Collective Intelligent Strategy for Improved Segmentation of COVID-19 from CT

Surochita Pal Das*, Sushmita Mitra, and B. Uma Shankar

Machine Intelligence Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata, 700108, West Bengal, India.

December 21, 2022

Abstract

The devastation caused by the coronavirus pandemic makes it imperative to design automated techniques for a fast and accurate detection. We propose a novel non-invasive tool, using deep learning and imaging, for delineating COVID-19 infection in lungs. The Ensembling Attention-based Multi-scaled Convolution network (EAMC), employing Leave-One-Patient-Out (LOPO) training, exhibits high sensitivity and precision in outlining infected regions along with assessment of severity. The Attention module combines contextual with local information, at multiple scales, for accurate segmentation. Ensemble learning integrates heterogeneity of decision through different base classifiers. The superiority of EAMC, even with severe class imbalance, is established through comparison with existing state-of-the-art learning models over four publicly-available COVID-19 datasets. The results are suggestive of the relevance of deep learning in providing assistive intelligence to medical practitioners, when they are overburdened with patients as in pandemics. Its clinical significance lies in its unprecedented scope in providing low-cost decision-making for patients lacking specialized healthcare at remote locations.

Keywords— Ensembling, Deep learning, COVID-19 segmentation, Class imbalance, Multi-scaling

1 Introduction

The recent pandemic, called the novel coronavirus-disease-2019 (COVID-19), has been a major threat to world-health [29]; with medical systems collapsing around the globe. It resulted in an increasing demand for health services, encompassing finite components like beds, critical medical equipment, and healthcare workers (who also get regularly infected). Even the year 2022 has seen proliferation of newer strains of the virus affecting humankind. Some of the major COVID-19 complications, in case of serious level of infection, include acute respiratory distress syndrome (ARDS), pneumonia, multi-organ failure, septic-shock, and even death. Serious illness is more likely to

*corresponding Author.
email(s): pal.surochita@gmail.com; sushmita@isical.ac.in; uma@isical.ac.in
All the authors contributed equally to this research.

This work was supported by the J. C. Bose National Fellowship, sanction no. JCB/2020/000033 of S. Mitra

result in people with existing co-morbidities. Often there exist long term side-effects in post-COVID patients.

An early detection, diagnosis, isolation and prognosis, play a major role in controlling the spread of the disease. Computed tomography (CT) and X-rays are the commonly used imaging techniques for the lung. The CT scan uses X-rays to produce a 3D view comprising cross-sectional slices, for detecting existing anomalies. Occurrence of false negatives in the "gold standard" RT-PCR test results often lead to the chest CT scans being an useful supplement in projecting typical infection characteristics – like Ground-Glass Opacity (GGO) and/or mixed consolidations. It was reported [12] that Lung CT images are more sensitive (98%), as compared to RT-PCR (71%), in correctly predicting COVID-19.

Doctors reported difference in CT abnormalities related to COVID-19 patients in multiple studies [10, 17]. It was observed, even at early stages, that viral infections were indicated by clear patterns [5, 10]. In Ref. [17] the researchers assessed the effectiveness of chest CT in the diagnosis and treatment of COVID-19. The CT characteristics of COVID-19 were presented and compared with the manifestations of other viruses.

Abnormalities in CT may occur [10] before the appearance of clinical symptoms. Multifocal, unilateral, and peripherally based GGO are examples of classic patterns, which are also observed in symptomatic cases. Abnormalities like inter-lobular septal thickening, thickening of the surrounding pleura, round cystic alterations, nodules, pleural effusion, bronchiectasis, and lymphadenopathy were infrequently detected in the asymptomatic group.

Manually detecting COVID-affected regions from lung CT scans is time consuming and prone to inherent human bias. Thus automated or semi-automated Computer-Aided-Diagnosis (CAD) becomes necessary [4, 24]. An accurate, automated detection and delineation of the COVID-19 infection is of great importance since this results in an effective monitoring of its spread within the lungs. This helps in predicting the severity of the infection, as well as its prognosis.

Smart machines can imitate the human brain to some extent. Everything that makes a machine smarter falls under the umbrella called Artificial Intelligence (AI). Machine Learning (ML), which is a subset of AI, consists of a collection of algorithms and tools which enable a machine to understand patterns within the data without being explicitly programmed. ML uses this underlying structure to perform logical reasoning for a task. Deep Learning (DL), again, is a sub-domain of ML [14]. It aids a machine in learning hidden patterns within the data without any expert intervention, to make predictions – given high computational power and a massive volume of annotated data. A convolution neural network (CNN), which is a DL model, has been shown to perform effectively in analyzing visual images. A CNN model, which was designed to recognize objects in natural-images from the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), was found to be comparable in efficiency to humans [16].

The U-Net [27] is an encoder-decoder type of CNN architecture, designed for the segmentation of biomedical images in a fast and precise manner. The encoder arm causes the spatial dimension to be decreased, while increasing the number of channels. In contrast, the decoder arm decreases the channels while raising the spatial dimensions. Introduction of attention gates (AGs) [23] in the U-Net framework, help reduce the feature responses in irrelevant background regions, while providing more weight to region of interest (ROI). The network is guided towards learning only the relevant information in terms of the weighted local features. Incorporating dilated convolutions [13] allows feature extraction at multiple scales.

Multi-scalar approaches, which observe and evaluate a dataset at several scales, are popular in the machine learning domain. They capture the local geometry of neighbourhoods, which are characterised by a collection of distances between points or groups of closest neighbours. This is analogous to looking at a portion of a slide at

various microscopic resolutions; whereby, very small features can be detected at high resolution from a restricted region of the sample. As the majority portion of the slide is examined at a lower resolution, it allows one to examine the larger (global) features as well. Multi-scalar methods have been found to perform better than state-of-the-art techniques, with reduced sample sizes, in the medical domain [32].

An ensemble-based classifier system is designed by merging multiple diverse classifier models. Ensembling makes statistical sense in a number of situations. We regularly employ such an approach in our daily lives while seeking the advice of various experts prior to taking a major decision. For instance, we frequently seek the advice of numerous doctors before consenting to a medical procedure. The main objective is to reduce the regrettable choice of a needless medical procedure. The experts must differ from one another in some way for this mechanism to be successful. Individual classifiers, due to their inherent diversity, can produce various decision limits within the context of classification. This is commonly achieved by employing distinct training setup for each classifier. If adequate diversity is established, each classifier will commit a separate error, which may then be strategically combined to lower the overall error [25].

Advantage of ensemble learning lies in its inherent diversity. This can be introduced by embedding different training datasets, or features, or classifiers; or even differing initialization and/or parameters of the classifier(s) involved. According to Dietterich [9] there are three main justifications for employing an ensemble-based system, *viz.* statistical, computational, and representational. The computational criterion refers to the model selection problem. The statistical cause is connected to the insufficiency of available data to accurately represent a distribution. The representational cause addresses situations where the selected model is unable to accurately represent the desired decision boundary.

Numerous studies have been reported in recent times in the domain of COVID-19, using neural networks and data-driven algorithms. These include machine learning approaches for diagnosis of COVID-19 from X-Ray/ CT images [22, 26, 35]. A pre-trained deep-learning model, called DenseNet, was developed [35] for classifying 121 CT-images into COVID-19 positive and negative categories. Application of the ResNet-18 was made [37] to segment and classify lung-lesions of COVID-19, pneumonia infection, and normal ones.

A deep learning based AI system was designed [41] to detect and quantify lesions from chest CT. It can remove scan-level bias to extract precise radiomic features. The Unified CT-COVID AI Diagnostic Initiative (UCADI) [3] enables independent training at each host institution, under a combined learning framework, without data sharing. This was shown to outperform the local models, thereby advancing the prospects of utilizing combined learning for privacy-preserving AI in digital health.

Deep learning has been employed for evaluating the severity of COVID-19 infection [36]. Well-known deep models, like U-Net [27], Residual U-Net [39], Attention U-Net [23], have been used for screening COVID-19. There exist ensemble methods for segmentation of CT images [7, 11]. The Inception-V3, Xception, InceptionResNet-V2 and DenseNet-121 were ensembled [11] for a multiclass segmentation of GGO and Consolidation in COVID-19 CT data over the data CT-Seg (Table 1). Each of these models used the CNN as backbone, with pre-trained weights from ImageNet being further trained over the CT-Seg data. The split into training, validation and testing sets were 40, 10, 50 images, respectively, with pixel-level soft majority voting being employed for their aggregation.

A cascade of two U-Nets, with VGG backbone, was ensembled [7] to extract the lung region, followed by the delineation of the GGO and consolidation regions. Multiclass segmentation of GGO and Consolidation was performed over CT-Seg, Seg-nr.2 and Kaggle-COVID-19 datasets (Table 1) along with some private dataset; while the training data contained parts of CT-Seg and Seg-nr.2, the remaining parts of the datasets

were used for testing the model. The training process of each network in the ensemble differed due to random weight initialization, and data augmentation with shuffling.

Domain Extension Transfer Learning (DETL) was employed [6] for the screening of COVID-19, with characteristic features being determined from chest X-Ray images. In order to get an idea about the COVID-19 detection transparency, the authors employed the concept of Gradient Class Activation Map (Grad-CAM) for detecting the regions where the model paid more attention during the classification. The results are claimed to be strongly correlated with clinical findings.

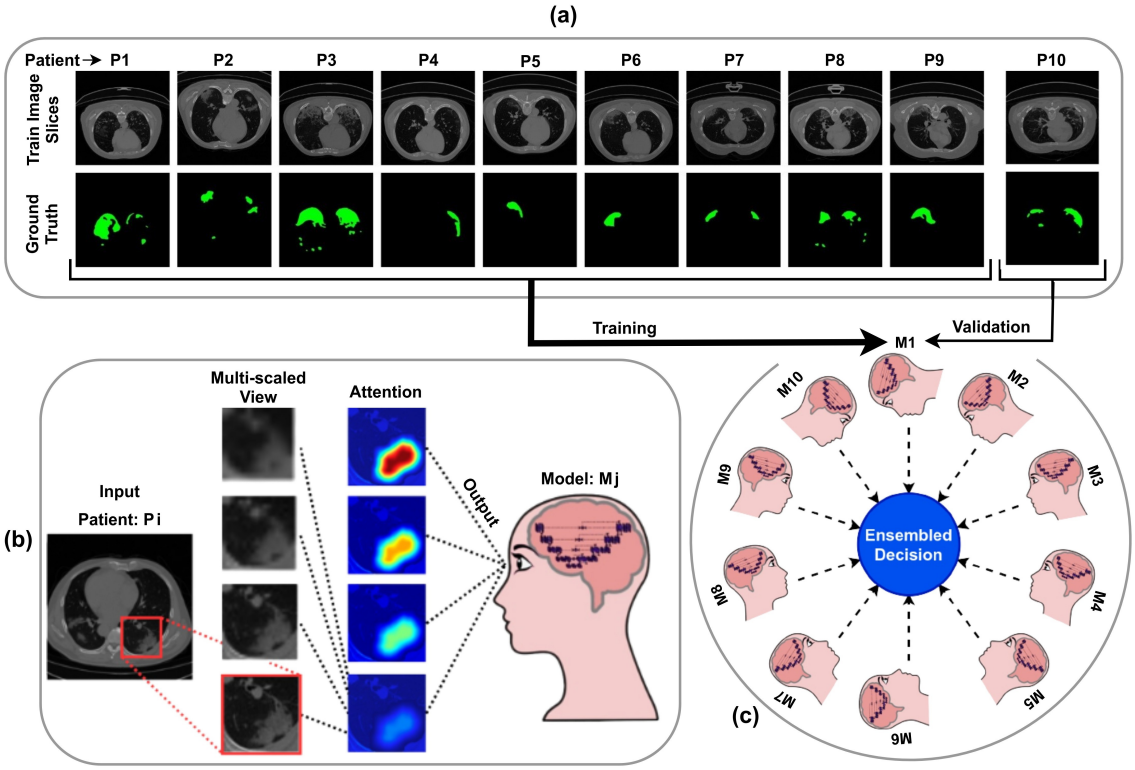


Figure 1: Schematic representation of Ensembling with LOPO the Attention-based Multi-scaled CNNs. (a) Leave-One-Patient-Out scheme *LOPO*. (b) Attention-based Multiscaled view in *AMC*. (c) Ensembling of the *AMC* Models

