
On Automated Analysis of Lung Images with Deep Learning for Healthcare

A thesis submitted to Indian Statistical Institute
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science

by
Surochita Pal
Senior Research Fellow

Under the supervision of
Prof. Sushmita Mitra



Machine Intelligence Unit
Indian Statistical Institute,
Kolkata, India

June, 2025

Declaration of Authorship

I, Surochita Pal, declare that this thesis titled, "**On Automated Analysis of Lung Images with Deep Learning for Healthcare**" and the work presented in it is the original work done by me. I confirm that:

- I have completed this work while pursuing my Ph.D. in Computer Science at the Indian Statistical Institute, Kolkata. This thesis has not been submitted anywhere else for the award of any degree.
- Whenever I have referred to the work of others, I have always provided clear attribution and reference.
- I have acknowledged all the main sources of help.
- In the case of joint works with other co-authors, I have obtained a no-objection certificate from each co-author to include the work in my thesis. Each co-author has acknowledged my adequate contribution to merit the inclusion of the joint work in my thesis.

Signature: 

Date: 30.06.2025

(Surochita Pal)

❧ Certificate from Supervisor ❧

This is to certify that the work contained in the thesis entitled "**On Automated Analysis of Lung Images with Deep Learning for Healthcare**", submitted by **Surochita Pal** for the award of the degree of Doctor of Philosophy in Computer Science to Indian Statistical Institute, Kolkata, is a record of the bonafide research works carried out by her under my supervision and guidance.

I consider that the thesis has reached the standards and fulfilling the requirements of the rules and regulations relating to the nature of the degree. The contents embodied in the thesis have not been submitted as a whole or as a part for the award of any degree or diploma or any other academic award anywhere given before.

Signature: *Sushmita Mitra*

Date: 30.06.2025

Prof. Sushmita Mitra
Professor, Machine Intelligence Unit,
Indian Statistical Institute, Kolkata,
203 Barrackpore Trunk Road,
Kolkata, West Bengal , India - 700108

To my Friends and Family . . .

Acknowledgements

Completing this Ph.D. thesis has been a long and challenging journey, and I am deeply grateful to all those who have supported and inspired me along the way. This journey has not only helped me grow academically but also taught me to stay positive. It has shown me the value of patience, hard work, and self-belief, even during difficult times.

First and foremost, I would like to express my profound gratitude to my supervisor, Prof. Sushmita Mitra. Your unwavering support, immense knowledge, and insightful guidance have been the cornerstone of my research. You have not only been an exceptional mentor but also a source of inspiration and motivation. Your patience and encouragement have enabled me to push my boundaries and achieve my goals.

I would like to extend my heartfelt thanks to Prof. B. Uma Shankar for his invaluable support in writing and drafting articles. His meticulous feedback, insightful suggestions, and dedication to refining each draft have significantly enhanced the quality of our work. His expertise and encouragement have been crucial in shaping this research, and I am deeply grateful for his unwavering assistance throughout this process.

I sincerely thank Prof. Sarbani Palit, Prof. Arijit Bishnu, and Prof. Ashish Ghosh for their invaluable feedback and constructive criticism, which have greatly enriched my work. I am deeply grateful for the time and effort you invested in supporting my research.

I am incredibly thankful to my colleagues and friends, especially those in the Research Group, including Ritabanda, Pallabi, Shramana, and Riddhasree, for their camaraderie, intellectual discussions, and collaborative spirit. I am also grateful to Ankitadi, Soumi, Sayan, Chayan, Debendra, and Koustab for their constant encouragement and companionship throughout this journey.

Special thanks to the administrative staff of the Institute, especially those in my Unit and at the Dean's Office, for their assistance and support throughout my studies. Your efficiency and dedication have greatly facilitated my work. Partial financial support was provided by J. C. Bose National Fellowship, grant no. JCB/2020/000033 of my supervisor during the final months of writing.

I am deeply grateful to my friends, especially my virtual family, for providing me with much-needed emotional support. Your friendship has been a constant source of strength and joy.

To my family: Maa (mother), Mama (maternal uncle), Mami (maternal aunt), and my beloved cousin Diya, your unwavering love, sacrifices, and belief in me have been my anchor. Your steadfast support through every challenge is something for which I will be eternally grateful.

Lastly, I would like to extend my heartfelt gratitude to anyone I may have unintentionally overlooked. Your kind wishes and support have been truly invaluable, and I deeply appreciate your contributions. Thank you all for your invaluable contributions to this journey.

Surochita Pal

Abstract

Automated detection and diagnosis of lung diseases through medical image analysis offers a noninvasive alternative to invasive procedures, especially considering the challenges and potential risks associated with repeat lung operations. Noninvasive image-guided diagnostic techniques, such as lung imaging, have become essential in clinical practice. This thesis focuses on the development of a computer-aided system aimed at enhancing the classification, detection, and segmentation of lung diseases, specifically caused by COVID-19 and lung tumors, leveraging advanced computational methods. Novel segmentation algorithms, such as *EFMC* and *WDU-Net*, are devised based on encoder-decoder architectures within deep convolution networks. These algorithms undergo rigorous validation against ground truth or manual segmentation by radiologists, ensuring their accuracy and reliability. The *EFMC* algorithm employs a selective focus mechanism with multi-resolution blocks, allowing for precise delineation of COVID-19 affected regions in lung CT scans. Its performance is validated through extensive comparison with expert annotations, demonstrating its effectiveness in capturing subtle abnormalities while accurately segmenting lung anomalies. Similarly, *WDU-Net* integrates weighted deformable convolution. Here the deformable convolution modules enhance its ability to capture irregular shapes and features in COVID-19 and lung tumors. Validation against manual segmentation reveals its robustness and accuracy in segmenting COVID-19 and lung tumors from CT images; thereby, showcasing its potential for aiding clinical diagnosis and treatment planning. Next automated classification of lung tumors is devised, in the multi-modal PET-CT framework, using the innovative *DEMF* model. The network leverages deep convolution networks, in conjunction with dimensionality reduction, to efficiently detect and classify lung abnormalities. This demonstrates superior performance in lung cancer classification across multimodal images. Finally, the *DGMC* is developed to enhance diagnosis and classification of diseases, by co-learning from multimodal images. Utilizing a novel multihead classifier, the *DGMC* can efficiently distinguish between COVID-19, tumors, and healthy slices of the lung. The input signal encompasses CT, along with EIT-processed CT scans, in order to provide a multimodal flavour. It captures granular details of the infection, while visualizing the activation regions. Together, these advancements represent significant progress in the automated analysis of lung diseases, by providing valuable tools for the early detection and diagnosis in clinical settings.

Contents

Declaration of Authorship	iii
Certificate from Supervisor	v
Acknowledgements	ix
Abstract	xi
1 Introduction and Scope of Thesis	1
1.1 Introduction	2
1.2 Motivation	5
1.3 Preliminaries	8
1.3.1 Image analytics	9
1.3.2 Deep learning	13
1.3.3 Class imbalance and data augmentation	16
1.3.4 Role of multimodal imaging	17
1.4 Background Literature	20
1.4.1 Classification	20
1.4.2 Segmentation	23
1.5 Performance Evaluation	25
1.5.1 Quantitative evaluation	25
1.5.2 Severity prediction	28
1.6 Datasets and Software Packages	28
1.6.1 Datasets	29
1.6.2 Software	29
1.7 Scope of Thesis	31
1.7.1 Multi-resolution deep ensemble learning for segmentation [128]	32
1.7.2 Weighted deformable deep net for efficient segmentation [112], [113], [126], [129]	32
1.7.3 Deep ensemble with multimodal fusion for classification [127]	33
1.7.4 Multimodal deep classification of lung diseases [111] . . .	33
1.7.5 Conclusion and Future Scope	34
2 Multi-Resolution Deep Ensemble Learning for Segmentation	35
2.1 Introduction	36
2.2 Motivation	37
2.3 Methodology	38
2.3.1 Data used	39
2.3.2 Data preparation	39
2.3.3 The model	40
2.4 Experimental Results	45
2.4.1 Implementation details	45
2.4.2 Output	46

2.4.3	Qualitative analysis	47
2.4.4	Comparative study	49
2.4.5	Ablations	52
2.4.6	Severity assessment	54
2.5	Conclusions and Discussion	55
3	Weighted Deformable Deep Net for Efficient Segmentation	57
3.1	Introduction	58
3.2	The Weighted Deformable <i>U</i> -Net	60
3.2.1	Deformable convolution	60
3.2.2	Weight generation	61
3.2.3	Network architecture	63
3.2.4	Loss functions	65
3.3	Implementation and Results	65
3.3.1	Data preparation	66
	Training	66
	Testing	66
3.3.2	Implementation	67
3.3.3	Analysis and discussion	68
	Qualitative	69
	Quantitative	70
	Ablations	71
	Complexity	72
3.3.4	Comparative study	73
	Superiority over <i>FMC</i> -Net	74
3.4	Conclusions	76
4	Deep Ensemble with Multimodal Fusion for Classification	77
4.1	Introduction	78
4.2	Methodology	79
4.2.1	Fusion of PET-CT modalities	79
	Principal Component Analysis (PCA)	80
	Autoencoder	81
4.2.2	Ensemble classification	81
4.3	Implementation and Experimental Results	82
4.3.1	Classification	83
4.3.2	Qualitative analysis	84
4.3.3	Ablations	85
4.4	Conclusions and Discussion	88
5	Multihead Deep Classification of Lung Diseases	91
5.1	Introduction	92
5.2	Methodology	94
5.2.1	Augmentation of CT samples	95
	Generator architecture	96
	Discriminator architecture	96
	Loss functions	97
5.2.2	EIT for processed CT	97

5.2.3	Multihead classification	99
	Multihead gradient module	100
	n -ary classification	101
5.3	Implementation and Results	101
5.3.1	Data preparation	102
5.3.2	Augmentation by CTGAN	102
5.3.3	Generating EIT-processed CT scans	102
5.3.4	Multihead classification	103
5.3.5	Visualization and explainability	106
5.3.6	Ablations	107
5.4	Conclusions and Discussion	109
6	Conclusions and Future Scope	111
6.1	Conclusions	112
6.2	Limitations	114
6.3	Future Scope	115
	List of Publications	117
	Bibliography	119

List of Figures

1.1	Proposed Computer Aided Diagnostic System	5
1.2	Some challenges associated with medical image analysis.	8
1.3	Schematic architecture of the CNN	14
1.4	Sample CT images of the lung, afflicted with (a) COVID-19, and (b) tumor. The ground truth of affliction is marked in red.	18
1.5	Sample images of lung tumors, using (a) CT and (b) PET scans, with the ground truth marked in red.	19
1.6	Sample images of the lung, with (a) CT scans and (b) corresponding EIT. The ground truth of affliction is marked in red. First two columns correspond to COVID-19 infection of the lung, and last two columns display lung tumors.	19
1.7	A sample confusion matrix.	27
1.8	Handling the challenges.	31
2.1	Schematic representation of <i>EFMC</i> , demonstrating the (a) Leave-One-Patient-Out scheme <i>LOPO</i> , with (b) selective Focus-based Multi-resolution view. (c) Ensemble of the <i>EFMC</i> Models	38
2.2	<i>Row 1</i> : Patch extraction from training CT slices, depicting no COVID-infected regions. 1(a) Non-infected patches, and 1(b) corresponding annotated masks. <i>Row 2</i> : Patch extraction from training CT slices, depicting COVID-infected regions. 2(a) Overlapping patches, and 2(b) mapping to corresponding annotated masks.	40
2.3	The (a) extracted patches, and (b) corresponding annotation, for sample patches which are 1: infected, and 2: non-infected	41
2.4	Patch extraction from test CT slices.	41
2.5	(a) The <i>FMC-Net</i> , with a detailed representation of the (b) <i>MR</i> -block, and (c) selective Focusing Module (<i>FM</i>).	42
2.6	<i>D</i> -dilated convolutions, with $D = 1, 2, 3, 5$	42
2.7	Sample learning curves, for Model M3, over training and validation sets; (a) loss: <i>FL</i> , and (b) accuracy: <i>DSC</i>	45
2.8	Confusion Matrix for classification: (a) Kaggle-COVID-19: Part-2 (b) CT-Seg, (c) Seg-nr.2, and (d) MosMed	46
2.9	Visualization of features extracted after each Block of the <i>FMC-Net</i> from Fig 2.5	48
2.10	Sample segmentation output by base classifiers, in the uniform framework of ensemble with <i>LOPO</i> . (a) Original CT scan, with (b) Annotated masks. Segmentation obtained by (c) <i>U</i> -Net, (d) Attention <i>U</i> -Net, (e) <i>FMC-Net</i> , (f) R2UNet, (g) Inception-Net, (h) Xception-Net, and (i) DenseNet, where regions Green: True Positive, red: False Negative, Yellow: False Positive	49

2.11	Confusion matrix for a sample <i>S4</i> from Fig. 2.10 segmentation mask, by the base classifiers, in the uniform framework of ensemble with <i>LOPO</i> . Models (a) <i>U</i> -Net, (b) Attention <i>U</i> -Net, (c) <i>FMC</i>-Net , (d) R2UNet, (e) Inception-Net, (f) Xception-Net, and (g) DenseNet. The <i>DSC</i> is marked for all models used . . .	50
2.12	Comparative performance evaluation of base classifiers, in the uniform framework of ensemble with <i>LOPO</i> , over the test datasets	51
2.13	Role of different components of <i>EFMC</i> over the test datasets .	53
2.14	Visualization of infection, in the eight sample slices from Fig. 2.10, as predicted by <i>EFMC</i> . (a) Input CT scan, (b) corresponding annotation, and (c) JET ColorMap on the prediction .	55
3.1	Illustration of sampling locations, as modeled in basic and deformable convolution.	61
3.2	Feature maps generated with and without incorporating the <i>WG</i> modules.	62
3.3	The WDU-Net framework, with detailed representation of (b) Weight Generation (<i>WG</i>), and (c) Deformable Convolution (<i>DC</i>) blocks.	63
3.4	Patch extraction from training CT slices, depicting <i>Row 1</i> . No tumor-affected regions: (a) Non-infected patches, and (b) corresponding masks; and <i>Row 2</i> . Tumor-affected regions: (a) overlapping patches, and (b) mapping to corresponding annotated masks.	66
3.5	Patch extraction from test CT slices.	67
3.6	Comparative study on sample slices from the test datasets, using different models, with corresponding <i>DSC</i> being inscribed on each slice. (a) Original CT scan, with (b) Annotated masks. Segmentation obtained by (c) <i>U</i> -Net, (d) <i>WU</i> -Net (without <i>DC</i>), (e) <i>DU</i> -Net (without <i>WG</i>), (f) WDU-Net, (g) <i>U</i> -Net++, (h) MultiResUNet, (i) <i>U</i> -Mamba, (j) Swin-UMamba with regions indicating Green: <i>TP</i> , Red: <i>FN</i> , Yellow: <i>FP</i>	69
3.7	Comparative performance on unseen test data with related deep models, using ensemble with <i>FAS</i> loss (with the same testing framework used in all models)	70
3.8	Comparative study on a sample test patch, using different models, indicating the corresponding generated <i>FP</i> and <i>FN</i> regions. (a) Original CT scan, with (b) Annotated masks. Segmentation by (c) <i>U</i> -Net, (d) <i>U</i> -Net with <i>WG</i> -Module, (e) <i>U</i> -Net with <i>DC</i> -Module, and (f) WDU-Net, with regions labeled Green: <i>TP</i> , Red: <i>FN</i> , Yellow: <i>FP</i>	71
3.9	Effect of introducing Deformable Convolution (<i>DC</i>) module at Block 8 (of Fig. 3.3) on the feature maps generated at different stages of training of the WDU-Net	71
3.10	Sample segmentation output. (a) Original CT scan, with (b) Annotated masks. Segmentation obtained by (c) <i>FMC</i> -Net, and (f) WDU-Net, with regions Green: True Positive, red: False Negative, Yellow: False Positive	74

3.11	Comparative performance on unseen test data in terms of <i>FAS</i> loss	75
4.1	Workflow of ensemble classification DEMF, with multimodal image fusion	79
4.2	Image fusion strategy	80
4.3	Extracting lung region from full body scans. CT of (a) full body, (b) lung area, and PET of (c) full body, (d) lung, from coronal view	83
4.4	Training samples from (a) CT, (b) PET, and (c) the corresponding fused images, encompassing cancer-affected sample slices along with healthy ones	84
4.5	Sample augmentation of fused training data	85
4.6	Visualization with Grad-CAM, over sample test set images from (a) CT, (b) PET, and (c) Fused versions, with the corresponding (d) attention map generated by DEMF	86
4.7	Comparative study of image fusion strategies, in terms of <i>Accuracy</i> over each test set, on the performance of (f) ensemble DEMF and the (a) VGG16, (b) ResNet, (c) Inception-Net, (d) Xception-Net, (e) Efficient-Net.	87
4.8	Ablation of multimodal fusion strategies, in terms of <i>Accuracy</i> by different classifiers	87
4.9	Visualization over sample test cases, using our DEMF with Grad-CAM, in terms of the (a) CT image, with (b) its visualization; the (c) PET image, with (d) its visualization; and the (e) Fused image, with (f) its visualization	88
4.10	Effect of the imaging modalities, in terms of <i>Accuracy</i> by the classifiers	89
5.1	The workflow of <i>DGMC</i>	94
5.2	The <i>CTGAN</i> architecture	95
5.3	Mapping between modalities, from extracted mask I in CT to $I_{filtered}$ in EIT	98
5.4	Formation of processed CT, with $C_{thr} = 0.5$	99
5.5	Multihead classifier	100
5.6	Sample augmented COVID-19-infected and tumor-affected CT slices, as generated by CTGAN	103
5.7	(a) Sample CT images, with (b) portion of Healthy Lung (regions without infection), (c) the corresponding synthetic EIT version, and (d) the EIT-processed CT image, for all the three classes	104
5.8	Visual attention of classifier <i>DGMC</i> , as generated through Grad-Cam, highlighting the COVID-19 infected and tumour affected regions in the eight sample CT scans.	106
5.9	Comparative performance of different classifiers, in terms of <i>AUC</i> , on the six test datasets (of COVID-19 and Tumor slices) with input as (a) only CT images, (b) CT and CTGAN-generated images, and (c) CT and CTGAN-generated images with corresponding EIT-processed CT.	107

5.10 Confusion matrices generated by different classifiers, with input as (a) only CT images, (b) CT and CTGAN-generated images, and (c) CT and CTGAN-generated images plus EIT-processed CT version.	108
---	-----

List of Tables

1.1	Details of the lung datasets used	30
2.1	Distribution of infected and non-infected Slices & Patches extracted for training	40
2.2	The <i>FMC-Net</i> module of Fig. 2.5(a)	43
2.3	The <i>MR</i> -block module of Fig. 2.5(b)	43
2.4	Selective Focusing Modules (<i>FM</i>) of Fig. 2.5(c)	43
2.5	Description of Ensemble models, with <i>DSC</i> on corresponding validation sets	46
2.6	Predicted binary classification of Infected & Non-Infected slices, in the test datasets, by <i>EFMC</i>	47
2.7	Comparative computational analysis of base model parameters, along with <i>DSC</i> obtained	50
2.8	Comparative analysis of performance with recent literature on deep learning for segmentation of COVID-19 regions from lung CT	52
2.9	Effect of various loss functions on <i>FMC-Net</i> , measured in terms of <i>DSC</i> , over the test datasets	53
2.10	Effect of varying Dilation rate <i>D</i> on <i>FMC-Net</i> with <i>FL</i> , measured in terms of <i>DSC</i> over the test datasets	54
2.11	Prediction of affected region by <i>EFMC</i> , considering two random samples slices from each test dataset	54
3.1	Distribution of the infected and non-infected Slices & Patches, extracted from each bin for training	67
3.2	Ensemble WDU-Net model M_i , with the performance evaluated on the corresponding validation set B_i	68
3.3	Effect of various loss functions on <i>DSC</i> for the WDU-Net, with ensemble, over the test datasets	72
3.4	Computational complexity of the various modules of WDU-Net and existing models	73
3.5	Comparative analysis of recent literature on deep learning in segmentation of lung CT	73
3.6	Effect of various loss functions, measured in terms of <i>DSC</i>	75
4.1	Train-test split of the cancerous & healthy slices of the lung	83
4.2	Comparative study of the individual classifiers and their ensemble, for the fused multimodal image data	85
5.1	Comparative performance of different classifiers, in terms of the evaluation metrics and network size	105
5.2	Impact of multihead classifier module, as evaluated in terms of <i>Accuracy</i> , on the input images	109

List of Abbreviations

<i>DSC</i>	Dice Similarity Coefficient
<i>JSC</i>	Jaccard Similarity Coefficient
<i>Precision</i>	Precision Metric
<i>Recall/Sensitivity</i>	Recall Metric
<i>ROC</i>	Receiver Operating Characteristic
<i>AUC</i>	Area Under the ROC Curve
<i>HD95</i>	Hausdorff Distance 95%
<i>Accuracy</i>	Accuracy Metric
<i>F1</i>	F1 -Score
<i>Macro – F1</i>	Macro F1 -Score
<i>AP</i>	Average Precision
<i>mAP</i>	mean Average Precision
<i>ROI</i>	Region Of Interest
<i>TP</i>	True Positive
<i>TN</i>	True Negative
<i>FP</i>	False Positive
<i>FN</i>	False Negative
<i>TPR</i>	True Positive Rate
<i>FPR</i>	False Positive Rate
<i>DL</i>	Dice Loss
<i>CE</i>	Cross Entropy
<i>CEDL</i>	Cross Entropy Dice Loss
<i>IoUL</i>	Intersection over Union Loss
<i>TL</i>	Tversky Loss
<i>FL</i>	Focal Loss
<i>FTL</i>	Focal Tversky Loss
<i>ASL</i>	Asymmetric Similarity Loss
<i>DC</i>	Deformable Convolution
<i>HU</i>	Hounsfield Unit

List of Symbols

I	: 2-dimentional Image
μ	: Mean
σ	: Standard Deviation
$\emptyset$	: Null
$\cup$	: Union
$\cap$	: Intersection
C	: Classes
B	: Bias
g	: Activation function
p_d	: Dropout probability
Lr	: Learning rate
Gr	: Gradient
y	: Ground Truth Mask
$\hat{y}$	: Predicted Mask
$\mathcal{D}$	: Euclidean distance
p	: Estimated Probability
p_t	: Probability depending on value of y
$\mathcal{C}$	: Channels
K	: Kernel height and width
H	: Image height
W	: Image width
ω	: Learned weight
$\mathcal{M}$	: Output channels
L	: Levels (Depth of model)
E	: Encoder arm
D	: Decoder arm
$\mathcal{C}$	: Convolution
<i>Sigmoid</i>	: Sigmoid operation
$\oplus$	: element-wise addition
$\otimes$	: element-wise multiplication
L_G	: Generative Loss
L_D	: Discriminative Loss
T	: Feature map blocks or Tensors
L_{CE}	: Categorical Cross Entropy Loss



Chapter 1



Introduction and Scope of Thesis





1.1 Introduction

An image is a two-dimensional representation of visual data, typically captured through cameras or sensors, composed of pixels arranged in rows and columns. Each pixel holds information about the color and intensity at its specific location, collectively forming a visual scene or object. Image processing (IP) [54], [75] refers to the set of techniques and algorithms employed to manipulate, analyze, and extract valuable information from them. It encompasses tasks like noise reduction, feature extraction, object recognition, and enhancement, to improve the quality and relevance of visual data. IP acts as a critical gateway, enabling machines to understand and interpret the visual world; making it an indispensable component in applications ranging from autonomous vehicles and medical imaging to surveillance systems and augmented reality. IP performs some defined transformations on an image, to generate desired information or extract features, such that automated algorithms can be employed for subsequent decision-making.

Today medical imaging [163] is a specialized component of healthcare, comprising of various technologies and techniques to create visual representations of the internal structures and functions of the human body. These images provide invaluable insights for diagnosing, monitoring, and planning treatment for a wide range of medical conditions. Medical imaging comprises of multiple modalities [136], like X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, Positron Emission Tomography (PET), Electrical Impedance Tomography (EIT), histopathology, among others. Analyzing medical images is imperative for several reasons. First and foremost, it aids healthcare professionals in accurate disease diagnosis and assessment; thereby, enabling them to identify abnormalities, tumors, fractures, and other conditions that may not be visible through physical examinations alone. Medical image analysis [20], [48] also plays a crucial role in diagnosis, prognosis, treatment planning, and monitoring disease progression. This allows physicians to tailor personalized treatment strategies and track the effectiveness of medical interventions over time. Furthermore, it facilitates non-invasive procedures; thereby, reducing the need for invasive surgeries while minimizing associated

patient discomfort and risks. Automated analysis of medical images is an indispensable tool in modern medicine [31], [36], [47], enhancing patient care with better prognosis.

- Supervised learning is a category of ML where the algorithm is trained on a labeled dataset. It implies that the model is provided with input data and corresponding target outputs, allowing it to learn the relationship between the inputs and outputs. Common tasks in supervised learning include classification and regression, where the model can predict discrete categories or continuous values, respectively.
- Unsupervised learning, in contrast, involves models trained with unlabeled data. The goal is to uncover inherent patterns or structures within the data without predefined target labels. Key tasks in unsupervised learning include (i) clustering, where data points are grouped into similar clusters, and (ii) segmentation, implying clustering of pixels in an image.

Feature selection involves choosing a subset of relevant features (attributes or variables) from a dataset to reduce dimensionality and improve model efficiency. Feature extraction, on the other hand, creates new attributes by transforming or combining the existing ones to represent data more effectively. These techniques are crucial for enhancing model performance and interpretability, while reducing dimensionality.

Deep Learning (DL) [55], [91] is a subset of ML, a branch of AI that focuses on training computer systems to perform tasks without explicitly programming them. It aids a machine in learning hidden patterns within the data, without expert intervention, for end-to-end predictions. DL requires high computational power and a massive volume of annotated data. On the other hand, ML encompasses a broader range of algorithms that learn patterns from data for decision-making. It is also called shallow learning, and needs features to be provided as input. What sets DL apart is its use of artificial neural networks (ANNs), particularly deep neural networks, which are inspired by the structure and functioning of the human brain. These networks consist of multiple layers of interconnected nodes (neurons) that process data hierarchically; allowing them to automatically learn intricate patterns and representations from large datasets. Convolution neural networks (CNNs) have been shown to perform effectively in analyzing visual images [86]. DL has revolutionized various fields, including medical imaging, by enabling automated interpretation and decision-making from complex data. DL provides cutting-edge solutions to various medical imaging problems [92].

- Automated detection and diagnosis: Deep models, like CNNs, excel at identifying patterns and features in medical images [8]. They can autonomously detect anomalies, lesions, or specific structures, and significantly reduce the burden on radiologists while improving the speed and accuracy of diagnosis.
- Risk assessment: Deep learning algorithms can assess the risk associated with certain medical conditions by analyzing images [172]. For instance,

they can predict the likelihood of cancer recurrence or cardiovascular events, by evaluating factors like tumor characteristics or plaque buildup in blood vessels; thus enabling early intervention with personalized treatment plans.

- **Image enhancement:** It can enhance the quality of medical images by reducing noise, improving resolution, and optimizing contrast [155]. Clearer images aid in more accurate diagnosis and better visualization of anatomical structures.
- **Segmentation and localization:** Deep models can precisely segment and localize specific structures or Regions Of Interest (ROIs) within medical images [52]. This is particularly valuable in surgical planning and radiation therapy, where precise targeting becomes critical.
- **Monitoring disease progression:** By analyzing sequential medical images, captured over time, deep algorithms can track disease progression and treatment efficacy [167]. This information helps healthcare providers make informed decisions about treatment adjustments and patient care plans.
- **Personalized medicine:** DL can analyze medical images in conjunction with other patient data, such as genetic information and medical histories, to tailor precise treatments and interventions targeted at individual patients [168]. Such personalized approach can lead to more effective therapies, with improved outcomes.
- **Early detection:** They can identify subtle changes from images, that may be indicative of the onset of a disease condition [156]. Early detection can be crucial in improving patient outcomes and reducing healthcare costs.

