The First International Conference on
Machine Learning and Knowledge Discovery (MLKD 2024)
Amirkabir University of Technology, December 18-19, 2024

A Novel Multi-Task Learning Architecture for COVID-19 Detection and Lung Infection Segmentation in CT Scans

Shirin Kordnoori[1] Maliheh Sabeti[1,*] Hamidreza Mostafaei[2] Saeed Seyed Agha Banihashemi[3]

Abstract

In response to the critical need for accurate detection of COVID-19 and segmentation of lung infections in CT scans, this study introduces a novel Multi-Task Learning framework. The proposed model, built upon the U-Net architecture, features a shared encoder, a dedicated segmentation decoder, and a multi-layer perceptron for classification. Preprocessing involves image enhancement techniques such as median filtering and morphological opening, combined in pairs to boost task performance. Additionally, the integration of the Convolutional Block Attention Module (CBAM) aids in extracting critical features. The model achieved a classification accuracy of 96.22% and a segmentation accuracy of 95.84% when evaluated on a benchmark dataset. The approach was also successfully applied to U-Net++ and ResUnet architectures, underscoring the potential for improving multi-task learning structures. This work sets a new benchmark for COVID-19 detection and advances medical image analysis.

Keywords: Deep Multi-task learning model, Chest CT scans, Shared representations

1 Introduction

The global spread of COVID-19 has created an urgent demand for effective diagnostic tools [1]. While PCR testing remains a standard, it has limitations in speed and accuracy, highlighting the critical role of medical imaging, especially CT scans, for COVID-19 detection. Challenges like image noise and artifacts can impact the reliability of these scans. To address these challenges, deep learning models, especially those using artificial intelligence, have shown great promise in improving the analysis of medical images.

In medical image processing, segmentation [2] and classification [3] play complementary roles. Classification helps in diagnosing the overall health status of the lungs, while segmentation pinpoints the specific areas of infection. In cases where patient conditions are critical, segmentation becomes essential for accurate localization of infections, aiding medical professionals in their diagnosis.

*Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran, sabeti@shirazu.ac.ir

Multi-Task Learning (MTL) frameworks have emerged as powerful tools in handling both segmentation and classification tasks. By sharing representations and improving feature extraction, these architectures enable models to perform multiple tasks simultaneously, which leads to enhanced diagnostic precision. However, previous works in this field have often overlooked the potential of integrating preprocessing techniques and attention mechanisms into these models [4].

In this study, a novel multi-task deep learning model is presented for COVID-19 diagnosis using chest CT scans. The model incorporates image preprocessing, attention mechanisms like the Convolutional Block Attention Module (CBAM) [5], and carefully balances the performance of segmentation and classification tasks. Our proposed model achieves high classification accuracy while providing precise segmentation results, offering a reliable solution for diagnosing COVID-19 infections in medical imaging.

Key contributions of this work include:

1. Unified Multi-Task Learning: The integration of classification and segmentation tasks in a single framework enhances model generalization and stability.
2. Image Processing: The application of preprocessing techniques improves image quality, helping the model achieve higher accuracy.
3. Attention Mechanisms: CBAM enhances feature extraction, allowing for more focused and accurate analysis.
4. Versatility: We propose the foundational architecture based on U-Net but have also tested these components within U-Net++ [6] and ResUNet [7] in multi-task mode to demonstrate their effectiveness.

By building upon U-Net and enhancing it with multi-task learning, attention modules, and image preprocessing, this study introduces a more robust and comprehensive approach to medical image analysis for COVID-19 detection.

2 Related Works

In this section, we provide a detailed overview of the current literature on multi-task learning frameworks, with

a particular focus on their applications in the medical field. These models have played a crucial role in advancing medical image analysis. Table 1 summarizes various studies on multi-task architectures, outlining their approaches and key outcomes. From the comparison of these studies, it is evident that most have primarily emphasized task operations, often overlooking the integration of image processing techniques. Recent research [15, 16, 17] highlights similar limitations, underscoring the need for further exploration in this area to improve the performance and efficiency of multi-task learning models.

Table 1: Summary of related works in multi-task learning.