2 Results

We propose a novel Ensembled Attention-based Multi-scaled Convolution networks (*EAMC*), using *LOPO* learning, based on CT images of TEN patients and trained using TEN base-classifiers. As each classifier takes only nine samples (patients) for training, and starts from scratch, it does result in a completely new classifier with different set of parameters. The remaining ONE sample is left for validation in each case (as elaborated in Section 2.2). These TEN trained classifiers are ensembled to segment the COVID-infection region (ROI) through majority voting, over four different test datasets collected from various publicly available sources (Table 1). The workflow of the *EAMC* is visualized in Fig. 1.

The objective of the model is to segment a COVID-19 infected CT lung image into its Regions of Interest (ROI) [*i.e.*, GGO and Consolidation], and background (containing all other regions in the image). This is not an easy task for a vanilla *U*-Net. Therefore, an *AG* is carefully incorporated to focus on the ROI, whereas the multi-scalar dilation provides the necessary local and neighbourhood information representation for

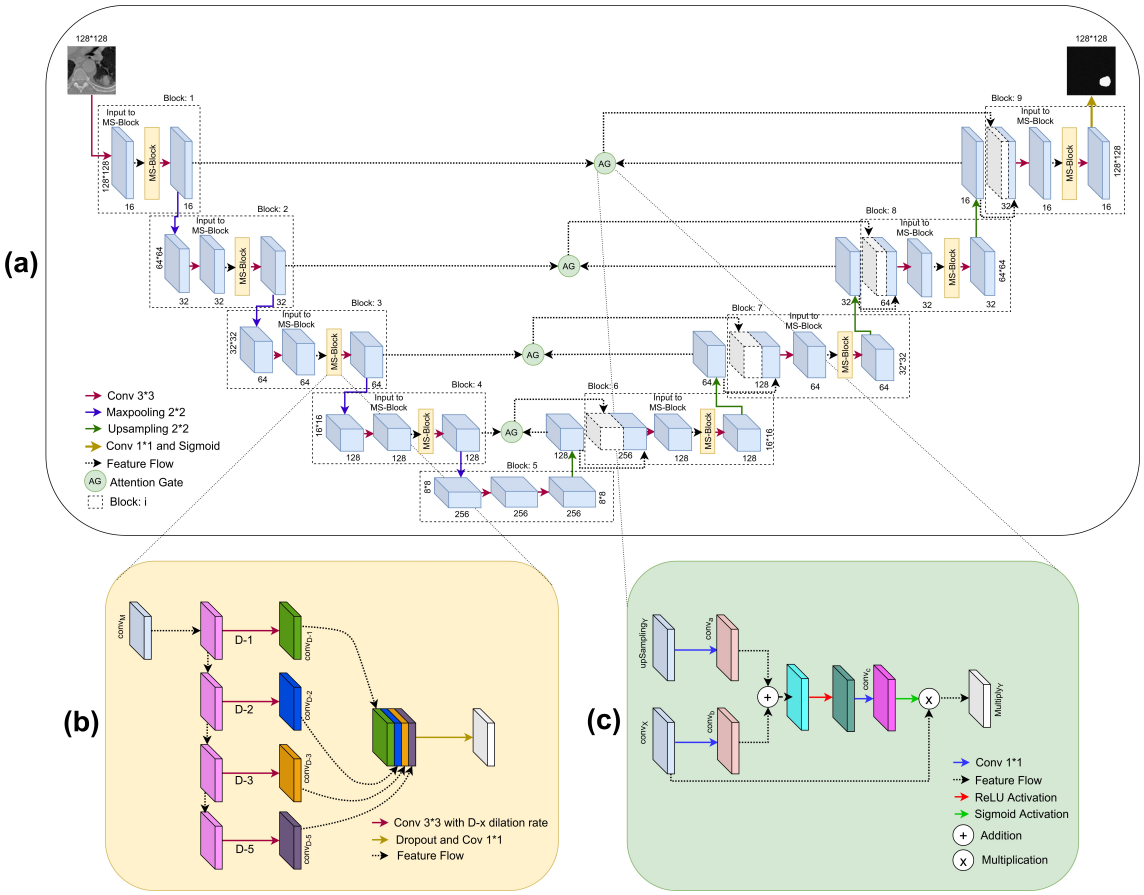


Figure 2: (a) The *AMC-Net* framework, with detailed representation of (b) of *MS*-block, and (c) Attention Gate (*AG*)

delineating the ROI. Focal Loss (*FL*) [eqn. (3)] is used as the loss function to compensate for class imbalance. Finally we implement ensembling by *LOPO* for training with limited annotated data. The details about the training dataset are provided in Section 5. Validation is successfully performed on unseen, independently-compiled data from multiple publicly-available data sources. The proposed *EAMC* of Fig. 1, using LOPO ensemble learning on *AMC-Net* of Fig. 2, demonstrates good generalization capability. It is efficient, accurate, consistent and robust on the unseen data. The architecture of *AMC-Net* is summarized in Tables 7-9.

2.1 Data used

Four datasets, with details as provided in Table 1, were used in this study. The Kaggle-COVID-19 data comprises 20 patient samples, of which ten (having at least 200 but not more than 301 slices) were kept for training. This was named as Kaggle-COVID-19:Part-1 dataset. The remaining ten samples, each containing < 200 or > 301 slices, were retained for testing. This was termed the Kaggle-COVID-19:Part-2 dataset. The data is available on the Kaggle platform[1] in annotated form [19]. The other three datasets, of Table 1, were also clubbed together for testing. The unseen test datasets were thus used only for evaluating the generalization performance of *EAMC*.

Table 1: Breakup of Infected & Non-Infected samples and slices, in the training and test datasets

Sr.	Dataset	Total patients	Infected slices	Non-Infected slices	Comment
1	Kaggle-COVID-19[a]: Part-1	10	1351	1230	Training
	Kaggle-COVID-19[a]: Part-2	10	493	446	Testing
2	CT-Seg[b]	>40	100	0	Testing
3	Seg-nr.2[b]	9	372	457	Testing
4	MosMed[d]	47[§]	761	1166	Testing

[a]Kaggle-COVID-19:
`https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans` [19]
[b]CT-Seg & Seg-nr.2 (Two Datasets): `http://medicalsegmentation.com/covid19/`
[c]MosMed: `https://mosmed.ai/en/` [21]
[§]Relevant 47 samples are used (actual available samples being 50).

Table 2: Description of Ensemble models, with DSC on corresponding validation sets

Model No.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Training	P/{P10}	P/{P9}	P/{P8}	P/{P7}	P/{P6}	P/{P5}	P/{P4}	P/{P3}	P/{P2}	P/{P1}
Validation	P10	P9	P8	P7	P6	P5	P4	P3	P2	P1
DSC	0.8882	0.8952	0.8617	0.8814	0.8432	0.8867	0.8944	0.8821	0.8787	0.8615
where P = { Pi: $i \in \mathbb{N} \wedge i \in [1, 10]$ }										
Mean validation DSC = 0.8773, with Standard deviation = 0.0158										

2.2 Implementational details

The ten classifier models considered are $M1$, $M2$, ..., $M10$; with the validation datasets defined as $P10$, $P9$, ..., $P1$, and described in Table 2. In each case, the rest of the corresponding patient's dataset is used for training by the AMC-Net model of Fig. 2. For example, in case-1, $M1$ is trained with $P1$ to $P9$ patient datasets and validated on $P10$ patient dataset. Similarly in case-2, $M2$ is trained with $P1$ to $P8$ and $P10$ patient datasets and validated on $P9$ patient dataset, and so on for all the models $M3$ to $M10$. Each model, $M1$ to $M10$, is trained with different parameters, pertaining to initialization, learning rate, dropout probability, etc. The batch size was kept uniform at 16, using the Adam optimizer [15] over 70 epochs. Values of learning rate and dropout probability were set at 0.001 and 0.2, respectively, after several experiments. Run time augmentation (rotaion $\pm$ 10°; horizontal shift Range $\pm$ 0.2; vertical shift Range $\pm$ 0.2; zoom range $\pm$ 0.2) was employed during training.

The implementation was made in the Tensorflow framework, running behind wrapper library Keras using python version 3.6, Keras version 2.2.4, and Tensorflow-GPU version 1.13.1, with dedicated GPU (NVIDIA TESLA P6 having capacity of 16GB).

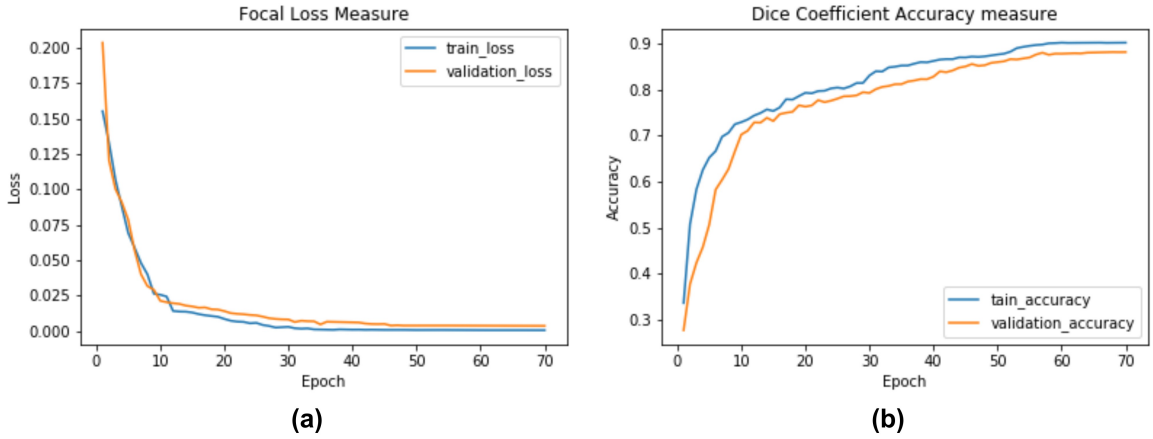


Figure 3: Sample learning curves, for Model M3, over training and validation sets; (a) loss: FL, and (b) accuracy: DSC

2.3 Experimental outputs

Sample learning curves, depicting loss [*Focal Loss (FL)* of eqn. (3)] and accuracy [*Dice Score Coefficient (DSC)* of eqn. (9)], are illustrated in Fig. 3. The validation set corresponding to each model, with the resultant (DSC), are presented in Table 2. The accuracy and robustness of output in each case is evident, showing an average DSC of 0.8773 with a standard deviation (SD) of 0.0158 on the validation datasets.