Multimodal medical image analysis with DL constitutes a powerful approach to harness the power of ANNs for extracting useful insights simultaneously from multiple modalities. In the realm of healthcare, various imaging modalities like CT, PET, EIT offer distinct perspectives on a patient's condition. DL models, typically CNNs, are often employed to integrate and analyze such diverse data. This allows a more comprehensive understanding of complex medical situations. Such a holistic approach enhances diagnostic accuracy, enables more robust disease characterization, and provides clinicians with a comprehensive view of patient health. Furthermore, it facilitates the identification of subtle correlations between different modalities; thereby, contributing to earlier disease detection and their personalized management. Multimodal medical imaging, with DL, is poised to revolutionize clinical decision-making – offering a more comprehensive assessment of patient conditions for improving overall outcome in healthcare [114].

This thesis presents several computational strategies, encompassing DL, for the identification, localization, and delineation of some lung diseases, from multimodal medical images, along with their analysis. The workflow of the computer aided diagnostic system is depicted in Fig. 1.1. The rest of the chapter is organized as follows. The motivation of this research is explained in Section

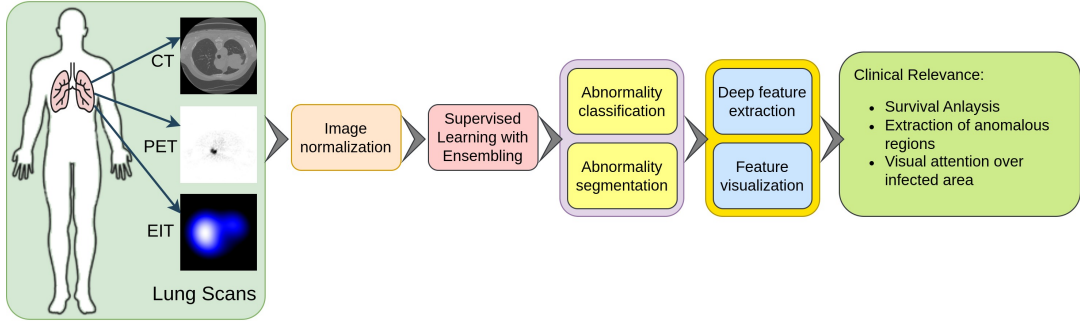


FIGURE 1.1: Proposed Computer Aided Diagnostic System

1.2. Section 1.3 introduces the preliminaries of image analytics, DL, and multimodal imaging characteristics, with reference to the lung. A short survey of related state-of-the-art literature, encompassing automated detection, demarcation, diagnosis and prognosis of some lung diseases (like COVID-19 and cancer), is provided in Section 1.4. The different performance evaluation metrics, employed in this research, are explained in Section 1.5. The datasets and software packages, used here, are outlined in Section 1.6. Finally the scope of the thesis is highlighted in Section 1.7.

1.2 Motivation

The recent global havoc caused by the pandemic COVID-19 focused our attention on the early detection and precise management of its progression for minimization of damage to human life. The same holds for long-term, non-infectious afflictions like cancer. We chose to explore the role of AI, particularly ML and DL, for an automated and accurate handling of images towards effective management of some lung diseases. The afflictions considered here are COVID-19 and cancer tumors, both posing significant health hazards with severe complications for individuals. Although these are two distinct medical conditions, they share the commonality of affecting the respiratory system and lungs. While COVID-19 is an infectious respiratory disease, caused by a virus, lung cancer is a non-communicable condition primarily associated with abnormal cell growth. COVID-19 and lung tumors underscore the critical importance of early detection, timely medical intervention, and ongoing individualized care, to mitigate their potentially devastating impacts on the health and well-being of individuals.

COVID-19 [4], or "Coronavirus Disease 2019", is a highly contagious infectious disease caused by the novel coronavirus SARS-CoV-2. It was first identified in Wuhan, China, in late 2019 and then spread globally, leading to a pandemic. COVID-19 primarily affects the respiratory system, with symptoms ranging from mild respiratory distress to severe pneumonia. Common symptoms include fever, cough, and difficult breathing. In severe cases, the disease can lead to acute respiratory distress syndrome (ARDS), which requires mechanical ventilation. COVID-19 has a wide range of clinical presentations and can affect multiple organs, including the lungs. This viral infection can be

especially dangerous for people with comorbidities like diabetes, high pressure, heart ailments, and cancer. Preventive measures like vaccination, wearing of masks, and social distancing are essential in controlling the spread of the virus.

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 [16], [130]. 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 [118]. This helps in predicting severity of infection, along with prognosis.

Lung cancer [56] is the leading cause of cancer-related death in the developed countries, with about 80% of lung cancer patients being clinically symptomatic. Less than 10% of the patients are asymptomatic when the cancer is detected as an incidental finding. Nearly 60% of patients with lung cancer are found to have distant metastases upon diagnosis, and in some cases these are detected before the primary tumor. Primary lung cancers are broadly classified as non-small cell lung carcinoma (NSCLC) or small cell lung carcinoma (SCLC), with the former constituting around 85% of the total cases and the latter being characterized by rapid growth rate with metastasis [28]. Approximately half of NSCLC cases are localized at diagnosis, and treated either by surgical resection alone or a combination therapy with/ without resection. The five-year survival rate for lung cancer is just 17.8%, which is much lower than that of other major malignancies. Common symptoms of lung tumors include persistent cough, chest pain, shortness of breath, and coughing up blood. Early detection and treatment are critical for improving outcomes, and diagnosis typically involves medical imaging, tissue biopsy, and staging to determine the extent of the disease. It can be life-threatening, and often requires extensive treatment like surgery, chemotherapy, radiation therapy, and/or immunotherapy.

The ability to predict the radiologic response to treatment depends on the accurate demarcation of the tumor [63]. The precise segmentation of the gross tumor volume and the adjacent organs-at-risk is advantageous for radiation therapy planning. This is (i) critical, because subsequent feature extraction depends on its accuracy, (ii) challenging, as many tumors have indistinct borders, and (iii) contentious, since there exist ongoing debates regarding the relative merits of ground truth vs reproducibility along with manual vs automated segmentation issues. In order to overcome the problems of inherent human bias and uncertainty in manual segmentation, the need for an automated or semi-automated CAD becomes apparent. It can also serve to improve the accuracy in automated detection, for assisting doctors in diagnosing faster and on time.

The imaging modalities typically involved in the study of lung images include, among others, CT¹, PET², and EIT [22]. They play a pivotal role in the detection, characterization, and assessment of lung abnormalities – through a comprehensive view of their anatomical, functional, and physiological aspects. When used in conjunction with other modalities it leads to a comprehensive diagnostic approach, by enabling healthcare professionals make informed decisions about the treatment and management of the diseases of the lung.

¹Computed Tomography (CT): <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct>

²Positron Emission Tomography (PET): <https://www.radiologyinfo.org/en/info/pet>

- CT scans help create detailed cross-sectional images of the lungs, allowing for the identification of abnormalities based on their size, shape, and location. Typically X-rays are used to produce a 3D view comprising cross-sectional slices, for detecting existing anomalies. They are excellent for visualizing anatomical details, making them valuable in diagnosing a wide range of medical conditions. In the multimodal framework, CT provides structural information that serves as a reference for the other modalities. Moreover, CT is instrumental in evaluating the extent of disease spread and can guide treatment planning.
- PET is a functional imaging technique that provides information about metabolic and physiological processes within the body. It involves the injection of a radiotracer, which emits positrons that interact with electrons in the body, producing gamma rays. PET scans capture the distribution of the radiotracer and can identify areas highlighted with abnormal metabolic activity. PET is frequently used in cancer diagnosis, staging, and monitoring of treatment response. This is because malignant cells often exhibit heightened metabolic rates. PET scans can aid in distinguishing between benign and malignant lung lesions; thereby, aiding in its accurate staging. Additionally, in the context of COVID-19, PET imaging can identify regions of inflammation and infection in lungs.
- EIT is a functional imaging method that uses electrical measurements to monitor changes in tissue properties, such as conductivity. It is non-invasive and involves the application of low-intensity electrical currents at multiple electrodes placed on the body surface. EIT creates real-time images of internal electrical conductivity, by measuring voltage changes resulting from these currents. It is valuable for monitoring lung ventilation, detecting abnormalities in lung function, and assessing conditions like ARDS and pneumonia. EIT is often used in critical care settings to monitor lung function in real-time, particularly in patients with mechanical ventilation. It can aid in the assessment of the severity of respiratory conditions (including those related to COVID-19).

The primary aim of this thesis is to design and develop an advanced CAD system for assisting medical practitioners through automated detection and diagnosis. It should enhance the diagnostic and prognostic capabilities for both COVID-19 and lung tumors in patients, even in the presence of class imbalance. The challenges addressed here are depicted in Fig. 1.2. Our objectives within this research are directed at improving existing prognostic criteria in a noninvasive manner, for enhancing patient care. These are summarized below.

1. Development of a non-invasive DL tool to detect, classify and segment COVID-19 lung infections in CT, with high precision, through ensemble learning.
2. Design of computerized lung cancer detection, segmentation and prognosis, by developing a new deep model based on varying geometric shapes of tumors from CT scans.

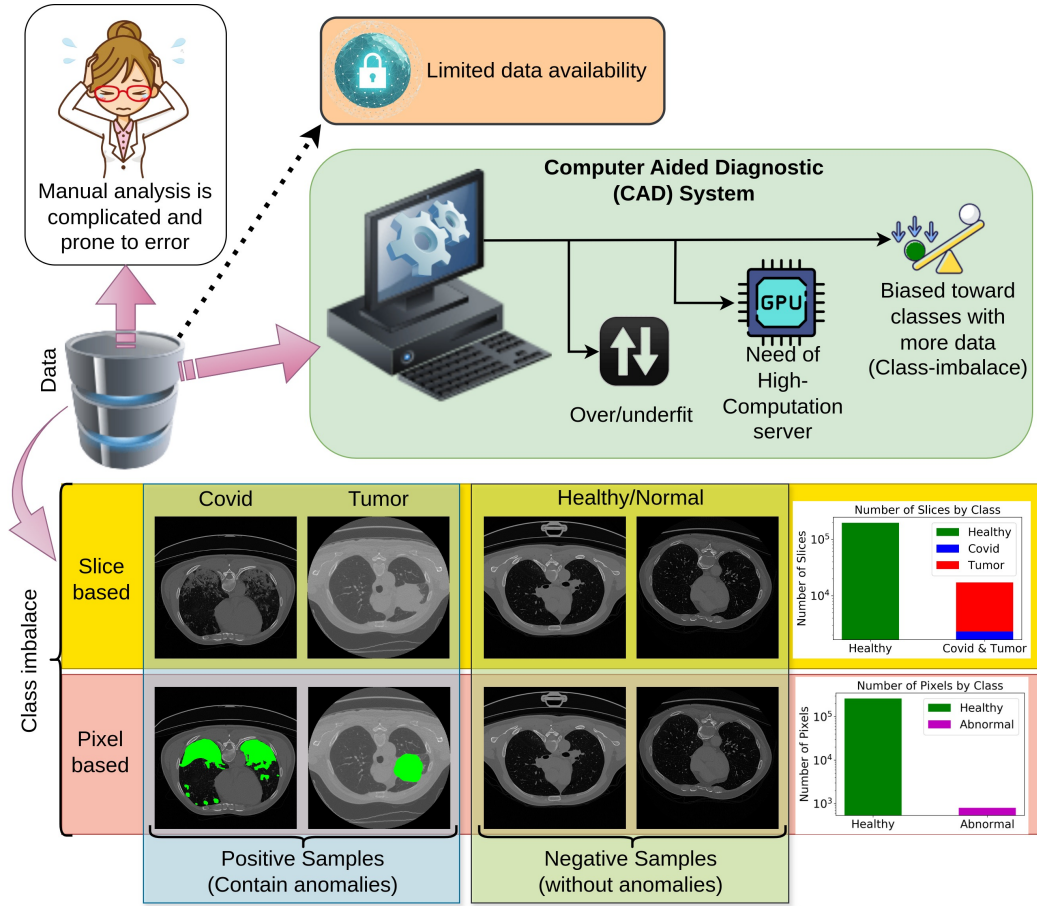


FIGURE 1.2: Some challenges associated with medical image analysis.

3. Classification of cancerous lung slices from CT and PET images, with dimensionality reduction through PCA and Autoencoder, for ensemble classification and visualization in the presence of data scarcity.
4. Formulation of a novel generalized detection framework for lung infections, from CT and EIT-processed CT scans, for deep multihead classification of some lung diseases *viz.* COVID-19, tumor, with respect to healthy scans.

1.3 Preliminaries

In this section we elaborate on a few of the basic components, which serve as the primary concepts and functions used in this thesis. This encompasses topics from (i) image processing and analysis, dealing with normalization, segmentation, classification, and ensemble, (ii) deep learning, and (iii) the role of multimodal imaging in handling the lung diseases considered here.

1.3.1 Image analytics

The goal of image processing [54], [75] is to enhance perceptual acuity for improved analysis. This thesis involves the use of some major components, including image normalization, segmentation, feature visualization, and classification.

Normalization: Medical images often exhibit variations in brightness, contrast, and intensity due to differences in the acquisition devices, protocols, and patient characteristics. Image normalization mitigates these variations, ensuring that images are on a consistent scale; thus simplifying analysis and interpretation. It is a fundamental step for empowering medical professionals and researchers to make more accurate and consistent assessments – towards improving patient care and advancing medical knowledge.

First and foremost, it enhances the comparability of images across different patients and imaging systems, making it invaluable in multi-center studies and longitudinal assessments. Standardized pixel values facilitate the development and application of algorithms for quantitative analysis, such as the measurement of lesion size or density. It also supports the accuracy and reproducibility of image-based diagnoses and quantitative evaluations, which are critical in fields like radiology and oncology. Furthermore, image normalization reduces the impact of variations caused by differences in illumination and imaging conditions; thereby, making it easier for computer algorithms to detect patterns, features, and anomalies in the image data. Besides, normalized images are more robust for tasks like object recognition; as they enable classifiers to focus on essential image content rather than being sensitive to variations in lighting or contrast.

Normalization involves transforming the pixel values of an image to conform to a standard scale or distribution. This can be expressed as [17]

$$I_{\text{normalized}}(x, y) = \frac{I(x, y) - \mu}{\sigma}, \quad (1.1)$$

where $I_{\text{normalized}}(x, y)$ represents the normalized pixel value at coordinates (x, y) , $I(x, y)$ is the original pixel value, μ denotes the mean intensity of the image, and σ is the standard deviation.

Segmentation: This is a crucial component of medical image processing, involving unsupervised learning. It is used to analyze complex anatomical structures or lesions in an image. Segmentation enables the precise identification and differentiation of various anatomical structures and/or pathological regions; which is vital for subsequent disease diagnosis and treatment planning. For example, it can delineate different organs, lesions, or tumors within an image; thereby, providing clinicians with valuable insights into the patient's condition. Secondly, segmentation facilitates quantitative analysis by allowing the calculation of volumes, areas, or other morphological measurements for each segmented region. This can, again, be critical for tracking disease progression or evaluating treatment efficacy. Lastly, it supports the development of automated diagnostic tools and AI-based systems to assist healthcare professionals by providing accurate and consistent image interpretations. The synergy of pattern recognition

and segmentation contributes significantly to advancing our understanding of complex medical images for improved clinical decision-making. This ultimately enhances the efficiency and reliability of medical image analysis – leading to more effective patient care and improved decision-making.

Segmentation involves partitioning an image into distinct regions [17]. Let R denote the complete spatial region encompassed by an image I . Image segmentation can be conceived as a procedure that divides R into n disjoint sub-regions, denoted as $R_1, R_2, \dots, R_n$, such that

- (a) $\bigcup_{i=1}^n R_i = R$,
- (b) R_i is a connected set, for $i = 0, 1, 2, \dots, n$,
- (c) $R_i \cap R_j = \emptyset$ for all i and $j, i \neq j$,

where $\emptyset$ denotes the null set, and the symbols $\cup$ and $\cap$ signify set union and intersection, respectively. The two regions, R_i and R_j , are said to be *adjacent* if their union constitutes a connected set. Otherwise, these regions are described as *disjoint*. Each region corresponds to a specific segment or category.

Classification: Here the objective is to categorize input data into predefined classes or labels, based on learned patterns and features. Classification, employing ML and DL, is designed to extract discriminative patterns from data; thereby, enabling their automated and accurate assignment to specific classes. These algorithms learn to recognize intricate patterns that distinguish one class from another, arriving at decisions based on the detected features. Classifiers [42] constitute a fundamental component of learning, and lead to categorization of images into specific classes through supervised learning. While segmentation is unsupervised, classification provides a systematic and objective approach to diagnosing and interpreting medical images through training. By categorizing images into classes, a healthcare professional can quickly and accurately identify the presence of diseases, anomalies, or specific conditions within an image. Secondly, classification enhances the efficiency and consistency of image analysis by reducing the variability in interpretations that can arise from human subjectivity. It forms the basis for the development of automated diagnostic tools and related AI systems, where DL algorithms can be trained to recognize specific patterns or features within images based on ground truth (or annotations).

Let C denote the set of possible classes, with $f : I \rightarrow C$ being the classification function that assigns an image I to a class $c \in C$. This can be expressed as $c = f(I)$. The supervised learning comprises training, validation, and testing components, towards building, optimizing, and assessing the performance of the classifier. During training the classifier adjusts the model parameters to minimize the error of prediction w.r.t. the actual target values, based on the training set. The learning iteratively minimizes the error, through tuning of model hyperparameters, to capture any patterns or relationships inherent in the data. Validation is typically performed on a separate validation set, to determine the generalizability of the learning. Finally the learning performance is evaluated on an unseen test set. This can offer insights into the robustness and reliability of the classifier in real-world scenarios.

Ensemble classification: An ensemble-based classifier system is designed by merging multiple diverse classifier models [115]. This makes statistical sense in a number of situations, particularly with experts differing in decision-making. Individual classifiers, due to their inherent diversity, may have various decision limits within the context of classification. This is commonly due to the distinct training setup in each classifier. If adequate diversity can be established, each classifier may make different decisions; which may then be strategically combined, through ensemble, to reduce the overall error [138].

Advantage of ensemble learning lies in its inherent diversity. This can be introduced by embedding different training datasets, or features, or classifiers; or even varying initialization and/or parameters of the classifier(s) involved. According to Dietterich [39] 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 factor addresses situations where the selected model is unable to accurately represent desired decision boundaries.

Bootstrap Aggregating (Bagging) [25] involves training multiple instances of the same base model on randomly sampled subsets with replacements (bootstrap samples) of the training data. Individual model predictions are then averaged (for regression) or majority-voted (for classification) to produce the final ensemble prediction. Random forests, based on decision trees, employ bagging. Boosting [25] allows a series of weak learners to be sequentially trained, with each one focusing on samples that the previous learners failed to classify. These predictions are weighted on their accuracy, with the final decision aggregating the weighted individual predictions. Some of the popular boosting algorithms include Adaptive Boosting (AdaBoost), Gradient Boosting, and XGBoost. Stacked Generalization (Stacking) combines predictions from multiple diverse models, not necessarily of the same type. A meta-model is trained on top of the base models to learn how to best combine their predictions. It leverages the strengths of various models and is often used to further improve their predictive performance. Gradient Boosting, including variations like XGBoost, LightGBM, and CatBoost, builds a strong predictive model by sequentially adding weak models. It minimizes a cost function by adjusting the model parameters in a direction reducing the prediction error. Each new model is trained to correct the errors of the previous ones, resulting in a highly accurate ensemble. Voting is a straightforward ensemble strategy, where the final prediction is determined by majority vote (in classification) or average (in regression) over a number of individual predictions. Voting can be hard (majority vote) or soft (weighted average), depending on how the final prediction is computed.

Ensemble essentially strives to harness the collective wisdom of multiple models, by leveraging their complementary strengths and mitigating their individual weaknesses. It is widely used in ML to improve predictive accuracy, robustness, and the overall performance of models across various domains.

Feature visualization: Feature extraction and feature selection are integral components of pattern recognition and ML, particularly when dealing with high-dimensional data. Feature extraction [42] involves projecting the original data into a more compact and informative representation space, by generating a set of transformed features that capture relevant information from the raw data. This process reduces the dimensionality of the data while retaining its discriminative power. Some of the well-known techniques include Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and various DL approaches like autoencoders (AEs).

On the other hand, feature selection [42] focuses on identifying and retaining only the most relevant features from the original set. This process also aims to reduce dimensionality, enhance model interpretability, and mitigate the risk of overfitting. Feature selection methods assess the importance of each feature and select a subset of them for model training. Popular categories include filter methods, wrapper methods, and embedded methods, each with its individual criteria and algorithms.

Once the essential attributes have been identified, visualization allows the exploration and interpretation of their significance. It helps reveal patterns, relationships, and interactions among selected features. It can be essential for understanding the decision boundaries of ML models; by providing insights into how features contribute to classification or regression outcomes. Feature visualization is used in medical image processing, for highlighting and rendering significant patterns or structures (features) within medical images. It aids in the extraction of relevant information from complex data, leading to more accurate diagnoses and informed medical decisions [71].

Let $F(x, y)$ denote the feature map, which highlights specific features within the image I , with $F(x, y) \in \{0, 1\}$. Pixels in the feature map can be binary values (0 or 1), indicating the presence or absence of a particular feature at coordinates (x, y) . It simplifies the interpretation of complex medical images by emphasizing important structures or anomalies; thereby, reducing the cognitive burden on healthcare professionals. It enhances the efficiency of image analysis through its clearer and more focused view. Also, it supports more precise and targeted treatment planning with clinicians being able to easily identify ROIs for intervention. For instance, in radiation therapy, it aids in precise tumor targeting while sparing healthy tissue. Thirdly, feature visualization provides some interpretation about the black box in DL models; and evaluate the effectiveness of deep networks for predicting diagnoses based on relevant features. Ultimately, feature visualization enhances the accuracy, efficiency, and reliability of medical image processing – resulting in better patient care and medical decision-making.

Gradient-weighted Class Activation Mapping (GradCAM) [151] is a powerful visualization technique used in DL, particularly involving CNNs. It provides insights about the regions of an input image which were most influential in the decision-making process of the model. GradCAM is especially useful for understanding where a neural network focuses its attention during learning, and why it makes specific predictions; thereby, enhancing its explainability in the context of the typical blackbox nature. Basically GradCAM computes

the gradients of the class score with respect to the feature maps, in the final convolutional layer of the network. These gradients highlight the regions of the image that had significant impact on the prediction. Overlaying these gradients onto the original image, one can generate a GradCAM view which visually represents those parts of the image that were crucial for the decision made by the network.

Color-map [203] is often used in combination with GradCAM to make the heat map visually more interpretable. Here the heat map is color-coded to further accentuate the ROIs. For example, the highest-impact regions can be highlighted in red, while less influential areas get shaded by cooler colors like blue or green. Such color-coded visualization enhances the interpretability of the feature map; thereby, making it easier for researchers and domain experts to understand which parts of the input have contributed to a particular decision.

Both GradCAM and color-map are valuable tools for enhancing the understandability of DL models, allowing practitioners to trust the predictions through insights gained about model behavior; thereby, making more informed decisions when working on complex multimodal image data.

1.3.2 Deep learning

Machine learning (ML) and deep learning (DL) are subsets of AI that focus on enabling computers to learn and make decisions from data. Both ML and DL algorithms are trained on data to recognize patterns and relationships. DL, a subfield of ML, utilizes deep neural networks (involving multiple layers) to model complex data. It learns end-to-end on the data, without requiring expert-extracted features to be supplied at the input; thereby, making it particularly effective for tasks like image and speech recognition. These technologies have demonstrated their potential in revolutionizing industries by enhancing their ability to process and analyze vast amounts of data – leading to improved decision-making and automation across a wide range of applications.

The beauty of DL lies in its capacity to automatically learn and discover representations and features within the data, while allowing for superior performance in various applications including medical image processing [78], [142], [171], [194]. For example, CNNs are adept at image classification tasks, enabling the accurate detection and categorization of diseases from medical images; be it at identifying tumors in mammograms or detecting anomalies from X-ray images. A schematic diagram of the CNN is provided in Fig. 1.3, along with a description of the various layers involved.

- **Input layer:** Receives the raw image as input to the deep neural network. Mathematically, here it represents the input features as I .
- **Convolutional layer:** Consists of multiple filters (kernels) that perform convolution operations on the input data. The mathematical representation for the output feature map at layer l is

$$F^l = W^l * I^{l-1} + B^l, \quad (1.2)$$

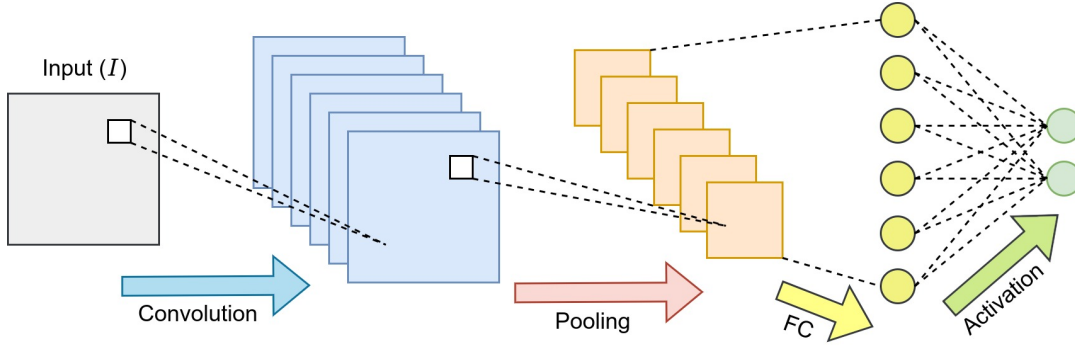


FIGURE 1.3: Schematic architecture of the CNN

where W^l are the convolutional filters, I^{l-1} is the input from the previous layer, and B^l represents the biases. The weight update at each iteration follows

$$\omega' = \omega - Lr \times Gr, \quad (1.3)$$

where ω' is the updated network weight vector, Lr is learning rate, Gr is the gradient of the loss function w.r.t. the weight vector, and ω is the corresponding weight vector at the previous iteration.

- **Activation Layer:** Introduces non-linearity into the network. The output of a neuron or unit in the activation layer is computed as

$$A^l = g(F^l), \quad (1.4)$$

where g is the activation function, like Rectified Linear Unit (ReLU) or sigmoid.

- **Pooling layer:** Reduces the spatial dimensions of the feature maps and, thereby, decreases the computational complexity. Max-pooling is a common operation, which typically selects the maximum value within a pooling window.
- **Fully Connected layer (FC):** Here all neurons are connected to the previous layer, which allows for complex patterns to be learned. The output of a neuron in a fully connected layer is computed as

$$FC^l = g^l(W^l A^{l-1} + b^l). \quad (1.5)$$

The learning rate is a hyperparameter that determines the step size at which the network weights are updated during training. It plays a crucial role in the performance as well as the speed of convergence of the model. Typical values for the learning rate are between 0.1 and 0.0001, but the optimal value depends on the specific problem and architecture.

Dropout is a regularization technique, typically applied at the fully connected layers and activation layers, to prevent overfitting of the model. It involves randomly "dropping out" a fraction of neurons during training, with probability p_d , such that they get excluded during the forward and backward

passes. This helps prevent the network from relying too heavily on specific neurons, and forces it to learn more robust and generalized features. We have

$$\text{Neuron Output} = \begin{cases} \frac{\text{Original Output}}{1-p_d} & \text{active,} \\ 0 & \text{inactive,} \end{cases} \quad (1.6)$$

where $1 - p_d$ is the probability of a neuron being retained. A typical dropout probability lies between 0.2 and 0.5. Adjusting the dropout probability provides a balance between regularization, and allowing the network to learn complex patterns from the training data.

Optimizers play a crucial role in training deep neural networks, during the minimization of the loss function. The primary purpose of optimizers is to adjust the weights and biases of the model to minimize the difference between the predicted output and the actual target values. Some of the well-known optimizers include [146] Stochastic Gradient Descent (SGD), Root Mean Square Propagation (RMSprop), Adaptive Gradient Algorithm (Adagrad), and Adaptive Moment Estimation (Adam) [82].