Research article	Year	Tasks and Results
Kordnoori et al. [9]	2023	Classification, Segmentation Acc: 0.97 (DS1), 0.96 (DS2) Dice: 88.86±0.05 (DS1), 87.97±0.02 (DS2)
El-Bana et al. [10]	2020	Classification, Segmentation Acc: 99.4% AUC: 4.5% improvement
Malhotra et al. [11]	2022	Classification, Semantic Segmentation Sens: 96.80% Spec: 90%
Ortiz et al. [12]	2022	Infection Segmentation, Classification (with metadata) AUC: 0.80
Li et al. [13]	2022	Classification, Segmentation Acc: 94% Sens: 93.32 Dice: 69.95
Munusamy et al. [14]	2021	Classification, Segmentation Acc: 99% Dice: 74.5
Polat [15]	2022	Lung Segmentation (three tasks) Dice: 0.98

3 Proposed Model

In this section, we present the architecture of the model designed for the precise and efficient analysis of COVID-19 lung infections. The model consists of two key stages: pre-processing and a U-Net-based network that incorporates integrated attention modules. Each of these stages is critical for enhancing the overall performance and accuracy of the model, thus offering a comprehensive solution for segmentation and classification tasks.

Our proposed framework, depicted in Figure 1, utilizes a simplified U-Net architecture for several compelling reasons. First, U-Net demonstrates superior performance compared to other architectures when faced with limited data availability. Second, the U-Net family encompasses a wide range of variations, allowing us to evaluate our proposed approach across different frameworks. Finally, it is easier to extend U-Net to a multi-task structure than to adapt other architectures.

Furthermore, we will extend the proposed framework to develop multi-task models based on U-Net++ and ResUNet architectures. The results of these experiments will be presented in the Experimental Results section, where we will compare the outcomes of the different network architectures in multi-task scenarios.

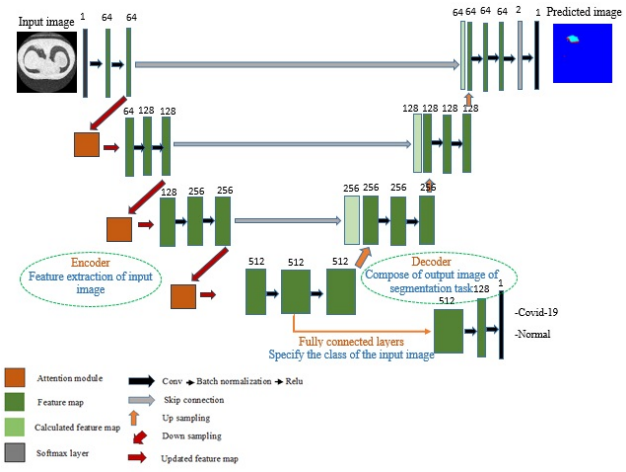

Figure 1: Illustration of the proposed model architecture.

3.1 Encoder-Decoder Structure

The proposed multi-task model combines classification and segmentation tasks within a shared encoder framework, using CBAM modules to enhance feature representation. The encoder extracts hierarchical features via convolutional layers, with filter sizes ranging from 64 to 512, and a kernel size of 3×3. The decoder handles segmentation, while a multi-layer perceptron performs infection classification. This structure enables efficient

lung infection detection and identification of affected regions by generating rich feature representations.

Each convolutional layer C^m produces feature maps by convolving the input with learned weight matrices $W_i^{(1)}$ and bias terms $b_i^{(1)}$, and applying a non-linear activation function:

$$C_i^{(m)} = f(b_i^{(1)} + W_i^{(1)} \cdot x), \quad i = 1, \ldots, F^{(1)} \tag{1}$$

Here, $C_i^{(m)}$ is the resulting feature map, and $W_i^{(1)}$ are the learned weight matrices.

The next layers compute feature maps through summing products of weights and feature maps:

$$C_i^{(n)} = f(b_i^{(n)} + \sum_{s=1}^{P^{(n-1)}} W_i^{(n)} \cdot P_s^{(n-1)}), \quad i = 1, \ldots, F^{(n)} \tag{2}$$

3.1.1 Classification Task

The classification task is implemented using a multilayer perceptron with two dense layers (512 and 128 neurons) and dropout (0.5) to prevent overfitting. The final layer has two neurons with softmax activation to predict class probabilities. The categorical cross-entropy loss function CE is defined as:

$$CE = -\sum_{i=1}^{C} t_i \log(f(S)_i) \tag{3}$$

with softmax probabilities given by:

$$f(S)_i = \frac{e^{S_i}}{\sum_{j=1}^{C} e^{S_j}} \tag{4}$$

3.1.2 Segmentation Task

The segmentation task uses sigmoid activation to enable pixel-wise classification for lung infection detection. The dice coefficient, with smoothing factor ϵ, measures the similarity between predicted and true masks:

$$\text{dice\_coef} = \frac{2|x \cap y| + \epsilon}{|x| + |y| + \epsilon} \tag{5}$$

Segmentation helps improve the model's precision and ensures accurate medical image analysis.

