

# ATTENTION-BASED CONVOLUTIONAL NEURAL NETWORK FOR CT SCAN COVID-19 DETECTION

A. Rondinella\*, F. Guarnera\*, O. Giudice\*, A. Ortis\*, F. Rundo†, S. Battiato\*

\*Department of Mathematics and Computer Science, University of Catania, Catania, Italy

†STMicroelectronics, ADG R&D Power and Discretes Division, Catania, Italy

*alessia.rondinella@phd.unict.it, francesco.guarnera@unict.it, giudice@dmf.unict.it,  
ortis@dmf.unict.it, francesco.rundo@st.com, battiato@dmf.unict.it*

## ABSTRACT

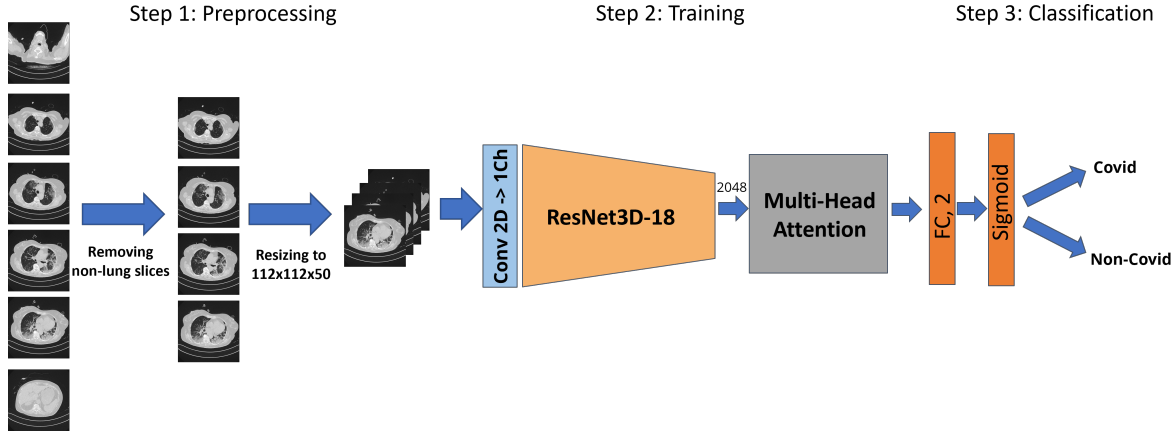
The accurate detection of Covid-19 from chest Computed Tomography (CT) images can assist in early diagnosis and management of the disease. This paper presents a solution for Covid-19 detection, presented in the challenge of 3rd Covid-19 competition, inside the "AI-enabled Medical Image Analysis Workshop" organized by IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP) 2023. In this work, the application of deep learning models for chest CT image analysis was investigated, focusing on the use of a ResNet as a backbone network augmented with attention mechanisms. The ResNet provides an effective feature extractor for the classification task, while the attention mechanisms improve the model's ability to focus on important regions of interest within the images. We conducted extensive experiments on a provided dataset and achieved a macro F1 score of 0.78 on the test set, demonstrating the potential to assist the diagnosis of Covid-19. Our proposed approach leverages the power of deep learning with attention mechanisms to address the challenges of Covid-19 detection in the early detection and management of the disease. In both test and validation set, the proposed method outperformed the baseline of the challenge, ranking fifth in the competition.

**Index Terms**— Covid-19 detection, Deep Learning, Computed Tomography classification, Medical imaging

## 1. INTRODUCTION

The outbreak of the Covid-19 pandemic had a devastating impact on the world, causing widespread illness and death [1]. Early detection of the disease is crucial for the timely management and control the spread of the virus. Medical imaging has emerged as a valuable tool in the fight against Covid-19, particularly in the detection of lung abnormalities associated with the disease [2]. Chest computed tomography (CT) has been widely used in the diagnosis and monitoring of Covid-19 patients, as it provides detailed information about the extent

