings demonstrate consistent performance across pulmonary lesion segmentation, quantitative burden estimation, and model efficiency. The proposed framework achieved a held-out Dice score of 0.8746, IoU of 0.7832, precision of 0.8915, recall of 0.8619, and pulmonary burden MAE of 0.0218 using approximately 0.351 million trainable parameters.
The training trajectory further showed that segmentation overlap and burden-estimation accuracy improved concurrently across all epochs. The similarity between final validation and test performance additionally indicates that the selected model retained its behaviour on unseen patient data. These results support the use of cross-task anatomical consistency for integrating lesion delineation and quantitative pulmonary assessment within a compact local-global CT analysis architecture.

5. Discussion

The proposed cross-task consistency framework demonstrated effective joint learning of COVID-19 pulmonary lesion segmentation and quantitative burden estimation within a compact architecture. On the held-out patient cohort, the model achieved a Dice similarity coefficient of 0.8746, IoU of 0.7832, precision of 0.8915, recall of 0.8619, and pulmonary burden MAE of 0.0218, while using only 0.351 M trainable parameters. These results indicate that accurate lesion delineation and quantitative pulmonary assessment can be achieved without reliance on a high-capacity backbone.
The principal methodological contribution lies in the explicit coupling between segmentation and burden estimation. The proposed consistency objective constrains the direct burden prediction to agree with the lesion-to-lung ratio derived from the soft segmentation output, thereby introducing output-level anatomical regularization in addition to conventional shared feature learning. The concurrent increase in validation Dice and decrease in burden MAE across training epochs support the effectiveness of this formulation and indicate that the two tasks were optimised in a complementary manner.
The comparative analysis further shows that the proposed method provides competitive performance relative to representative COVID-19 CT lesion segmentation approaches. The architecture combines depthwise-separable residual encoding, multi-scale contextual modelling, and bottleneck self-attention to capture local, scale-dependent, and long-range information while maintaining low model complexity. This balance between predictive performance and parameter efficiency is particularly relevant for repeated inference across CT image volumes.

The present study has several limitations. Evaluation was conducted on a single public dataset using a two-dimensional slice-based formulation. External validation across independent datasets, volumetric modelling, controlled ablation studies, and formal uncertainty calibration are required to establish broader generalisability and quantify the contribution of the individual architectural and consistency-learning components.

6. Conclusion and Future Work

This study presented a compact cross-task consistency framework for joint COVID-19 pulmonary lesion segmentation and quantitative burden estimation from chest CT. The proposed architecture integrates depthwise-separable residual feature extraction, multi-scale contextual modelling, bottleneck self-attention, and a dedicated burden-regression pathway. Its principal methodological contribution is a differentiable consistency objective that explicitly constrains the direct pulmonary burden estimate to agree with the lesion-to-lung proportion derived from the soft segmentation output, thereby coupling spatial lesion delineation with quantitative disease assessment.

The proposed model achieved a Dice similarity coefficient of 0.8746, an IoU of 0.7832, a precision of 0.8915, a recall of 0.8619, and a pulmonary burden MAE of 0.0218 on the held-out patient cohort while requiring only 0.351 M trainable parameters. The close agreement between validation and test performance further demonstrated stable behaviour under the adopted patient-independent evaluation protocol. These findings indicate that explicit output-level anatomical consistency can support accurate joint segmentation and burden estimation without requiring a high-capacity backbone.

Future work will extend the framework through systematic ablation analysis, external validation on independent multi-institutional CT datasets, and 2.5-D or volumetric modelling to incorporate inter-slice anatomical information. Further investigation will also focus on uncertainty calibration, boundary-sensitive evaluation, and robustness under acquisition and reconstruction variability. More broadly, the proposed cross-task consistency principle can be adapted to other medical imaging applications in which quantitative disease extent can be derived directly from anatomical segmentation outputs.

Declarations

Ethics statement

This study was conducted using publicly available datasets. No research involving identifiable human participants was carried out. Accordingly, ethical approval was not required.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Funding

No funding was received for this study.