3.2 Pre-Processing Phase

Our pre-processing strategies blend technical expertise with clinical relevance, focusing on optimizing the input data's quality. To achieve this, we implement a combination of median filtering and morphological opening algorithms.

This strategic integration serves multiple purposes, including enhancing overall image quality and facilitating the accurate detection of lung infections. Specifically, we employ intensity normalization to establish a consistent baseline relative to the average intensity of healthy lung tissue, ensuring that variations in image intensity are adjusted to a standardized level for more reliable analysis.

All images were standardized to a fixed size of 224×224 pixels. The technical advancements achieved through these pre-processing steps enhance our model's capabilities for precise COVID-19 detection and versatile analysis of various medical conditions, making it a valuable tool for healthcare professionals.

3.3 Dataset

This study integrates four distinct datasets to form a new dataset categorized into two classes: "healthy" and "infected." The datasets included in this research are the "MedSeg dataset" [18], "COVID-19 CT Lung and Infection Segmentation Dataset" [19], and "MosMedData: Chest CT Scans with COVID-19" [20]. Furthermore, the "COVID-19 Lung CT Scans" dataset [21], which contains images of healthy lung CT scans, was merged with datasets [18]-[20] to enrich our final dataset.

For the distribution of the datasets into training, validation, and testing subsets, we implemented a ratio of 60%, 20%, and 20%, respectively. To enhance the dataset and promote diversity in the image slices, various data augmentation techniques were applied. These techniques included rotations of 90° and 270°, as well as horizontal and vertical translations, zooming, and shearing. Table 2 provides detailed information on the number of cases in each class before and after data augmentation.

Table 2: Number of cases in the datasets before and after data augmentation for each class. (Class 1: infected, Class 2: healthy)

Dataset	Data Type	Class 1 (Infected)	Class 2 (Healthy)	Total
Combination of [18] and [21]	Raw data	470	470	940
	Augmented data	3760	3760	7520
Combination of [19] and [21]	Raw data	1351	1351	2702
	Augmented data	10808	10808	21616
Combination of [20] and [21]	Raw data	785	785	1570
	Augmented data	6280	6280	12560
Final	Augmented data	20848	20848	41696

3.4 Evaluation Metrics

The proposed model was rigorously trained, validated, and tested using COVID-19 datasets. For assessing performance in the segmentation task, several metrics were utilized, including the Dice coefficient, mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). In the classification task, key evaluation metrics comprised the area under the ROC curve (AUC), accuracy, sensitivity, and specificity.

The mean squared error (MSE) can be expressed mathematically as follows:

$$\text{MSE} = \frac{1}{MN} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} ||f(i,j) - g(i,j)||^2 \tag{6}$$

In this equation, f represents the pixel matrix of the original image, g denotes the pixel matrix of the degraded image, m indicates the number of pixel rows, n signifies the number of pixel columns, and |||| indicates the Euclidean distance.

The peak signal-to-noise ratio (PSNR) is calculated using the following formula:

$$\text{PSNR} = 20 \log_{10} \left(\frac{\text{Max}_f}{\sqrt{\text{MSE}}} \right) \tag{7}$$

Here, Max_f denotes the maximum signal value from the original image.

The structural similarity index measure (SSIM) incorporates evaluations of luminance (l), contrast (c), and structure (s) between the images f and g, weighted by α, β, and γ respectively:

$$\text{SSIM}(f,g) = l(f,g)^\alpha \cdot c(f,g)^\beta \cdot s(f,g)^\gamma \tag{8}$$

For the classification task, the accuracy is computed as follows:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Examples}} \tag{9}$$

Sensitivity is defined as:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{10}$$

And specificity is determined by:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \tag{11}$$

Where:

- **True Positives (TP)**: The number of actual positive samples correctly identified by the model. This metric indicates how well the model has recognized true positive cases (e.g., COVID-19 infections).
- **True Negatives (TN)**: The number of actual negative samples correctly identified by the model. This metric reflects how well the model has recognized true negative cases (e.g., healthy lungs).
- **False Positives (FP)**: The number of actual negative samples incorrectly identified as positive. This metric shows the model's errors in identifying positive cases.
- **False Negatives (FN)**: The number of actual positive samples incorrectly identified as negative. This metric highlights the model's inability to recognize true positive cases.

4 Experimental Results

In this section, we present the model's results, organized by the classification and segmentation tasks. To enable a more comprehensive comparison, we also provide the results of a simple multi-task model that uses the image processing techniques of median filtering and morphological opening but does not include CBAM modules. This allows us to better assess the impact of CBAM on the proposed model. The outputs for the classification and segmentation tasks are displayed in Tables 3 and 4 , respectively.