Several well-known DL models have been pivotal in advancing the fields of classification and segmentation in image analysis and computer vision. These include Residual Network (ResNet) [64], DenseNet [70], Visual Geometry Group (VGG) [158], Inception-Net [165], Xception-Net [29], Efficient-Net [84], MobileNet [23], Residual Attention Network and Multiscale Dilated Inception Network (RA-MDIN) [175], HyEx-Net [60], SE-ResNeXt-50 [140], *U*-Net [144], *U*-Net++ [206], Fully Convolutional Network (FCN) [102], R2UNet [5], Attention *U*-Net [124] and MultiResUNet [73]. The ResNet, with its residual connections, addresses the vanishing gradient problem; thereby, enabling the training of very deep networks. DenseNet maximizes feature reuse and gradient flow in DL by densely connecting each layer to every other layer. VGG, known for its simplicity and effectiveness, has been a popular choice for image classification tasks. Inception-Net employs multiple filter sizes in parallel, for enhancing feature extraction. Xception-Net balances model depth and computational efficiency, using depth-wise separable convolutions, to optimize feature extraction for image classification tasks. Efficient-Net balances the accuracy in performance with associated computational complexity, by efficiently scaling the network depth, width, and resolution. MobileNet uses depthwise separable convolutions to reduce computational complexity and model size while maintaining competitive performance, making it efficient for deployment on resource-constrained devices. RA-MDIN combines Residual Attention Network and Multiscale Dilated Inception Network, with model parameters optimized using Hybrid Beluga Whale Glowworm Swarm Optimization (HBWGSO). HyEx-Net integrates EfficientNet and Xception architectures to extract deep spatial features from combined inputs. SE-ResNeXt-50 is fused with convolutional layers, incorporating quadri-histogram equalization, data augmentation, and hyperparameter tuning to enhance feature representation and model robustness.

While the FCN introduced end-to-end segmentation using convolutional layers, the *U*-Net has been a game-changer in segmentation with its hallmark encoder-decoder architecture. The innovative *U*-Net++ version optimized skip

connections, to gather features at different semantic scales within the decoder sub-networks. The R2UNet combines the *U*-Net architecture with a recurrent neural network (RNN) for enhanced performance in medical image analysis. The Attention *U*-Net additionally incorporates attention mechanism to enhance the *U*-Net architecture; thereby, allowing it to focus on and prioritize important image features for improved segmentation accuracy. The MultiResUNet introduced multiple residual blocks, into the *U*-Net framework, for better segmentation in medical images.

The transfer learning paradigm involves leveraging the learned weights, of a pre-trained neural network, on a large dataset for a related task; followed by its fine-tuning on a target segmentation task with a smaller dataset. By transferring knowledge from the source task, the model can effectively learn hierarchical features and spatial relationships, while accelerating convergence and improving segmentation performance on the target task. This approach is particularly beneficial when labeled data is limited [132], since the pre-trained model has already captured generic features for various related visual tasks.

On the whole, DL has significantly improved the accuracy and efficiency of classification, detection, and segmentation tasks in medical imaging applications [119]. The family of *U*-Net architectures allow for a precise delineation of anatomical structures or lesions. This is critical for treatment planning and surgical guidance. Furthermore, DL techniques are leveraged to enhance image quality, reconstruct missing data, and remove noise. These are factors relevant for accurate diagnosis and treatment, made possible through its ability to decipher intricate patterns in images; particularly, when trained on large labeled datasets. DL has ushered a new era of progress in medical image analysis – by significantly elevating diagnostic accuracy, resulting in enhanced patient care.

1.3.3 Class imbalance and data augmentation

The term class imbalance in ML refers to a situation where the distribution of instances across different classes is significantly uneven. Such imbalance can pose challenges during model training, as the model may become biased towards the majority class with unacceptable performance on the minority classes. In scenarios where one class significantly outnumbers the others, the model tends to prioritize accuracy on the majority class while potentially neglecting the minority ones. Addressing class imbalance is crucial to ensure that the model learns to make accurate predictions across all classes. Strategies to mitigate class imbalance include resampling techniques, such as oversampling the minority class, undersampling the majority class, or using more advanced methods like generating synthetic samples. Proper handling of class imbalance is essential for building models towards reliable predictions across diverse class distributions.

As DL is data hungry in an environment with a scarcity in availability of annotated medical imaging data, it becomes imperative to devise ways to generate more training data. Data augmentation is an important technique in DL for enhancing the diversity and size of training datasets. It involves applying various transformations to the existing training data, for generating new

instances while preserving the underlying semantic content. Common augmentations include random rotations, flips, zooms, shifts, and changes in brightness or contrast. Presenting the model with such extensive range of variations, data augmentation helps improve its generalization capability and robustness. This is particularly beneficial when working with limited datasets, as it mitigates the risk of overfitting by exposing the model to a broader range of scenarios. Data augmentation is widely employed in computer vision tasks, such as image classification and object detection, to enhance the ability of a DL model to handle diverse real-world scenarios during inference.

1.3.4 Role of multimodal imaging

Medical imaging offers a noninvasive way of generating images of different regions in the human body, to assist medical practitioners to arrive at an effective decision regarding the detection, diagnosis and prognosis of a disease. Automated medical image analysis can effectively handle large volumes of image data, while overcoming any inherent human bias. A good visualization of different components in a lung region, encompassing blood vessels and tissue characteristics, can be associated with corresponding imaging features which become better visible when focusing on particular attributes. This highlights the utility of employing multiple imaging modalities while identifying and extracting heterogeneous affected region(s) from the lung scans. A multimodal approach, encompassing a combination of two or more of the imaging modalities like X-Rays, CT, MRI, PET, EIT, holds potential in enhancing the visualization, accuracy, efficiency, and viability of associated decision-making [109], [150].

CT scans provide high-resolution, three-dimensional images to help visualization, diagnosis and evaluation of various medical conditions. The scans can be reconstructed along multiple planes (axial, sagittal, and coronal), allowing healthcare professionals to view the lung from different perspectives. As early detection enables timely intervention with improved patient outcomes, the powerful modality of CT is often used in the diagnosis, monitoring, and treatment planning for a wide range of medical conditions of the lung. This includes cancer and COVID-inflicted pneumonia.

EIT measures the electrical current and voltage at multiple electrodes placed on the body surface, to reconstruct the internal impedance distribution. The advantages of EIT (over other modalities) include (i) Absence of any ionizing radiation or injection of contrast agents; making it particularly beneficial for frequent monitoring and bedside applications. (ii) Rapid update of the reconstructed images. (iii) Cost effectiveness, as compared to other imaging modalities. These make EIT a viable option in resource-limited settings, or for the routine real-time monitoring in critical care units. However, the spatial resolution of EIT is lower as compared to the other imaging modalities, and the image quality can be affected by external factors like electrode placement and movement artifacts.

PET imaging provides valuable insights into the functional and metabolic processes occurring within the human body. It involves the administration of a

small amount of a radiotracer (a radioactive substance labeled with a positron-emitting isotope) into the patient's body. As the radiotracer decays, it emits positrons which interact with electrons in the body to generate pairs of gamma rays detected by PET scanners. The advantages of PET over other imaging modalities are numerous. It excels at offering functional information, by revealing details about tissue activity, glucose metabolism, blood flow, and receptor binding. PET is known for early disease detection, making it particularly valuable for identifying cancers at their inception. It provides comprehensive whole-body scans, aids in personalized medicine by tailoring treatment strategies, and plays a pivotal role in medical research and drug development. Moreover, PET imaging is non-invasive, safe, and can be combined with other imaging modalities in hybrid setup. For example, PET-CT imaging offers a holistic view of the condition of a patient.

Let us summarize here the display characteristics of images of COVID-19 and lung cancer samples, corresponding to the three imaging modalities considered in this thesis *viz.* CT, PET and EIT.

1. CT scans can reveal characteristic features of COVID-19, such as ground-glass opacities (GGOs) and consolidation in the lungs. These findings help in diagnosing and assessing the extent of lung damage. CT scans are also essential for the detection and staging of lung tumors, through their effective visualization of the size, location, and potential spread of the tumor. Fig. 1.4 illustrates sample CT scans corresponding to these two types of lung affliction.

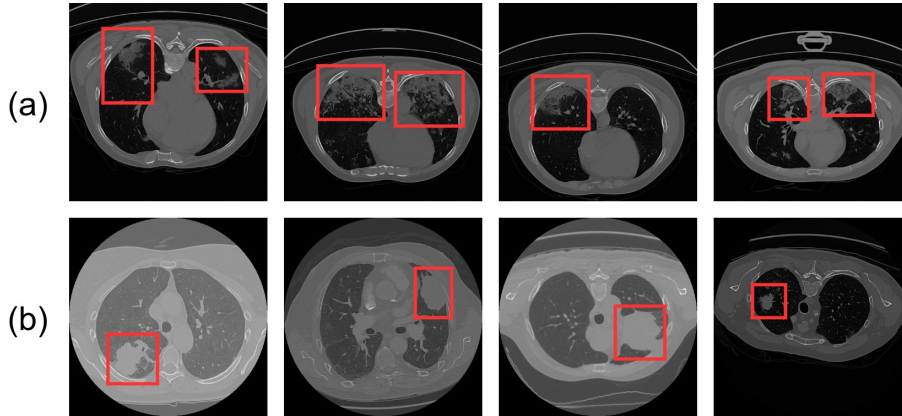


FIGURE 1.4: Sample CT images of the lung, afflicted with (a) COVID-19, and (b) tumor. The ground truth of affliction is marked in red.

2. PET scans are valuable for cancer staging. They reflect areas of increased metabolic activity, which is often indicative of malignant tumors. A combination of PET and CT images, as in Fig. 1.5, provides better visualization of metabolic activity. Although not commonly employed in COVID-19 diagnosis, their role can be explored to understand metabolic changes in the affected tissues.
3. EIT can help assess lung ventilation conditions, and potentially identify areas affected by the disease as measured in terms of reduced ventilation.

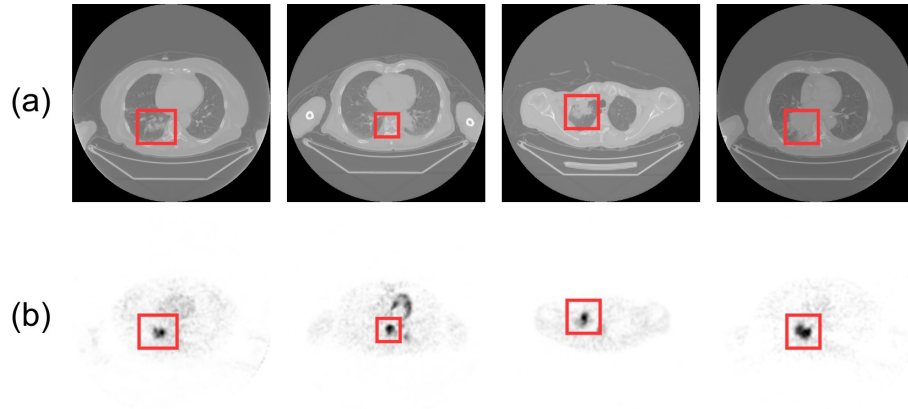


FIGURE 1.5: Sample images of lung tumors, using (a) CT and (b) PET scans, with the ground truth marked in red.

It can assist in monitoring lung function in patients with lung tumors, and evaluate the impact of tumor growth and/or treatment progression. Sample EIT images of the afflicted lung are displayed in Fig. 1.6.

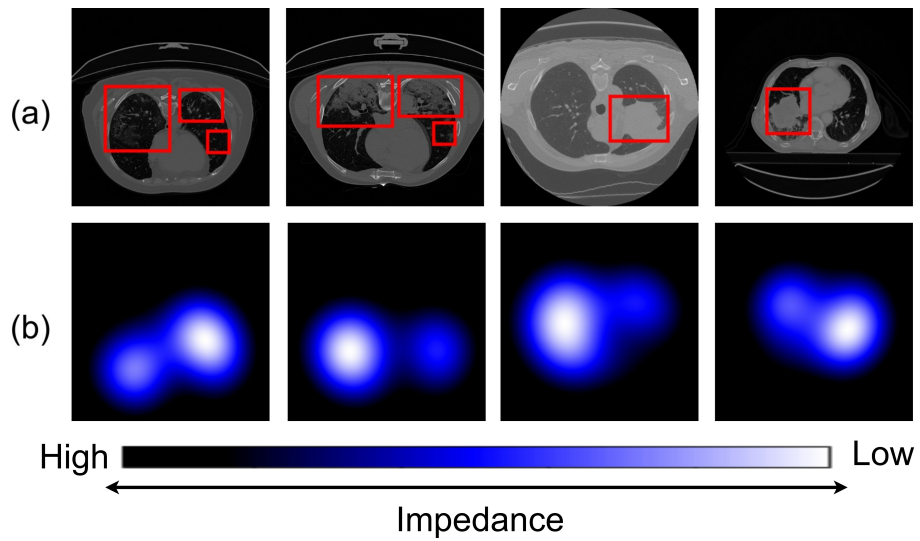


FIGURE 1.6: Sample images of the lung, with (a) CT scans and (b) corresponding EIT. The ground truth of affliction is marked in red. First two columns correspond to COVID-19 infection of the lung, and last two columns display lung tumors.

The integration of these modalities, often referred to as multimodal imaging or hybrid imaging, allows healthcare professionals to gather complementary information. For instance, while diagnosing lung tumors, a PET-CT scan combines the metabolic information from PET with the anatomical detail provided by CT. This approach can help in precisely locating tumors and assessing their activity. Combining CT and EIT can offer real-time insights into lung function and structural changes caused by both these diseases.



1.4 Background Literature

ML offers a consistent and reproducible way for the rapid evaluation of high volumes of screening and/or diagnostic imaging studies. It excels at the detection and precise delineation of affected lung tissues from the slices. DL models, in particular, have the capability to autonomously learn intricate patterns and representations directly from pixel values in medical images; thereby, mitigating manual errors often introduced through human intervention.

In this section we provide a brief overview on the state-of-the-art literature on the major functionalities of automated image analysis, as considered in this thesis; with particular emphasis on afflictions of the lung. This encompasses the use of DL in classification, segmentation, including ensemble learning and multi-modal imaging approaches; focused particularly in the domain of lung imaging applications. The studies underscore the wide-ranging application of fusion techniques in medical imaging, with each modality or approach contributing towards improved data integration and interpretation.

1.4.1 Classification

The advancement of the role of DL in image classification, particularly through CNNs, ushered in a period of significant progress and success. This journey began with the introduction of AlexNet [86], with its success in the ImageNet [37] classification competition; thereby, bolstering the explosive growth of DL. The Network in Network (NIN) architecture [97] emerged by incorporating global average pooling to mitigate risks of overfitting. The subsequent years witnessed notable advancements with GoogLeNet [165] and VGG-Net, both resulting in enhanced accuracy on the ImageNet dataset. GoogLeNet continued to evolve with versions like v2 [74], v3 [166], and v4 [164], entailing further improvement in performance. Addressing the limitations of fixed input size requirements in CNNs led to the introduction of the spatial pyramid pooling (SPP) model [65] with enhanced input data robustness. ResNet offered a solution to potential model degradation issues, while emphasizing the continuous advancements in DL technology.

Among the various DL algorithms applied to pulmonary nodules, CNNs have emerged as the most prevalent [96], [147]. Their application in lung cancer based CAD systems has yielded impressive results. Google introduced a CNN model [9], that outperformed six experienced radiologists. Additionally, alternative DL techniques have been successful in addressing pulmonary nodule challenges. Deep belief networks (DBN) have been utilized in the context of pulmonary nodule CAD [69], and AE helped distinguish between benign and malignant nodules [88]. A sliced RNN model was introduced [181] to enhance training efficiency by simultaneous training of the layers over reduced training time. A two-phase framework employed a 2D CNN for automatic detection of nodules from CT readings [190], incorporating a Region-based CNN and deconvolution. Here three models were trained for three kinds of slices, and the network results were combined through voting for the final classification outcome. A 3D CNN

with multi-scale prediction strategy was suggested [57] for pulmonary nodule detection. A 3D-ResNet was employed for the classification of CT images for presence of pulmonary nodules [170]. Texture radiomic features were utilized to develop CNN-based predictive models for analyzing malignant and benign lung nodules [133]. An automatic lung cancer prognosis system was designed to evaluate lung nodules, based on their morphological features [180].

In recent times, a surge of research endeavors has emerged in the realm of COVID-19; with a notable focus on the application of neural networks and data-driven algorithms [61]. These studies encompass a spectrum of ML approaches dedicated to diagnosing COVID-19 from CT images. A pre-trained DenseNet was developed [179] for classifying 121 CT-images into COVID-19 positive and negative categories. Additionally, a DL-based multi-view fusion model has been implemented to classify COVID-19 from CT scans [186], while a multi-objective differential evolution-based CNN has been utilized for the same purpose [159]. A deep multitask learning approach has been used to perform binary classification and estimate the percentage of affected lung tissue in COVID-19 from CT studies [53]. A hybrid COVID-19 framework was designed for rapid and automatic diagnosis of COVID-19 from X-Ray images, combining deep learning, genetic algorithms, weighted sum, and majority voting strategies [14]. Furthermore, a semi-supervised deep transfer learning framework [157] has demonstrated success in pulmonary nodule diagnosis from CT scans.

An AI system, rooted in DL, was intricately engineered to detect and quantify lesions within chest CT scans [204]. Moreover, the Unified CT-COVID AI Diagnostic Initiative (UCADI) [12] pioneered an innovative paradigm wherein independent training at each host institution was facilitated within a collaborative learning framework, thereby obviating the necessity for data sharing. This ground-breaking approach not only upheld privacy but also showcased the potential for combined learning to advance the field of privacy-preserving AI in digital health.

Using ensemble learning: They demonstrate remarkable adaptability to evolving data patterns, owing to their inherent model diversity; thus ensuring their enduring relevance. Their robust stability, efficient resource utilization, and state-of-the-art performance collectively establish ensemble deep models as a formidable approach for a wide array of applications [189].

In the realm of medical imaging, a deep ensemble classifier has been harnessed for the automatic detection of lung cancer from CT images [153]. Likewise, the ensemble classifier [154] proficiently extracted features to classify lung tumors based on CT scans. Pulmonary nodule classification has benefited from deep ensemble learning-based feature fusion [117], showcasing the versatility of ensemble. A combination of VGG-Net, DenseNet, ResNet, and CNN, has been effectively employed [148] for the classification of lung cancer and COVID-19. Moreover, a sophisticated ensemble multimodal DL technique has been devised for the classification of COVID-19, effectively combining X-ray images with audio data extracted from associated coughs [89]. These applications underscore the remarkable utility and adaptability of ensemble-based DL methodologies.

Using multimodal imaging: The fusion of multiple imaging modalities proves invaluable, as it equips deep models with enhanced features that contribute to more informed decision-making. Multimodal fusion amplifies the capabilities of DL across various applications, enhancing feature extraction, model robustness, interpretability, and overall performance; particularly, in the presence of intricate and multifaceted data. By amalgamating heterogeneous data from various imaging modalities such as MRI, CT, and PET, it produces a holistic and comprehensive data representation which, in turn, unveils concealed patterns while bolstering model accuracy and resilience. This fusion process effectively mitigates the impact of noise and enhances the model ability to withstand uncertainties and variations. Furthermore, it leverages the strengths of individual modalities to achieve superior feature extraction, augmenting the capacity of the model to discriminate and capture complex relationships. The process also elevates spatial and temporal resolution, thereby benefiting tasks like medical image segmentation and super-resolution.

In a notable study, VGG19 was applied to a diverse dataset comprising chest X-ray and CT scans, facilitating the classification of lung diseases such as COVID-19, pneumonia, and lung cancer [72]. Additionally, a hybrid approach combining residual blocks with a bidirectional gated recurrent unit was implemented to effectively classify COVID-19 from X-rays and CT scans [3]. A framework based on deep unsupervised learning was suggested [192] for the classification of lung diseases, using chest CT and X-ray images. Another approach involved leveraging transfer learning on a multifaceted dataset encompassing X-Ray, Ultrasound, and CT scans, for the purpose of COVID-19 detection [66]. Complementary information was leveraged in weakly paired multimodal MRI images of the chest, for tumor detection [205]. The COVID-DSNet was developed for diagnosing typical pneumonia, viral pneumonia, and COVID-19 from CT, CXR, and hybrid CT+CXR images [143]. A dual-input CNN classified the health status of patients, encompassing asthma, Chronic Obstructive Pulmonary Disease, Interstitial Lung Disease and infection, from a combined dataset comprising respiratory sounds and EIT recordings [135]. Automated heart and lung detection was performed [46] through unsupervised learning on EIT and CT images. An EIT reconstruction algorithm was then used to comprehensively analyze the data. This strategy showcased the potential of leveraging multimodal images to enhance the accuracy and robustness of diagnostic models, particularly in the context of a disease as complex as COVID-19. A filtering technique based on gradient structural similarity was introduced for the integration of CT and MRI data [50]. Similarly, a fusion approach for CT, MRI, SPECT, and PET images has harnessed a modified shark smell optimization method in combination with a hybrid wavelet-homomorphic filter [191]. In another study, a CNN was tailored to handle integrated CT, MRI, and PET images [188]. Moreover, a method grounded in PCA was developed to effectively process a comprehensive set of X-ray, CT, MRI, and SPECT images [18].

1.4.2 Segmentation

Medical image segmentation has witnessed significant advancements through the effective utilization of prominent models like the FCN and *U-Net*. ResNet-18 has been harnessed for the segmentation and classification of lung lesions associated with COVID-19, pneumonia infection, and normal lung structures [196]. Furthermore, the domain has explored the utility of 2D, 2.5D, and 3D *U-Net* models in a multitude of studies [27], [59], [177].

Utilization of ensemble methods has been reported [21], [44] for CT image segmentation. Notably, a combination of Inception-V3, Xception, Inception, ResNet-V2, and DenseNet-121 models was employed for multiclass segmentation targeting GGO and Consolidation regions within COVID-19 CT data, specifically using the CT-Seg dataset (Table 1.1). These models leveraged a common CNN backbone, with initial weights from ImageNet further fine-tuned on the CT-Seg dataset. In terms of data distribution, the dataset was divided into training (40 images), validation (10 images), and testing (50 images) sets. The aggregation of results was achieved through pixel-level soft majority voting.

Efforts to enhance the segmentation of COVID-19-related lesions in lung CT scans have led to the introduction of the Dual Multiscaled Dilated Fusion Network (DMDF-Net) [125]. When applied to the MosMed dataset of Table 1.1, the model demonstrated accurate and refined segmentation capabilities. Furthermore, a novel domain adaptation-based self-correction model has been developed for the precise segmentation of COVID-19 infection in CT images, harnessing the combined resources of both the CT-Seg and Mosmed datasets [79]. Additionally, the field has witnessed the presentation of a lightweight neural network with attention-guided features designed for lung lesion segmentation from CT scans, based on the Kaggle dataset [160]. A cascade comprising two *U-Nets* with a VGG backbone was employed to sequentially extract the lung region, followed by delineation of the GGO and consolidation regions [21]. This multistage segmentation approach was applied across diverse datasets, including CT-Seg, Seg-nr.2, Kaggle-COVID-19, and a private dataset, as outlined in Table 1.1. Training data included segments of CT-Seg and Seg-nr.2, while the remaining portions of these datasets were utilized for model testing. Notably, each network within the ensemble underwent distinct training procedures due to random weight initialization and data augmentation with shuffling.

Lung tumor segmentation, a critical application, has been addressed through various adaptations of the *U-Net* architecture. The deep residual *U-Net* model has harnessed multi-view learning for enhanced segmentation [173], while the CoLe-CNN model [137] introduced an adaptive loss function to capture nodule context. A deep residual separable CNN with maximum intensity projection-based pre-processing was employed [43] for lung tumor segmentation. The 3D *U-Net* was complemented with a 3D conditional random field for improved accuracy in volumetric segmentation [185].

A pair of multiple-resolution, residually-connected CNNs (MRRN) simultaneously combined features across multiple levels and image resolutions, through residual connections, to detect and segment NSCLC lung tumor volumes [77] from multiple publicly-available datasets. The segmentation accuracy, compared to expert delineations, was evaluated in terms of performance metrics.

The Adaptive Deep Belief Network [80] was used to segment 3D NSCLC lung tumor images from CT images. An efficient lung tumor detector was developed [106], based on the RetinaNet. A multi-scale feature fusion module aggregated the network layers, for simultaneously increasing the semantic information from shallower prediction layers. A dilated and lightweight algorithm to combined contextual information, at each layer, for effectively localizing tiny tumors. Residual blocks were integrated with convolution layers in ResDSda_ *U*-Net [76], for multi-scale feature extraction in pulmonary nodule segmentation. The detection and segmentation of lung nodules was also performed in DB-NET [24], incorporating the Mish nonlinearity function with mask class weights. An encoder decoder deep architecture was developed with selective convolution kernels and grouped convolutions helped to efficiently segment lung nodules [183].

The Geographic Attention model (GA-Net) was designed [33] for efficiently segmenting body composition tissues. The rapid multi-crop attention network **chen2021multi** was a novel approach capable of modeling both 2D and 3D contexts from neighboring axial slices, providing richer information. A two-stage framework [122] has been designed, encompassing a 3D multi-scale *U*-Net at the coarse stage. This was followed by a 2.5D multi-scale separable *U*-Net in the refinement stage, for delivering highly precise tumor delineation. The Dual-branch residual network (*DB*-ResNet) [26] has introduced an intensity-pooling layer and multi-scaling for improved segmentation. The multi-scale squeeze-and-excitation *U*-Net has effectively incorporated conditional random fields for tumor segmentation [195].

A teacher-student framework in Ref. [49] has enabled automated tumor segmentation through a student model, guided by additional pseudo-annotated data from a teacher network. The diverse landscape of DL models has further expanded with the introduction of the Recurrent 3D-Dense *U*-Net [81] and the dilated hybrid-3D convolutional net [67]. These models collectively represent a robust foundation for advancement in medical image segmentation.

Now the field has ventured into innovative avenues, including the application of transformers. The Multi-scale Aggregation-based Parallel Transformer Network (MAPTransNet) [35] has introduced a parallel transformer mechanism to effectively capture the global context within multi-scale encoder feature maps. The segmentation results generated by MAPTransNet are further utilized as ROIs for subsequent survival analysis, exemplifying the synergy between segmentation and downstream clinical applications. Hybridization of convolution layers with the State Space Model in *U*-Mamba [104] allowed the capture of local features, along with long-range dependencies, for biomedical image segmentation. The Swin-*U*Mamba [99], integrating Swin Transformer [101] and Mamba blocks, utilizes the strengths of both Transformers and structured state space models. Transfer learning has been instrumental in lung tumor segmentation [123].

Exploring understandability: While DL models are acknowledged for their proficiency in achieving accurate predictions, they often necessitate substantial

volumes of annotated training data, which can be resource-intensive. Moreover, these models are occasionally perceived as “black boxes” due to the limited transparency of their internal workings. Nonetheless, concerted efforts are underway to enhance the interpretability and explainability of DL in the realm of medical image analysis [108]. One notable approach is the development of integrated segmentation and classification techniques that facilitate explainable diagnoses of conditions like COVID-19 [182]. An explainable pipeline, based on deep learning, was developed for automated COVID-19 detection and lesion categorization from CT scans [134]. The COVID-19 Multi-Task Network (COMiT-Net) was designed [107] for automated screening, with feature prediction and explainable semantic segmentation, in chest x-ray images.

Methods such as Grad-CAM have emerged to enable researchers to visualize and elucidate the regions within an image that exert the most significant influence on the model decision. The Domain Extension Transfer Learning (DETL) was leveraged to screen for COVID-19, with characteristic features derived from chest X-ray images [19]. To shed light on the transparency of COVID-19 detection, the authors adopted the concept of Grad-CAM to identify regions where the model exhibited heightened attention during classification. Importantly, the study asserted a strong correlation of the model results with clinical findings.

DL has also found application in assessing the severity of COVID-19 infection [184]. The ongoing improvement of these models holds the potential to bolster the accuracy and efficiency of anomaly detection, ultimately resulting in earlier diagnoses and improved prognoses for various medical conditions.



1.5 Performance Evaluation

In this section we present the different measures used for quantitatively evaluating the performance of segmentation and classification in medical imaging. Besides, severity prediction in COVID-19 from lung CT scans is also included.