Next the explainability of the *AMC-Net* architecture was analysed. As evident from Fig. 4, each level of the encoder and decoder arms of the network architecture demonstrate extraction of meaningful features at different levels of abstraction. The abstraction level of the extracted features are dependent on the level of the Block. While some bright objects do get highlighted at the initial stages, as the Block depth increases the Attention and Multi-Scalar mechanisms of the $AMC - Net$ assist in highlighting the ROIs (like, GGO and Consolidation) present in the CT patches and suppress the background.

For example, after Block 1 there are some features highlighting healthy lung tissue (yellow box) in the figure. There are also relevant edges corresponding to lung parenchyma (orange box). As the Block level increases, the attention over the lesion boundary (blue box) and lesion (green box) in Fig. 4 become more prominent. Eventually the final feature which prominently emerges is the lesion.

2.4 Comparative study

Using the *AMC-Net* as the base classifier, with ten classifiers $M1$ to $M10$ ensembled by $LOPO$, results were generated on the four test datasets, *viz.* Kaggle-COVID-19: Part2, CT-seg., Seg-nr.2, and MosMed (as described in Table 1 and Section 5) in terms of the performance metrics defined in eqns. (9)-(11).

The comparative analysis of output generated by the $EAMC$, on different test sets, is presented in Fig. 5. In all cases the test slices were divided into non-overlapping patches (as elaborated in Section 5.1). The segmentation output is aggregated using majority voting.

It is observed that all the metrics provided a consistently better performance over the MosMed data. This is perhaps because the average intensity of the CT scans in MosMed data is higher than that of the rest of the CT datasets used; thereby, providing

[1] `https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans`

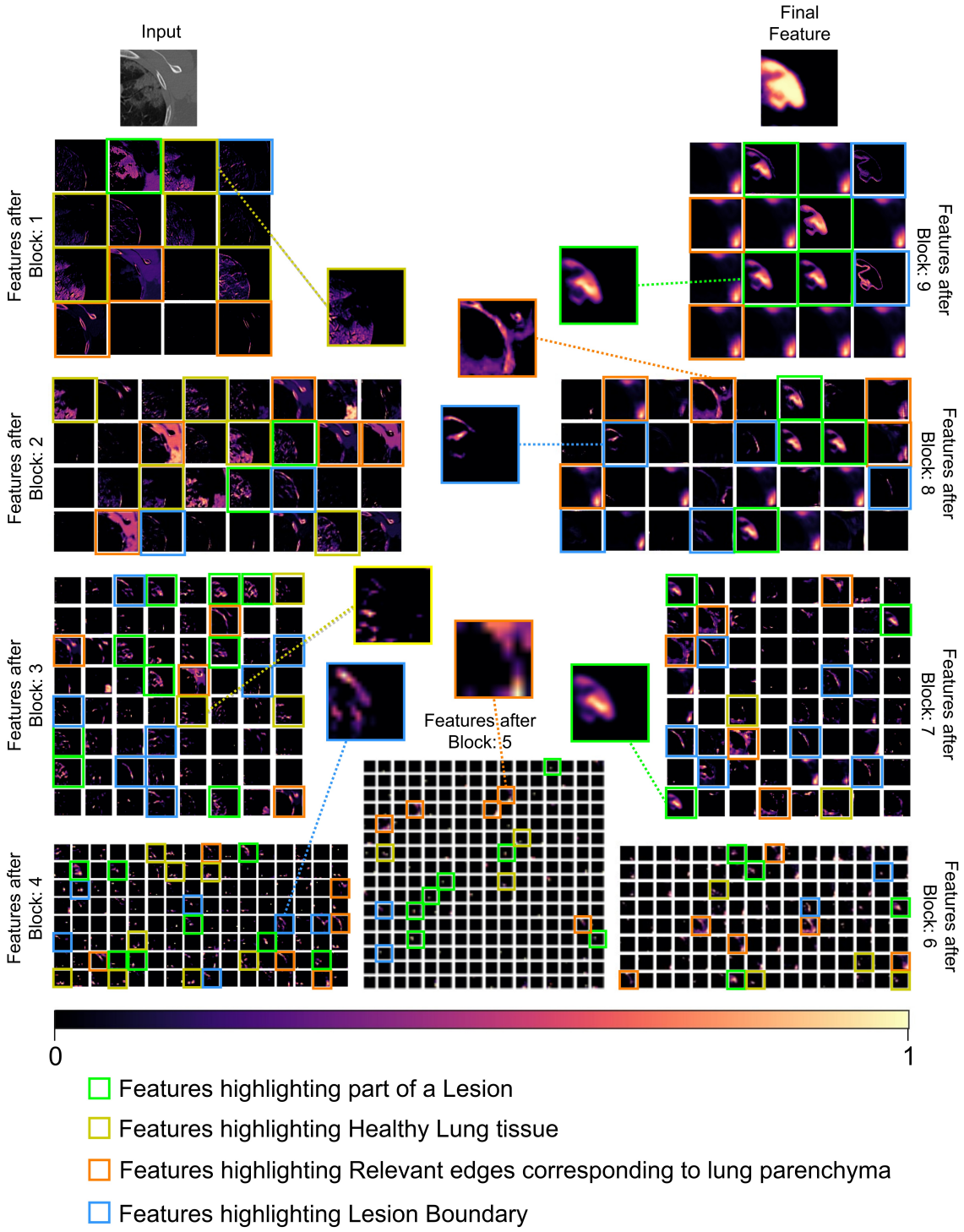
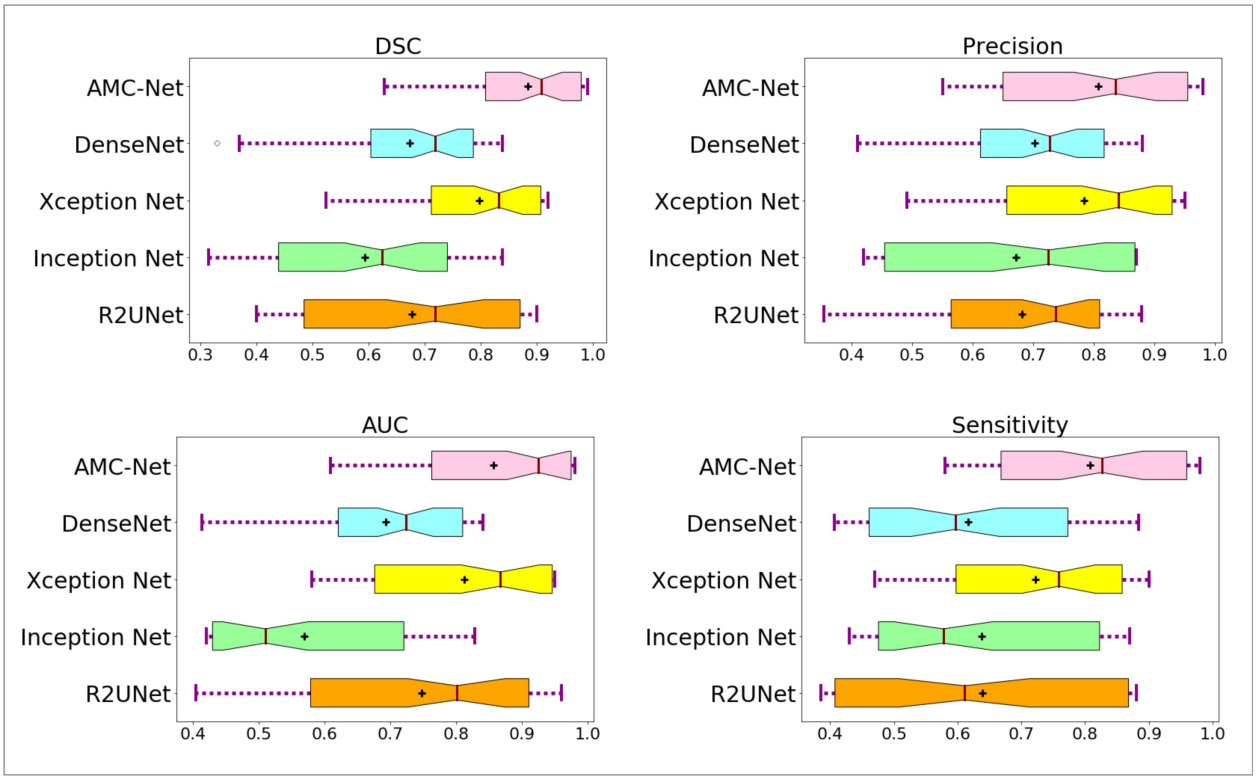


Figure 4: Visualization of features extracted after each Block of the *AMC-Net* from Fig 2

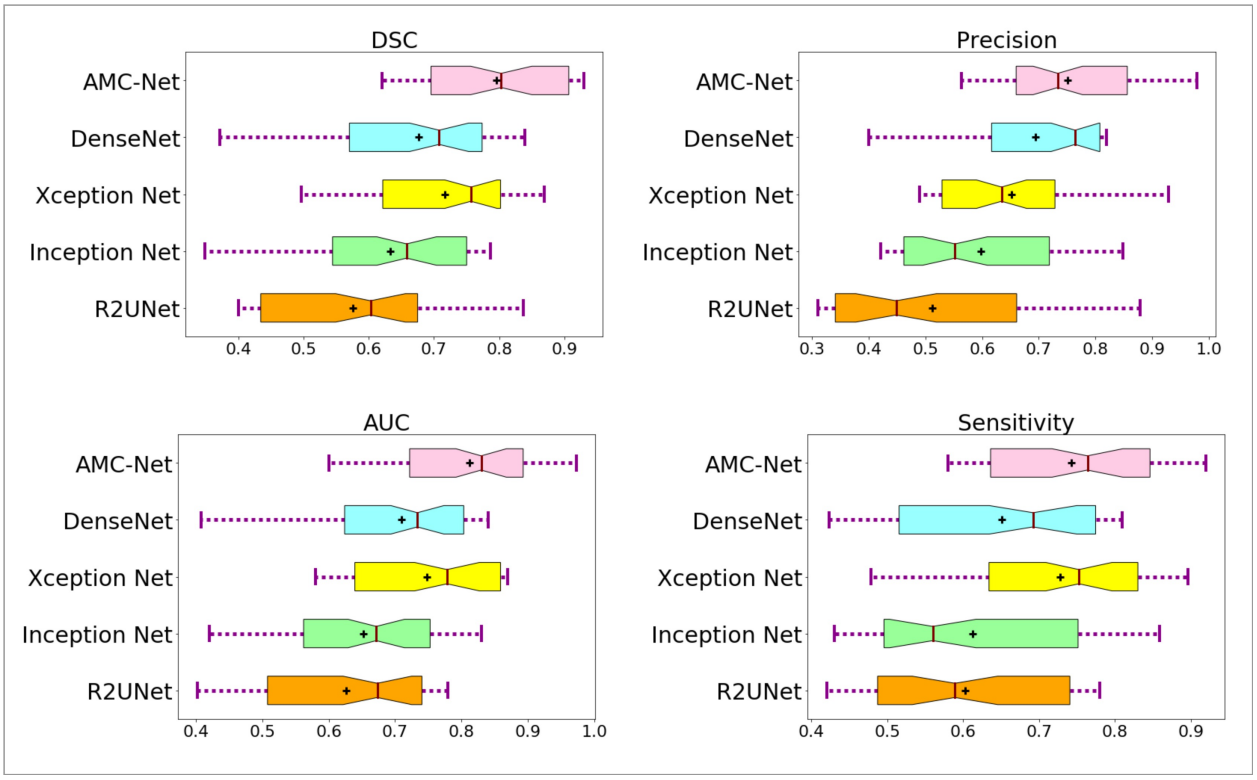
a better contrast between the ROI and background. Moreover the Hounsfield range is not distorted here, such that there is less noise present.