Table 3: Segmentation results.

Experiment	Acc	MSE	PSNR	SSIM	Dice ± Std
Simple Multi-Task Model	95.66	0.04	36.65	0.94	88.05 ± 0.01
Proposed Model	95.84	0.03	36.70	0.95	88.23 ± 0.01

Table 4: Classification results.

Experiment	Acc	Sens	Spec	AUC	Precision	F1	MCC
Simple Multi-Task Model	95.62	95.61	95.63	95.85	95.63	95.55	91.28
Proposed Model	96.22	96.22	96.22	96.66	96.22	96.23	92.01

Table 3 highlights the segmentation task results, while Table 4 presents the outcomes of the classification task. These tables clearly demonstrate the substantial improvements achieved in both tasks.

In Table 3, the integration of the CBAM module significantly enhances segmentation performance. The improvements are reflected in higher segmentation accuracy, fewer errors, better image quality, and improved

structural similarity. The Dice coefficient shows a notable increase to 88.23 ± 0.01, demonstrating the effectiveness of these enhancements compared to the simple multi-task model.

Moving to Table 4, the CBAM module's inclusion in the proposed model leads to superior classification performance over the simple multi-task model. This improvement highlights the critical role of the shared CBAM module in aligning the attention mechanisms for both segmentation and classification tasks in the multi-task framework. The attention harmonization enables the model to more effectively evaluate features relevant to both tasks, resulting in more comprehensive and accurate feature representations, and ultimately, better overall performance.

A paired t-test was conducted to compare the classification results in Table 4. With a significance level of 0.05, the p-value was much smaller than 0.05, leading us to reject the null hypothesis (H0). This indicates a statistically significant difference between the classification results of the two experiments. The t-statistic of -20.42 and the p-value of 8.97×10^{-7} further confirm the effectiveness of the proposed model's enhancements using the CBAM module.

For detailed information on the implementation settings and hyperparameters used in the multi-task model, refer to Table 5. This table outlines the key configurations that influence the model's training and performance, offering clarity on the hyperparameters employed throughout the implementation.

Table 5: Hyperparameter configurations and implementation details of the multi-task model.

Hyperparameter	**Value(s)**
Mini-Batch Size	32
Learning Rate	2e-4
Number of Epochs	100
Convolutional Kernel Size	3×3
Gradient Decay Rate	0.96
Loopy Belief Propagation (for CRF)	0.8

Figure 2 displays the confusion matrix for the proposed model, providing a comprehensive analysis of its classification performance across all datasets. The matrix is based on a 20% test subset from each dataset, giving insights into how the model distinguishes between classes.

Figure 3 shows the output images of the segmentation task. The black regions represent healthy areas, while the red regions denote infected areas. Each row corresponds to a test set image, demonstrating the model's performance through the pre-processing stage (using median filtering and morphological opening) and the application of the CBAM module for better segmentation of infection.

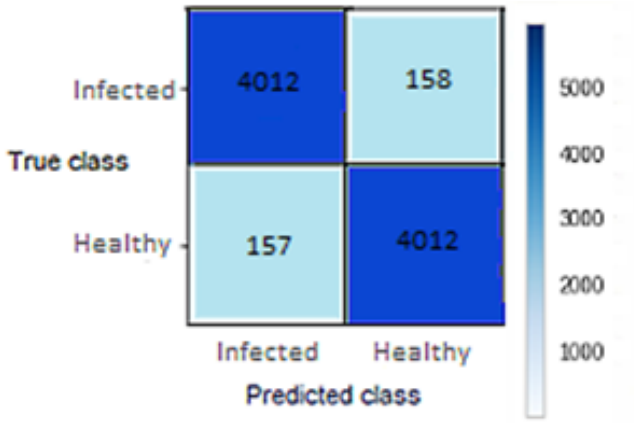

Figure 2: Confusion matrix for the proposed model in the classification task, offering a detailed evaluation of class distinctions.

To improve the interpretability and transparency of the classification task, we employed the Grad-CAM technique to produce localization maps for test set images identified as COVID-19 cases. This method highlights the key regions that influenced the model's prediction. Figure 4 displays examples of CT images alongside their corresponding Grad-CAM maps. The model successfully identified these cases as COVID-19, accurately marking abnormal areas within the CT scans and providing insights into the decision-making process.

5 Comparison and Discussion

In this section, we carry out an in-depth performance analysis of our proposed model and compare it with previous research. The main goal is to thoroughly evaluate the model's effectiveness, position it against state-of-the-art studies, and identify areas for potential improvements. The comparison is conducted in the following stages:

1. We analyze the outputs of our model against results from prior studies, considering both the segmentation and classification tasks.