1.5.1 Quantitative evaluation

The frequently used performance metrics [110] for segmentation are based on the misclassification error, obtained in terms of the confusion matrix. This encompasses the True positive (TP), False positive (FP), True negative (TN), and False negative (FN) components. We evaluate

Dice Similarity Coefficient

$$DSC = \left(\frac{2 * TP}{2 * TP + FP + FN} \right), \quad (1.7)$$

Jaccard Similarity Coefficient

$$JSC = \left(\frac{TP}{TP + FP + FN} \right), \quad (1.8)$$

Precision

$$Precision = \left(\frac{TP}{TP + FP} \right), \quad (1.9)$$

Recall/Sensitivity

$$Recall/Sensitivity = \left(\frac{TP}{TP + FN} \right). \quad (1.10)$$

Area Under the Curve

The *AUC* measures the ability of a model to distinguish between positive and negative classes by calculating the area under the Receiver Operating Characteristic (*ROC*) curve in terms of a numerical integration. The *ROC* curve is a plot of the true positive rate (*TPR*) vs the false positive rate (*FPR*) at different classification thresholds. With $TPR = \frac{TP}{TP+FN}$, $FPR = \frac{FP}{FP+TN}$, we have

$$AUC = \int_0^1 TPR(FPR^{-1}(t))dt, \quad (1.11)$$

where $FPR^{-1}(t)$ is the inverse function of the *FPR* at threshold t , and *TPR* is the true positive rate at threshold t . In practice, *AUC* ranges from 0.0 to 1.0, with higher values indicating better classification performance. An *AUC* of 0.5 indicates a model that is no better than random guessing, while an *AUC* of 1.0 corresponds to a perfect classifier [178].

Hausdorff Distance 95

The HD95 [43] measures the maximum distance between points p_i in the predicted mask $\hat{y}$ and their closest points g_j in the ground truth mask y , while considering only the 95% of points that are closest. It additionally presents insight into the accuracy and spatial correspondence of segmentation results, and is expressed as

$$HD95(\hat{y}, y) = \max \mathcal{D}(p_i, g_j) : i = \{1, 2, \dots, |\hat{y}|\} \text{ and } j = \{1, 2, \dots, k\}, \quad (1.12)$$

where $\lceil k = 0.95 \times N \rceil$ and $\mathcal{D}$ corresponds to the Euclidean distance.

Accuracy

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN} \right), \quad (1.13)$$

F1-score

For a single class i , the precision (P_i), recall (R_i), and $F1$ -score ($F1_i$), are defined as $P_i = \frac{TP_i}{TP_i + FP_i}$, $R_i = \frac{TP_i}{TP_i + FN_i}$, with

$$F1_i = \frac{2 \times P_i \times R_i}{P_i + R_i}, \quad (1.14)$$

where TP_i , TN_i , FP_i , FN_i , correspond to the respective values for class i .

Macro-F1 score

The *Macro* – $F1$ score, for multi-class classification, is the unweighted average of the $F1$ scores for all classes. Given C classes, it is expressed as

$$Macro - F1 = \frac{1}{C} \sum_{i=1}^C F1_i. \quad (1.15)$$

Mean Average Precision

The Average Precision (AP_i) is computed for each class, from the area under the *Precision* – *Recall* curve. The mean Average Precision (mAP) becomes

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i. \quad (1.16)$$

These metrics can help evaluate highly imbalanced inputs, to produce generalized outputs in n -ary ($n > 1$) classification. Higher values of these indices typically imply a better quality of classification.

Confusion matrix

This is a table that summarizes the performance of a segmentation/classification model by showing the counts of TP , TN , FP , and FN predictions, in terms of pixels/slices; thereby, helping evaluate how well a model classifies pixels/slices.

True labels	Negative (0)	TN	FP
	Positive (1)	FN	TP
		Negative (0)	Positive (1)
		Predicted labels	

FIGURE 1.7: A sample confusion matrix.

1.5.2 Severity prediction

Grading of severity in a lung infection is a crucial aspect of image analysis, which aims to assess the extent, progression, and potential impact of a disease (or condition) on a patient. The predictive process involves the utilization of advanced image analysis techniques, and often integrates clinical data to provide a comprehensive evaluation of disease severity. The primary goal is to assist healthcare professionals in making informed decisions regarding disease prognosis, patient management, treatment planning, and disease monitoring. Various methodologies, including ML, DL, and computer vision techniques, are typically employed. These use features extracted from images, such as lesion size, shape, and texture, to quantify disease progression. For instance, in lung cancer, the prediction of tumor stage and aggressiveness based on radiological images aids in choosing the appropriate treatment strategy; involving one or more of the following *viz.* surgery, chemotherapy, radiation therapy. In the case of COVID-19, the severity prediction models analyze chest X-rays or CT scans to assess the extent of lung involvement. This is valuable in determining the need for hospitalization or intensive care.

The methodology of grading the severity of COVID-19 infection, as developed by the Russian Federation [116], 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 the volume of damaged lung tissue. For example

- 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.

The advantages of severity prediction in medical images are numerous. It provides a quantitative and objective assessment of disease progression; thereby, reducing subjectivity in diagnosis and treatment decisions. Moreover, these predictive models can facilitate early intervention, personalized treatment plans, and an optimized allocation of healthcare resources. Hence, by harnessing the power of AI and image analysis, severity prediction from images play a pivotal role in improving patient care and outcomes across various medical conditions.



1.6 Datasets and Software Packages

In this section we outline some of the datasets and software, used in this thesis, for studying the segmentation of lung regions affected by COVID-19 and tumors, along with the classification between healthy slices, tumor-affected slices,

COVID-19-infected slices based on multimodal imaging. The models were implemented in the Tensorflow framework with a dedicated GPU (NVIDIA TESLA V100 having memory capacity of 16GB), running behind the wrapper library Keras with Python 3.6.8, Keras 2.2.4, and Tensorflow-GPU version 1.13.1.

1.6.1 Datasets

Two major categories of lung diseases are considered in this thesis. Here we elaborate on the publicly available datasets, from these two kinds.

COVID-19: The Kaggle-COVID-19 data consists of 20 patient CT samples, of which ten (having at least 200, but not more than 301 slices) were kept as a separate set. This was termed the Kaggle-COVID-19:Part-1 dataset. The remaining ten samples, each containing < 200 or > 301 slices, were kept as a separate set. This was named the Kaggle-COVID-19:Part-2 dataset. The data is available on the Kaggle platform [105]. The CT-Seg and Seg-nr.2 were taken from the medical segmentation platform³. Here the CT-Seg data contains CT scan slices from more than 40 patients, and the Seg-nr.2 data contains nine patient samples. The MosMed data contains 50 patient CT samples, and is available in the Mosmed platform [116].

Lung tumor: The NSCLC-radiomics [2] data comprises of 422 patient CT samples of lung tumors, taken from TCIA platform; of which 263 annotated samples were retained for training, while the rest were omitted because of improper annotation. The NSCLC-radiogenomics [13] data contains 89 patient CT samples of lung tumor with annotations, taken from TCIA platform. Decathlon [7] data contains 63 patient CT samples with lung tumor annotations, taken from the decathlon platform. The Moffitt [131] data, contains 440 patient CT samples with lung tumor annotation. The Figshare [41] data contains CT and PET samples from seven patients. The Y-PET/CT [174] consists of four PET and CT volume data of the lung. The Lung-PET-CT-Dx [94] contains 356 CT and PET-CT volumes of the lung, with 133 volumes containing both PET and CT images.

The Hounsfield Unit (HU) range (for the CT samples) was initially maintained in the range $[-1024, 3071]$, with the pixel values being normalised to provide appropriate comparison across various data sources used in the thesis.

1.6.2 Software

The algorithms developed in this research, for detection, segmentation, and classification of lung infections (*viz.* COVID-19 and tumours), were crafted using a combination of programming tools. MATLAB⁴, along with Python⁵, served

³<http://medicalsegmentation.com/covid19/>

⁴A versatile numerical computing environment and proprietary programming language developed by MathWorks: <https://www.mathworks.com/products/matlab.html>

⁵An interpreted, high-level, general-purpose language: <https://www.python.org/>

TABLE 1.1: Details of the lung datasets used

Disease	Sr.	Dataset	Total patients	Data Modality	Platform
COVID-19	1	Kaggle-COVID-19 ^a : Part-1	10	CT	Kaggle
		Kaggle-COVID-19 ^a : Part-2	10	CT	Kaggle
	2	CT-Seg ^b	>40	CT	Medical Segmentation
	3	Seg-nr.2 ^b	9	CT	Medical Segmentation
	4	MosMed ^c	50	CT	Mosmed
Lung tumor	1	NSCLC radiomics ^d	263	CT	TCIA
	2	NSCLC radiogenomics ^e	89	CT	TCIA
	3	Decathlon ^f	63	CT	Decathlon
	4	LIDC-IDRI ^g	875	CT	TCIA
	5	Moffitt ^h	440	CT	Moffitt
	6	Rider ⁱ	31	CT	TCIA
	7	Figshare ^j	7	CT, PET	Figshare
	8	Y-PET/CT ^k	4	CT, PET	Deep Blue Data
	9	Lung-PET-CT-Dx ^l	133	CT, PET	TCIA

^aKaggle-COVID-19: <https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans> [105]
^bCT-Seg & Seg-nr.2 (Two Datasets): <http://medicalsegmentation.com/covid19/>
^cMosMed: <https://mosmed.ai/en/> [116]
^dNSCLC radiomics: <https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics> [2]
^eNSCLC radiogenomics: <https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics> [13]
^fDecathlon: <https://drive.google.com/drive/folders/1HqEgzS8BV2c7xYNrZdEAnrHk7osJJ--2> [7]
^gLIDC-IDRI: <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254> [10]
^hMoffitt: [131]
ⁱRider: <https://wiki.cancerimagingarchive.net/display/Public/RIDER+Lung+CT> [199]
^jFigshare: <https://figshare.com/s/8370d1d7e5cbc195a1af> [41]
^kY-PET/CT: https://deepblue.lib.umich.edu/data/concern/data_sets/st74cq55j?locale=en [174]
^lLung-PET-CT-Dx: <https://www.cancerimagingarchive.net/collection/lung-pet-ct-dx/> [94]

as the primary development platforms. DL models were constructed using TensorFlow⁶, and Keras⁷, both integrated into the Python ecosystem. Additional software packages included pillow⁸, matplotlib⁹, and scikit-image¹⁰. Visualization was facilitated with the software tools ITK-SNAP¹¹ and 3D Slicer¹².

⁶A symbolic math library widely utilized for ML: <https://www.tensorflow.org/>

⁷An open-source neural network library: <https://keras.io/>

⁸Designed for image manipulation and analysis in Python: <https://pypi.org/project/Pillow/>

⁹For data visualization: <https://matplotlib.org/>

¹⁰An open-source Python library tailored for image processing and analysis: <https://scikit-image.org/>

¹¹An interactive application enabling users to navigate three-dimensional medical images: <http://www.itksnap.org/pmwiki/pmwiki.php>

¹²A freely available open-source software package designed for image analysis and scientific visualization: <https://www.slicer.org/>

1.7 Scope of Thesis

The objective of this thesis is to advance the field of medical image analysis by developing innovative non-invasive tools, based on deep learning and multi-modal imaging, for the delineation and classification of diseases related to the lung. More specifically, we present efficient detection and diagnostic models for the classification and segmentation of affected regions – resulting from diseases like COVID-19 and lung cancer. The imaging modalities considered here include CT, PET and EIT scans. We aim to contribute towards the development of more accurate and efficient methods for the diagnosis, prognosis, and classification of diseases, utilizing cutting-edge techniques to address the challenges of limited data and class imbalance. The goal is to ultimately offer valuable support to medical practitioners through enhancement of the field of automated medical applications. The effectiveness of our methods is established on the publicly accessible datasets of medical images of Table 1.1, through comparison with those of state-of-the-art related literature. The findings from our investigations are described in subsequent chapters, and also summarized in Fig. 1.8.

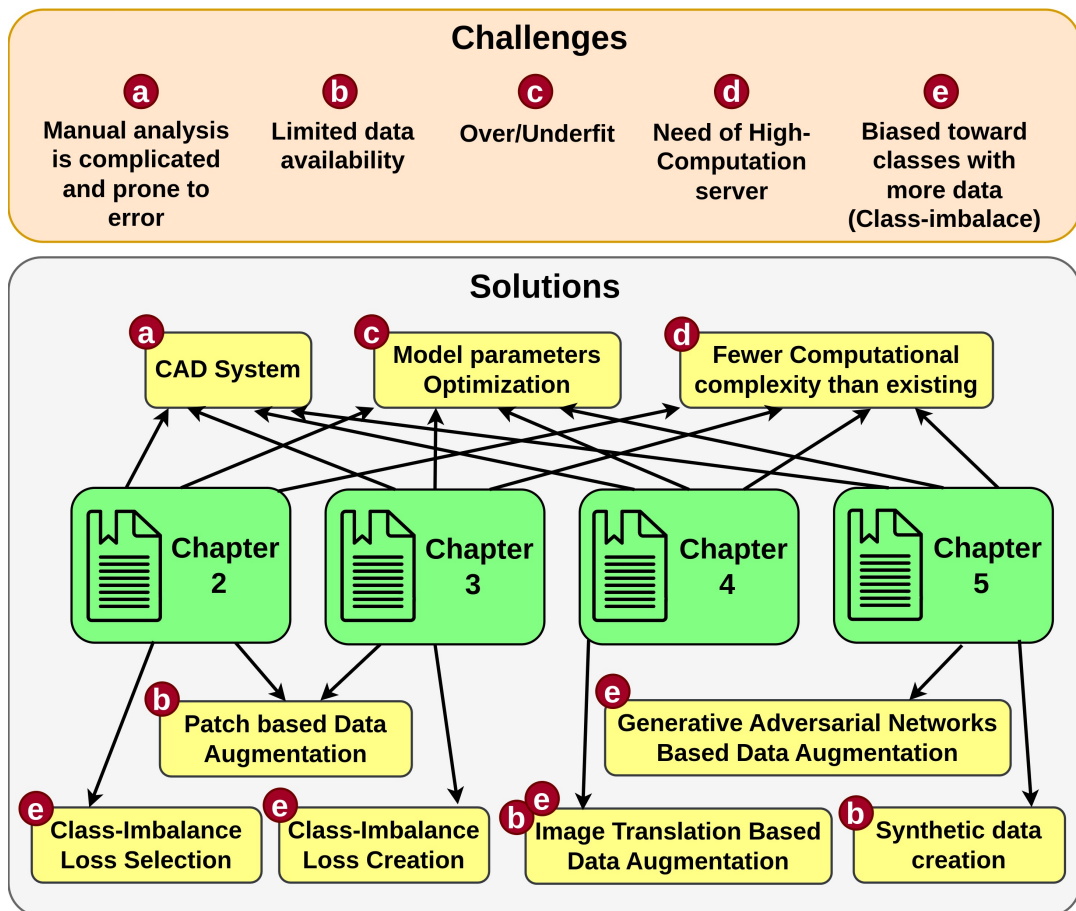


FIGURE 1.8: Handling the challenges.

1.7.1 Multi-resolution deep ensemble learning for segmentation [128]

The early detection, diagnosis, isolation, and prognosis of COVID-19 are pivotal in containing its spread. CT scans have emerged as a valuable supplement to the gold standard RT-PCR test, given the potential for FN s in the latter. These scans are found to be sensitive in detecting COVID-19 characteristics like Ground-Glass Opacity (GGO) and mixed consolidations. Furthermore, CT scans can reveal abnormalities even before clinical symptoms manifest with classic patterns like GGO. However, manual detection of COVID-affected regions in CT scans is time-consuming and subject to human bias; thereby, making automated or semi-automated CAD essential. Such tools enable accurate detection and delineation of the infection, while facilitating effective monitoring, severity prediction, and prognosis. DL, particularly CNNs, empower machines to learn hidden patterns in data for making them instrumental in image analysis (including CT scans).

Chapter 2 describes a new non-invasive tool for effectively delineating COVID-19 infection in lungs. The Ensemble of selective Focus-based Multi-resolution Convolution network (*EFMC*), employing Leave-One-Patient-Out (*LOPO*) training, exhibits high sensitivity and precision in outlining infected regions along with assessment of infection severity. The selective focus mechanism combines contextual with local information, at multiple resolutions, for accurate segmentation. Ensemble learning integrates heterogeneity of decision through different base classifiers. Unlike many existing methods that rely on single models, this strategy combines the outputs of multiple segmentation algorithms; thereby, harnessing collective intelligence to overcome the limitations of individualistic approaches. The superiority of *EFMC*, even with severe class imbalance, is established in terms of the performance evaluation metrics; through its comparison with existing state-of-the-art learning models over the four publicly-available COVID-19 datasets. The qualitative and quantitative results are suggestive of the relevance of DL in providing assistive intelligence to medical practitioners, when they are overburdened with patients; particularly, as in pandemic situations.

1.7.2 Weighted deformable deep net for efficient segmentation [112], [113], [126], [129]

Due to the increasing number of lung cancer cases, its high mortality rate, along with the availability of huge volumes of CT scans every day, the need for automated handling becomes imperative. This calls for efficient detection and diagnosis, through the design of new techniques for enhanced accuracy. Deep learning, particularly CNNs, shows promise in automating complex image analysis and enhancing diagnostic accuracy. Accurate tumor segmentation is vital for radiation therapy planning, and automated CAD systems can reduce human bias.

Chapter 3 introduces an innovative Weighted Deformable *U*-Net (WDU-Net) for precise and efficient delineation of the tumor region from CT scans. The

Deformable Convolution (*DC*) modules enable effective learning of deformable geometric transformations corresponding to arbitrary geometric shapes of ROIs. This is augmented by the Weight Generation (*WG*) module for suppressing unimportant features while highlighting relevant ones. A new Focal Asymmetric Similarity (*FAS*) loss function helps handle class imbalance. Ablation studies and comparison with state-of-the-art models help establish the effectiveness of WDU-Net, with ensemble learning, over five publicly-available lung cancer datasets. The results demonstrate fast learning with good generalization. The ensemble reduces over- and under- segmentation, yielding high accuracy with robust segmentation outcome.

1.7.3 Deep ensemble with multimodal fusion for classification [127]

Chapter 4 describes the Deep Ensemble Multimodal Fusion (*DEMF*) network, a novel approach for combining CT and PET scans to enhance lung cancer diagnosis. It utilizes PCA to retain essential features while filtering out irrelevant information from both CT and PET images. Subsequently, an Autoencoder reconstructs a meaningful fused representation from the combined features; thereby, resulting in an efficient fusion of anatomical (CT) and functional (PET) information. This multimodal fusion improves diagnosis precision, reduces false positives/negatives, and assists in therapy monitoring. The contributions include development of the *DEMF* network, ensemble of classifiers with majority voting, visualization of activation regions, and ablation studies to demonstrate its effectiveness on a publicly available dataset.

The output demonstrates superiority in lung cancer classification through various experiments, producing efficiency and effective generalization with limited data. Ablation studies and visualization confirm the model's ability to emphasize cancerous regions in the images. This methodology shows promise for extension to various diseases and diverse imaging modalities.

1.7.4 Multimodal deep classification of lung diseases [111]

Using both CT and EIT versions, involving real-time functional insights, offers a more comprehensive understanding of the ROI, aiding in anomaly detection and understanding their impact on tissue functionality. Chapter 5 develops a novel Deep Generative Multimodal Classifier (*DGMC*) network for multi-class detection of diseases from lung CT images.

Data augmentation addresses the issue of data scarcity, using generative adversarial network (GAN). The research contribution includes the efficient generation of high-resolution CT images with infected regions, the transformation of CT images into corresponding EIT versions to generate EIT-processed CT scans, and the design of a multihead classifier for identifying COVID-19 infected, lung tumor affected, and healthy slices. Effective visualization of activation regions is provided, along with comprehensive ablation and comparative studies with state-of-the-art models in literature. Results demonstrate the superiority of our *DGMC* over seven publicly available datasets.

While this study focuses on lung diseases, its principles can be extended to other imaging modalities and diseases, showcasing the versatility and potential for its broader applications. Additionally, the ongoing generation of EIT datasets holds promise for further research and applications.

1.7.5 Conclusion and Future Scope

Chapter 6 provides concluding remarks and outlines the potential future directions for this research.





Chapter 2



Multi-Resolution Deep Ensemble Learning for Segmentation





2.1 Introduction

Multi-resolution approaches, which observe and evaluate a dataset at several resolutions, 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 [176].

This chapter deals with the segmentation of COVID-19 infected lung CT scans. It is well-known that an early detection, diagnosis, isolation and prognosis, play a major role in controlling the spread of COVID-19 infections. 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 [45] that Lung CT images are more sensitive (98%), as compared to RT-PCR (71%), in correctly predicting COVID-19. Researchers assessed the effectiveness of chest CT in the diagnosis and treatment of COVID-19 [95]. The CT characteristics of COVID-19 were presented and compared with the manifestations of other viruses. Abnormalities in CT may occur [40] even before the appearance of clinical symptoms. Multifocal, unilateral, and peripherally based GGO are some such examples of classic patterns.

We developed a novel Ensemble Focus-based Multi-resolution Convolution network (*EFMC*), using Leave-One-Patient-Out (*LOPO*) learning [128]. It enhances the segmentation accuracy of detecting COVID-19 regions from CT scans. The strategy leverages an ensemble technique to combine the outputs of multiple segmentation algorithms. The approach differs from many existing methods which rely on a single algorithm for decision. Our strategy harnesses collective intelligence, by combining the results of multiple algorithms; thereby, addressing the limitations of the individual algorithms. We have evaluated this strategy against individual segmentation algorithms, in terms of well-established metrics like *DSC*, *Precision*, *Sensitivity*, *AUC*, and *Accuracy* [of Section

1.5.1]. The results of the experiments demonstrate that our *EFMC* outperforms the individual algorithms. It is also observed that our model could infer better with limited data for training. This comparative analysis provides empirical evidence of the advantages of the proposed *EFMC*. The findings have implications for more accurate diagnosis and treatment of COVID-19 based on CT imaging.

The contribution of the research [128] is summarized as follows.

- (i) Use of limited data for training the DL model.
- (ii) A multi-resolution selective-focus framework for image segmentation.
- (iii) Ensemble of classifiers through the new strategy LOPO, having good generalization.
- (iv) Comparison on efficiency and accuracy, with state-of-the-art methods.
- (v) Easy and accurate analysis of severity of infection, based on segmentation.

Rest of the chapter is organized as follows. Section 2.2 summarizes the motivation of this study. Section 2.3 discusses the methodology used. Section 2.4 provides the details of implementation and results obtained; along with ablation studies, comparisons, and severity analysis of COVID-19 infection. Section 2.5 concludes the chapter, while highlighting the benefits of the research.

2.2 Motivation

The state-of-the-art literature in segmentation of COVID-19 infection regions in CT scans demonstrates an accuracy in terms of *DSC* of around 0.8123 [160], which is low for all practical purposes. New models and techniques are therefore required to reduce the over-segmentation and/or under-segmentation, and the errors/artifacts. Typically this may be realised with abundant labelled data (which is not feasible in reality) and availability of high-performance computing resources. We have used a modified CNN, called *EFMC*, which uses multi-resolution with focus on the ROI to accurately segment it. This is enhanced by dividing the available small set of data into ten sets, to introduce high diversity among the constituent classifiers. Patches are extracted from each patient sample to make the computation more efficient, both in terms of required memory and speed. In other words, the required computing resources get reduced. This is very important.

It has been demonstrated that an ensemble of diverse classifier provides better accuracy with consistent results [21], [44]. Therefore the *EFMC* overcomes the limitations on data and high-performance computing by reducing the data requirement for training the model in the DL framework; along with a high accuracy of 0.9584 in terms of *DSC*. Given that other medical imaging domains also suffer from insufficiency of annotated data, our *EFMC* framework can be extended there as well.

It is observed that most of the segmentation literature explore only those CT scans having infected regions within; they do not perform any computation over the slices which were non-infected (negative examples). This strategy is prone

to FP when applied on non-infected patient CT scans. We, on the other hand, have also focused on the non-infected samples. The accuracy was computed over all the CT slices (infected and non-infected). This adds another dimension to existing methods, by reducing the chance of obtaining false-positive cases.

2.3 Methodology

The novel *EFMC* model employs *LOPO* learning on CT images of TEN patients; and is trained using TEN base-classifiers. As each classifier takes only NINE patient's samples for training, and starts from scratch, it does result in a completely new classifier with different set of parameters. The remaining ONE patient sample is left for validation in each case (as elaborated in Section 2.4.1). These TEN trained classifiers are ensemble to segment the COVID-infection region (ROI) through majority voting, over four different test datasets collected from various publicly available sources (Table 1.1). The workflow of the *EFMC* is visualized in Fig. 2.1. First we provide the details of data and patch preparation for training and testing.

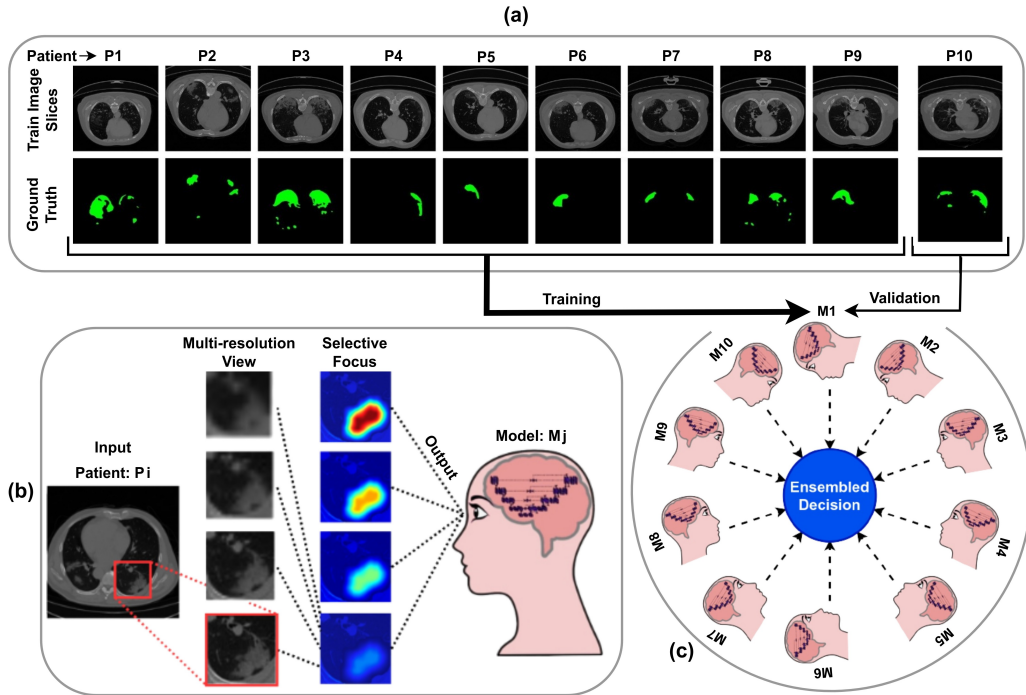


FIGURE 2.1: Schematic representation of *EFMC*, demonstrating the (a) Leave-One-Patient-Out scheme *LOPO*, with (b) selective Focus-based Multi-resolution view. (c) Ensemble of the *EFMC* Models

2.3.1 Data used

Four datasets, with details as provided in Table 1.1, were used in this study. The Kaggle-COVID-19:Part-1 dataset was used for training. Kaggle-COVID-19:Part-2, CT-Seg, Seg-nr.2, MosMed were kept separately for testing. The unseen test datasets were thus used only for evaluating the generalization performance of the *EFMC*.

2.3.2 Data preparation

Instead of initially extracting the lung part from the full CT slice [184], 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.

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 exists no infected region on a slice, it was considered a “non-infected” slice. Random 128×128 pixels patches were extracted from here. In other words, any random patch from here did not carry any infection; and was labeled as “non-infected” (See Fig. 2.2, green bounding box).

When there exists a region of infection in any slice then it was “infected” slice. Twenty random 128×128 pixels bounding boxes were drawn over the ROI to extract the patches (See Fig. 2.2, yellow bounding box). As all of these patches carry infection, they were labeled as “infected”. Next all twelve 128×128 pixels boundary patches (inside the 512×512 pixels axial slice) were considered (See Fig. 2.2, red bounding box). The patches with the ground truth are also illustrated in Fig. 2.2. If the patch consists of any infected region then it is termed an “infected” patch; whereas, the patch which does not consist any infected region is called a “non-infected” patch. (See Fig. 2.2)

The distribution of infected and non-infected slices and/or patches, after patch extraction to create the training set, is displayed in Table 2.1. Representative extracted patches, along with corresponding annotated masks, are presented in Fig. 2.3.(1)-(2) for the infected and non-infected slices, respectively.

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 of size 512×512 pixels were extracted from each test CT volume. Sixteen 128×128 pixels non-overlapping patches were obtained from each slice, as shown in Fig. 2.4.