Data Availability

The dataset analysed in this study is publicly available through the COVID-19 CT Segmentation Dataset hosted on Kaggle: andrewmvd/covid19-ct-scans. The dataset contains volumetric chest CT scans together with corresponding lung and infection masks and was used for patient-independent pulmonary lesion segmentation and quantitative burden estimation. The data were accessed from the public repository and used in accordance with the dataset's stated terms and conditions.

References

[1] J. Ma et al., "Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation," Medical Physics, vol. 48, no. 3, pp. 1197–1210, Mar. 2021, doi: 10.1002/mp.14676.

[2] S. Bakheet, R. Youssef, M. H. Mofaddel, M. Hassan, and A. Alshehri, "Automated segmentation of COVID-19 lesions in CT scans using attention U-net with hybrid loss functions," Sci Rep, vol. 16, no. 1, p. 1551, Jan. 2026, doi: 10.1038/s41598-025-26090-1.

[3] A. A. L. and V. C. S. S., "Cascaded 3D UNet architecture for segmenting the COVID-19 infection from lung CT volume," Sci Rep, vol. 12, no. 1, p. 3090, Feb. 2022, doi: 10.1038/s41598-022-06931-z.

[4] D. S, L. Padma Suresh, and A. John, "A Deep Transfer Learning framework for Multi Class Brain Tumor Classification using MRI," in 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India: IEEE, Dec. 2020, pp. 283–290. doi: 10.1109/ICACCCN51052.2020.9362908.

[5] S. Divya, L. Padma Suresh, and A. John, "Enhanced deep-joint segmentation with deep learning networks of glioma tumor for multi-grade classification using MR images," Pattern Anal Applic, vol. 25, no. 4, pp. 891–911, Nov. 2022, doi: 10.1007/s10044-022-01064-5.