2. To further demonstrate the efficiency of our proposed approach in creating impactful multi-task architectures, we integrated the same pre-processing techniques and CBAM module into more complex networks, such as U-Net++ and ResUNet. For classification, we introduced a multi-layer perceptron branch, transforming these models into a multi-task structure.

Table 8 offers a detailed comparison of the segmentation performance of our multi-task model with state-of-the-art techniques, while Table 9 highlights its classification task results. All comparisons were performed using

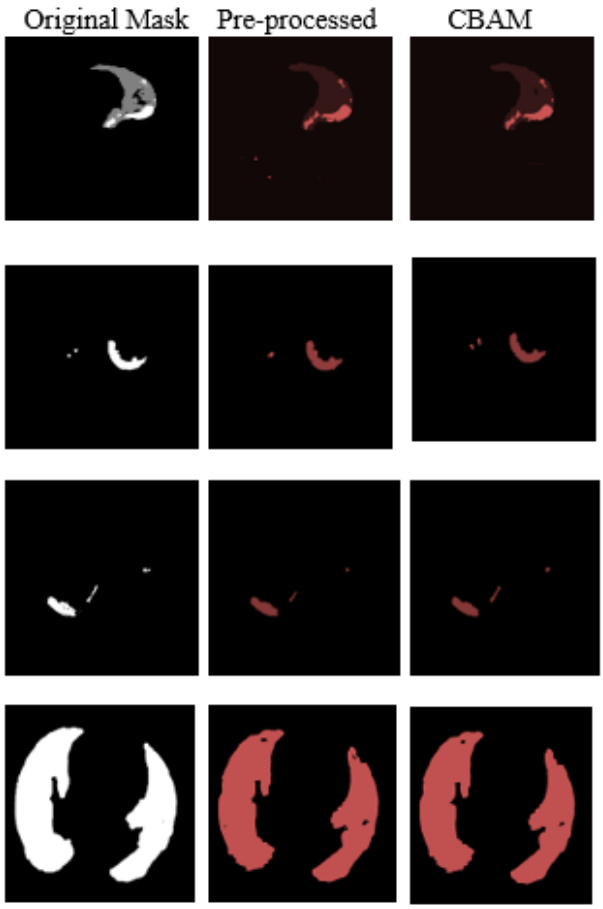

Figure 3: Output images of the segmentation task. The black regions represent healthy areas, while the red regions indicate infected areas. Each row corresponds to a test set image, showing the performance of the proposed model during the pre-processing stage and with the inclusion of the CBAM module for improved segmentation.

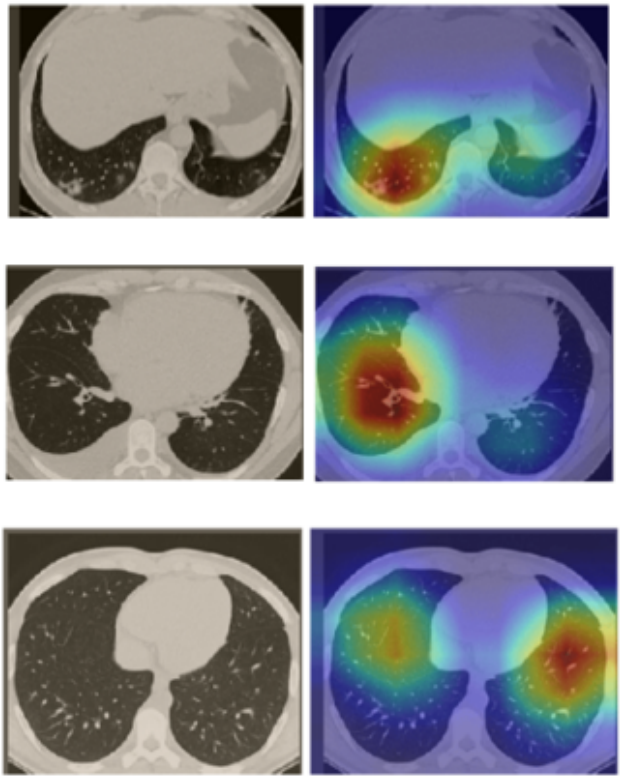

Figure 4: Grad-CAM visualizations for selected CT images. (Column 1: Original images of COVID-19 affected cases. Column 2: Grad-CAM visualizations showing areas of abnormalities in the CT scans.)

the same dataset and metrics for fairness, and we reproduced the results of previous studies for consistency. In Table 8, the proposed model delivers competitive results compared to other models in segmentation. While all models considered are multi-task models, our approach outperforms models from [10], [13], and [14]. Pairwise t-tests were conducted to compare Dice scores between our model and others. With a significance level of 0.05, the p-values were consistently below this threshold, allowing us to reject the null hypothesis (H0) in each case. Thus, we conclude that the Dice scores of our proposed model differ significantly from other models in the segmentation task.