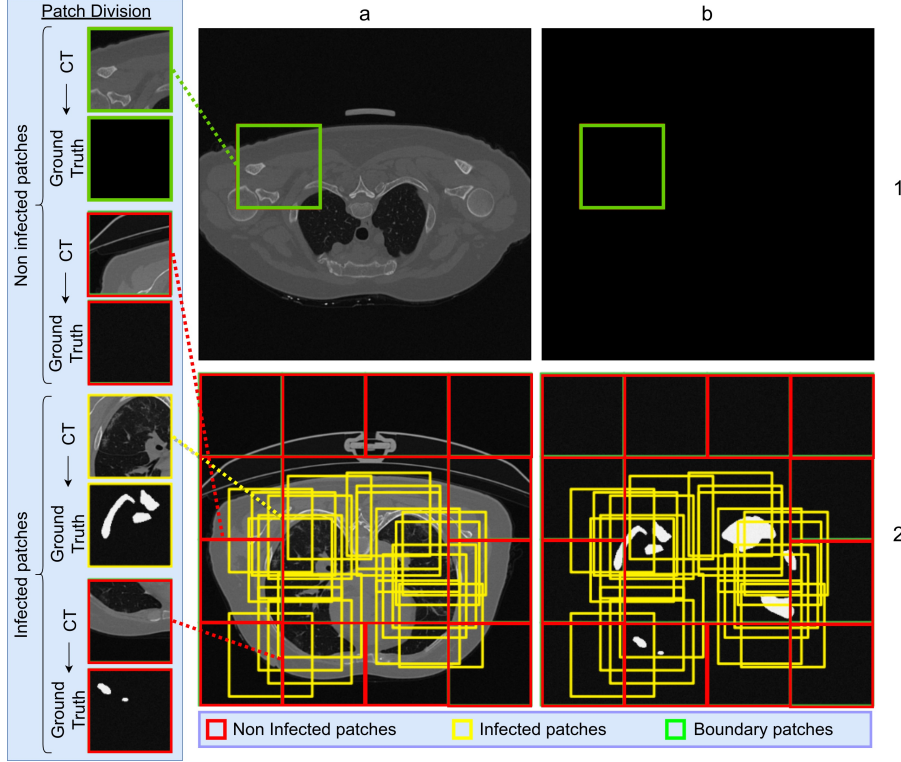


FIGURE 2.2: *Row 1:* Patch extraction from training CT slices, depicting no COVID-infected regions. 1(a) Non-infected patches, and 1(b) corresponding annotated masks. *Row 2:* Patch extraction from training CT slices, depicting COVID-infected regions. 2(a) Overlapping patches, and 2(b) mapping to corresponding annotated masks.

TABLE 2.1: 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

2.3.3 The model

The architecture of our *EFMC* is novel. It consists of an ensemble of selective Focus-modulated Multi-Resolution (*MR*) blocks along the encoding and decoding paths of a *U-Net*. Incorporating dilated convolutions [51] allows feature extraction at multiple resolution. The dilated convolutions in the *MR*-block improve the overall performance, and makes it robust to generalization. Focal loss function [98] is employed for effectively handling class imbalance in the data. The Leave-One-Patient-Out (LOPO) effectively trains a network from

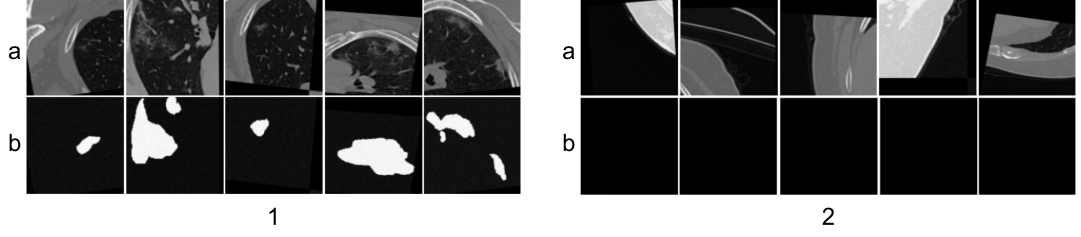


FIGURE 2.3: The (a) extracted patches, and (b) corresponding annotation, for sample patches which are 1: infected, and 2: non-infected

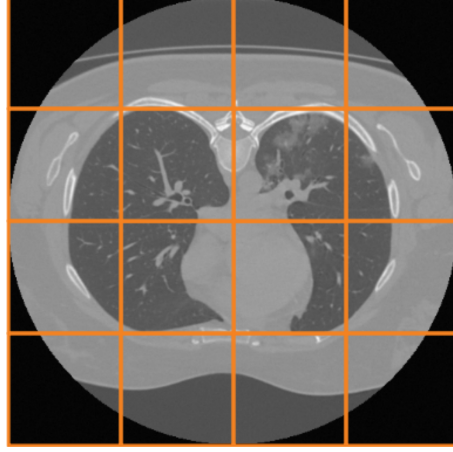


FIGURE 2.4: Patch extraction from test CT slices.

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 selective Focus-modulated *MR*-blocks constitute the *FMC-Net* modules, serving as the base classifier for our *EFMC*. This is illustrated in Fig. 2.5, with elaborated description of the individual modules being provided in Tables 2.2-2.4. The *MR*-block extracts multi-scalar features using a concatenation of four dilated convolutional layers, having dilation rates $D = 1, 2, 3, 5$. A dilated convolution, as in Fig. 2.6, inserts holes into the standard convolution map; thereby, expanding its receptive fields. Thus dilated convolutions enlarge a receptive field, without loss of information, while retaining the kernel size. Finally 1×1 convolutions are employed on the concatenated multi-resolution feature map. The *MR*-block is shown in Fig. 2.5(b) and Table 2.3.

The *FMC-Net* contains nine convolutional blocks, four max-pooling layers, four Upsampling convolutional layers and eight *MR*-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 *MR*-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 *MR*-blocks are added after the ordinary convolutions in the first four encoder and the last four decoder layers. They help obtain multi-resolution contextual information, to reduce the error along the segmentation boundary for improved accuracy. The COVID-19 infection lesions are typically hard to

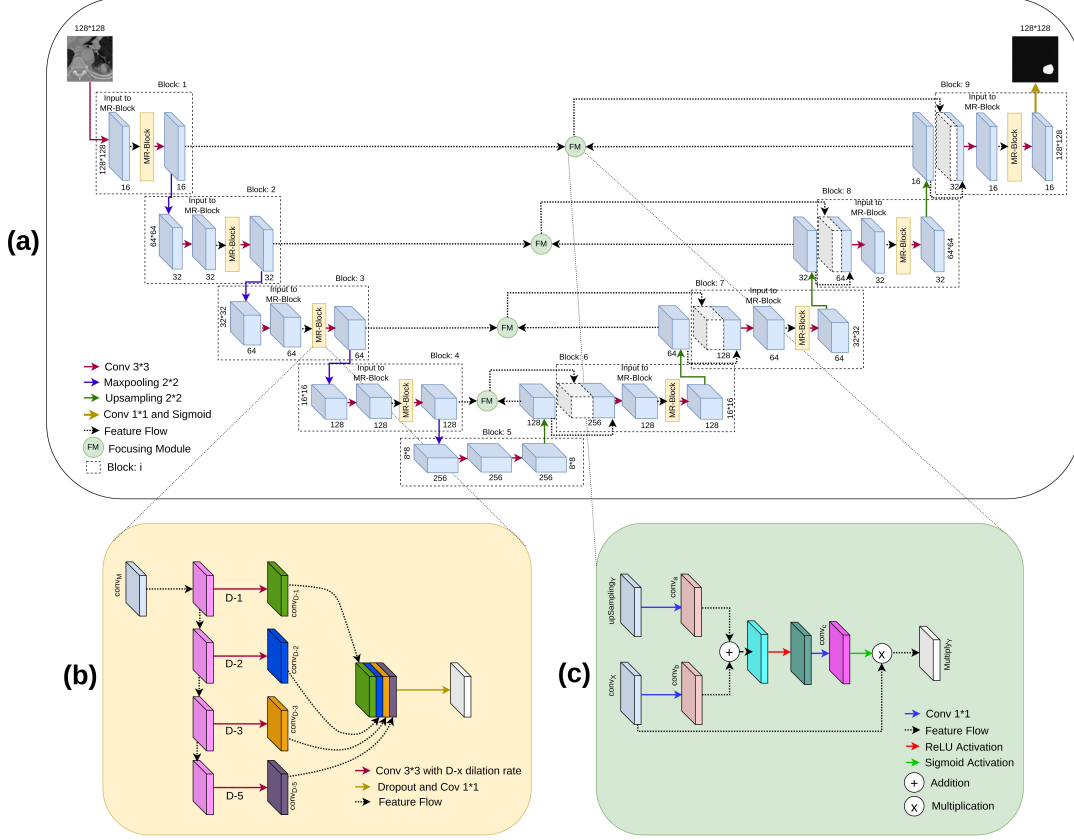


FIGURE 2.5: (a) The *FMC-Net*, with a detailed representation of the (b) *MR-block*, and (c) selective Focusing Module (*FM*).

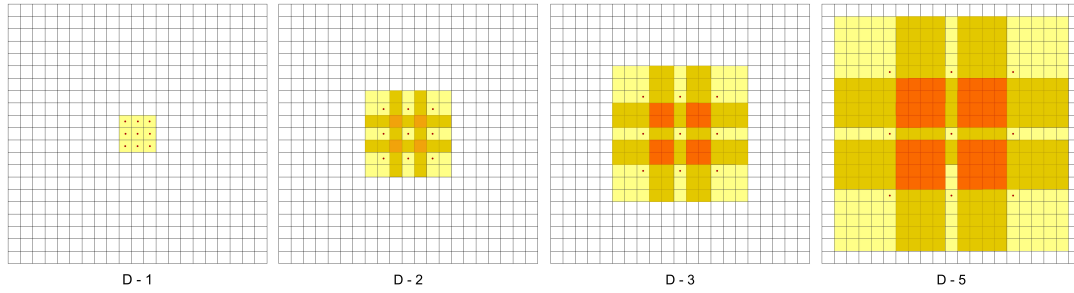


FIGURE 2.6: D -dilated convolutions, with $D = 1, 2, 3, 5$

segment; mainly due to their uneven distribution and varying dimensions. The high-level semantic feature maps in the decoder, concatenated through the selective focusing mechanism, access 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 selective Focusing Module (*FM*) of Fig. 2.5(c) provides 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 focus. Adaptive selection of spatial information is achieved by emphasizing pixels from the ROI, while suppressing the less relevant ones. Four focusing modules are introduced for adaptive feature refinement. A sequential spatial focusing module is embedded into each decoding

TABLE 2.2: The *FMC-Net* module of Fig. 2.5(a)

Block: i	Input: (128 * 128 * 1) CT image
1	conv ₁ (3*3), 16 MR-Block (conv ₁), 16 (Described in Table 2.3) conv ₂ (3*3), 16 maxpooling ₁ (2*2)
2	conv ₃ (3*3), 32 MR-Block (conv ₃), 32 conv ₄ (3*3), 32 maxpooling ₂ (2*2)
3	conv ₅ (3*3), 64 MR-Block (conv ₅), 64 conv ₆ (3*3), 64 maxpooling ₃ (2*2)
4	conv ₇ (3*3), 128 MR-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) FM (conv ₈ , Upsampling ₁), 64 (Described in Table 2.4)
6	concat ₁ (Upsampling ₁ , Multiply ₁) conv ₁₁ (3*3), 128 MR-Block (conv ₁₁), 128 conv ₁₂ (3*3), 128 Upsampling ₂ (2*2) FM (conv ₆ Upsampling ₂), 32
7	concat ₂ (Upsampling ₂ , Multiply ₂) conv ₁₃ (3*3), 64 MR-Block (conv ₁₃), 64 conv ₁₄ (3*3), 64 Upsampling ₃ (2*2) FM (conv ₄ , Upsampling ₃), 16
8	concat ₃ (Upsampling ₃ , Multiply ₃) conv ₁₅ (3*3), 32 MR-Block (conv ₁₅), 32 conv ₁₆ (3*3), 32 Upsampling ₄ (2*2) FM (conv ₂ , Upsampling ₄), 8
9	concat ₄ (Upsampling ₄ , Multiply ₄) conv ₁₇ (3*3), 16 MR-Block (conv ₁₇), 16 conv ₁₈ (3*3), 16 conv ₁₉ (1*1), 1 Sigmoid activation
	Output: (128 * 128 * 1)

TABLE 2.3: The *MR*-block module of Fig. 2.5(b)

MR-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 2.4: Selective Focusing Modules (*FM*) of Fig. 2.5(c)

FM (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

block to avoid overfitting, while accelerating the training of the *EFMC*.

The activation function at the final layer of the *FMC-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 They represent the computation of error over each batch, during backpropagation training, and reflect the adjustment of network weights. Some of the loss functions, as outlined below, are explored in the ablation studies.

These include the Dice Loss (*DL*) [197]. Let the ground truth segmentation mask be $y \in \{0,1\}$, with the corresponding predicted mask being $\hat{y}$ with

estimated probability $p \in [0,1]$.

$$DL(y, \hat{y}) = \left(1 - \frac{2y\hat{y} + 1}{y + \hat{y} + 1}\right), \quad (2.1)$$

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 [98] is defined as a combination of *DL* and the cross-entropy (*CE*) loss for binary classification, thereby incorporating the benefits from both. We have

$$CEDL(y, \hat{y}) = -\{y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\} + DL(y, \hat{y}). \quad (2.2)$$

The *IoU* loss [202], or Jaccard Index loss, is computed from Jaccard Index which is the ratio between the overlap of the positive instances between two sets, and their mutual combined values. *IoU* loss is expressed as

$$IoUL(y, \hat{y}) = \left(1 - \frac{y\hat{y} + 1}{y + \hat{y} + y\hat{y} + 1}\right). \quad (2.3)$$

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 use

$$TL(y, \hat{y}) = \left(1 - \frac{y\hat{y} + 1}{y\hat{y} + \beta(1 - y)\hat{y} + \alpha y(1 - \hat{y}) + 1}\right). \quad (2.4)$$

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

$$FTL(y, \hat{y}) = (1 - TL)^{1/\gamma}. \quad (2.5)$$

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

Let p_t be the probability depending on the value of y . The Focal Loss (*FL*) [98] overcomes class imbalance in datasets, when positive patches are relatively scarce. Now the cross entropy (*CE*) loss for binary classification is defined as

$$CE(p, y) = \begin{cases} -\log p, & \text{if } y = 1, \\ -\log(1 - p), & \text{otherwise.} \end{cases} \quad (2.6)$$

For convenience, let

$$p_t = \begin{cases} p, & \text{if } y = 1, \\ 1 - p, & \text{otherwise,} \end{cases} \quad (2.7)$$

such that $CE(p, y) = CE(p_t) = -\log(p_t)$. The α -balanced focal loss is

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t), \quad (2.8)$$

with the choice of weighting factor $\alpha = 0.8$ and focusing parameter $\gamma = 2$ being

made after several experiments on train data. It was found to be the best choice from the loss functions enumerated here.

Diversity is introduced in this ensemble of the *FMC-Nets*, by varying the training datasets through *LOPO*; thereby, changing the initialization of the networks, and modulating the choice of parameters of the *EFMC* system. The performance of the models is evaluated in terms of the metrics *DSC*, *Precision*, *Sensitivity*, and *AUC*.

2.4 Experimental Results

The objective of the model is to segment a COVID-19 infected CT lung image into its ROIs, [*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 *FM* is carefully incorporated to focus on the ROI, whereas the multi-scalar dilation provides the necessary local and neighbourhood information representation for delineating the ROI. The *FL* [eqn. (2.8)] is used as the loss function to compensate for class imbalance. Finally we implement ensemble by *LOPO* for training with limited annotated data. The details about the training dataset are provided in Section 2.3.2. Validation is successfully performed on unseen, independently-compiled data from multiple publicly-available data sources. The proposed *EFMC* of Fig. 2.1, using *LOPO* ensemble learning on *FMC-Net* of Fig. 2.5, demonstrates good generalization capability. It is efficient, accurate, consistent and robust on the unseen data. The architecture of *FMC-Net* is summarized in Tables 2.2-2.4.

2.4.1 Implementation details

The training of each model was closely monitored from the learning curves, over each fold. Representative learning curves, depicting loss [*FL* of eqn. (2.8)] and accuracy [*DSC*] of eqn. (1.7)], are illustrated in Fig. 2.7.

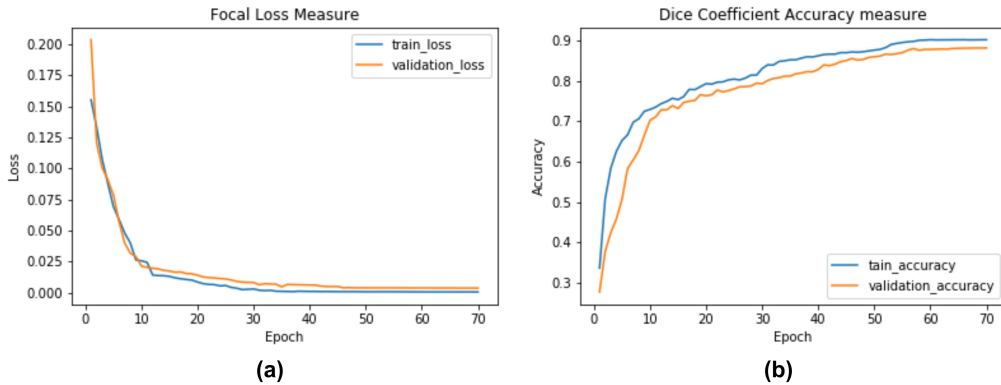


FIGURE 2.7: Sample learning curves, for Model M3, over training and validation sets; (a) loss: *FL*, and (b) accuracy: *DSC*

TABLE 2.5: 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 $P = \{ P_i: i \in \mathbb{N} \wedge i \in [1, 10] \}$										
Mean validation DSC = 0.8773, with Standard deviation = 0.0158										

The *EFMC-Net* is *FMC-Net* with the new ensemble strategy. Let there be TEN sample $P1, P2, \dots, P10$, as defined in Table 2.1. The ten models considered are $M1, M2, \dots, M10$; with the validation datasets defined as $P10, P9, \dots, P1$, and described in Table 2.5. In each case, the rest of the corresponding patient's dataset is used for training by the *FMC-Net* model of Fig. 2.5. 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 initialization. For each of the models, the batch size was kept uniform at 16, using the Adam optimizer 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 (rotation range $\pm 10^\circ$; horizontal shift range ± 0.2 ; vertical shift range ± 0.2 ; zoom range ± 0.2) was employed during training. A ten-fold training needed 10 hours approximately for each fold, with total number of trainable parameters (3.346×10^6). The model requires high-resolution CT images at input, with appropriate hardware and software system support.

2.4.2 Output

The training of each model was closely monitored from the learning curves, over each fold. Representative learning curves, depicting loss and accuracy, are illustrated in Fig. 2.7. The validation set corresponding to each model, with the resultant DSC , are presented in Table 2.5. 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 in the framework of 10-fold cross validation. This resulted in ten distinct classifiers to makeup the ensemble.

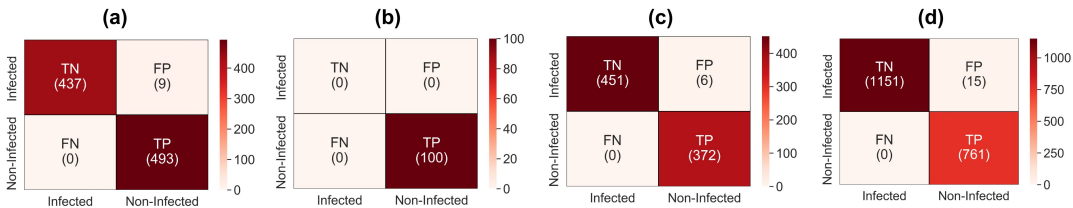


FIGURE 2.8: Confusion Matrix for classification: (a) Kaggle-COVID-19: Part-2 (b) CT-Seg, (c) Seg-nr.2, and (d) MosMed

TABLE 2.6: Predicted binary classification of Infected & Non-Infected slices, in the test datasets, by *EFMC*

Sr.	Dataset	Original		Predicted		Classification Accuracy (%) [eqn. (1.13)]
		Infected slices	Non-Infected slices	Infected slices	Non-Infected slices	
1	Kaggle-COVID-19: Part-2	493	446	502	437	99.04
2	CT-Seg	100	0	100	0	100
3	Seg-nr.2	372	457	378	451	99.27
4	MosMed	761	1166	776	1151	99.22

Table 2.6 and Fig. 2.8 presents the binary classification accuracy between COVID and Non-COVID slices by *EFMC*. It is observed that the accuracy of any model in correctly identifying the infected (+) and non-infected (−) slices is very high over all four datasets considered in this study. The classification accuracy is computed by eqn. (1.13).

Next the explainability of the *FMC-Net* architecture was analysed. As evident from Fig. 2.9, each level of the encoder and decoder arms of the network architecture demonstrates the 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 focus and Multi-resolution mechanisms of the *FMC-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 focus over the lesion boundary (blue box) and lesion (green box) in Fig. 2.9 become more prominent. Eventually the final feature which prominently emerges is the lesion.

2.4.3 Qualitative analysis

Sample qualitative results in Fig. 2.10, corresponding to the models compared in Fig. 2.12, help establish the robustness and effectiveness of *EFMC* under ensemble by *LOPO* on the backbone *FMC*-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*, *S4* from CT-Seg; (iii) *S5*, *S6* from Seg-nr.2; and (iv) *S7*, *S8* from MosMed. It is observed that the *FMC*-Net, of column (e) in the figure, performed the best. Results were corroborated with the confusion matrix of Fig. 2.11, providing an indication of the distribution of misclassified pixels for a sample segmentation mask *S4*. The confusion was seen to be the least in case of the *FMC*-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 *S4* as an example, it is clearly evident that state-of-the-art models like *U*-Net,

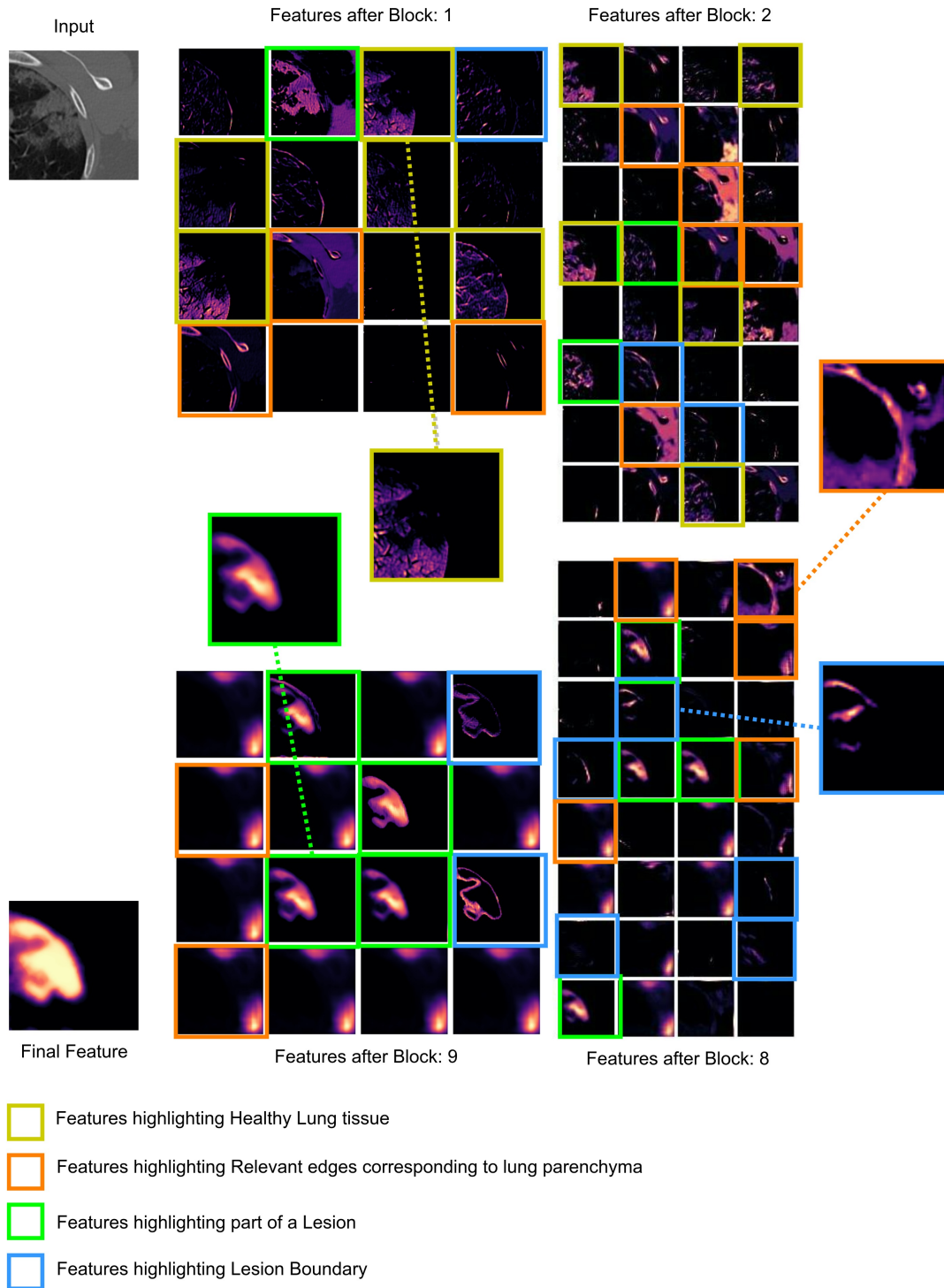


FIGURE 2.9: Visualization of features extracted after each Block of the *FMC-Net* from Fig 2.5

Attention *U-Net*, Inception-Net, Xception-Net, DenseNet demonstrate under-segmentation, while model R2UNet is indicative of over-segmentation. On the other hand, the *FMC-Net* [column (e), Fig. 2.12] exhibited significantly lower under- and/or over-segmentation for the same sample *S4*. The corresponding confusion matrix in Fig. 2.11 corroborates these findings. This model takes a

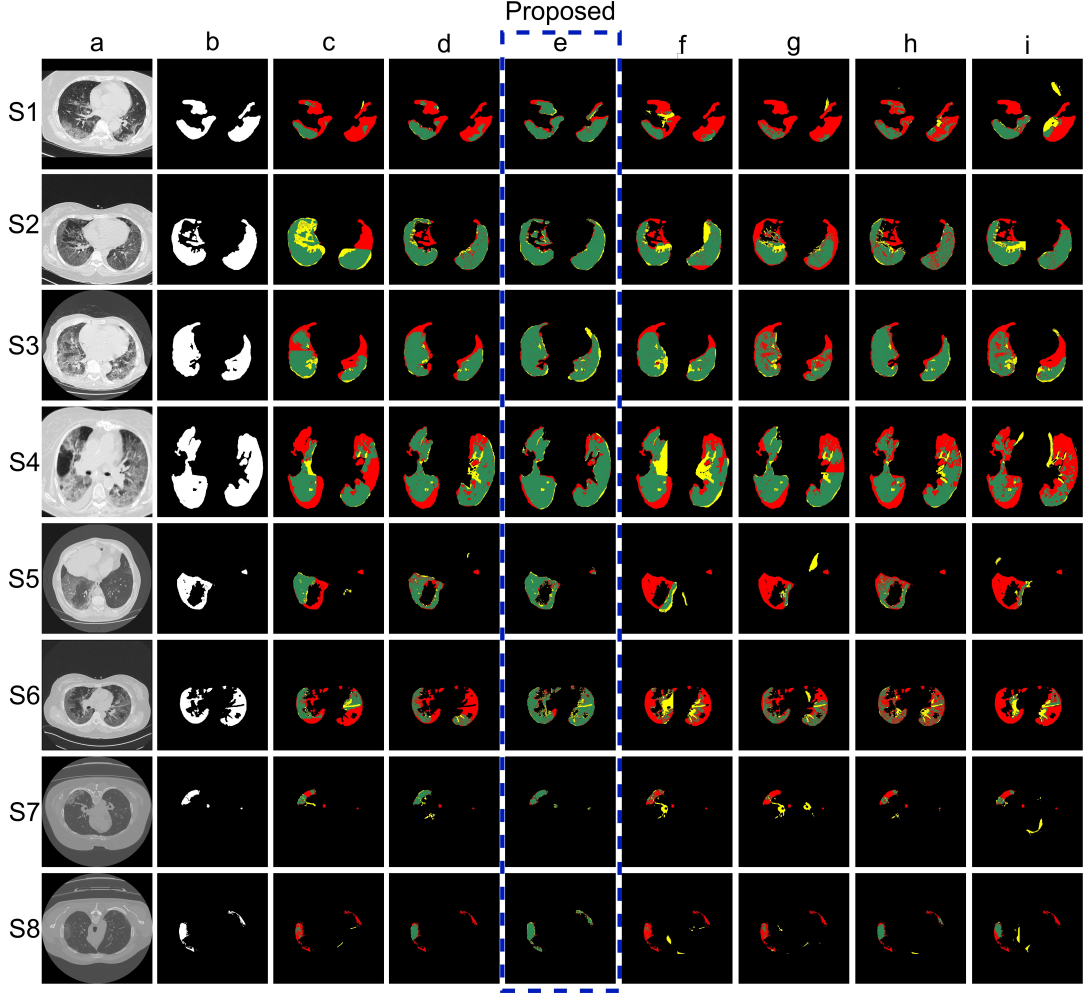


FIGURE 2.10: Sample segmentation output by base classifiers, in the uniform framework of ensemble with *LOPO*. (a) Original CT scan, with (b) Annotated masks. Segmentation obtained by (c) *U-Net*, (d) Attention *U-Net*, (e) ***FMC-Net***, (f) *R2UNet*, (g) *Inception-Net*, (h) *Xception-Net*, and (i) *DenseNet*, where regions Green: True Positive, red: False Negative, Yellow: False Positive

few seconds to delineate an entire scan, whereas human experts spend longer time and effort while being prone to error with the mechanical activity involved.