In order to explore the effectiveness of *AMC*-Net as the base classifier in *EAMC*, we also compared four state-of-the-art models like R2UNet [2], Inception Net [11], Xception Net [11], and DenseNet [11] (which have been extensively employed in COVID-19 segmentation literature). Ensembling of classifiers, in each case, was by *LOPO* during training (involving same hyper-parameters). Testing was performed on the four datasets, as elaborated earlier. A comparative study of the evaluation metrics [eqns. (9)-(11)] is provided in Fig. 5. It is found that all the metrics resulted in higher values

for the ensembled *AMC-Net*, indicating a better generalization in segmentation over each test dataset.

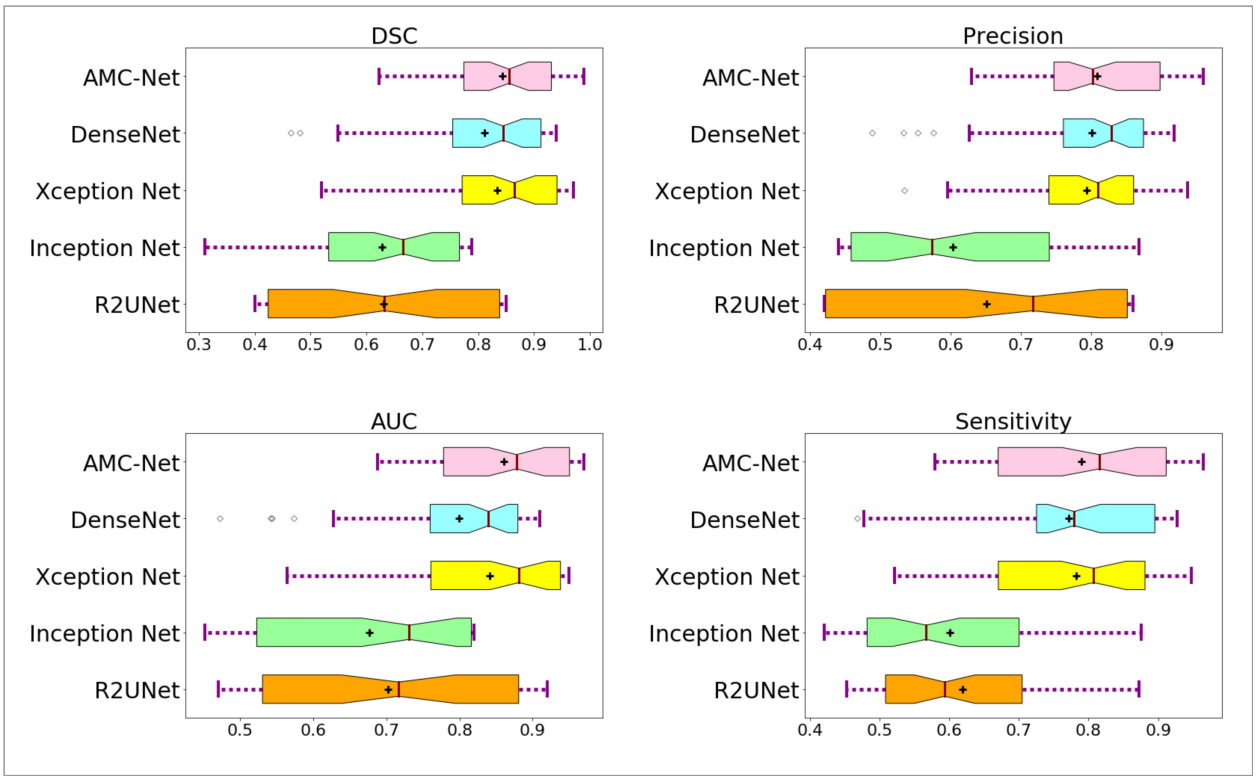


(a) Comparison over Kaggle Data

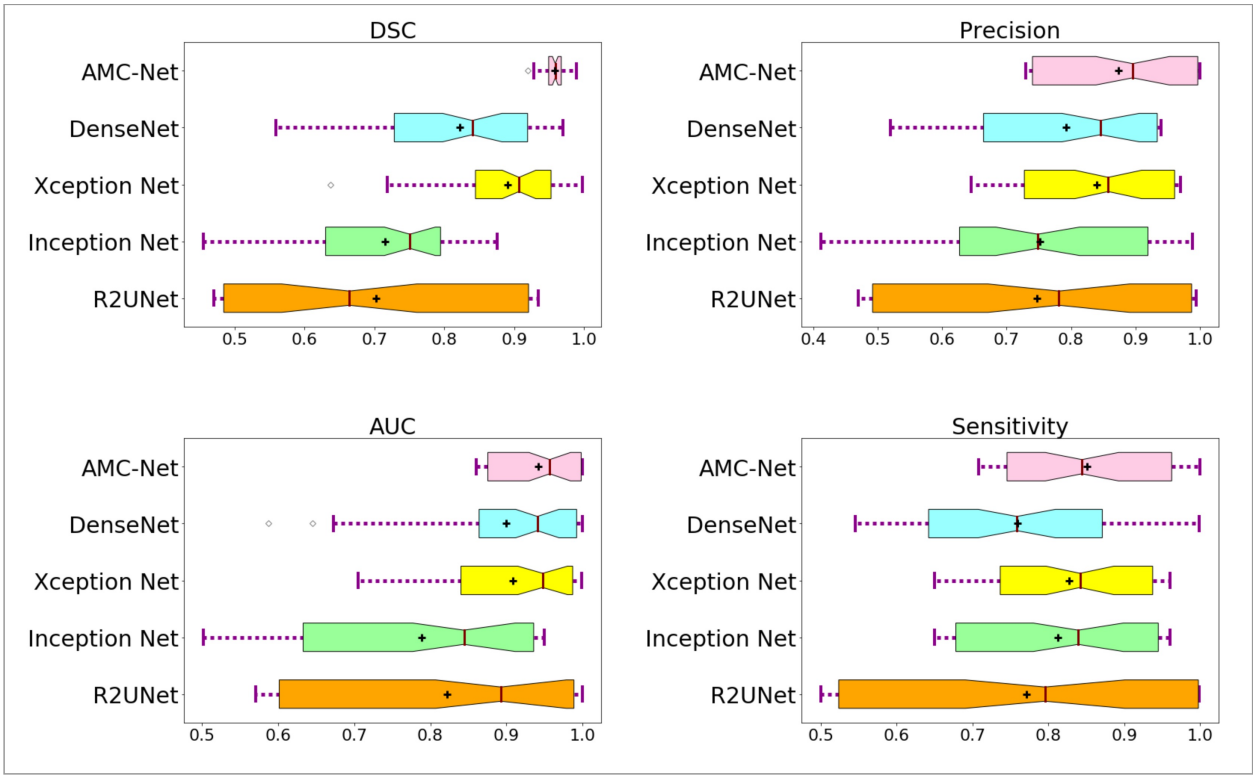


(b) Comparison over CT-Seg Data

Figure 5: Comparative performance evaluation of base classifiers, in the uniform framework of ensembling with *LOPO*, over the test datasets (continued)



(c) Comparison over Seg-nr.2 Data



(d) Comparison over MosMed Data

Figure 5: Comparative performance evaluation of base classifiers, in the uniform framework of ensembling with *LOPO*, over the test datasets

Summarized below is the performance of our *EAMC* over the four test datasets under consideration.

- Kaggle-COVID-19: DSC 0.8840 $\pm$ 0.103, $Precision$ 0.8072 $\pm$ 0.148, AUC 0.8562 $\pm$ 0.126, $Sensitivity$ 0.8082 $\pm$ 0.145;
- CT-Seg: DSC 0.7955 $\pm$ 0.107, $Precision$ 0.7500 $\pm$ 0.121, AUC 0.8123 $\pm$ 0.099, $Sensitivity$ 0.7431 $\pm$ 0.114;
- Seg-nr.2: DSC 0.8431 $\pm$ 0.101, $Precision$ 0.8089 $\pm$ 0.096, AUC 0.8613 $\pm$ 0.091, $Sensitivity$ 0.7905 $\pm$ 0.125;
- MosMed: DSC 0.9584 $\pm$ 0.014, $Precision$ 0.8733 $\pm$ 0.112, AUC 0.9417 $\pm$ 0.056, $Sensitivity$ 0.8514 $\pm$ 0.107.

An investigation into related results on COVID-19 segmentation with deep networks, as reported in literature [7, 8, 19, 28, 31, 34] using some of the same test datasets, led to interesting conclusions with respect to our $EAMC$. The Seg-nr.2 test set yielded DSC of 0.673 [19], 0.620 [28]. The model [28] employed generative adversarial network (GAN) with U-Net as backbone, and reported a $Sensitivity$ of 0.672. Their results on the MosMed test set show DSC 0.584, with $Sensitivity$ 0.768 . Using the Kaggle-COVID-19 dataset the authors reported [34] DSC 0.7103, $Sensitivity$ 0.6860. Combination of the test sets Kaggle-COVID-19 and Seg-nr.2 was also reported. The authors in Ref. [8], using the R2UNet as backbone, obtained a DSC of 0.851. On the other hand, using the basic U-Net as backbone [7] attained DSC 0.80. The Medseg (combination of CT-Seg and Seg-nr.2) dataset resulted in DSC 0.77 [31]. This establishes the effectiveness of our $EAMC$, the ensembled classifier using AMC-Net, in terms of these compared performance metrics over all test datasets.

2.5 Ablations

The first set of experiments were performed by evaluating the role of each of the components in AMC-Net, *viz.* MS-block and AG, when used in $EAMC$. The traditional U-Net is thus the baseline, with Attention U-Net incorporating only the AG. The U-Net with only the MS-block is termed the MSU-Net. The AMC-Net is the U-Net with both AG and MS-block. All four models were trained using the same ensembled $LOPO$ framework. Comparative results on the four test datasets of Table 1, evaluated in terms of the performance metrics of eqns. (9)-(11), and Area Under the ROC Curve (AUC), are presented in Fig. 6. Here It is observed that the proposed AMC-Net performs the best, over all the metrics in all test datasets, as compared to the rest (lacking one or more of its modules). This helps justify the effectiveness of the proposed $EAMC$, which ensembles with LOPO a set of AMC-Net models.