This multi-task approach allows the model to capture generalized features that apply to various scenarios. As a result, the proposed model excels at identifying both common and specific characteristics of the target object, which leads to higher segmentation accuracy. In contrast, models trained solely for a single segmentation task may be constrained by dataset-specific biases and variations, limiting their adaptability to new data. Our multi-task strategy, which leverages shared knowledge and diverse data, enables the model to achieve better segmentation accuracy and robustness, as demonstrated by its superior performance compared to models trained on individual datasets.

Table 9 provides a comparison of the classification task performance between our model and state-of-the-art approaches. Once again, the proposed model shows significant improvements in accuracy and other metrics, showcasing its superior generalization capability and performance consistency across multiple datasets.

In the final step, we advanced the complexity of our network by integrating U-Net++ and ResUNet architectures, which are more intricate than the basic U-Net. We further enhanced these architectures by adding a multi-layer perceptron branch, transforming them into multi-task models that included both segmentation and classification tasks. Pre-processing, attention modules,

Table 6: A Comparison of Segmentation Task Performance between the proposed model and State-of-the-Art Research.

Reference	Acc	MSE	PSNR	SSIM	Dice ± Std
[8]	95.18	0.04	33.06	0.91	85.43 ± 0.03
[10]	95.77	0.04	33.54	0.92	86.07 ± 0.02
[13]	95.72	0.04	32.54	0.91	85.10 ± 0.02
[14]	94.60	0.05	33.11	0.92	85.54 ± 0.01
Proposed	95.84	0.03	36.70	0.95	88.23 ± 0.01

Table 7: A Comparison of Classification Task Performance between the proposed model and State-of-the-Art Research.

Reference	Acc	Sens	Spec	AUC	Precision	F1	MCC
[7]	94.49	94.52	95.57	94.85	94.50	93.76	90.41
[9]	95.34	95.37	96.42	95.70	95.35	94.61	91.26
[10]	95.08	95.11	96.16	95.44	95.09	94.35	91.00
[11]	95.49	95.52	96.57	95.85	95.50	94.76	91.41
[12]	95.03	95.06	96.11	95.39	95.04	94.30	90.95
[13]	93.91	93.94	94.99	94.27	94.20	93.18	89.83
[15]	96.05	96.08	97.13	96.41	96.06	95.92	91.97
Proposed	96.22	96.22	96.22	96.66	96.22	96.23	92.01

and post-processing were also applied to refine these models.

Our objective was to investigate the impact of more complex encoder-decoder architectures on multi-task performance. By utilizing U-Net++ and ResUNet with additional multi-layer perceptron branches, we aimed to explore the advantages and limitations of these more sophisticated designs. Tables 8 and 9 present the segmentation and classification results, highlighting the improved performance of the U-Net++, ResUNet, and U-Net models under the multi-task structure.

The results underscore the strength of multi-task learning in improving both segmentation accuracy and classification performance through shared knowledge across tasks. It is notable that the simple U-Net multi-task architecture serves as the baseline, without the enhancements of pre-processing and CBAM modules. These comparisons emphasize the benefits of the proposed approach in CT image analysis, demonstrating its versatility and effectiveness in both diagnostic and analytical tasks.

This study has a few limitations, outlined as follows:

1. **Requirement for labeled data with masks**: Our method relies on datasets where both labels and corresponding masks are available simultaneously. This dependency may restrict its applicability in scenarios where such datasets are not readily accessible.

2. **Higher performance in related tasks**: The effectiveness of our multi-task architecture is more pronounced in tasks that are closely related. For tasks with weaker correlations, the method may require further refinement and optimization to enhance its performance.

Table 8: Comparison of the proposed multi-task structure across different network architectures in the segmentation task.

Mode	Model	Acc	MSE	PSNR	SSIM	Dice ± Std
Simple Multi-Task	U-net	95.24	0.04	35.64	0.94	87.33 ± 0.01
	U-net++	95.58	0.04	35.97	0.95	87.67 ± 0.01
	RESUNET	95.13	0.04	35.53	0.94	87.24 ± 0.01
Proposed Method	U-net	95.84	0.03	36.70	0.95	88.23 ± 0.01
	U-net++	96.18	0.03	36.93	0.96	88.36 ± 0.02
	RESUNET	95.74	0.03	36.47	0.95	88.07 ± 0.01

Table 9: Comparison of the proposed multi-task structure across different network architectures in the classification task.