2.4.4 Comparative study

Using the *FMC-Net* as the base classifier, with ten classifiers *M1* to *M10* ensemble 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 1.1 and Section 2.3.2) in terms of the performance metrics defined in [eqns. (1.7), (1.9) - (1.11)]

The comparative analysis of output generated by the *EFMC*, on different test sets, is presented in Fig. 2.12. In all cases the test slices were divided into non-overlapping patches (as elaborated in Section 2.3.2). 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

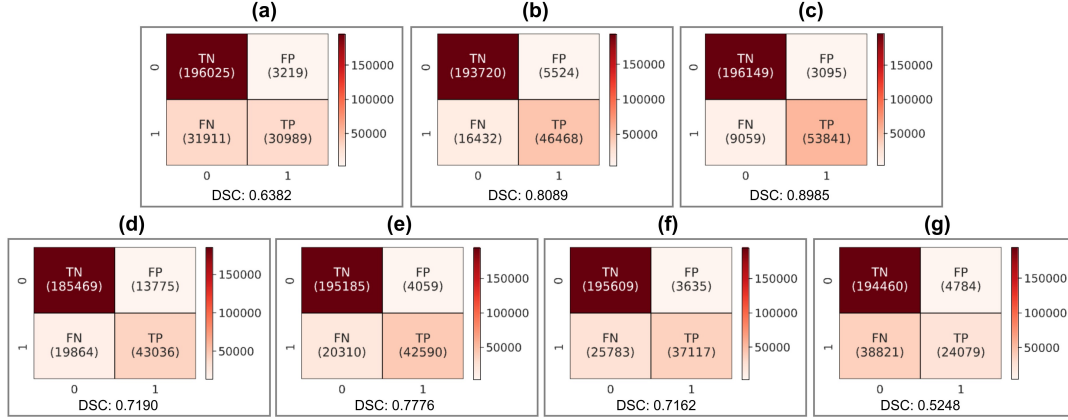


FIGURE 2.11: Confusion matrix for a sample $S4$ from Fig. 2.10 segmentation mask, by the base classifiers, in the uniform framework of ensemble with *LOPO*. Models (a) *U-Net*, (b) *Attention U-Net*, (c) ***FMC-Net***, (d) *R2UNet*, (e) *Inception-Net*, (f) *Xception-Net*, and (g) *DenseNet*. The *DSC* is marked for all models used

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 a better contrast between the ROI and background. Moreover, the original Hounsfield range is preserved, meaning the tissue-specific contrast is maintained. As a result, the image retains anatomical fidelity with reduced noise or distortion.

TABLE 2.7: Comparative computational analysis of base model parameters, along with *DSC* obtained

Model	<i>U-Net</i>	<i>FMU-Net</i>	<i>MRU-Net</i>	<i>FMC-Net</i>	<i>R2UNet</i>	<i>Inception-Net</i>	<i>Xception-Net</i>	<i>DenseNet</i>
#parameters	1.96M	1.99M	3.31M	3.34M	6.00M	11.99M	2.05M	4.26M
<i>DSC on</i>	Kaggle	0.6082	0.7492	0.7477	0.8840	0.6772	0.5924	0.7972
	CT-Seg	0.5124	0.7054	0.6924	0.7955	0.5762	0.6327	0.7172
	Seg-nr.2	0.5921	0.7180	0.7173	0.8431	0.6304	0.6275	0.8331
	MosMed	0.6896	0.8300	0.8224	0.9584	0.7029	0.7145	0.8908

The model complexity, in terms of the number of parameters, is enumerated in Table 2.7. Although our *EFMC* involves slightly more parameters than *U-Net*, *Attention U-Net*, *MRU-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 *EFMC* is better than the state-of-the-art methods, as evident from the *Box-and-Whisker plot* of Fig. 2.12. A representative study for *DSC* on all the four test datasets is also provided in Table 2.7.

In order to explore the effectiveness of *FMC-Net* as the base classifier in *EFMC*, we also compared four state-of-the-art models like *R2UNet*, *Inception-Net*, *Xception-Net*, and *DenseNet* (which have been extensively employed in COVID-19 segmentation literature). Ensemble 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. (1.7), (1.9) - (1.11)]. is provided in Fig. 2.12. It is

found that all the metrics resulted in higher values for the ensemble *FMC-Net*, indicating a better generalization in segmentation over each test dataset.



FIGURE 2.12: Comparative performance evaluation of base classifiers, in the uniform framework of ensemble with *LOPO*, over the test datasets

Summarized below is the performance of our *EFMC* 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 [21], [32], [100], [105], [152], [178] using some of the same test datasets, led to interesting conclusions with respect to our *EFMC*. The Seg-nr.2 test set yielded DSC of 0.673 [105], 0.620 [152]. The model [152] 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

TABLE 2.8: Comparative analysis of performance with recent literature on deep learning for segmentation of COVID-19 regions from lung CT

Dataset	Methodology	Performance Metric	
		<i>DSC</i>	<i>Sensitivity</i>
Kaggle	RMT_VAT [178]	0.7103	0.6860
	Prior wavelet knowledge [160]	0.8123	
	2.5D DPE-Net [100]	0.7589	
	EFMC-Net	0.8840	0.8082
CT-Seg	Domain adaptation [79]	0.7633	
	EFMC-Net	0.7955	0.7431
Seg-nr.2	Data-efficient learning [105]	0.6730	
	GAN with Unet [152]	0.6200	0.6720
	EFMC-Net	0.8431	0.7905
MosMed	GAN with Unet [152]	0.5840	0.7680
	DMDF-Net [125]	0.7291	
	Domain adaptation [79]	0.6066	
	2.5D DPE-Net [100]	0.7589	
	EFMC-Net	0.9584	0.8514

dataset the authors reported [178] *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. [32], using the R2UNet as backbone, obtained a *DSC* of 0.851. On the other hand, using the basic *U*-Net as backbone [21] attained *DSC* 0.80. The DMDF-Net [125] presented findings on Mosmed data, reporting a *DSC* of 0.7291. A domain adaptation technique [79] achieved a *DSC* of 0.7633 on CT-Seg and 0.6066 on the Mosmed dataset. Incorporation of prior wavelet knowledge [160] led to a *DSC* of 0.8123 on the Kaggle dataset. A 2.5D DPE-Net [100] was developed, utilizing a 2.5D dual-path encoder network model with a multiscale dynamic weighted deep supervision mechanism. It achieved a *DSC* of 0.7589 on the combined Kaggle and MosMed datasets. These outcomes reiterate the effectiveness of our *EFMC*, the ensemble classifier using *FMC*-Net, in terms of these compared performance metrics over all test datasets. This is summarized in Table 2.8.

2.4.5 Ablations

The first set of experiments were performed by evaluating the role of each of the components in *FMC*-Net, *viz.* *MR*-block and *FM*, when used in *EFMC*. The traditional *U*-Net is thus the baseline and with *FMU*-Net incorporating only the *FM*. The *U*-Net with only the *MR*-block is termed the *MRU*-Net. The *FMC*-Net is the *U*-Net with both *FM* and *MR*-block. All four models were trained using the same ensemble *LOPO* framework. Comparative results on the four test datasets of Table 1.1, evaluated in terms of the performance metrics of eqns. (1.7), (1.9) - (1.11) are presented in Fig. 2.13. It is observed that the proposed *FMC*-Net performs the best, over all the metrics in all test datasets, as compared to the rest (lacking one or more of its component modules). This

helps justify the effectiveness of the *EFMC*, which ensembles with LOPO a set of *FMC*-Net models.

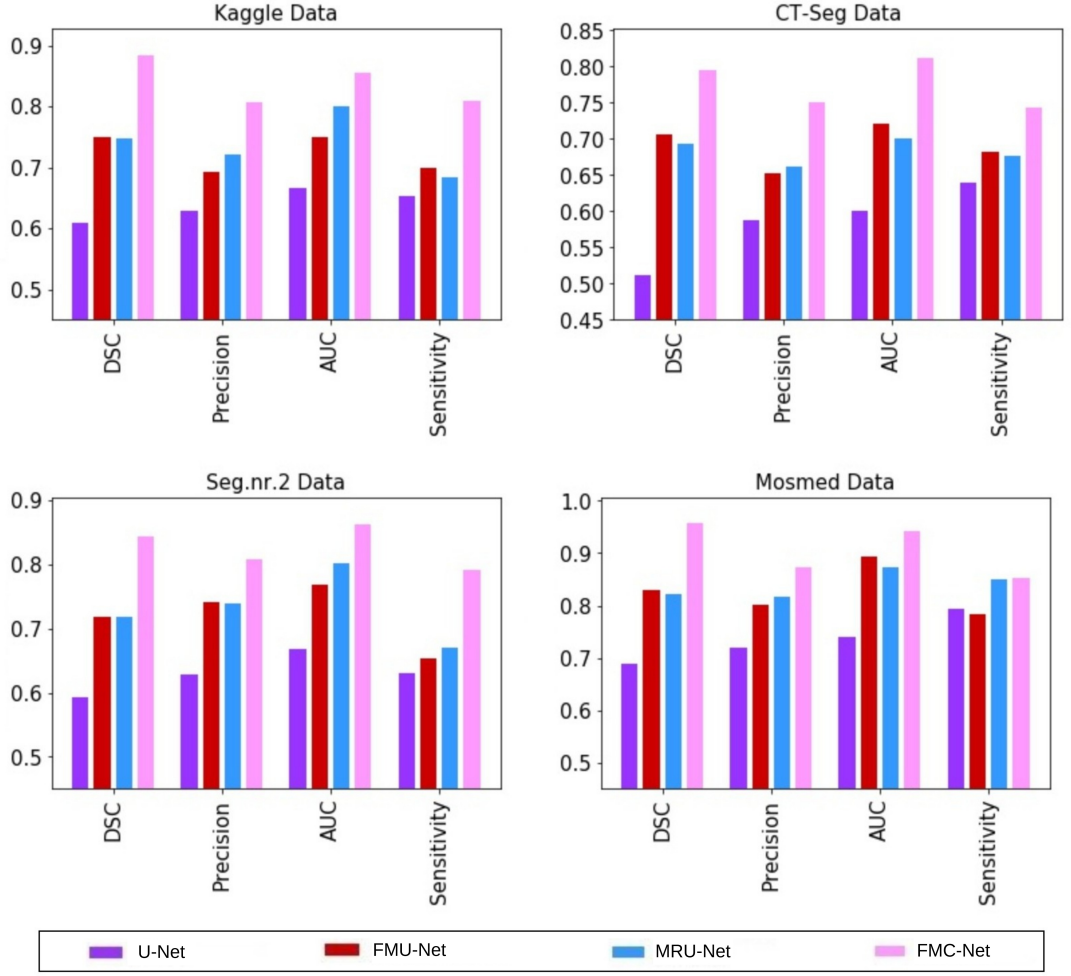


FIGURE 2.13: Role of different components of *EFMC* over the test datasets

TABLE 2.9: Effect of various loss functions on *FMC*-Net, measured in terms of *DSC*, over the test datasets

Loss function	Kaggle-COVID-19: Part-2	CT-Seg	Seg-nr.2	MosMed
<i>DL</i>	0.7521 ± 0.174	0.7238 ± 0.201	0.7218 ± 0.136	0.8465 ± 0.251
<i>CEDL</i>	0.7272 ± 0.142	0.7377 ± 0.193	0.7996 ± 0.147	0.8129 ± 0.180
<i>IoUL</i>	0.7937 ± 0.137	0.7190 ± 0.165	0.7660 ± 0.125	0.8706 ± 0.197
<i>FL</i> (eqn.2.8)	0.8840 ± 0.103	0.7955 ± 0.107	0.8431 ± 0.101	0.9584 ± 0.014
<i>TL</i>	0.7103 ± 0.176	0.7201 ± 0.130	0.7542 ± 0.203	0.8360 ± 0.071
<i>FTL</i>	0.8238 ± 0.192	0.7725 ± 0.144	0.8285 ± 0.194	0.8973 ± 0.214

Next we explored the effect of the different loss functions on the performance of the base model *FMC*-Net, without any ensemble. Results are provided in Table 2.9, over all the four test datasets. In the present study we have considered the Dilation rate as $D = 1, 2, 3$, and 5 . However in order to see the other combinations, we also considered different combination of Dilation

rates. Here, we present the results on two such combinations, viz. $D = 1, 2, 3$, and 4, and $D = 1, 2, 4$ and 8, as given in Table 2.10. It is observed that the combination $D = 1, 2, 3$, and 5, performed the best, in terms of average DSC . Here the ROI (GGO and consolidation) covers a minuscule portion of a CT slice. The 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 ensemble in $EFMC$ enhances the corresponding performance (as in Fig. 2.13) in terms of DSC . This is also depicted in Table 2.7 with respect to U -Net, FMU -Net, MRU -Net and compared with our FMC -Net.

TABLE 2.10: Effect of varying Dilation rate D on FMC -Net with FL , measured in terms of DSC over the test datasets

D in MR -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	0.8840 ± 0.103	0.7955 ± 0.107	0.8431 ± 0.101	0.9584 ± 0.014

2.4.6 Severity assessment

Table 2.11 quantifies the infected region in the sample test images, comprising the eight slices (two each from the four sets of test data) of Fig 2.10, along similar lines.

Fig. 2.14 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 the JET Colormap.

TABLE 2.11: Prediction of affected region by $EFMC$, considering two random samples slices from each test dataset

Dataset	Sample no. in Fig. 2.10	Ground truth (%)	Prediction (%)
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

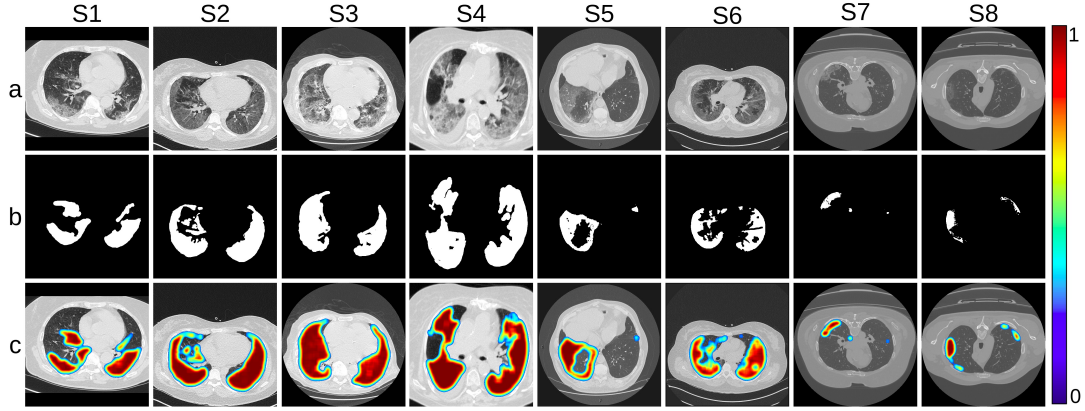


FIGURE 2.14: Visualization of infection, in the eight sample slices from Fig. 2.10, as predicted by *EFMC*. (a) Input CT scan, (b) corresponding annotation, and (c) JET ColorMap on the prediction

2.5 Conclusions and Discussion

A novel ensembling method by *LOPO* was developed for a collective, efficient delineation of COVID-19 affected region in the lung; a gradation of the severity of the disease was also generated with very limited training data. We used only TEN patient samples, with patch based learning, to achieve very high accuracy. Multi-resolution selective focus, with deep supervision, enabled enhanced accuracy, in terms of improved sensitivity and precision in segmentation of ROI, for our *EFMC* model. 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. It yielded an average *DSC* of 0.8773 on training data, with a standard deviation of 0.0158 under 10-fold cross validation. Moreover, the resulting score was in the range 0.7955 to 0.9584 over the four test datasets.

Our research, using deep convolution 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, based on the grading obtained. The qualitative output demonstrated that the *EFMC* generated 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 was effectively addressed by the loss function discussed. Our patch-based method performed exceptionally well, in terms of accuracy and loss over training and validation, for an effective management of overfitting in DL. Use of CT images, obtained from several other sources, established the robustness of our methodology in handling imaging differences at the source. Table 2.8 shows how good

our model performed in comparison to state-of-the-art methods. Ablation studies also helped reinforce its effectiveness. It can be incorporated into a CAD system to help radiologists efficiently analyse CT images.

Radiologists' assessments of CT scans are often subjective, and may be somewhat influenced by their experience. DL models provide benefits in resolving these issues. Therefore DL models might be beneficial for emerging nations like India, or other underdeveloped nations where there is an uneven distribution of medical resources between urban and rural areas. These models can fill the gap in CT diagnosis between national hospitals and primary healthcare centres. They can also help in providing a prompt and accurate diagnosis of COVID-19 to radiologists, particularly assisting the junior radiologists.

Our approach holds promise for extension to other medical imaging domains, as most of these also suffer from insufficiency of annotated data. The future research will concentrate on testing and benchmarking of the model with more data, along with porting to mobile devices. As and when any other versions of the SARS-COV virus become prevalent, then such incremental data has to be collected and the model should be trained for improved performance on such new variations.

In the following chapter we introduce the new deformable deep network for an efficient and robust segmentation of lung tumor regions in CT scans.





Chapter 3

Weighted Deformable Deep Net for
Efficient Segmentation



3.1 Introduction

Classification of tumor stage is a cornerstone of providing uniform consistent care for patients with cancer worldwide [38]. NSCLCs can be centrally located masses, invading the mediastinal structures, or peripherally situated lesions that invade the chest wall [141]. Tumors can have margins which are smooth, lobulated or irregular and spiculated. They can be uniformly solid or can have central necrosis and cavitation. Sometimes the tumor resembles an infective pathology and is seen as an area of consolidation, a ground-glass opacity, or a combination of both. CT currently being the primary means for screening and monitoring lung cancer in clinics, an efficient segmentation of the tumor region becomes extremely important from CT scans.

Improving the specificity and sensitivity of lung cancer screening is imperative because of the high clinical and financial costs of missed diagnosis, late diagnosis, and unnecessary biopsy procedures resulting from false negatives and false positives. Deep learning approaches offer the exciting potential to automate more complex image analysis, detect subtle holistic image findings, and unify methodologies for image evaluation [187], [201]. CNNs have been used [9] in an end-to-end approach, based on a patient's current and prior CT volumes, to predict the risk of lung cancer. While the model's localization directs follow-up for specific lesion(s) of greatest concern, the consideration of a larger context in the form of ROI around candidate regions in 3D improves its prediction ability. State-of-the-art performance is claimed on the National Lung Cancer Screening Trial (NLST) cases, with performance on-par or superior to that of experienced radiologists. Investigations with CT images of NSCLC helped establish the effectiveness of semi-automated segmentation vis-a-vis the manual one [130]. Use of extracted quantitative imaging biomarkers [103], and a single-click ensemble segmentation [58] are available in literature.

However as lung tumor sizes typically fluctuate between 1.05 to 28.14cm^3 , it often adversely affects the performance of deep models like the *U-Net* at a single spatial scale. Assigning weight(s) to feature(s), which focus on the ROI, is expected to improve the representational capacity of a model through an adaptive suppression of the less important feature maps. Typical CNNs rely on data augmentation to learn geometric transformations, because they are fundamentally invariant to big and unknown transformations. Hence it is

particularly desirable to be able to dynamically change the receptive field in fully convolutional networks (FCNs).

A deformable convolution (*DC*) [34], unlike the basic convolution, is not constrained by predefined geometrical structures of the kernels. It can model a learnable and dynamic receptive field. It allows the sampling grid to be deformed in free form through the addition of 2D offsets to the grid sampling points of an ordinary convolution. Additional convolution layers are employed to learn the offsets from the preceding feature maps. As a result, the deformation is local, dense, and adaptively conditioned on the input attributes. This enables effective simulation of large and unknown transformations. They have the benefit of an adaptive receptive field, which changes depending on the size of the object and is learned from the data. Although *DC*s are known to perform well in object detection, their role in semantic segmentation of medical images is yet to be investigated.

Although there exists literature [11], [93], [193] reporting the use of *DC* in medical image segmentation (not specifically focusing on lung tumor CT scans), there is a general absence of any form of optimization for exploring the full potential of *DC* in improving outcome. Specifically, the effectiveness of *DC* has not been comprehensively analyzed, either quantitatively or qualitatively. In contrast, our research focuses on optimizing *DC* specifically for the task of segmentation in lung CT. We have conducted extensive experiments that not only optimize the parameters of the *DC*, but also include exhaustive quantitative and qualitative ablation studies to assess their impact. These analyses provide deeper insights into how *DC* contributes towards improving the performance of segmentation models. Our results indicate that the integration of the optimized *DC* component significantly enhances segmentation accuracy, reduces error rates, and effectively addresses issues related to spatial bias in the segmentation process. Such thorough examination and optimization underscore the potential of *DC* as a powerful tool in advancing medical image segmentation; particularly, by offering improvements that have not been fully realized in existing literature.

We have extended the *DC* framework to design a novel deep *WDU*-Net for the efficient and accurate delineation of lung tumors from CT slices. The model is implemented in terms of 2D slices, mainly because of the limited availability of annotated data and computational power. The effectiveness of the model is validated through ablation studies to evaluate generalization in performance, along with comparison with state-of-the-art methods, over multiple diverse datasets in terms of several performance metrics.

The contribution of the research [112], [113], [126], [129] is outlined below.

- (i) The proposed *WDU*-Net consists of Deformable Convolution modules and Weight Generation modules, for effective segmentation of the ROI.
 - The *DC* module incorporates deformable convolutions to model irregular shape boundaries.
 - The *WG* module estimates the weights of the pixels in the channels.
- (ii) A Focal Asymmetric Similarity (FAS) loss is introduced for improved performance.

- (iii) The new ensemble inferencing combines the results of ten classifiers for accurate and robust demarcation of ROI.
- (iv) The WDU-Net performs much better than classical as well as state-of-the-art methods, as evaluated over five test datasets.

The rest of the chapter is organized as follows. The Weighted Deformable U -Net (WDU-Net) is introduced in Section 3.2, with network architecture details, loss functions, and performance evaluation indices. The experimental setup, along with qualitative and quantitative results on five public datasets, are presented in Section 3.3. Comparison with state-of-the-art literature is also provided. Section 3.4 concludes the chapter.



3.2 The Weighted Deformable U -Net

Here we introduce the architectural details of the novel Weighted Deformable U -Net (WDU-Net), which encompasses the Deformable Convolution (DC) modules with the new Weight Generation (WG) modules. An outline of the loss functions and performance evaluation metrics is also provided.

3.2.1 Deformable convolution

The fixed geometric structure of conventional convolution modules often constrains their capability in modeling large-scale unknown geometric transformations. The input feature map gets sampled at pre-determined locations, with all the activation units of a layer having the same receptive field size. This becomes problematic at the higher levels of the CNN when encoding the semantics across spatial locations. Adaptive determination of the scales or receptive field sizes often becomes desirable for visual recognition involving fine localisation. This is because different locations may correspond to objects possessing varying scales or deformation. A deformable convolution (DC) augments the ordinary grid sampling points in the standard convolution by adding 2D offsets. This enables the sampling grid to be deformed freely, with the offset learning module including a few extra parameters and computations. Here the receptive field of the operator can be adapted to the target ROI by deforming (or warping) the kernel throughout. We depict the framework geometrically in Fig. 3.1.

Let us mathematically formulate the operation. Given an input feature map X of size $H \times W \times \mathcal{C}$, where H represents the height, W corresponds to the width, and $\mathcal{C}$ refers to the number of input channels, the basic convolution operation (with a kernel size of $K \times K$) produces an output feature map $\mathcal{F}$ of size $H \times W \times \mathcal{M}$, with $\mathcal{M}$ indicating the number of output channels. Let $\omega = \{\omega^i \mid 1 \leq i \leq \mathcal{M} ; \omega^i \in (K \times K)\}$ be the set of learnable kernel weights of size $K \times K \times \mathcal{M}$. The output feature $\mathcal{F}$, corresponding to pixel location (i, j) and output channel m , is computed as

$$\mathcal{F}_{i,j}^m = \sum_{c=1}^{\mathcal{C}} \sum_{p=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \sum_{q=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \omega_{p,q}^m X_{\{i+p\},\{j+q\}}^c, \quad (3.1)$$

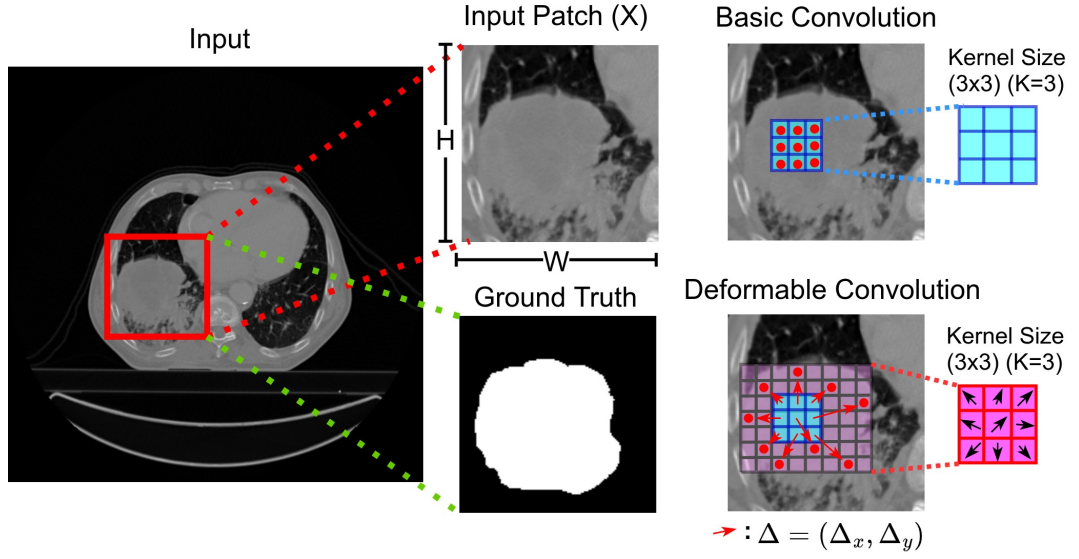


FIGURE 3.1: Illustration of sampling locations, as modeled in basic and deformable convolution.

where $c = \{1, \dots, \mathcal{C}\}$ refers to the input channels, $m = \{1, \dots, \mathcal{M}\}$ corresponds to the output channels, and K is assumed to be odd. Now let $\Delta = \{(\Delta_x^n, \Delta_y^n) \mid 1 \leq n \leq K^2\}$ be the set of paired learnable offsets of size $H \times W$ for deformable convolution (DC). The schematic representation of the DC module is presented in Fig. 3.3(c). The output feature $\mathcal{F}$, corresponding to pixel location (i, j) and output channel m , becomes

$$\mathcal{F}_{i,j}^m = \sum_{c=1}^{\mathcal{C}} \sum_{p=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \sum_{q=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \omega_{p,q}^m X_{\{(i+p)+\Delta_x^{(i+p)}\}, \{(j+q)+\Delta_y^{(j+q)}\}}^c. \quad (3.2)$$

As observed from Fig. 3.1, here Δ_x and Δ_y are the learned offsets. They represent the amount by which the position of the pixel location can be shifted along the abscissa and ordinate, respectively. Thus the DC allows the receptive field of the network to adapt to the features of the input. This can be particularly useful for tasks such as object detection and segmentation, where the objects of interest may have arbitrary shapes and sizes. The DC blocks are incorporated in the WDU-Net architecture of Fig. 3.3 (green blocks) for better modeling of different geometric shapes of the tumor region(s). This can be validated from the results depicted in Fig. 3.8 and 3.9.