[6] Y. Yan, B. Hou, and J. Sun, "MDSTransUNet: Multi-scale deep supervised transformer U-Net for COVID-19 lung tissue and infection segmentation," Neurocomputing, vol. 655, p. 131413, Nov. 2025, doi: 10.1016/j.neucom.2025.131413.
[7] S. Gupta et al., "Four Transformer-Based Deep Learning Classifiers Embedded with an Attention U-Net-Based Lung Segmenter and Layer-Wise Relevance Propagation-Based Heatmaps for COVID-19 X-ray Scans," Diagnostics, vol. 14, no. 14, p. 1534, Jul. 2024, doi: 10.3390/diagnostics14141534.
[8] N. Enshaei et al., "COVID-rate: an automated framework for segmentation of COVID-19 lesions from chest CT images," Sci Rep, vol. 12, no. 1, p. 3212, Feb. 2022, doi: 10.1038/s41598-022-06854-9.
[9] X. Zhang, J. Wang, J. Wei, X. Yuan, and M. Wu, "A Review of Non-Fully Supervised Deep Learning for Medical Image Segmentation," Information, vol. 16, no. 6, p. 433, May 2025, doi: 10.3390/info16060433.
[10] H. Polat, "Multi-task semantic segmentation of CT images for COVID-19 infections using DeepLabV3+ based on dilated residual network," Phys Eng Sci Med, vol. 45, no. 2, pp. 443–455, Jun. 2022, doi: 10.1007/s13246-022-01110-w.
[11] J. Ding, J. Chang, R. Han, and L. Yang, "CDSE-UNet: Enhancing COVID-19 CT Image Segmentation With Canny Edge Detection and Dual-Path SENet Feature Fusion," International Journal of Biomedical Imaging, vol. 2025, no. 1, p. 9175473, Jan. 2025, doi: 10.1155/ijbi/9175473.
[12] I. Bakkouri and K. Afdel, "MLCA2F: Multi-Level Context Attentional Feature Fusion for COVID-19 lesion segmentation from CT scans," SIViP, vol. 17, no. 4, pp. 1181–1188, Jun. 2023, doi: 10.1007/s11760-022-02325-w.
[13] Y. Qiu, Y. Liu, S. Li, and J. Xu, "MiniSeg: An Extremely Minimum Network Based on Lightweight Multiscale Learning for Efficient COVID-19 Segmentation," IEEE Trans. Neural Netw. Learning Syst., vol. 35, no. 6, pp. 8570–8584, Jun. 2024, doi: 10.1109/TNNLS.2022.3230821.
[14] Y. Yan, B. Hou, and J. Sun, "MDSTransUNet: Multi-scale deep supervised transformer U-Net for COVID-19 lung tissue and infection segmentation," Neurocomputing, vol. 655, p. 131413, Nov. 2025, doi: 10.1016/j.neucom.2025.131413.
[15] H. He, C. Sun, Z. Wang, Y. Dan, and Editorial Office, "Linear Transformer Based U-Shaped Lightweight Segmentation Network," JACIII, vol. 29, no. 6, pp. 1319–1328, Nov. 2025, doi: 10.20965/jaciii.2025.p1319.
[16] C.-F. Li et al., "MultiR-Net: A Novel Joint Learning Network for COVID-19 segmentation and classification," Computers in Biology and Medicine, vol. 144, p. 105340, May 2022, doi: 10.1016/j.compbiomed.2022.105340.
[17] D. Baldi et al., "Severity assessment of COVID-19 disease: radiological visual score versus automated quantitative CT parameters using a pneumonia analysis algorithm," Front. Med., vol. 12, p. 1606771, Sep. 2025, doi: 10.3389/fmed.2025.1606771.
[18] X. Wang, J. Yu, B. Zhang, X. Huang, X. Shen, and M. Xia, "LightAWNet: Lightweight adaptive weighting network based on dynamic convolutions for medical image segmentation," J Applied Clin Med Phys, vol. 26, no. 2, p. e14584, Feb. 2025, doi: 10.1002/acm2.14584.
[19] B. Schott, V. Santoro-Fernandes, Ž. Klaneček, S. Perlman, and R. Jeraj, "Uncertainty quantification for deep learning-based metastatic lesion segmentation on whole body PET/CT," Phys. Med. Biol., vol. 70, no. 11, p. 115009, Jun. 2025, doi: 10.1088/1361-6560/add9df.
[20] M. Jun et al., "COVID-19 CT Lung and Infection Segmentation Dataset." Zenodo, Apr. 20, 2020. doi: 10.5281/ZENODO.3757476.
[21] D. Müller, I. Soto-Rey, and F. Kramer, "Robust chest CT image segmentation of COVID-19 lung infection based on limited data," Informatics in Medicine Unlocked, vol. 25, p. 100681, 2021, doi: 10.1016/j.imu.2021.100681.
[22] Q. Zhang, X. Ren, and B. Wei, "Segmentation of infected region in CT images of COVID-19 patients based on QC-HC U-net," Sci Rep, vol. 11, no. 1, p. 22854, Nov. 2021, doi: 10.1038/s41598-021-01502-0.
[23] X. Zhao et al., "D2A U-Net: Automatic segmentation of COVID-19 CT slices based on dual attention and hybrid dilated convolution," Computers in Biology and Medicine, vol. 135, p. 104526, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104526.
[24] J. Zhang, X. Ding, D. Hu, and Y. Jiang, "Semantic segmentation of COVID-19 lesions with a multiscale dilated convolutional network," Sci Rep, vol. 12, no. 1, p. 1847, Feb. 2022, doi: 10.1038/s41598-022-05527-x.
[25] S. Bakheet, R. Youssef, M. H. Mofaddel, M. Hassan, and A. Alshehri, "Automated segmentation of COVID-19 lesions in CT scans using attention U-net with hybrid loss functions," Sci Rep, vol. 16, no. 1, p. 1551, Jan. 2026, doi: 10.1038/s41598-025-26090-1.