The second task was to explore the effect of the different loss functions of eqns. (3)-(8) on the performance of the base model AMC-Net, without any ensembling. Results are provided in Table 3, over all the four test datasets. Here the ROI (GGO and consolidation) covers a minuscule portion of a CT slice. The Focal loss FL is found to be the best because of its capability in handling class imbalance; thereby, reducing misclassification error. As it predicts the outcome as probability, it can better distinguish between grades of severity in outcome. A mechanism of down-weighting the easier samples while emphasizing the more challenging ones, helps FL focus on the smaller ROIs while suppressing the background regions.

Note that ensembling in $EAMC$ enhances the corresponding performance (as reported in Fig. 6) in terms of DSC.

Finally investigations were pursued with different Dilation rates in the convolution layers of the MS-block. It was observed that the combination $D = 1$, 2, 3, and 5, provides the best results over the test data, in terms of average DSC.

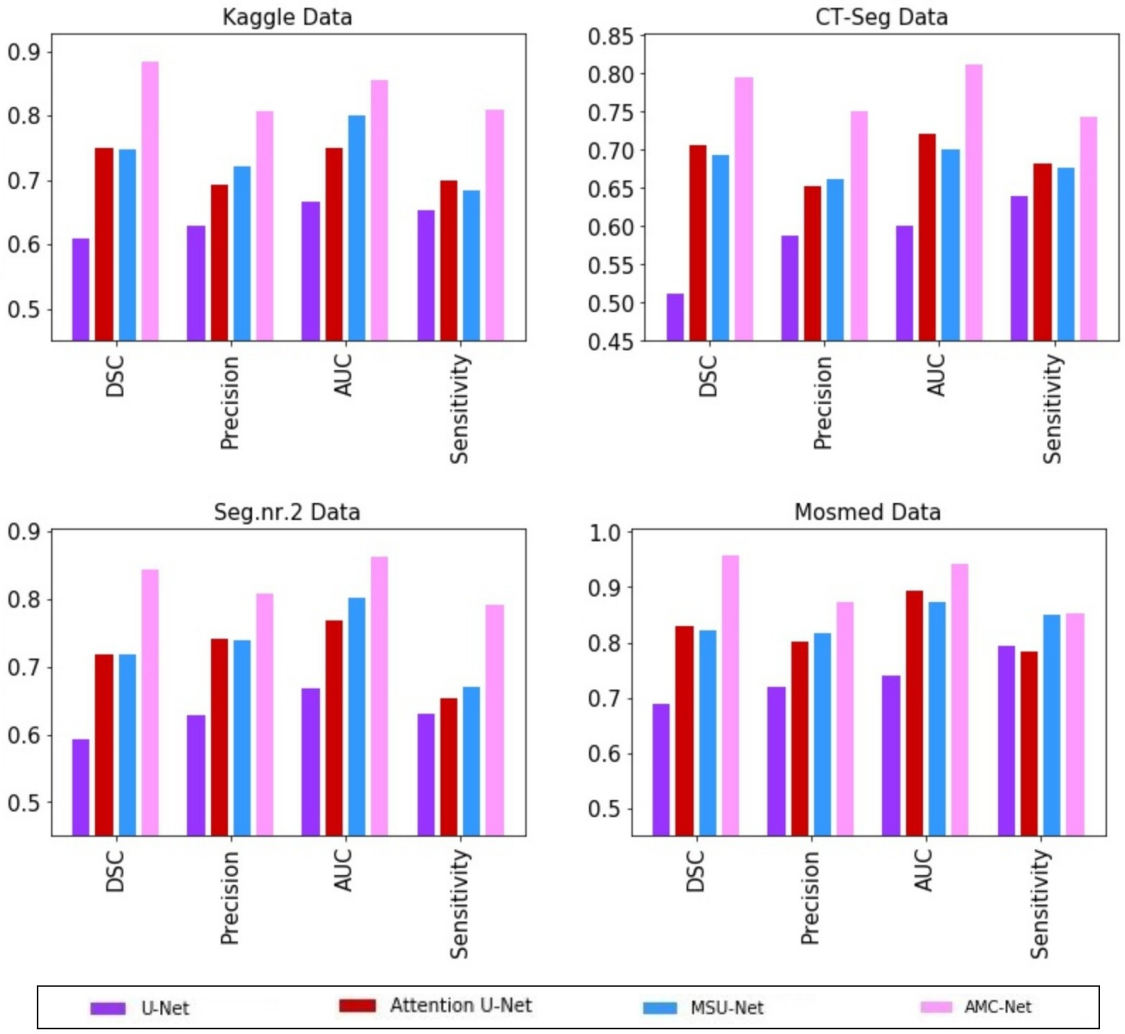


Figure 6: Role of different components of *AMC*-Net, as base classifier in *EAMC*, over the test datasets

2.6 Severity assessment

The methodology of grading the severity of COVID-19 infection, as developed by the Russian Federation [20], is based on individually computing the volume ratio of lesions in each lung and using the maximal value to assess the overall severity score. The range of the score is divided into five categories based on volume of damaged lung tissue.

- CT-0: not consistent with pneumonia (including COVID-19), *ie*, normal
- CT-1: infection involvement of ≤ 25 %
- CT-2: infection involvement of 25-50 %
- CT-3: infection involvement of 50-75 %
- CT-4: infection involvement of 75-100 %

Patients with CT-3 (severe pneumonia) or higher are typically hospitalized, with CT-4 (acute pneumonia) required to be admitted to an intensive care unit.

Table 5 quantifies the infected region in the sample test images, comprising the eight slices (two each from the four sets of test data) of Fig 8, along similar lines.

Fig. 7 depicts the qualitative assessment of the severity of infection, over the same eight sample slices, based on the color mask of the segmented output generated by JET Colormap [33]. The visualization uses the predicted probability values of the various

Table 3: Effect of various loss functions on AMC-Net, measured in terms of DSC, over the test datasets

Loss function	Kaggle-COVID-19: Part-2	CT-Seg	Seg-nr.2	MosMed
DL	0.7521 ± 0.174	0.7238 ± 0.201	0.7218 ± 0.136	0.8465 ± 0.251
$CEDL$	0.7272 ± 0.142	0.7377 ± 0.193	0.7996 ± 0.147	0.8129 ± 0.180
IoU	0.7937 ± 0.137	0.7190 ± 0.165	0.7660 ± 0.125	0.8706 ± 0.197
FL	**0.8706 ± 0.129**	**0.7821 ± 0.151**	**0.8315 ± 0.132**	**0.9513 ± 0.096**
TL	0.7103 ± 0.176	0.7201 ± 0.130	0.7542 ± 0.203	0.8360 ± 0.071
FTL	0.8238 ± 0.192	0.7725 ± 0.144	0.8285 ± 0.194	0.8973 ± 0.214

Table 4: Effect of varying Dilation rate D on AMC-Net, measured in terms of DSC over the test datasets

D in MS-block	Kaggle-COVID-19: Part-2	CT-Seg	Seg-nr.2	MosMed
1,2,3,4	0.8017 ± 0.182	0.7018 ± 0.176	0.7552 ± 0.150	0.8910 ± 0.101
1,2,4,8	0.7482 ± 0.149	0.6617 ± 0.168	0.7001 ± 0.152	0.8527 ± 0.107
1,2,3,5	**0.8706 ± 0.129**	**0.7821 ± 0.151**	**0.8315 ± 0.132**	**0.9513 ± 0.096**

Table 5: Prediction of affected region, by $EAMC$, considering two samples slices from each test dataset

Dataset	Sample no. in Fig. 8	Ground truth (%)	Prediction (%) by $EAMC$
Kaggle-COVID-19: Part-2	S1	57.63	57.71
	S2	85.28	84.71
CT-Seg	S3	89.11	88.12
	S4	78.82	79.34
Seg-nr.2	S5	60.91	61.15
	S6	62.71	62.18
MosMed	S7	11.62	10.98
	S8	13.74	14.02

regions, to illustrate the grading of severity. Here red corresponds to the most severe, yellow indicates moderate, and blue refers to the least severe infections. Such analysis can be of assistance in predicting the grading of infection in a sample patient, along with an estimation of the expected prognosis.

3 Discussion

Medical imaging provides useful assistance for the safe, efficient, and early detection, diagnosis, isolation, and prognosis of diseases. It enables non-invasive examination of the interior organs, bones and tissues, allowing for accurate assessment of disease severity. Particularly, with the advancement in CT imaging technology, very high resolution images serve as suitable diagnostic tool in the medical domain. The recent COVID-19

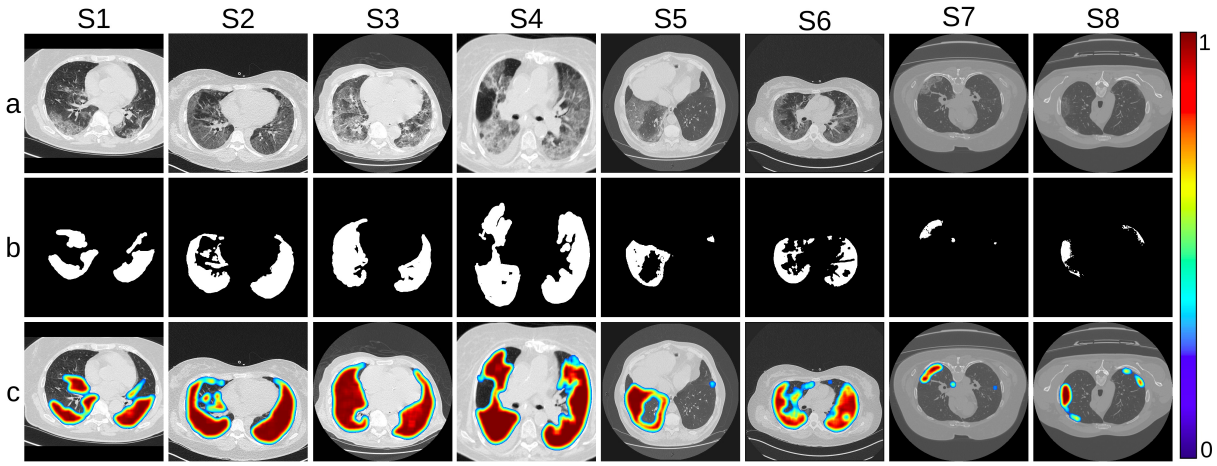


Figure 7: Visualization of infection, in the eight sample slices from Fig. 8, as predicted by *EAMC*. (a) Input CT scan, (b) corresponding annotation, and (c) JET ColorMap on the prediction

pandemic demonstrated that CT images are often more accurate as compared to the standard RT-PCR tests, which can exhibit False Positives. The CT images can provide a lot more information, like the severity of infection and the presence and distribution of pathologies like GGO and Consolidation. Therefore, CT images are gradually becoming a primary tool for the detection and prognosis in COVID-19. As the increased number of cases for diagnosis made the job over-burdening, the need for automated and more accurate segmentation, detection and analysis became evident.