3.2.2 Weight generation

These modules are introduced for a weighted combination of feature maps along the encoder pathway with the corresponding ones in the decoder pathway. Let $\mathbf{E} = \{E_i \mid 1 \leq i \leq L\}$, where L is number of levels in the network and E_i corresponds to the feature map from the encoder pathway at level i . Analogously, for the decoder pathway we have $\mathbf{D} = \{D_i \mid 1 \leq i \leq L\}$, where D_i is the feature

map from the decoder pathway at the same level i . We have E_i and D_i of size $H \times W \times C$.

The novel weight generation module employs the convolution operator of eqn. (3.1) on both E_i and D_i to reduce their channel dimensions. The WG module structure is elaborated in Fig. 3.3(b), in the global framework of the WDU-Net architecture of Fig. 3.3(a). Having $conv_i^a = C_{1 \times 1}(D_i)$ and $conv_i^b = C_{1 \times 1}(E_i)$, where $C_{1 \times 1}$ denotes the 1×1 convolution, the weight matrix is computed as

$$S_i = C_{1 \times 1 \times 1}\{ReLU(conv_i^a \oplus conv_i^b)\}. \quad (3.3)$$

Here $C_{1 \times 1 \times 1}$ indicates a $1 \times 1 \times 1$ convolution, and $\oplus$ is element-wise addition. The pointwise convolution operation, with a kernel size of $1 \times 1 \times 1$, essentially amounts to multiplying the input tensor by a scalar weight. Despite its smaller spatial extent, it can have a significant impact through the weighted sum of input values; for allowing channel-wise interactions. This is often used in neural network architectures to adjust the number of channels (or dimensions) and to introduce non-linearities. The normalized weightage matrix becomes

$$w_i = Sigmoid(S_i), \quad (3.4)$$

with a sigmoid operation ($Sigmoid$) being applied along the spatial dimensions. The weighted feature map G_i is computed by element-wise multiplication $\otimes$ of the normalized weight matrix with the feature map E_i from the encoder pathway. We have

$$G_i = w_i \otimes E_i. \quad (3.5)$$

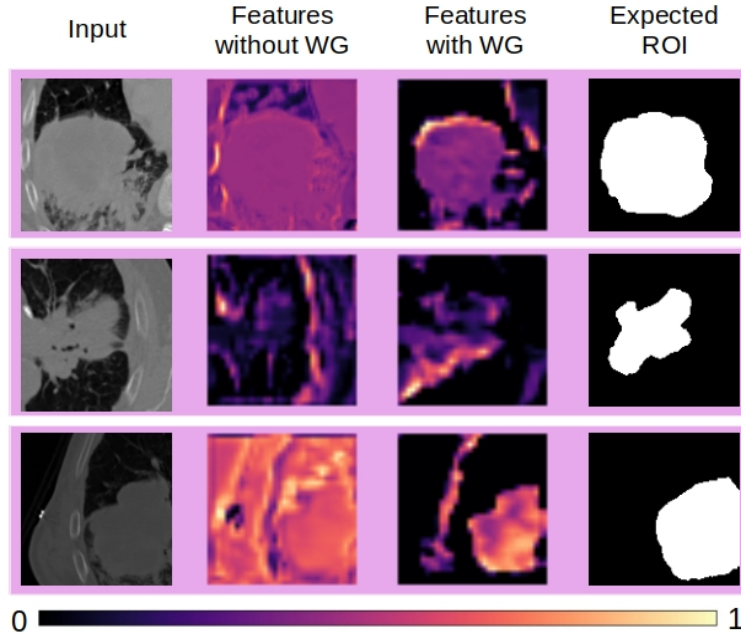


FIGURE 3.2: Feature maps generated with and without incorporating the WG modules.

Incorporating weight generation (WG) module into the network architecture of Fig. 3.3 (blue circle) achieves dynamic assignment of importance to different

spatial locations in the feature maps. This allows more precise and accurate information fusion between the encoder and decoder pathways. Thereby, the network is able to capture finer details for improving the segmentation performance; particularly, in those regions of the input image that require higher focus and refinement. The claim can be supported by Fig. 3.2, where we provide sample intermediate feature maps, generated with and without incorporating the *WG* modules. It is evident that the feature maps generated with *WG* provide a better match with the expected ROI depicted in last column of the figure.

3.2.3 Network architecture

Let us now introduce the WDU-Net. The unique architecture is composed of a group of *WG* and *DC* blocks, which are placed along the encoding and decoding arms of the *U*-Net framework. The architecture is illustrated in Fig. 3.3. The segmentation performance gets enhanced and is robust to generalisation through the use of the *DC* blocks in lieu of down- and/or up-sampling. The *WG* mechanism of Fig. 3.3(b) assigns each pixel the necessary weight during decoding of the *DC* in Fig. 3.3(c).

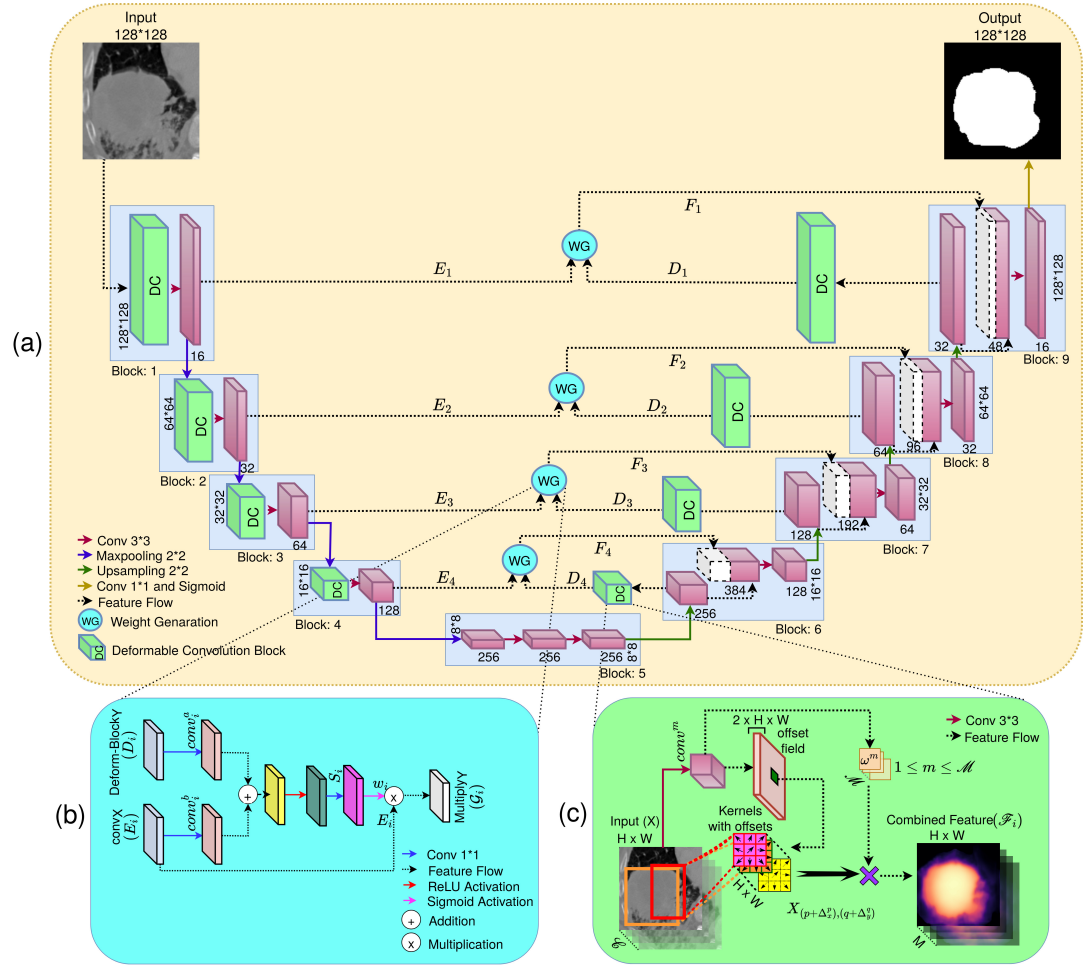


FIGURE 3.3: The WDU-Net framework, with detailed representation of (b) Weight Generation (*WG*), and (c) Deformable Convolution (*DC*) blocks.

The offsets of the DC are derived by applying a convolutional layer over the same input feature map, as shown in Fig. 3.3(c). Spatial resolution and dilation of the convolution kernel are identical to those of the current convolutional layer [*i.e.*, also 3×3 with dilation 1]. The spatial resolution of the output offset field matches with that of the corresponding input feature map. The channel dimension $2N$ is equivalent to $N \times 2D$ offsets. Both offsets and output feature-generating convolution kernels are concurrently learned during training.

The WDU-Net architecture comprises nine convolution blocks, four max-pooling layers, four upsampling convolution layers, and eight deformable convolution blocks. Initially, CT image patches of size 128×128 pixels are fed at the input. With a stride of 1, the patches filter through four sets of iterations encompassing 2×2 max-pooling layers, 3×3 convolution layers, and DC block layers. The 2×2 Upsampling layers of the decoder aid in regaining the final resolution of the image(s).

The DC blocks are introduced in the first four encoder layers and the final four decoder layers. They assist in gathering ROI-specific data, thereby lowering any error at the segmentation boundary while increasing the accuracy Fig. 3.8. The position of the lung tumor infection lesions often overlap with the bone, bronchiole, and liver structures; besides, they resemble the bronchioles and are hence difficult to segment. Therefore, the high-level semantic feature maps in the decoder are concatenated through the WG module to focus on the lower-level details in the retrieved feature maps (of the encoder).

The up-sampled images are merged with their equivalent encoded representations to enhance the significance of a pixel through the WG module. Adaptive selection of spatial information is accomplished by highlighting the pixels from the ROI, while suppressing the less important ones. The last layer of the WDU-Net uses sigmoid activation function to generate a probabilistic ROI at the output.

The gradient of the predicted segmentation mask is computed w.r.t. the model parameters, by back propagating the error using the chain rule. Let the weight matrix at the i th level be W_i , with model parameters θ . The gradient of the weight matrix, at each level of the decoder network, is given by

$$\frac{\partial W_i}{\partial \theta} = \frac{\partial \phi_{wgt}}{\partial W_i} \frac{\partial f_{dec}}{\partial Z_i} \frac{\partial Z_i}{\partial \theta} + \frac{\partial \phi_{wgt}}{\partial W_i}, \quad (3.6)$$

where $\frac{\partial \phi_{wgt}}{\partial W_i}$ is the gradient of the weight generation operation w.r.t. the weighted mask, $\frac{\partial f_{dec}}{\partial Z_i}$ is the gradient of the decoder block w.r.t. the DC block, and $\frac{\partial Z_i}{\partial \theta}$ is the gradient of the DC block w.r.t. the model parameters.

The gradient of the DC block w.r.t. the model parameters, at each level of the encoder network, becomes

$$\frac{\partial Z_i}{\partial \theta} = \frac{\partial \phi_{conv}}{\partial Z_i} \frac{\partial Z_{i-1}}{\partial \theta} + \frac{\partial \phi_{conv}}{\partial \theta}, \quad (3.7)$$

where $\frac{\partial \phi_{conv}}{\partial Z_i}$ is the gradient of the basic convolution w.r.t. DC , and $\frac{\partial \phi_{conv}}{\partial \theta}$ is the gradient of convolution w.r.t. model parameters.

3.2.4 Loss functions

Let the ground truth segmentation mask (for N pixels) be $y \in \{0,1\}$, with the corresponding predicted mask being $\hat{y}$ having estimated probability $p \in [0,1]$, and p_t is the probability depending on the value of y . The focal loss (FL) of eqn. (2.8) overcomes class imbalance in datasets, where positive number of pixels are relatively insufficient. A weighting factor $\alpha = 0.7$ and focusing parameter $\gamma = 2$ were selected experimentally.

The asymmetric similarity loss (ASL) [62] adjusts the weights between FP and FN (thereby, achieving a good balance between precision and recall) while training a network over highly imbalanced data. Asymmetric similarity loss is defined as

$$\mathcal{L}_{as}(y, p) = \frac{(1 + \beta^2) \sum_{i=1}^N p_i y_i}{(1 + \beta^2) \sum_{i=1}^N p_i y_i + \beta^2 \sum_{i=1}^N (1 - p_i) y_i + \sum_{i=1}^N p_i (1 - y_i)}, \quad (3.8)$$

with the choice of hyper-parameter $\beta = 1.5$ being determined over experiments on train data.

We combine the merits of the loss functions of eqns. (2.8) and (3.8) as the new **Focal asymmetric similarity loss** (FAS) for improved segmentation of highly imbalanced data, like the ROI being very small in size with respect to the background region. This is defined as

$$\begin{aligned} \mathcal{L}_{fas} &= \lambda \mathcal{L}_{as} + (1 - \lambda) \mathcal{L}_f, \\ \Rightarrow \mathcal{L}_{fas} &= \frac{\lambda(1 + \beta^2) \sum_{i=1}^N p_i y_i}{(1 + \beta^2) \sum_{i=1}^N p_i y_i + \beta^2 \sum_{i=1}^N (1 - p_i) y_i + \sum_{i=1}^N p_i (1 - y_i)} \\ &\quad - \alpha(1 - \lambda)(1 - p_t)^\gamma \log(p_t), \end{aligned} \quad (3.9)$$

where the choice of hyper-parameter $\lambda = 0.65$ was made experimentally on train data.

A few other existing loss functions [1] were also compared. These include the DL , $CEDL$, $IoUL$, TL , and FTL , of eqns. (2.1)-(2.5). It is evident that the performance of a model is directly affected by the selection of the appropriate loss function. This is experimentally elaborated in Section 3.3.3.

The frequently used performance metrics for segmentation are based on the misclassification error, obtained in terms of the confusion matrix which encompasses the TP , FP , TN , and FN . We evaluate with DSC , JSC , $Precision$, $Recall$, AUC , and $HD95$, of eqns. (1.7)-(1.12).



3.3 Implementation and Results

Six publicly available lung tumor datasets, *viz.* NSCLC radiomics, NSCLC radiogenomics, Decathlon, LIDC-IDRI, Moffitt, and Rider, as outlined in Table 1.1, were used for data preparation related to the training and testing of the models. The performance of WDU-Net was evaluated on the five test datasets,

viz. NSCLC radiogenomics, Decathlon, LIDC-IDRI, Moffitt, and Rider, obtained from different sources, and which were not used during training. This criterion justifies the generalization capability of our model.

3.3.1 Data preparation

As described in Section 2.3.2, patches were extracted in such a way that the class imbalance between the cancerous regions and the background region could be minimized.

Training

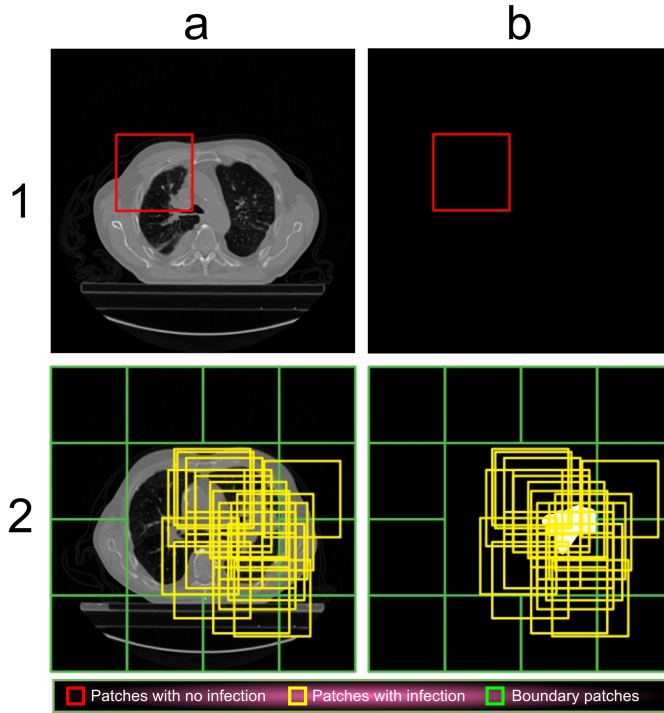


FIGURE 3.4: Patch extraction from training CT slices, depicting
Row 1. No tumor-affected regions: (a) Non-infected patches, and (b) corresponding masks; and
Row 2. Tumor-affected regions: (a) overlapping patches, and (b) mapping to corresponding annotated masks.

Let there be a set $B = 263$ samples, divided into subsets $B1, B2, \dots, B10$, as ten bins. The distribution of the infected and non-infected slices and/or patches for the training is provided in Table 3.1.

Testing

Patches were analogously extracted from the test image. However unlike during training, here the overlap between the patches was kept at a minimum to avoid any possible missing edge regions among the patches. All non-overlapping patches were extracted, along with those overlapping patches having a 25%

TABLE 3.1: Distribution of the infected and non-infected Slices & Patches, extracted from each bin for training

Bin No.	Number of samples	Slices		Patches	
		Infected	Non-infected	Infected	Non-infected
B1	26	602	2275	12095	9444
B2	26	517	2463	10427	8580
B3	26	468	2797	9360	8413
B4	26	332	2897	6651	6870
B5	26	375	2534	7557	6977
B6	26	500	2477	10004	8473
B7	26	497	2361	9962	8303
B8	26	432	2736	8640	7920
B9	26	447	2674	8961	8017
B10	29	479	3026	9604	8750
Total:	263	4649	26240	93261	81747

overlap from each of their corresponding four neighbouring patches. This is normally sufficient to cover the entire lung region in a CT sample.

Axial slices (512×512 pixels) were taken from each test CT volume. Each slice produced sixteen 128×128 pixels of non-overlapping patches and nine 128×128 pixels overlapping patches, as shown in Fig. 3.5.



FIGURE 3.5: Patch extraction from test CT slices.

3.3.2 Implementation

The training set, constituting 263 patients, was randomly divided into 10 bins followed by 10-fold cross validation. Training used ten versions of our WDU-Net, and is represented here as $M1, M2, \dots, M10$. While training with one model, one of the ten bins was left for testing. The detail of the data splitting

for each of the ten versions of the training models (WDU-Net) is shown in Table 3.2. The corresponding validation datasets are designated as $B_1, B_2, \dots, B_{10}$, and elaborated in the table. As an illustration, in instance 1, the model M_1 is trained using bins B_2 through B_{10} , and validated on the data in bin B_1 . Analogously, case 2 involves training M_2 with bins B_1 and B_3 through B_{10} , while validating upon bin B_2 . Here each model, from M_1 to M_{10} , is trained using a separate set of initialization, learning, dropout, and other parameters. The Adam optimizer was used, with a constant batch size of 16. After multiple studies, the learning rate and dropout probability were selected as 0.001 and 0.2, respectively.

TABLE 3.2: Ensemble WDU-Net model M_i , with the performance evaluated on the corresponding validation set B_i

Model	Training	Validation	Performance evaluation metric					
			DSC	JSC	$Precision$	AUC	$Recall$	$HD95$
M1	B/{B1}	B1	0.8707	0.6934	0.8032	0.8905	0.8213	14.0562
M2	B/{B2}	B2	0.8318	0.7129	0.7841	0.8219	0.8567	9.5277
M3	B/{B3}	B3	0.9047	0.7735	0.8537	0.8927	0.8629	9.3801
M4	B/{B4}	B4	0.8920	0.7310	0.8217	0.8834	0.8874	12.3549
M5	B/{B5}	B5	0.8516	0.6813	0.8109	0.8269	0.8130	8.2713
M6	B/{B6}	B6	0.8817	0.6720	0.8573	0.9017	0.8505	10.6405
M7	B/{B7}	B7	0.8681	0.7703	0.8148	0.8852	0.8280	7.2780
M8	B/{B8}	B8	0.8591	0.7418	0.8336	0.8470	0.8517	8.3924
M9	B/{B9}	B9	0.8959	0.8014	0.8539	0.8802	0.8833	12.1390
M10	B/{B10}	B10	0.9065	0.7972	0.8725	0.8934	0.8848	11.3424
Mean $\Rightarrow$			0.8762	0.7375	0.8306	0.8723	0.8540	10.3383
std. dev. $\Rightarrow$			0.0232	0.0448	0.0268	0.0277	0.0254	2.0305
where B = { Bi: $i \in \mathbb{N} \wedge i \in [1, 10]$ }								

The WDU-Net was trained on CT images of 263 patients, through ensemble of ten base-classifiers using ten sets of data. Each classifier starts from scratch and uses only nine bins (Table 3.2) of the data during training. Therefore, each time it creates an entirely new classifier with a unique set of parameters. Each scenario leaves one bin for validation. The five independently-collected, publicly available test datasets (Table 1.1) were then used for testing the ten trained WDU-Net models, whose outputs were ensemble to segment the lung tumor infection region via majority voting. The performance, evaluated in terms of different metrics, was observed to be consistent, accurate and reliable across ten models; depicting good generalisation with 10-fold training.

3.3.3 Analysis and discussion

The goal is to segment a CT lung image into the ROI (*ie.*, lung tumor) from the background (containing all other regions in the image). This is often difficult for a conventional U -Net. The deformable convolution DC circumvents the problem by appropriately incorporating the necessary local and neighbourhood information for precise delineation of the ROI, while the WG module helps to assign weightage on it. The loss function FAS [eqn. (3.9)] handles

class imbalance within image patches. We explored the effect of the different modules, additively included in the WDU-Net model. These comprise of the vanilla *U-Net*, *WU-Net* (with *WG* but without *DC*), *DU-Net* (with *DC* but without *WG*), and our proposed WDU-Net. In all the tables, provided here, the best output scores are marked in bold; with all scores being appended by the corresponding standard deviation (std. dev).

Qualitative

A comparative visual study on one slice, from each test dataset, is presented in Fig. 3.6. The sample slices from the five test datasets are selected as *S1* from NSCLC Radiogenomics, *S2* from Decathlon, *S3* from LIDC-IDRI, *S4* from Moffitt, and *S5* from Rider. It is observed that the red (under-segmented) and yellow (over-segmented) portions are minimum in our WDU-Net (column (f)), whereas the mismatch of *U-Net* in column (c) is maximum compared to the ground truth. The results are depicted for different ROI sizes, covering *very small* as well as the *larger* ones; In all cases the proposed WDU-Net performed the best.

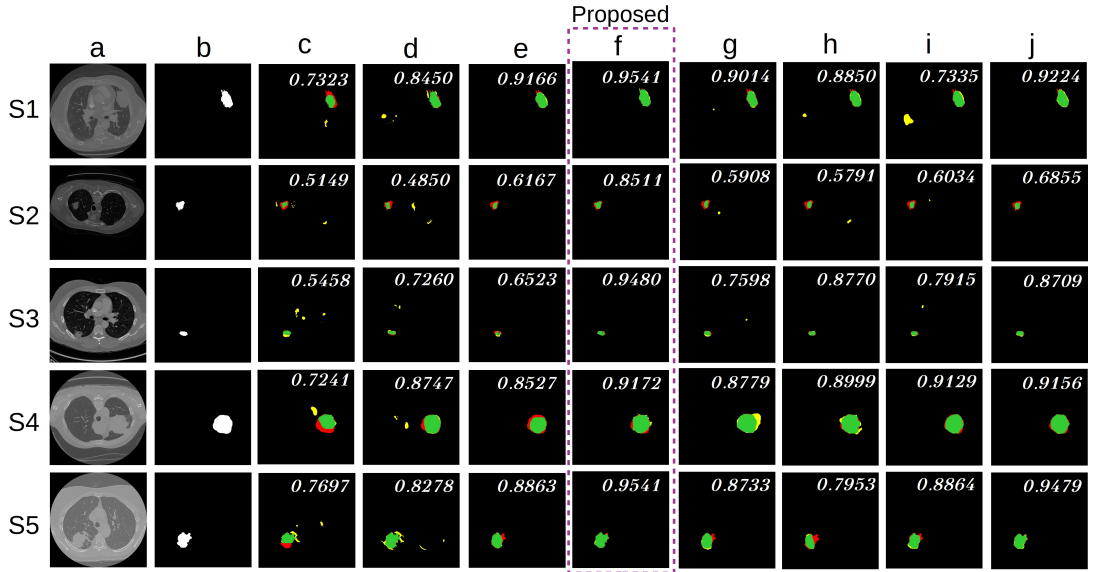


FIGURE 3.6: Comparative study on sample slices from the test datasets, using different models, with corresponding *DSC* being inscribed on each slice. (a) Original CT scan, with (b) Annotated masks. Segmentation obtained by (c) *U-Net*, (d) *WU-Net* (without *DC*), (e) *DU-Net* (without *WG*), (f) WDU-Net, (g) *U-Net++*, (h) *MultiResUNet*, (i) *U-Mamba*, (j) *Swin-UMamba* with regions indicating Green: *TP*, Red: *FN*, Yellow: *FP*

In order to understand the effect of deformable convolution, we consider a sample tumor patch with arbitrary geometric shape in Fig. 3.8. While part (a) of the figure represents the original input patch, part (b) corresponds to its ground truth. Parts (c)-(f) of the figure refer to the segmented outputs generated, upon application of the models *U-Net*, *U-Net* with the *WG*-module, *U-Net* with the *DC*-module and our WDU-Net, respectively. It is evident from Fig. 3.8(e) that the shape of the tumor patch has been accurately segmented

compared to parts (c) and (d) [having no deformable convolution]. This is supported by the counts of FP and FN , as is evident in the figure. In fact, there is a great reduction in FP s in the segmented shape. Adding weight to the deformable convolution further enhanced the accuracy of the model in part (f) of the figure, with a greater reduction in FN s, although at the cost of slightly higher FP s. This highlights the superiority of our WDU-Net in segmenting an arbitrary region.

Quantitative

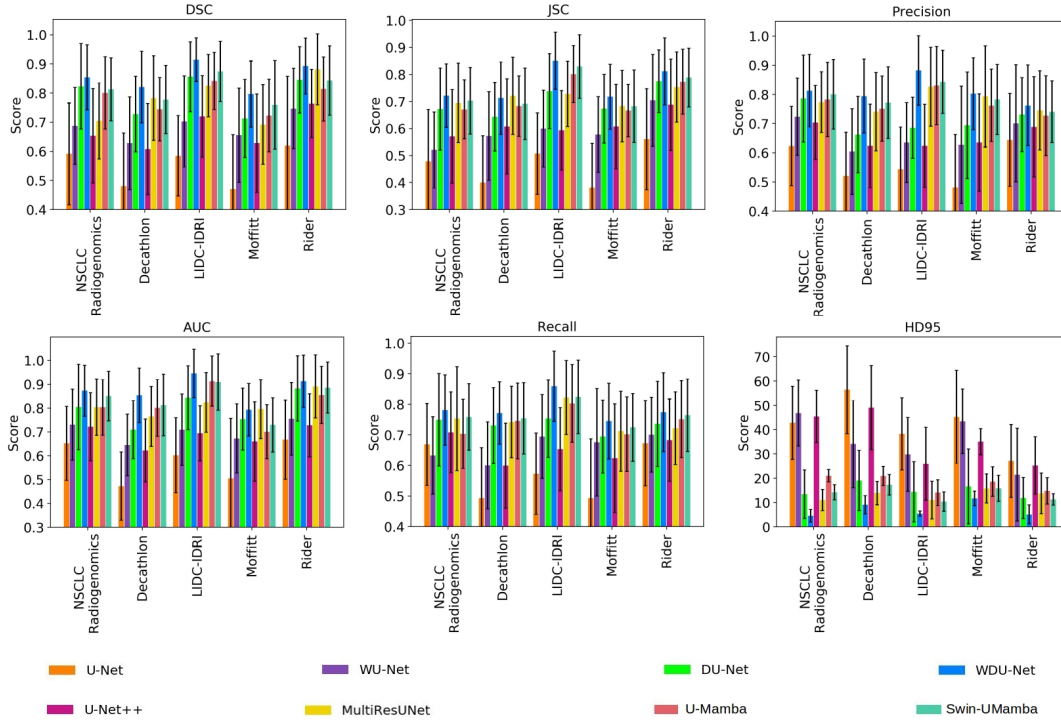


FIGURE 3.7: Comparative performance on unseen test data with related deep models, using ensemble with FAS loss (with the same testing framework used in all models)

Fig. 3.7 visualizes the performance of the WDU-Net, as the base model in this ensemble framework, along with those of the vanilla U -Net, WU-Net (without DC), DU-Net (without WG), U -Net++, MultiResUNet, Random forest, SVM, and Otsu thresholding. Here we provide results using six metrics, where DSC , JSC , $Precision$, AUC , $Recall$ signify improvement for higher values, while $HD95$ implies betterment with lower values. It is evident that the deep methods, like U -Net, WU-Net, DU-Net, U -Net++, MultiResUNet, WDU-Net, outperformed the classical methods like Otsu, SVM and Random forest. Furthermore, our WDU-Net reported the best performance.