and severity of lung involvement. Furthermore, CT imaging has been found to be more sensitive than chest X-rays in diagnosing Covid-19, and is useful for assessing the follow-up of the lungs of patients with respiratory complications. However, visual interpretation of CT scans by physicians can be complicated and time-consuming and requires expert knowledge and experience. The challenges in visually interpreting CT scans have led to develop artificial intelligence-based algorithms to support physicians and radiologists in the diagnosis of Covid-19. Furthermore, the potential benefits of deep learning in assisting with medical image analysis, including CT scans, suggest that it may be a valuable tool in the fight against Covid-19. In recent years, deep learning has shown great potential in the field of medical imaging. There has been a rise in research into employing deep learning algorithms to identify and diagnose diseases using medical imaging data, such as chest CT scans, and Covid-19 is no exception. In this paper, a fully automated solution for the first challenge of the 3rd Covid-19 competition [3] was presented; the challenge is a part of "AI-enabled Medical Image Analysis Workshop" organized by the 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023). The challenge is composed of two competitions: the first aims to differentiate COVID-19 cases from non-COVID ones (covid detection), while the second focuses on determining the severity of COVID-19 between four stages (covid-severity classification). We propose a framework for the detection of Covid-19 from chest CT scans. Our proposed approach is based on a modified ResNet backbone network with attention levels to focus on relevant regions of interest within the lung. The remainder of this paper is organized as follows: Section 2 provide a brief overview of the existing literature on the use of medical imaging for Covid-19 detection. In Section 3 we describe the dataset and present our proposed automated approach for the detection of Covid-19 from chest CT scans. Section 4 present the experimental setup and analysis of our approach, followed by a discussion. Finally, conclusions are summarized in Section 5.



**Fig. 1.** The proposed pipeline for Covid-19 detection from chest CT scans. First, the non-lung slices were removed, followed by slice resizing. The proposed network was then applied to the preprocessed slices to perform Covid-19 detection.

## 2. STATE OF THE ART

Medical image analysis has become increasingly reliant on deep learning techniques, with a wide range of applications in fields such as Covid-19 detection and general medical image analysis. However, the variability of medical images can pose a significant challenge for classification tasks, making it difficult to accurately identify and classify abnormalities. This is particularly true for CT scans, which can differ greatly in the number of scans and can be acquired by different scanners, making it complex to develop an approach that works for all scans. To address these challenges, research efforts are ongoing to develop more robust and adaptable deep learning algorithms for various tasks. For instance, a deep learning method was proposed in [4] to classify the computed tomography of patients with recurrent hip replacement infections, while authors in [5] propose a deep learning framework for automatically segment multiple sclerosis (MS) lesion starting from magnetic resonance (MR) images. The Covid-19 pandemic has led to a large demand for effective and efficient methods for the detection and diagnosis of the disease. Deep learning approaches have emerged as a promising tool for this task, owing to their ability to learn complex representations of medical images and extract relevant features for accurate detection [6][7][8]. Several studies have been conducted to explore the potential of deep learning in Covid-19 detection. For instance, authors in [9] propose a mixup data augmentation technique on Covid-19 to deducing the severity of infection at slice-level. Authors in [10] use a Convolutional Neural Network (CNN) to detect the presence of Covid-19 on both X-ray and CT scan images. Authors in [11] Polsinelli et al. propose a CNN that follow the SqueezeNet architecture to classify chest CT scan images into Covid-19, healthy or pneumonia. Also methods that leverage data augmentation techniques are becoming more and more popular, as they can significantly increase the variability of the data and im-

prove the performance of the models. For instance, in [12] a self-augmentation scheme based on reinforcement learning was proposed for chest X-rays. The authors then processed the resulting 3D scans using a CNN. Moreover, recent studies have shown promising results in COVID-19 detection from CT scans using the Vision Transformer [13], as demonstrated in [14][15], which proposed Visual Transformer-based frameworks for the detection of COVID-19 from chest X-rays. However, there are still challenges to be addressed in the use of deep learning for Covid-19 detection, including the need for large and diverse datasets and the interpretability and transparency of the models. Further research is needed to address these challenges and to develop more robust and reliable deep learning models for Covid-19 detection and diagnosis.

## 3. METHODOLOGY

Chest CT scans play a crucial role in the detection and diagnosis of Covid-19, as they can provide detailed information about the extent and severity of lung involvement in patients. Our framework for Covid-19 detection, with the preprocessing steps and the modified ResNet backbone network will be described in detail in Section 3.2.