Our research using deep convolutional networks in medical imaging analysis, demonstrated efficient extraction of valuable features. The results depict how observable features from the COVID-19 ROI, encompassing GGOs and consolidations, could be effectively retrieved from various blocks at different levels. The severity of the disease could also be assessed. The qualitative and quantitative statistics illustrate the superiority of our model with respect to related methods in literature. The qualitative output demonstrates that the proposed *EMAC* generates very few misclassified pixels corresponding to the ROI. The data revealed the presence of a substantial proportion of pixels from the background region, with a relatively smaller number from the ROI. Such a significant class imbalance is effectively addressed by the loss function discussed. Our patch-based method performs exceptionally well, in terms of accuracy and loss over training and validation, for an effective management of overfitting in deep learning. Use of CT images, obtained from several other sources, established the robustness of our methodology in handling imaging differences at the source.

A novel ensembling method by *LOPO* was developed for a collective, efficient delineation of COVID-19 affected region in the lung, along with a gradation of the severity of the disease, using very limited training data. Multi-scalar attention with deep supervision enabled enhanced accuracy, in terms of improved sensitivity and precision in segmentation of ROI, for the proposed model *EAMC*. The loss function helped focus on the imbalanced representation of the ROIs, in terms of GGOs and consolidations in the CT slices. While the training was performed on one set of annotated data, the testing set comprised of an assortment of data from different sources of publicly available sets. The superiority of the network was thus established in a broader generalization framework.

Sample qualitative results in Fig. 8, corresponding to the models compared in Fig. 5, help establish the robustness and effectiveness of *EAMC* under ensembling by *LOPO* on the backbone *AMC*-Nets. The output is evaluated in terms of the annotated masks. The eight samples explored were (i) $S1$, $S2$ from Kaggle-COVID-19: Part-2; (ii) $S3$,

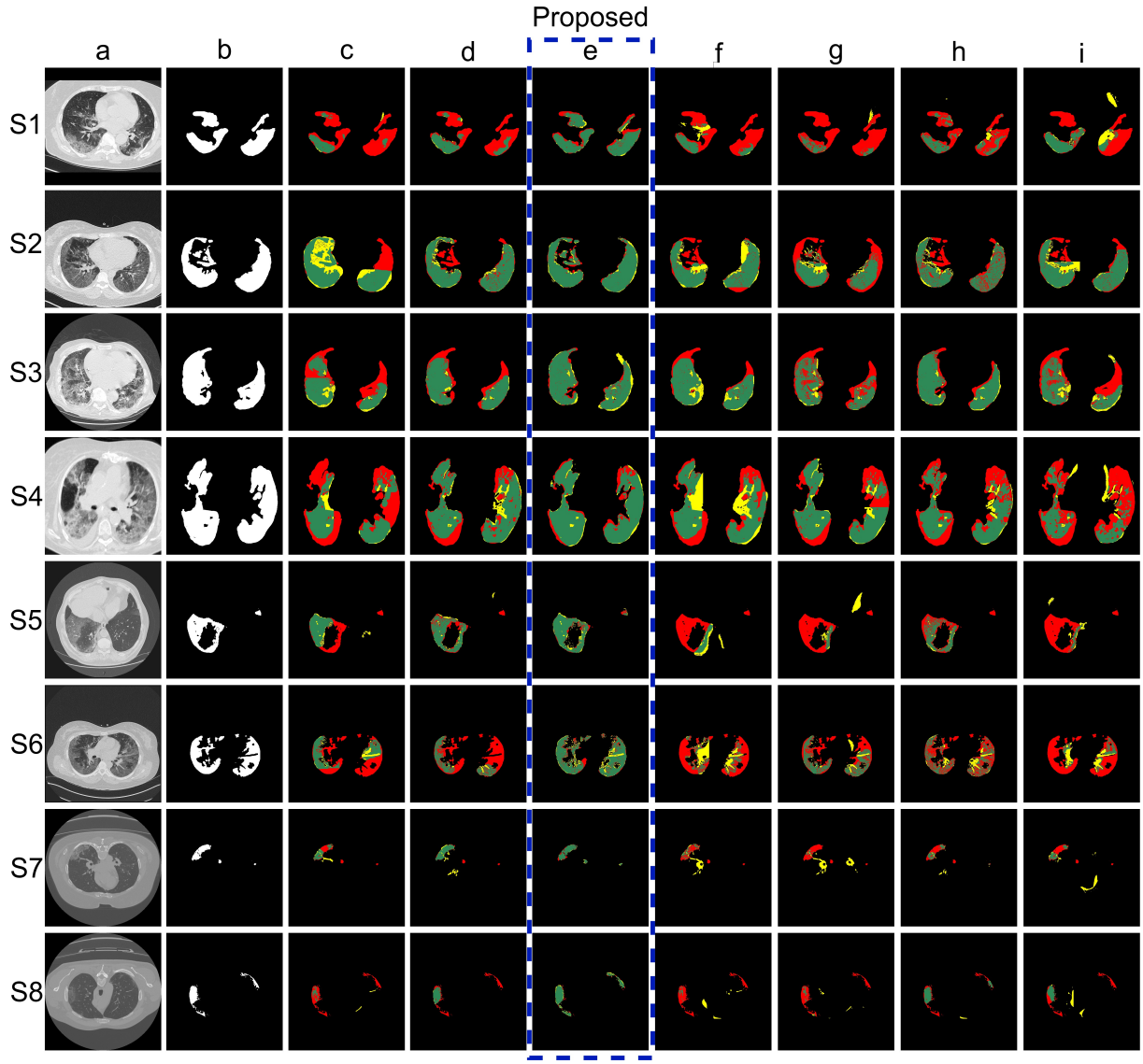


Figure 8: Sample segmentation output by base classifiers, in the uniform framework of ensembling with *LOPO*. (a) Original CT scan, with (b) Annotated masks. Segmentation obtained by (c) *U*-Net, (d) Attention *U*-Net, (e) *AMC*-Net, (f) R2UNet, (g) Inception Net, (h) Xception Net, and (i) DenseNet, where regions Green: True Positive, Red: False Negative, Yellow: False Positive

*S*4 from CT-Seg; (iii) *S*5, *S*6 from Seg-nr.2; and (iv) *S*7, *S*8 from MosMed. It is observed that the *AMC*-Net, of column (e) in the figure, performed the best. Results were corroborated with the confusion matrix of Fig. 9, providing an indication of the distribution of misclassified pixels for a sample segmentation mask *S*4. The confusion was seen to be the least in case of the *AMC*-Net [column (c)]. It resulted in the minimum over-and under-segmentation for all eight samples considered here.

Note that over-segmentation corresponds to higher count of FP pixels (yellow) and under-segmentation refers to higher FN (red). Considering sample *S*4 as an example, it is clearly evident that state-of-the-art models *U*-Net, Attention *U*-Net, Inception Net, Xception Net, DenseNet demonstrate under-segmentation, while model R2UNet is indicative of over-segmentation. On the other hand, our *AMC*-Net [column (e), Fig. 5] exhibited significantly lower under- and/or over-segmentation for the same sample *S*4. The corresponding confusion matrix in Fig. 9 corroborates these findings.

The model complexity, in terms of the number of parameters, is enumerated in

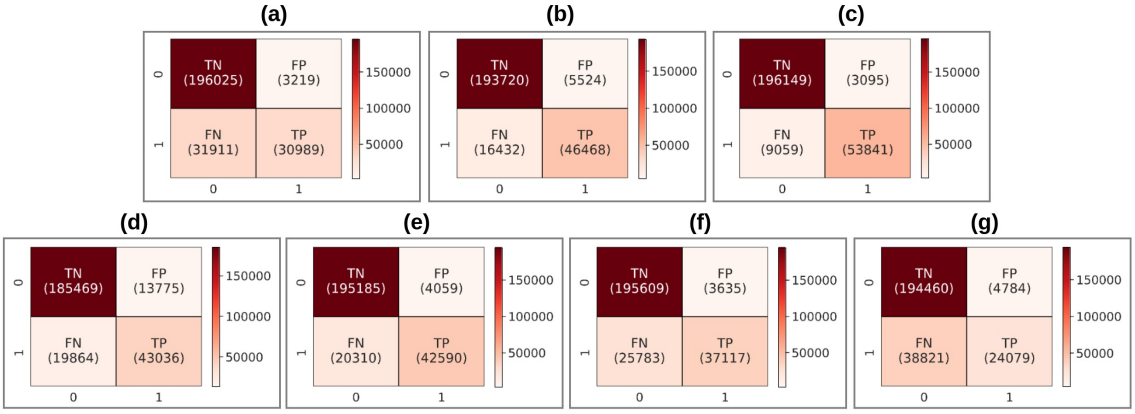


Figure 9: Confusion matrix for a sample $S4$ from Fig. 8 segmentation mask, by the base classifiers, in the uniform framework of ensembling with $LOPO$. Models (a) U-Net, (b) Attention U-Net, (c) AMC-Net, (d) R2UNet, (e) Inception Net, (f) Xception Net, and (g) DenseNet

Table 6: Comparative computational analysis of base model parameters, along with DSC obtained

	Model	U-Net	Attention U-Net	MSU-Net	AMC-Net	R2UNet	Inception Net	Xception Net	DenseNet
	#parameters	1.96M	1.99M	3.31M	3.34M	6.00M	11.99M	2.05M	4.26M
DSC on	Kaggle	0.6082	0.7492	0.7477	**0.8840**	0.6772	0.5924	0.7972	0.6729
	CT-Seg	0.5124	0.7054	0.6924	**0.7955**	0.5762	0.6327	0.7172	0.6766
	Seg-nr.2	0.5921	0.7180	0.7173	**0.8431**	0.6304	0.6275	0.8331	0.8105
	MosMed	0.6896	0.8300	0.8224	**0.9584**	0.7029	0.7145	0.8908	0.8220

Table 6. Although the proposed $EAMC$ involves slightly more parameters than U-Net, Attention U-Net, MSU-Net, Xception Net, it is less than that of R2UNet, DenseNet, and much lower as compared to Inception Net. Note that the overall comparative performance of $EAMC$ is better than the state-of-the-art methods, as evident in Fig. 5. A representative study for DSC on all the four test datasets is also provided in the table.

The technique holds promise in other medical image domains, including (but not limited to) detection of lesions in MRI images of the brain or pathologies in fundus images of the eye for screening diabetic retinopathy.

4 Methodology

The architecture of our $EAMC$ is novel. It consists of an ensemble of Attention-modulated Multi-Scalar (MS) blocks along the encoding and decoding paths of a U-Net. Incorporation of dilated convolutions in the MS-block, in lieu of down- and/or up-sampling, improves the overall performance and makes it robust to generalization. Focal loss function [18] is employed for effectively handling class imbalance in the data. The Leave-One-Patient-Out (LOPO) ensembling effectively trains a network from scratch (each time leaving one patient sample out). This scheme creates the necessary diversity in data, while training a completely new model with different parameters and scarce annotated samples.

The Attention-modulated MS-blocks form the AMC-Net modules, serving as the base classifier for our ensembled $EAMC$. This is illustrated in Fig. 2, with elaborated

description of the individual modules being provided in Tables 7-9. The MS-block extracts multi-scalar features using a concatenation of four dilated convolutional layers, having dilation rates $D = 1, 2, 3, 5$. A dilated convolution (Fig. 10) inserts holes into the standard convolution map, thereby expanding its receptive fields. Thus dilated convolutions can enlarge a receptive field, without any loss of information, while retaining the kernel size. Finally 1×1 convolutions are employed on the concatenated multi-scalar feature map. The MS-block is depicted in Fig. 2(b) and Table 8.