Mode	Model	Acc	Sens	Spec	AUC	Precision	F1	MCC
Simple Multi-Task	U-net	95.24	95.27	95.22	95.60	95.25	94.70	91.16
	U-net++	95.60	95.63	95.58	95.96	95.61	95.06	91.52
	RESUNET	95.15	95.18	95.13	95.51	95.22	94.61	91.07
Proposed Method	U-net	96.22	96.22	96.22	96.66	96.22	96.23	92.01
	U-net++	96.16	96.19	96.14	96.52	96.17	95.43	92.08
	RESUNET	94.94	94.97	94.92	95.35	94.95	94.21	91.33

6 Conclusion

In light of the global COVID-19 pandemic, the need for precise diagnostic tools has led to an increased reliance on artificial intelligence and CAD systems as essential medical aids. This study proposed a novel Multi-Task Learning (MTL) architecture designed for two main tasks: segmentation and classification of chest CT scan images. The proposed model, based on U-Net, incorporates a shared encoder for both tasks, with a dedicated decoder for segmentation and a multi-layer perceptron branch extending from the final encoder layer to handle classification.

The distinguishing feature of this MTL model is its focus on enhancing image quality, improving infection detection, and overall model performance. Preprocessing techniques were applied to improve the quality of input images, and a CBAM attention module in the shared encoder enhanced feature extraction. The proposed method was evaluated using multiple datasets.

For segmentation, the model achieved an accuracy of 95.84%, an MSE of 0.03, a PSNR of 36.70, an SSIM of 0.95, and a Dice coefficient of 88.23. In the classification task, the model achieved an accuracy of 96.22%, a sensitivity of 96.22%, a specificity of 96.22%, an AUC of 96.66%, an F1 score of 96.23%, and an MCC of 92.01. Comparisons with previous MTL models and pre-trained single-task models highlighted the competitiveness of the proposed approach. Furthermore, the benefits of preprocessing, attention, and post-processing were demonstrated when applied to U-Net++ and ResUnet architectures, showing their potential for improving multi-task models.

However, this study has certain limitations. The benefits of MTL are only realized when the tasks are closely related, as is the case in this study. Additionally, the proposed model requires inputs that contain both labels and masks, which may limit the selection of datasets. Future research should focus on addressing these challenges.

Moving forward, it is essential to refine and advance MTL architectures for medical image analysis, particularly in the context of COVID-19 applications. Exploring additional tasks, such as anomaly detection or disease severity assessment, could extend the model's capabilities. Additionally, investigating model simplification or pruning techniques is crucial to enhance computational efficiency while maintaining performance, particularly for architectures like U-Net++ and ResUnet. Optimizing hyperparameters and tailoring network architectures to various medical imaging modalities and datasets can also contribute to the development of more robust and adaptable models. The proposed method holds significant potential as a primary screening tool, helping primary healthcare providers refer patients to specialists efficiently, especially in resource-constrained settings. The adoption of such technology paves the way for broader global studies aimed at improving diagnostic rates across healthcare systems.

References

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 2020.

[2] R. Wang, C. Ji, Y. Zhang, Y. Li. Focus, Fusion, and Rectify: Context-Aware Learning for COVID-19 Lung Infection Segmentation. *IEEE Trans. Neural Netw. Learn. Syst.*, 33 (1) (2022).

[3] C.L. Alves, T.G.L.O. Toutain, P. de Carvalho Aguiar, A.M. Pineda, K. Roster, C. Thielemann, J.A.M. Porto, F.A. Rodrigues. Diagnosis of autism spectrum disorder based on functional brain networks and machine learning. *Scientific Reports*, 13(1):8072, 2023. doi: 10.1038/s41598-023-34650-6.

[4] S. Kordnoori, M. Sabeti, H. Mostafaei, S. Seyed Agha Banihashemi. LungXpertAI: A deep multitask learning model for chest CT scan analysis and COVID-19 detection. *Biomedical Signal Processing and Control*, Volume 99, 2025, 106866. https://doi.org/10.1016/j.bspc.2024.106866.

[5] S. Woo, J. Park, J.-Y. Lee, and I.S. Kweon. CBAM: Convolutional Block Attention Module. *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII*. Springer-Verlag, Berlin, Heidelberg, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.

[6] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Lecture Notes in Computer Science, Vol. 11045: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, DLMIA ML-CDS 2018*. Springer, Cham, 2018. https://doi.org/10.1007/978-3-030-00889-5_1.