### 3.1. Dataset

The COV19-CT-DB dataset [16][17] was used to evaluate the performance of our proposed approach for automated Covid-19 detection from chest CT scans. This dataset consists of 3D CT scans of the chest labelled by experts, corresponding to a large number of patients. In total, the dataset contains 3746 CT scans, with 1147 scans labelled as Covid and 2599 scans labelled as non-Covid. Each scan in this dataset contains a different number of axial slices, and these slices were provided as JPEG images of size  $512 \times 512$  after being appropriately preprocessed. Specifically, the images were ex-

Annotation	Train data	Validation data	Test data
Covid	922	225	-
Non-Covid	2110	489	-
Total	3032	714	4308

**Table 1.** Dataset composition into training and validation sets.

tracted from the DICOM file and the voxel intensity values were clipped using a window level of 350 Hounsfield Unit (HU), to enhance soft-tissue. The scans were then divided into training and validation sets, with 3032 scans used for training and 714 scans used for validation. The test set consists of 4308 CT scans, but no labels were provided for this set. This is because the test set is intended to be used for evaluating the performance of trained models on new, previously unseen data. The dataset composition is summarised in Table 1. This dataset was intended for both challenges of the 3rd Covid-19 competition. For the first challenge, the dataset was divided into Covid and non-Covid labels. For the second challenge, the same dataset was used, but with the Covid scans labelled according to four levels of severity: Mild, Moderate, Severe, and Critical.

### 3.1.1. Pre-processing

The COV19-CT-DB dataset was preprocessed with the aim to extract only the relevant scans for the Covid-19 detection task. Since not all scans in the dataset have the same length, we selected only the scans containing the lung region, which is of primary interest for Covid-19 detection. To extract these scans, a sequence of 50 consecutive scans per patient was selected, starting from the center of each CT. This allowed us to capture a consistent and representative portion of the lung region in each scan ignoring unrepresentative slices. We also ensured that each selected scan had the same resolution of  $112 \times 112$  pixels, by resizing all slices. This preprocessing step was also crucial in reducing the computational cost of having the model process the entire scan. By selecting only a representative subset of the scans and resizing them to a consistent size, the model's processing time was significantly reduced. These 3D sequence obtained represents the input for our model.

### 3.2. Implementation details

Our approach is based on a modified 3D ResNet [18], which is a widely used deep learning architecture for image classification tasks. Specifically, we used the 3D ResNet-18 architecture as our backbone, which consists of 18 layers and is well-suited for processing 3D medical images. To improve the performance of the baseline architecture for Covid-19 detection, we added attention mechanisms to the network. Specifically,

four heads of multi-head attention [19] were added to the architecture, allowing the network to focus on important regions of interest. The attention mechanism enhances the network's ability to identify and discriminate between Covid and non-Covid cases, by focusing on the regions of the lung that are most indicative of the disease. We also modified the ResNet input to make it work with a 1-channel input instead of 3-channel. Figure 1 shows the overall proposed pipeline.

## 4. EXPERIMENTS AND RESULTS

### 4.1. Experimental setup

As explained in Section 3 we trained our proposed network on the preprocessed COV19-CT-DB dataset, using a binary cross-entropy with logits loss function and the Adam optimizer with a learning rate of 0.0001. We used a batch size of 4 and trained the network for 50 epochs. To evaluate the performance of our proposed approach, we used initially the validation set from the challenge dataset. Subsequently, we evaluated the model's ability to generalize to new data by testing it on the test set, which was made available to us. We report the results in terms of Recall, Precision and Macro F1 score, which are commonly used evaluation metrics for binary classification tasks. Note that in the case of the test set, only values in terms of F1 scores will be reported, as we are not aware of the correct label and rely on the results published by the challenge.