Table 7: *AMC-Net* module of Fig. 2(a)

Block: i	Input: (128 * 128 * 1) CT image
1	$conv_1$(3*3), 16 MS-Block($conv_1$), 16 (Described in Table 8) $conv_2$(3*3), 16 $maxpooling_1$(2*2)
2	$conv_3$(3*3), 32 MS-Block($conv_3$), 32 $conv_4$(3*3), 32 $maxpooling_2$(2*2)
3	$conv_5$(3*3), 64 MS-Block($conv_5$), 64 $conv_6$(3*3), 64 $maxpooling_3$(2*2)
4	$conv_7$(3*3), 128 MS-Block($conv_7$), 128 $conv_8$(3*3), 128 $maxpooling_4$(2*2)
5	$conv_9$(3*3), 256 $conv_{10}$,(3*3), 256 $Upsampling_1$(2*2) AG($conv_8$, $Upsampling_1$), 64 (Described in Table 9)
6	$concat_1$ ($Upsampling_1$, $Multiply_1$) $conv_{11}$(3*3), 128 MS-Block($conv_{11}$), 128 $conv_{12}$(3*3), 128 $Upsampling_2$(2*2) AG ($conv_6$ $Upsampling_2$), 32
7	$concat_2$ ($Upsampling_2$, $Multiply_2$) $conv_{13}$(3*3), 64 MS-Block($conv_{13}$), 64 $conv_{14}$(3*3), 64 $Upsampling_3$(2*2) AG ($conv_4$, $Upsampling_3$), 16
8	$concat_3$ ($Upsampling_3$, $Multiply_3$) $conv_{15}$(3*3), 32 MS-Block($conv_{15}$), 32 $conv_{16}$(3*3), 32 $Upsampling_4$(2*2) AG ($conv_2$, $Upsampling_4$), 8
9	$concat_4$ ($Upsampling_4$, $Multiply_4$) $conv_{17}$(3*3), 16 MS-Block($conv_{17}$), 16 $conv_{18}$(3*3), 16 $conv_{19}$(1*1), 1 Sigmoid activation
	Output: (128 * 128 * 1)

Table 8: MS-block module of Fig. 2(b)

MS-Block($conv_M$), n (n = No. of features)
Input($conv_M$) -> $conv_{D-1}$(3*3), n Input($conv_M$) -> $conv_{D-2}$(3*3), n Input($conv_M$) -> $conv_{D-3}$(3*3), n Input($conv_M$) -> $conv_{D-5}$(3*3), n concat ($conv_{D-1}$, $conv_{D-2}$, $conv_{D-3}$, $conv_{D-5}$)

Table 9: Attention Gates (AG) of Fig. 2(c)

AG ($conv_X$, $Upsampling_Y$), n (n = No. of features)
Input($Upsampling_Y$) -> $conv_a$(1*1), n Input($conv_X$) -> $conv_b$(1*1), n $conv_a$ + $conv_b$ ReLU activation $conv_c$(1*1), 1 Sigmoid Activation -> SA $conv_X$ * SA -> $Multiply_Y$

The *AMC-Net* contains nine convolutional blocks, four max-pooling layers, four Upsampling convolutional layers and eight MS-blocks. First the CT image patches of size 128×128 are fed at the input. The patches percolate down four sets of iterations of 2×2 max-pooling layers, 3×3 convolutional and MS-block layers, involving a stride

of 1 in the encoder. The 2×2 Upsampling layers in the decoder help recover the final resolution of an image.

The MS-blocks are added after the ordinary convolutions in the first four encoder and the last four decoder layers. They help obtain multi-scalar contextual information, to reduce the error along the segmentation boundary for improved accuracy. The COVID-19 infection lesions are typically hard to segment; mainly due to their uneven distribution and varying dimensions. The high-level semantic feature maps in the decoder, concatenated through the attention mechanism, focus on the low-level details in the extracted feature maps (of the encoder) to accurately recover the details of the infected regions in the CT slices.

The Attention gates (AG) of Fig. 2(c) provide the necessary importance to each pixel during decoding. The upsampled images, along with their encoded versions at the same level, are combined to enhance the importance of a pixel through spatial attention. Adaptive selection of spatial information is achieved by emphasizing pixels from the regions of interest, while suppressing the less relevant ones. Four attention modules are introduced for adaptive feature refinement. A sequential spatial attention module is embedded into each decoding block to avoid overfitting, while accelerating the training of the $EAMC$.

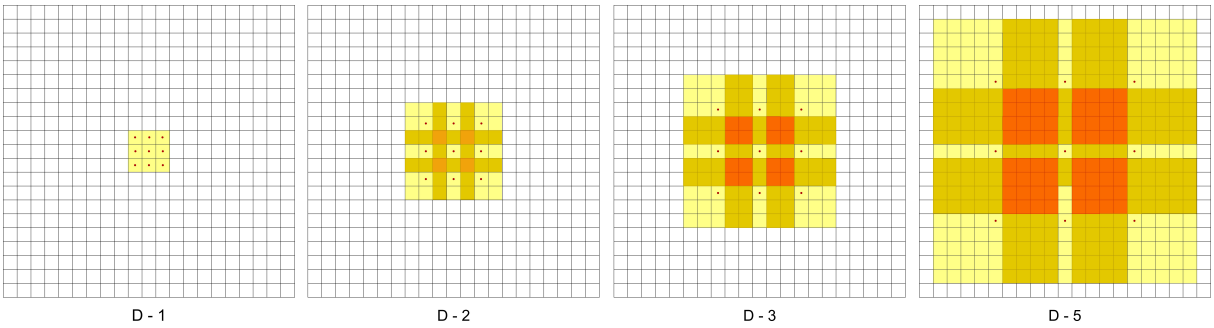


Figure 10: D-dilated convolutions, with $D = 1, 2, 3, 5$

The activation function at the final layer of the *AMC-Net* is the sigmoid. It generates a probabilistic output for the ROI. The choice of a loss function has direct impact on model performance. Loss functions represent the computation of error over each batch, during backpropagation training, and reflect the adjustment of network weights. It was found, after several experiments, that focal loss was the best choice. Let the ground truth segmentation mask be $\mathbf{y} \in \{\pm 1\}$, with the corresponding predicted mask being $\hat{\mathbf{y}}$ with estimated probability $p \in [0,1]$. Focal Loss (FL) [18] overcomes class imbalance in datasets, where positive patches are relatively scarce. The cross entropy (CE) loss for binary classification is defined as

$$CE(p, y) = \begin{cases} -\log p, & \text{if } y = 1, \\ -\log(1-p), & \text{otherwise.} \end{cases} \tag{1}$$

For convenience, let

$$p_{\mathrm{t}} = \begin{cases} p, & \text{if } y = 1, \\ 1-p, & \text{otherwise,} \end{cases} \tag{2}$$

such that $CE(p, y) = CE(p_{\mathrm{t}}) = -log(p_{\mathrm{t}})$. The α-balanced focal loss is defined as

$$FL(p_{\mathrm{t}}) = -\alpha(1-p_t)^{\gamma} log(p_t), \tag{3}$$

with the choice of weighting factor $\alpha = 0.8$ and focusing parameter $\gamma = 2$ being made after several experiments.

Some of the other loss functions, explored in the ablation studies, include the Dice Loss (DL) [38]

$$DL(y, \hat{y}) = \left(1 - \frac{2y\hat{y} + 1}{y + \hat{y} + 1}\right), \tag{4}$$

where 1 is added in the numerator and denominator to ensure that in edge case scenarios, such as when $\mathbf{y} = \hat{\mathbf{y}} = 0$, the function does not become undefined. The $CEDL$ loss [18] is defined as a combination of DL and the cross-entropy (CE) loss for binary classification, thereby incorporating the benefits from both. We have

$$CEDL(y, \hat{y}) = -\{y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) + DL(y, \hat{y})\}. \tag{5}$$

The IoU metric [40], or Jaccard Index, is computed as the ratio between the overlap of the positive instances between two sets, and their mutual combined values. It is expressed as

$$IoU(y, \hat{y}) = \left(1 - \frac{y\hat{y} + 1}{y + \hat{y} + y\hat{y} + 1}\right). \tag{6}$$

The Tversky loss (TL) [1] optimises the segmentation on imbalanced medical datasets. It adjusts the constants α and β to give special weightage to errors like FP and FN We have

$$TL(y, \hat{y}) = \left(1 - \frac{y\hat{y} + 1}{y\hat{y} + \beta(1 - y)\hat{y} + \alpha y(1 - \hat{y}) + 1}\right). \tag{7}$$

The Focal Tversky loss (FTL) [1] also focuses on the difficult samples, by down-weighting easier (or common) ones. It attempts to learn the harder examples, like small ROIs, with the help of the γ coefficient. It is defined as

$$FTL(y, \hat{y}) = (1 - TL)^{1/\gamma}. \tag{8}$$

A value of $\gamma = 2$ was employed, after several experiments.

Diversity is introduced in this ensembling of the *AMC-Nets*, by varying the training datasets through $LOPO$; thereby, changing the initialization of the networks, and modulating the choice of parameters of the $EAMC$ system. The performance of the models is evaluated in terms of the following metrics. We define the number of pixels, (i) correctly classified as positive by True Positive (TP), (ii) incorrectly classified as positive, by False Positive (FP), (iii) correctly identified as negative as True Negative (TN), and falsely classified as negative by False Negative (FN).
The metrics used are *Dice Score Coefficient*

$$\mathbf{DSC} = \left(\frac{2 * TP}{2 * TP + FP + FN}\right), \tag{9}$$

$$\mathbf{Precision} = \left(\frac{TP}{TP + FP}\right), \tag{10}$$

$$\mathbf{Sensitivity} = \left(\frac{TP}{TP + FN}\right), \tag{11}$$

and Area Under the *Receiver Operating Characteristic* Curve (AUC). The ROC curve typically plots the TP rate vs the FP rate, over different thresholds. Higher values of these indices imply a better quality of segmentation [30, 34].

5 Data Preparation

Pixel values in range [0, 255] were normalized, keeping the HU range in interval [-1024, 3071], to enable the model visualize and learn all the areas (like, infection, bone,

tissues) inside the CT scan images. Instead of initially extracting the lung part from the full CT slice [36], we directly detect the infected area from the entire image for subsequent segmentation of the COVID-infected ROI. Class imbalance between the infected and non-infected areas of the CT slices was considered, in terms of positive (infected) and negative (non-infected) patches over the data.