[7] F.I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 162, 2020, Pages 94-114. https://doi.org/10.1016/j.isprsjprs.2020.01.013.

[8] S. Kordnoori, M. Sabeti, H. Mostafaei, S.S.A. Banihashemi. An efficient deep multitask learning structure for COVID-19 disease. *IET Image Process.*, 00, 1–18 (2023). https://doi.org/10.1049/ipr2.12893.

[9] S. El-Bana, A. Al-Kabbany, M. Sharkas. A multitask pipeline with specialized streams for classification and segmentation of infection manifestations in COVID-19 scans. *PeerJ Computer Science*, 6:e303, 2020.

[10] A. Malhotra, S. Mittal, P. Majumdar, S. Chhabra, K. Thakral, M. Vatsa, R. Singh, S. Chaudhury, A. Pudrod, A. Agrawal. Multi-task driven explainable diagnosis of COVID-19 using chest X-ray images. *Pattern Recognit.*, 122:108243, 2022. doi: 10.1016/j.patcog.2021.108243.

[11] A. Ortiz, A. Trivedi, J. Desbiens, et al. Effective deep learning approaches for predicting

COVID-19 outcomes from chest computed tomography volumes. *Sci Rep*, 12, 1716, 2022. https://doi.org/10.1038/s41598-022-05532-0.

[12] C.-F. Li, Y.-D. Xu, X.-H. Ding, J.-J. Zhao, R.-Q. Du, L.-Z. Wu, W.-P. Sun. MultiR-Net: A Novel Joint Learning Network for COVID-19 segmentation and classification. *Computers in Biology and Medicine*, Volume 144, 2022, 105340. https://doi.org/10.1016/j.compbiomed.2022.105340.

[13] H. Munusamy, J.M. Karthikeyan, G. Shriram, et al. FractalCovNet architecture for COVID-19 Chest X-ray image Classification and CT-scan image Segmentation. *Biocybernetics and Biomedical Engineering*, Volume 41, Issue 3, 2021, Pages 1025-1038.

[14] H. Polat. Multi-task semantic segmentation of CT images for COVID-19 infections using DeepLabV3+ based on dilated residual network. *Phys Eng Sci Med*, 45, 443–455, 2022. https://doi.org/10.1007/s13246-022-01110-w.

[15] N. Ghassemi, A. Shoeibi, M. Khodatars, J. Heras, A. Rahimi, A. Zare, Y.-D. Zhang, R.B. Pachori, J.M. Gorriz. Automatic diagnosis of COVID-19 from CT images using CycleGAN and transfer learning. *Applied Soft Computing*, Volume 144, 2023, 110511. https://doi.org/10.1016/j.asoc.2023.110511.

[16] G. Bao, H. Chen, T. Liu, G. Gong, Y. Yin, L. Wang, X. Wang. COVID-MTL: Multitask learning with Shift3D and random-weighted loss for COVID-19 diagnosis and severity assessment. *Pattern Recognition*, Volume 124, 2022, 108499. https://doi.org/10.1016/j.patcog.2021.108499.

[17] Y.-D. Zhang, Z. Zhang, X. Zhang, S.-H. Wang. MIDCAN: A multiple input deep convolutional attention network for Covid-19 diagnosis based on chest CT and chest X-ray. *Pattern Recognition Letters*, Volume 150, Pages 8-16, 2021.

[18] MedSeg (2020) MedSeg Segmentation dataset, COVID-19 CT. 2020. Available at: `https://medicalsegmentation.com/covid19/`.

[19] Ma Jun, Ge Cheng, Wang Yixin, An Xingle, Gao Jiantao, Yu Ziqi, Zhang Minqing, Liu Xin, Deng Xueyuan, Cao Shucheng, Wei Hao, Mei Sen, Yang Xiaoyu, Nie Ziwei, Li Chen, Tian Lu, Zhu Yuntao, Zhu Qiongjie, Dong Guoqiang, & He Jian. COVID-19 CT Lung and Infection Segmentation Dataset (Version 1.0) [Data set]. Zenodo, 2020. Available at: `https://doi.org/10.5281/zenodo.3757476`.

[20] S. Morozov et al. MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset. *arXiv preprint*, arXiv:2005.06465, 2020.

[21] M. Aria, M. Ghaderzadeh, F. Asadi, R. Jafari. COVID-19 Lung CT Scans, 2021. Available at: `https://www.kaggle.com/dsv/1875670`. DOI: `https://doi.org/10.34740/KAGGLE/DSV/1875670`, 2021.