### 4.2. General results

The performance of the proposed approach were evaluated through the validation and test set of the COV19-CT-DB using the aforementioned metrics. Tables 2 and 3 show the results of the baseline approach, the results of a simple 3D Resnet-18 model and the results of three different configurations of our model verified on validation and test set, respectively. The tables shows the results obtained by applying the best models from the experiments carried out both the standard ResNet (ResNet3D-18) and the augmented ResNet with the multi-head attention module (ResNet3D-18 + MHA). For each of the two models, the predictions  $p$  at the margin (i.e. the predictions for which the sigmoid output returns  $0.4 \leq p \leq 0.6$ ) considered as "uncertain" were analysed, in order to make a final decision. For these

Architecture	Recall	Precision	Macro F1 Score
Baseline[3]	-	-	0.74
ResNet3D-18[18]	0.8847	0.9138	0.8973
ResNet3D-18 + MHA	0.8867	0.9233	0.9021
Mix ResNet3D-ResNet3D+MHA	0.8907	0.9188	0.9029
Mix ResNet3D+MHA-ResNet3D	<b>0.8905</b>	<b>0.9229</b>	<b>0.9043</b>

**Table 2.** Comparison results with different network configurations employing the validation set of COV19-CT-DB.

Architecture	F1 Covid	F1 Non-Covid	Macro F1 Score
Baseline[3]	-	-	0.76
ResNet3D-18[18]	0.6227	0.9269	0.7749
ResNet3D-18 + MHA	0.6371	0.9324	0.7847
Mix ResNet3D-ResNet3D+MHA	0.6237	0.9281	0.7759
Mix ResNet3D+MHA-ResNet3D	<b>0.6410</b>	<b>0.9330</b>	<b>0.7870</b>

**Table 3.** Comparison results with different network configurations employing the test set of COV19-CT-DB.

cases, a voting scheme was used: these scans were tested a second time with the other model, combined the prediction from both model and considered the scan as Covid if the combined prediction was greater than 1; otherwise, it was considered Non-Covid. These two further experiments are reported as Mix ResNet3D-ResNet3D+MHA and Mix ResNet3D+MHA-ResNet3D in Table 2 and Table 3. Specifically, Mix ResNet3D-ResNet3D+MHA reports the outcomes of using the standard ResNet3D-18 as the first model and the ResNet3D-18 + MHA as the second model, while Mix ResNet3D+MHA-ResNet3D displays the opposite approach. The baseline architecture uses a ResNet combined with a Recurrent Neural Network (RNN) and achieves a macro F1 score of 0.76. The current approach builds upon previous research efforts in developing deep neural architectures for predicting COVID19, following the work of [6][7]. As can be seen from the tables, our proposed approach with attention mechanisms outperformed the baseline model in terms of F1 score. Specifically, the network with attention achieved a Recall of 0.88, a Precision of 0.92 and a macro F1 score of 0.90 on the validation set. On the test set, the network with attention obtained a macro F1 score of 0.7848, which is lower than the validation set but still allowed us to rank fifth above the baseline in the challenge. These results demonstrate the high performance of our proposed approach in detecting COVID-19 from chest CT scans. Overall, our proposed approach provides a promising solution for the automated detection of Covid-19 from chest CT scans, which can be useful in the early detection and management of the disease.

## 5. CONCLUSION AND FUTURE WORKS

In this paper, a fully automated approach for the detection of Covid-19 from chest CT scans using deep learning techniques was presented. Our proposed framework employs a modified

Resnet backbone network with attention mechanisms to detect Covid-19 from 3D CT scans of the chest. The COV19-CT-DB dataset, consisting of 3D CT scans of the chest labeled as Covid and non-Covid, was used to train and validate our proposed approach. The experimental results show that our approach achieves a macro F1 score of 0.7870, which ranks us fifth in the COVID-19 Severity Detection Challenge. Future research will focus on investigating the potential of our framework on a larger dataset, as well as exploring the possibility of incorporating other imaging modalities for Covid-19 detection. Finally, we plan to investigate the potential of transfer learning to improve the performance of our proposed approach.

## 6. ACKNOWLEDGEMENTS

Alessia Rondinella is a PhD candidate enrolled in the National PhD in Artificial Intelligence, XXXVII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma.

Experiments were carried out thanks to the hardware and software granted and managed by iCTLab S.r.l. - Spinoff of University of Catania.