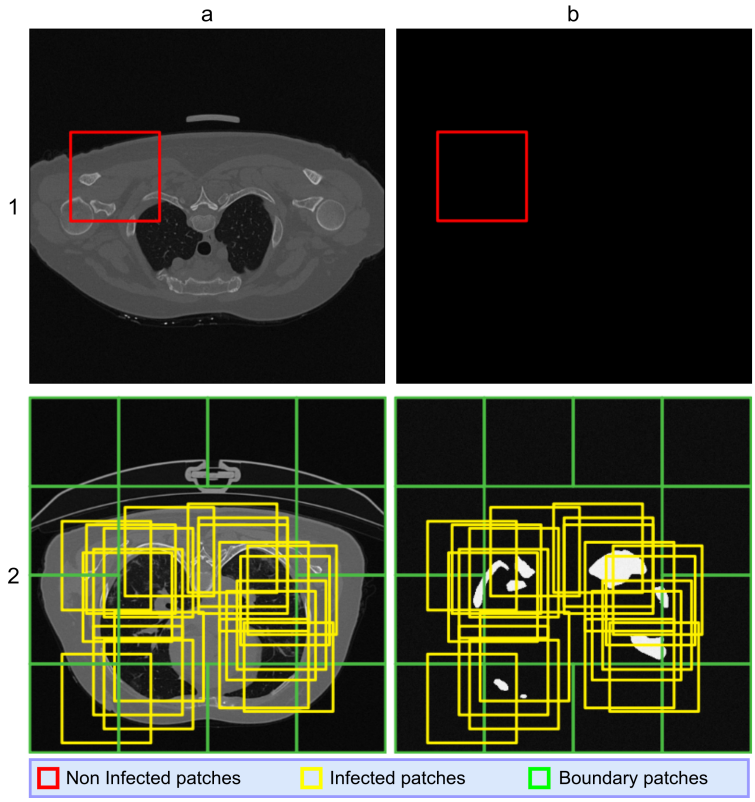


Figure 11: *Row 1:* Patch extraction from training CT slices, depicting no infected regions. 1(a) Non-infected patches, and 1(b) corresponding annotated masks.
Row 2: Patch extraction from training CT slices, depicting infected regions. 2(a) Overlapping patches, and 2(b) mapping to corresponding annotated masks.

5.1 Training

Availability of annotated training data, depicting infection masks, is scarce and leads to class imbalance. In order to circumvent this problem, we extracted overlapping patches to increase the training data while uniformly representing relevant ROI.

The ground truth corresponding to each axial slice, of each CT volume of the training data was checked. If there existed no infected region on a slice then it was labeled as "non-infected" (Fig. 11, *Row: 1*). Random 128×128 patches were extracted.

When there existed a region of infection in any axial slice, it was labeled as "infected" (Fig. 11, *Row: 2*). Twenty random 128×128 bounding boxes were drawn over the ROI to extract the patches. Next all twelve 128×128 boundary patches (inside the 512×512 axial slice) were considered.

The distribution of infected and non-infected slices and/or patches, after patch extraction to create the training set, is displayed in Table 10. Representative extracted patches, along with the corresponding annotated masks, are presented in Fig. 12.1 for the infected slices and Fig. 12.2 for the non-infected ones.

Table 10: Distribution of infected and non-infected Slices & Patches extracted for training

Patient No.	Sample Name	Slices		Patches	
		Infected	Non-infected	Infected	Non-infected
P1	coronacases_org_001	161	140	3534	1758
P2	coronacases_org_002	143	57	3114	1519
P3	coronacases_org_003	137	63	3198	1249
P4	coronacases_org_004	113	157	2341	1432
P5	coronacases_org_005	116	174	2342	1544
P6	coronacases_org_006	70	143	1503	880
P7	coronacases_org_007	93	156	2067	1065
P8	coronacases_org_008	216	85	4647	2350
P9	coronacases_org_009	111	145	2276	1421
P10	coronacases_org_010	191	110	4375	1847

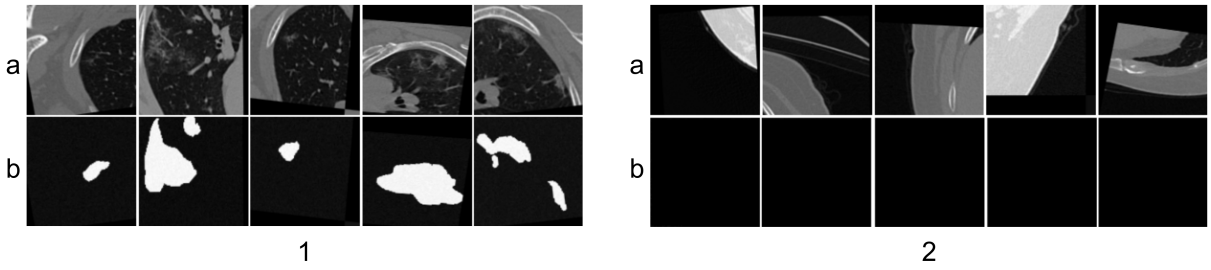


Figure 12: The (a) extracted patches, and (b) corresponding annotation (post-run-time augmentation), for sample slices which are 1: infected, and 2: non-infected

5.2 Testing

As only the ROI and background need to be separated for the test images, here the extraction of non-overlapping patches serve the purpose. Axial slices (512×512) were extracted from each test CT volume. Sixteen 128×128 non-overlapping patches were obtained from each slice, as depicted in Fig. 13.

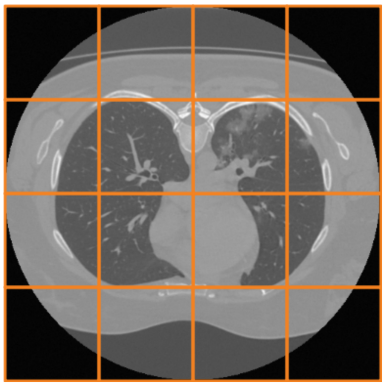


Figure 13: Patch extraction from test CT slices.

References

[1] N. Abraham and N. M. Khan. A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation. In *Proceedings of the IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, volume ISBI 2019, pages 683–687, 2019.

[2] M. Z. Alom, M. Hasan, et al. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *CoRR*, abs/1802.06955, 2018.

[3] X. Bai, H. Wang, et al. Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence. *Nature Machine Intelligence*, 3(12):1081–1089, 2021.

[4] S. Banerjee, S. Mitra, and B. Uma Shankar. Automated 3D segmentation of brain tumor using visual saliency. *Information Sciences*, 424:337–353, 2018.

[5] M. Barstugan, U. Ozkaya, et al. Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods. *CoRR*, abs/2003.09424, 2020.

[6] S. Basu, S. Mitra, and N. Saha. Deep Learning for Screening COVID-19 using Chest X-Ray Images. In *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 2521–2527. IEEE, 2020.

[7] T. Ben-Haim, R. M. Sofer, et al. A Deep Ensemble Learning Approach to Lung CT Segmentation for Covid-19 Severity Assessment. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 151–155. IEEE, 2022.

[8] R. Cong, Y. Zhang, et al. Boundary Guided Semantic Learning for Real-time COVID-19 Lung Infection Segmentation System. *IEEE Transactions on Consumer Electronics*, 68(4):376–386, 2022.

[9] T. G. Dietterich. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*, pages 1–15. Springer, 2000.

[10] X. Ding, J. Xu, et al. Chest CT findings of COVID-19 pneumonia by duration of symptoms. *European Journal of Radiology*, 127:109009, 2020.

[11] N. Enshaei, P. Afshar, et al. An Ensemble Learning Framework For Multi-Class Covid-19 Lesion Segmentation From Chest CT Images. In *Proceedings of the IEEE International Conference on Autonomous Systems (ICAS)*, volume ICAS 2021, pages 1–6, 2021.

[12] Y. Fang, H. Zhang, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*, 296(2):E115–E117, 2020.

[13] L. Geng, S. Zhang, et al. Lung segmentation method with dilated convolution based on VGG-16 network. *Computer Assisted Surgery*, 24(sup2):27–33, 2019.

[14] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[15] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[16] A. Krizhevsky, I. Sutskever, et al. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[17] Y. Li and L. Xia. Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management. *American Journal of Roentgenology*, 214(6):1280–1286, 2020.

[18] T.-Y. Lin, P. Goyal, et al. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 2017.

[19] J. Ma, Y. Wang, et al. Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. *Medical Physics*, 48(3):1197–1210, 2021.

[20] S. P. Morozov, Andreychenko, et al. MosMedData: data set of 1110 chest CT scans performed during the COVID-19 epidemic. *Digital Diagnostics*, 1(1):49–59, 2020.

[21] S. P. Morozov, A. E. Andreychenko, et al. MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset. *CoRR*, abs/2005.06465, 2020.

[22] Y. Oh, S. Park, et al. Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets. *IEEE Transactions on Medical Imaging*, 39(8):2688–2700, 2020.

[23] O. Oktay, J. Schlemper, et al. Attention U-Net: Learning Where to Look for the Pancreas. *CoRR*, abs/1804.03999, 2018.

[24] C. Parmar, E. Rios Velazquez, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PloS One*, 9(7):e102107, 2014.

[25] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45, 2006.

[26] M. Roberts, D. Driggs, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, 3(3):199–217, 2021.

[27] O. Ronneberger, P. Fischer, et al. U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, et al., editors, *Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015.

[28] S. Shabani, M. Homayounfar, et al. Self-supervised region-aware segmentation of COVID-19 CT images using 3D GAN and contrastive learning. *Computers in Biology and Medicine*, 149:106033, 2022.

[29] H. K. Siddiqi, P. Libby, et al. COVID-19–A vascular disease. *Trends in Cardiovascular Medicine*, 31(1):1–5, 2021.

[30] Y. Song, J. Liu, et al. COVID-19 Infection Segmentation and Severity Assessment Using a Self-Supervised Learning Approach. *Diagnostics*, 12(8):1805, 2022.

[31] W. Sun, X. Feng, et al. Weakly supervised segmentation of COVID-19 infection with local lesion coherence on CT images. *Biomedical Signal Processing and Control*, 79, Part-1:104099, 2022.

[32] G. Wang, W. Li, et al. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging*, 37(7):1562–1573, 2018.

[33] P. Wang, Z. Li, Y. Hou, and W. Li. Action recognition based on joint trajectory maps using convolutional neural networks. In *Proceedings of the 24th ACM International Conference on Multimedia*, pages 102–106, 2016.

[34] Q. Wang, X. Li, et al. A regularization-driven mean teacher model based on semi-supervised learning for medical image segmentation. *Physics in Medicine & Biology*, 67(17):175010, 2022.

[35] S. Wang, Y. Zha, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *European Respiratory Journal*, 56(2):2000775: Pages–9, 2020.

[36] D. Wu, K. Gong, et al. Severity and consolidation quantification of COVID-19 from CT images using deep learning based on hybrid weak labels. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3529–3538, 2020.

[37] K. Zhang, X. Liu, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell*, 181(6):1423–1433.e11, 2020.

[38] Y. Zhang, S. Liu, et al. Rethinking the Dice Loss for Deep Learning Lesion Segmentation in Medical Images. *Journal of Shanghai Jiaotong University (Science)*, 26(1):93–102, 2021.

[39] Z. Zhang, Q. Liu, et al. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

[40] D. Zhou, J. Fang, et al. IoU Loss for 2D/3D Object Detection. In *Proceedings of the International Conference on 3D Vision (3DV)*, volume 3DV, pages 85–94, 2019.

[41] L. Zhou, X. Meng, et al. An interpretable deep learning workflow for discovering subvisual abnormalities in CT scans of COVID-19 inpatients and survivors. *Nature Machine Intelligence*, 4(5):494–503, 2022.