## 7. REFERENCES

- [1] Marco Ciotti, Massimo Ciccozzi, Alessandro Terrinoni, Wen-Can Jiang, Cheng-Bin Wang, and Sergio Bernardini, "The covid-19 pandemic," *Critical reviews in clinical laboratory sciences*, vol. 57, no. 6, pp. 365–388, 2020.
- [2] Dimitrios Kollias, Anastasios Arsenos, and Stefanos Kollias, "Ai-mia: Covid-19 detection & severity

- analysis through medical imaging,” *arXiv preprint arXiv:2206.04732*, 2022.
- [3] Dimitrios Kollias, Anastasios Arsenos, and Stefanos Kollias, “Ai-mia: Covid-19 detection and severity analysis through medical imaging,” in *Computer Vision—ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*. Springer, 2023, pp. 677–690.
  - [4] Francesco Guarnera, Alessia Rondinella, Oliver Giudice, Alessandro Ortis, Sebastiano Battiato, Francesco Rundo, Giorgio Fallica, Francesco Traina, and Sabrina Conoci, “Early detection of hip periprosthetic joint infections through cnn on computed tomography images,” *arXiv preprint arXiv: https://arxiv.org/abs/2304.08942*, 2023.
  - [5] Alessia Rondinella, Elena Crispino, Francesco Guarnera, Oliver Giudice, Alessandro Ortis, Giulia Russo, Clara Di Lorenzo, Davide Maimone, Francesco Pappalardo, and Sebastiano Battiato, “Boosting multiple sclerosis lesion segmentation through attention mechanism,” *arXiv preprint arXiv: https://arxiv.org/abs/2304.10790*, 2023.
  - [6] Dimitrios Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and S Kollias, “Deep transparent prediction through latent representation analysis,” *arXiv preprint arXiv:2009.07044*, 2020.
  - [7] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and Stefanos D Kollias, “Transparent adaptation in deep medical image diagnosis,” in *TAILOR*, 2020, p. 251–267.
  - [8] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias, and Georgios Tagaris, “Deep neural architectures for prediction in healthcare,” *Complex & Intelligent Systems*, vol. 4, no. 2, pp. 119–131, 2018.
  - [9] Maria Ausilia Napoli Spatafora, Alessandro Ortis, and Sebastiano Battiato, “Mixup data augmentation for covid-19 infection percentage estimation,” in *Image Analysis and Processing. ICIAP 2022 Workshops: ICIAP International Workshops, Lecce, Italy, May 23–27, 2022, Revised Selected Papers, Part II*. Springer, 2022, pp. 508–519.
  - [10] Himadri Mukherjee, Subhankar Ghosh, Ankita Dhar, Sk Md Obaidullah, KC Santosh, and Kaushik Roy, “Deep neural network to detect covid-19: one architecture for both ct scans and chest x-rays,” *Applied Intelligence*, vol. 51, pp. 2777–2789, 2021.
  - [11] Matteo Polsinelli, Luigi Cinque, and Giuseppe Placidi, “A light cnn for detecting covid-19 from ct scans of the chest,” *Pattern recognition letters*, vol. 140, pp. 95–100, 2020.
  - [12] Francesco Rundo, Angelo Genovese, Roberto Leotta, Fabio Scotti, Vincenzo Piuri, and Sebastiano Battiato, “Advanced 3d deep non-local embedded system for self-augmented x-ray-based covid-19 assessment,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 423–432.
  - [13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
  - [14] Debaditya Shome, T Kar, Sachi Nandan Mohanty, Prayag Tiwari, Khan Muhammad, Abdullah AlTameem, Yazhou Zhang, and Abdul Khader Jilani Saudagar, “Covid-transformer: Interpretable covid-19 detection using vision transformer for healthcare,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 21, pp. 11086, 2021.
  - [15] Koushik Sivarama Krishnan and Karthik Sivarama Krishnan, “Vision transformer based covid-19 detection using chest x-rays,” in *2021 6th International Conference on Signal Processing, Computing and Control (IS-PCC)*. IEEE, 2021, pp. 644–648.
  - [16] Anastasios Arsenos, Dimitrios Kollias, and Stefanos Kollias, “A large imaging database and novel deep neural architecture for covid-19 diagnosis,” in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2022, p. 1–5.
  - [17] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian, and Stefanos Kollias, “Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, p. 537–544.
  - [18] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri, “A closer look at spatiotemporal convolutions for action recognition,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 6450–6459.
  - [19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.