In order to understand the effect of deformable convolution, we consider a sample tumor patch with arbitrary geometric shape in Fig. 3.8. While part (a) of the figure represents the original input patch, part (b) corresponds to its ground truth. Parts (c)-(f) of the figure refer to the segmented outputs generated, upon application of the models U -Net, U -Net with WG -Module, U -Net with DC -Module, and our WDU-Net, respectively. It is evident from

Fig. 3.8(e), that the shape of the tumor patch has been accurately segmented as compared to parts (c) and (d) [having no deformable convolution]. This is supported by the count of FP and FN as well, as evident in the figure. There is, in fact, a great reduction in FP s with the segmented shape coming out more-or-less nicely. Adding weight to the deformable convolution further enhanced the accuracy of the model in part (f) of the figure, with more reduction in FN s, *albeit* at the cost of slightly higher FP s. This highlights the superiority of our WDU-Net in segmenting an arbitrary region.

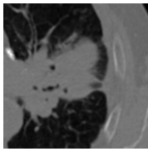
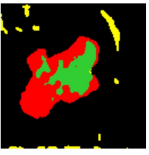
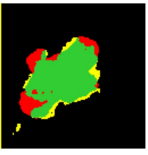
a	b	c	d	e	f
					
	Positive pixels: 3154	TP: 1105 FP: 422 FN: 2049	TP: 2543 FP: 465 FN: 611	TP: 2903 FP: 51 FN: 251	TP: 3117 FP: 144 FN: 37

FIGURE 3.8: Comparative study on a sample test patch, using different models, indicating the corresponding generated FP and FN regions. (a) Original CT scan, with (b) Annotated masks. Segmentation by (c) U -Net, (d) U -Net with WG -Module, (e) U -Net with DC -Module, and (f) WDU-Net, with regions labeled Green: TP , Red: FN , Yellow: FP

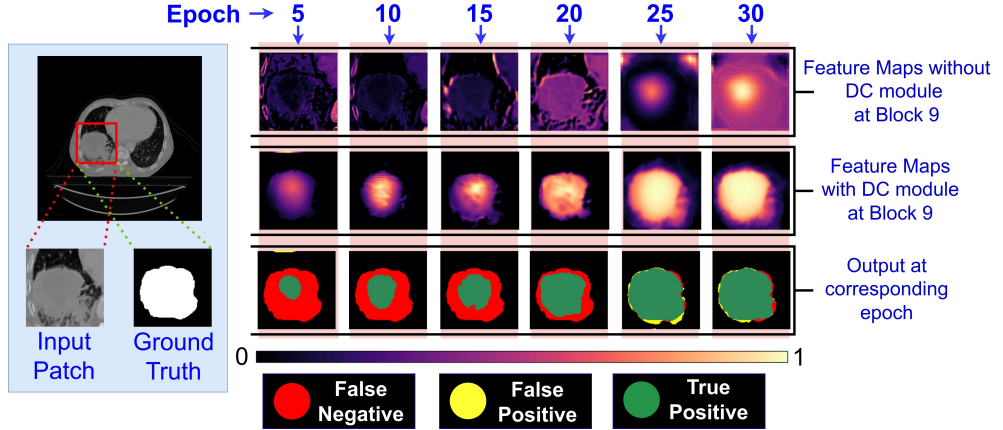


FIGURE 3.9: Effect of introducing Deformable Convolution (DC) module at Block 8 (of Fig. 3.3) on the feature maps generated at different stages of training of the WDU-Net

Ablations

Thus the ablation study comprised of two parts, *viz.* the effect of (i) using the DC module and WG module in the WDU-Net, and (ii) a combination of two loss functions, as evaluated w.r.t. the effect of each individual loss. The effect of the DC and WG modules is demonstrated in terms of the segmentation results of Fig. 3.8. It is clearly observed that while DC block can properly capture the irregular shapes of lesions, the WG block helps to appropriately focus on the

ROI. On the other hand, the vanilla *U*-Net is found to miss many important parts of the cancerous region. Analogously, the *U*-Net, *WU*-Net, *DU*-Net, are visually found to be weaker in performance as compared to our *WDU*-Net. The objective evaluation is reported in Fig. 3.7 in terms of the error-bars, where the average performance of *WDU*-Net is observed to be the highest among all others with lower variation across the data.

The effect of Deformable Convolution (*DC*) on the feature maps generated by our *WDU*-Net, at different stages of training, was also analysed. It is evident from Fig. 3.9 that the weighted *DC* helps the network converge faster on the ROI. Here the feature map is displayed after the 5th, 10th, 15th, 20th, 25th and 30th epochs, respectively. The ROI is observed to be gradually estimating the target ground truth, as depicted in the figure. It is also seen that the convergence to the ROI is faster and more accurate, as compared to the model when not incorporating *DC* module.

Table 3.3 presents a study on the effect of using different loss functions [eqns. (2.1)-(2.8)] on the performance of *WDU*-Net, as measured in terms of the *DSC* of eqn. (1.7) on all five test datasets. The loss function *FAS* is observed to be the best in handling class imbalance, such that the error in segmentation is minimum. The network can focus on the smaller ROIs while the background regions get suppressed. Simultaneously, the *FP*s and *FN*s get minimized through down-weighting of the easier samples as the model emphasises the more difficult ones; thereby, resulting in an improved precision-recall balance.

TABLE 3.3: Effect of various loss functions on *DSC* for the *WDU*-Net, with ensemble, over the test datasets

Loss	NSCLC Radiogenomics	Decathlon	LIDC-IDRI	Moffitt	Rider
DL	0.7431 $\pm$ 0.156	0.6714 $\pm$ 0.182	0.7392 $\pm$ 0.121	0.6841 $\pm$ 0.209	0.7562 $\pm$ 0.134
CEDL	0.6172 $\pm$ 0.139	0.5779 $\pm$ 0.174	0.7088 $\pm$ 0.134	0.6225 $\pm$ 0.176	0.7390 $\pm$ 0.129
IoU	0.7509 $\pm$ 0.126	0.6782 $\pm$ 0.162	0.7239 $\pm$ 0.117	0.7097 $\pm$ 0.172	0.7883 $\pm$ 0.106
FL	0.8198 $\pm$ 0.143	0.7027 $\pm$ 0.156	0.7597 $\pm$ 0.154	0.7129 $\pm$ 0.139	0.8297 $\pm$ 0.110
TL	0.7994 $\pm$ 0.162	0.6834 $\pm$ 0.138	0.7441 $\pm$ 0.120	0.6876 $\pm$ 0.135	0.7999 $\pm$ 0.104
FTL	0.8210 $\pm$ 0.119	0.7875 $\pm$ 0.124	0.8214 $\pm$ 0.133	0.7132 $\pm$ 0.117	0.8553 $\pm$ 0.107
ASL	0.7635 $\pm$ 0.129	0.6916 $\pm$ 0.134	0.7735 $\pm$ 0.127	0.6807 $\pm$ 0.131	0.8039 $\pm$ 0.118
FAS	0.8534 $\pm$ 0.112	0.8201 $\pm$ 0.123	0.9137 $\pm$ 0.074	0.7965 $\pm$ 0.114	0.8924 $\pm$ 0.096

The effect of loss functions in Table 3.3 demonstrates that the new combination *FAS* is better, as compared to *FL* and *ASL* considered individually. The *DC* and *WG* modules, with *U*-Net, help in reducing over-and-under segmentation by minimising the error. The *FAS* allows maintaining a proper balance between *Precision* and *Recall* in Table 2.5, such that the effect of class imbalance gets minimized.

Complexity

The model complexity, in terms of the number of parameters, is enumerated in Table 3.4. Note that the ensemble *WDU*-Net demonstrates the best overall performance among all variants, while maintaining an optimal number of parameters, as evident from the Table 3.4 and Fig. 3.7.

TABLE 3.4: Computational complexity of the various modules of WDU-Net and existing models

Model	$\Rightarrow$	<i>U</i> -Net	<i>WU</i> -Net	<i>DU</i> -Net	WDU-Net	MultiResUNet	<i>U</i> -Net++	<i>U</i> -Mamba	Swin- <i>U</i> Mamba
# parameters	$\Rightarrow$	1.96M	1.99M	3.92M	3.95M	7.26M	4.87M	60.02M	28.47M

3.3.4 Comparative study

Finally, the performance of some of the effective state-of-the-art methods, as reported in related literature on lung tumor segmentation, were compared in Table 3.5 with reference to that of our WDU-Net. The performance metrics used were *DSC*, *JSC*, and *Precision* of eqns. (1.7)-(1.9). In each case the test datasets were as indicated in the table. As mentioned in Section 3.1, there exist research reported on datasets LIDC-IDRI ([24], [76], [77], [137], [173], [183], [185], [206]), NSCLC Radiogenomics ([123]), NSCLC-Radiomics ([35], [49], [77], [195]), Decathlon ([43], [49]), and Moffitt [58]. It is observed from the bold entries in the table that our WDU-Net performed the best over all these state-of-the-art methods, across each individual dataset. It may be pointed out that with the Moffitt dataset, Ref. [58] reported a 3D similarity index of 0.7829. Although our results are based on 2D similarity indices, these are definitely comparable. The average *DSC* values are also presented in the table, on each of the test sets. The highest *DSC* value of 0.9137 was obtained in *LIDC – IDRI*, while 0.7965 was our lowest value in *Moffitt* (which is still reasonably good). It is evident that WDU-Net demonstrated its superior performance and efficacy in various contexts.

TABLE 3.5: Comparative analysis of recent literature on deep learning in segmentation of lung CT

Dataset	Methodology	Performance Metric			
		<i>AUC</i>	<i>DSC</i>	<i>JSC</i>	<i>Precision</i>
NSCLC Radiogenomics [13]	Transfer Learning [123]		0.7255		
	WDU-Net	0.8724	0.8534	0.7213	0.8125
LIDC-IDRI [10]	Unet++ [206]			0.7705	
	Residual Unet[173]		0.8755		
	CoLe-CNN [137]		0.8610	0.7660	
	3D-Unet [185]		0.8020		
	MRRN [77]		0.6800		0.6700
	ResDSda_U-Net [76]		0.8665	0.7673	0.8722
	Encoder-decoder Network [183]		0.7960	0.6650	
	DB-Net [24]		0.8889		
	WDU-Net	0.9449	0.9137	0.8503	0.8817
Decathlon [7]	Teacher-student Model [49]		0.7100		
	Residual separable CNN [43]		0.6490		0.7650
	WDU-Net	0.8531	0.8201	0.7124	0.7940
Moffitt [131]	Ensemble segmentation [58]	Similarity Index (3D): 0.7829			
	WDU-Net	0.7924	0.7965	0.7173	0.8021
Rider [199]	Multi-view Aggregation [103]			0.7004	
	WDU-Net	0.9120	0.8924	0.8113	0.7629
NSCLC Radiomics [2]	Teacher-student Model [49]		0.5389		
	Multi-scale Aggregation-based Parallel Transformer Network [35]		0.5353		
	Multi-scale squeeze-and-excitation U-Net [195]		0.7800	0.6440	
	MRRN [77]		0.7400		0.7300
	WDU-Net	0.8707	0.6934	0.8306	

Superiority over *FMC*-Net

Another set of experiments was performed to comparatively evaluate the effectiveness of WDU-Net with respect to *FMC*-Net of Chapter 2, over both COVID-19 and lung tumor datasets of Table 1.1.

Results depicted in Fig. 3.10 are based on annotated masks. Four of the samples from the COVID-19 data, and three from the lung tumor data, were investigated qualitatively. It is seen that WDU-Net yielded better segmentation.

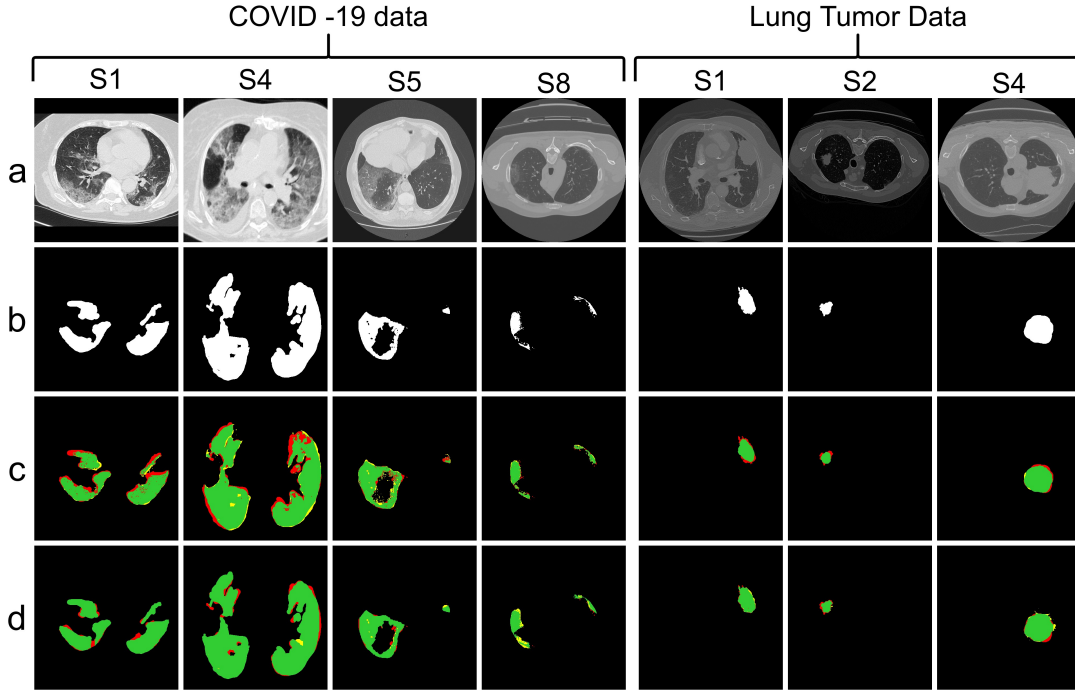
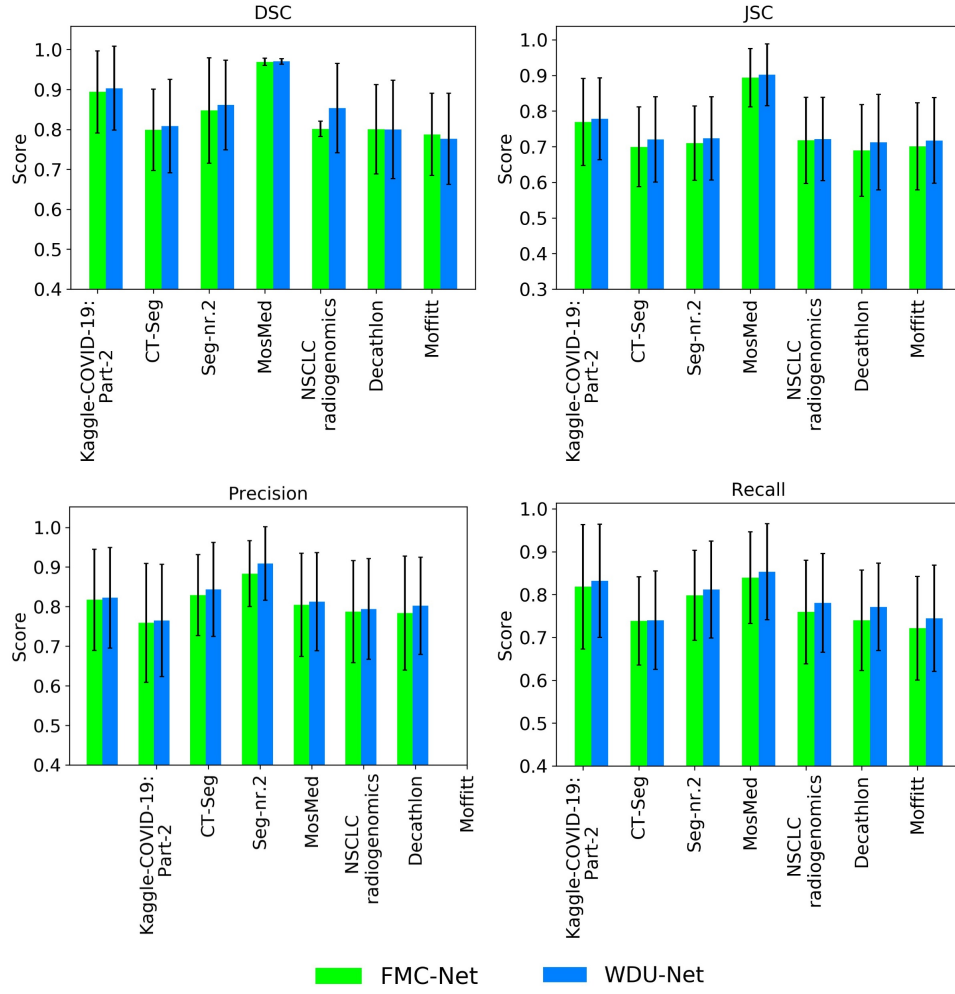


FIGURE 3.10: Sample segmentation output. (a) Original CT scan, with (b) Annotated masks. Segmentation obtained by (c) *FMC*-Net, and (f) WDU-Net, with regions Green: True Positive, red: False Negative, Yellow: False Positive

Figure 3.11 compares the quantitative performance of the two models. Results are presented in terms of the four metrics, *viz.* *DSC*, *JSC*, *Precision*, and *Recall* [eqns. (1.7)-(1.10)]. It is clearly observable that the WDU-Net exhibited better performance over most of the datasets used.

Table 3.6 illustrates the influence of different loss functions [eqns. (2.1)-(2.8)] on the performance of the two deep networks under comparison (in terms of *DSC*). The two datasets used correspond to COVID-19 and lung tumors, respectively. Again the WDU-Net is found to perform better in all cases.

It can be concluded that in datasets where the ROI exhibited a specific geometrical shape, both the WDU-Net and *FMC*-Net demonstrated equivalent performance. However, in cases where the ROI lacked any distinctive geometrical shape, the WDU-Net was found to outperforms the other deep networks examined in this study. In the context of COVID-19, infected regions comprising of ground glass opacity (GGO), consolidation, and pleural effusion, often exhibited a “crazy paving” pattern. Tumors, on the other hand, typically follow a solid pattern; which may not always adhere to a specific geometrical shape.

FIGURE 3.11: Comparative performance on unseen test data in terms of FAS lossTABLE 3.6: Effect of various loss functions, measured in terms of DSC

Dataset	Loss function	FMC -Net	WDU -Net
MosMed	DL	0.8465 ± 0.220	0.8472 ± 0.142
	IoU	0.8706 ± 0.137	0.8874 ± 0.105
	FL	0.9584 ± 0.012	0.9603 ± 0.015
	TL	0.8360 ± 0.056	0.8417 ± 0.045
	FTL	0.8973 ± 0.162	0.9032 ± 0.072
	ASL	0.8724 ± 0.101	0.9133 ± 0.035
	FAS	0.9694 ± 0.009	0.9703 ± 0.007
NSCLC Radiogenomics	DL	0.6549 ± 0.133	0.7431 ± 0.156
	IoU	0.6617 ± 0.138	0.7509 ± 0.126
	FL	0.7832 ± 0.140	0.8198 ± 0.143
	TL	0.7655 ± 0.125	0.7994 ± 0.162
	FTL	0.7283 ± 0.114	0.8210 ± 0.119
	ASL	0.7550 ± 0.102	0.7635 ± 0.129
	FAS	0.8015 ± 0.119	0.8534 ± 0.112

The WDU-Net, therefore, exhibited superior performance due to its adaptability to varying patterns and shapes of the ROI.



3.4 Conclusions

A novel Weighted Deformable *U*-Net (WDU-Net) has been designed for efficient segmentation of tumors from lung CT. It performed better, as compared to state-of-the-art models, in many aspects. The reason was three-fold. (1) The *DC* module captured the irregular shape of the tumor region, assisted by the *WG* module for suppressing unimportant features and highlighting the relevant ones. (2) The *FAS* loss function, a judicious combination of the Focal Loss and Asymmetric Similarity Loss, helped to effectively model class imbalance. (3) The patch-based training aided improved and balanced learning. Combining the outputs of ten ensemble classifiers, through majority voting, added another dimension to the inference process while arriving at a proper decision regarding the segmentation of the ROI. The WDU-Net was found to be superior, compared to existing methods, in the segmentation of lung cancer. This was demonstrated through several experiments.

The model WDU-Net was fast in learning, with limited data, and produced good generalization. It was observed through ablation studies that the *DC* and *WG* modules could help the network focus on the ROI. The training and testing times were also lower. Ensemble learning provided an effective strategy for combined decision-making. The high accuracy and robustness of the output was due to the reduction in over-and-under segmentation.

Deformable convolutions do introduce additional overhead due to their additional parameters for learning the spatial offsets. This leads to increased computational complexity, higher memory requirements, and longer training times. However, the trade-off lies in their improved adaptability in capturing fine-grained spatial relationships. This makes deformable convolutions valuable in object detection and semantic segmentation.

By advancing methods for continuous monitoring and analysis of human physical activities, this research could lead to better healthcare outcomes, especially for elderly individuals or those with chronic conditions. The ability to accurately detect and respond to physical states and movements in real-time can improve the quality of life by promoting proactive health management, and reducing healthcare costs. Additionally, the development of such technology can contribute to the broader adoption of smart healthcare systems, fostering a more connected and health-conscious society.

The subsequent two chapters focus on the development of classification models for efficiently detecting cancer of the lung, and distinguishing between tumors and COVID-19, in the framework of multimodal imaging. This includes CT, PET and synthetically generated EIT scans.



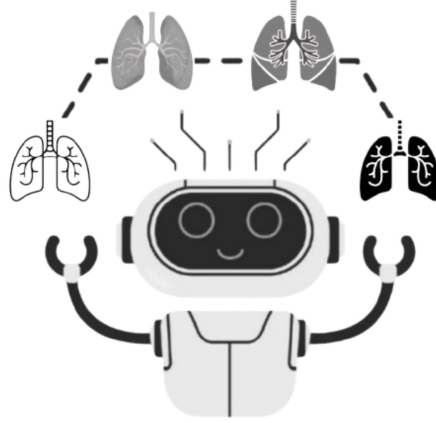


Chapter 4



Deep Ensemble with Multimodal Fusion for Classification





4.1 Introduction

Medical image analysis benefits from multimodal fusion through precise diagnosis, treatment planning, and improved model performance. Fusion strengthens discrimination capabilities, helping capture complex relationships for comprehensive data representation. Combining CT-PET modalities is found to enhance accuracy in cancer detection and prognosis [6]. Integrating MRI, CT, and PET data optimizes feature extraction, interpretability, and robustness, by revealing hidden patterns and reducing noise impact.

A gradient structural similarity-based filter was devised for multimodal analysis of CT/MRI data [50]. Analyzing tumor regions extraction in lung images, five optimization algorithms were assessed within the PET-CT framework using CNN [87]. Abnormal PET uptake in lungs was found significant for tumor detection, highlighting the value of multimodal representations from PET and CT scans, combining anatomical and functional information for more precise diagnosis with reduced false positives/negatives [109]. A fully automated pipeline for non-small cell lung cancer (NSCLC) detection and volumetric segmentation was validated on 1328 thoracic CT scans from eight institutions [139].

This chapter introduces a novel Deep Ensemble Multimodal Fusion (DEMF) network, employing a sequence of PCA and Autoencoder (PCAE), for integrating CT and PET scans. The PCA retains essential features from each modality, while filtering out irrelevant details and optimizing computational resources. The Autoencoder reconstructs fused representations for capturing meaningful patterns efficiently. The PCAE offers compact, interpretable representation by transforming data into uncorrelated principal components; thereby, enhancing efficiency and generalization over diverse applications. The DEMF shows robust performance with limited data availability. The main contributions are highlighted below.

- (i) Incorporation of PET-CT fusion mechanism PCAE to efficiently extract pertinent reduced features from PET and CT imagery of lung cancer.
 - The PCA individually extracts the important components of CT and PET modalities, as determined by the PCA graph. The corresponding images are generated in the reduced domains.

- The reconstructed CT and PET images are integrated by the autoencoder to generate the corresponding *fused CT-PET* image.
- (ii) Creation of an ensemble of foundational classifiers, using majority voting to integrate diverse decisions from the base classifiers. This approach accommodates variability in decisions, stemming from individual classifiers.
 - (iii) Streamlined visualization of activation regions within the framework of multimodal ensemble.
 - (iv) Ablation and comparative analyses, to showcase the efficacy of the new deep ensemble multimodal classifier model on publicly accessible datasets.

Section 4.2 describes the fusion technique with ensemble classification in the DEMF network. Section 4.3 provides the experimental implementation and results. Finally Section 4.4 concludes the chapter.

4.2 Methodology

Here we describe the different components of the DEMF network. Both the CT and PET modality images are already co-registered in the dataset. Initially the CT and PET image features are reduced and fused. Then the integrated image data is augmented (Fig. 4.5), followed by classification through ensemble. The overview of the entire procedure is illustrated in Fig. 4.1.

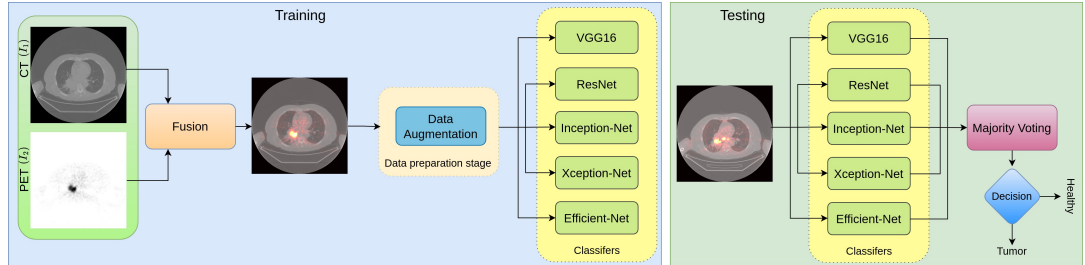


FIGURE 4.1: Workflow of ensemble classification DEMF, with multimodal image fusion

4.2.1 Fusion of PET-CT modalities

Image fusion in PCAE is accomplished through PCA, followed by an autoencoder. It leverages the merits of PCA for dimensionality reduction and feature extraction, along with the capabilities of an autoencoder to discover meaningful patterns from both modalities for efficiently reconstructing the fused representation. The process involves applying PCA to each modality for selecting principal components and then using an autoencoder to learn the relevant patterns, followed by reconstruction of the fused image. The schematic representation of the fusion of CT and PET image modalities is depicted in Fig. 4.2.

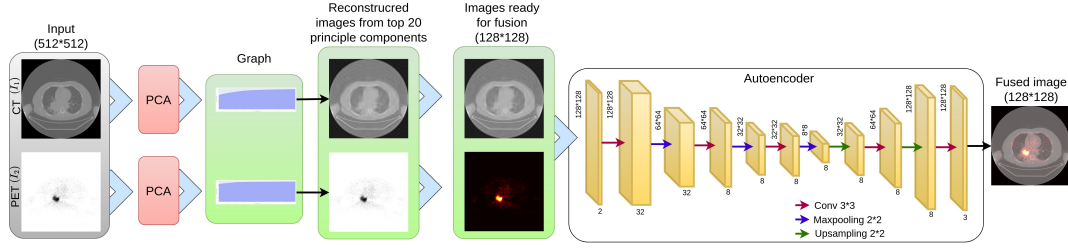


FIGURE 4.2: Image fusion strategy

Let us consider the two input images, I_1 and I_2 , represented as pixel matrices. They are integrated using PCA, followed by the autoencoder. Dimensionality reduction is performed on the top 20 principal components, which retain significant relevant information. Then the pair of 128×128 CT and PET images, reconstructed from the principal components, are supplied at the input of the autoencoder of Fig. 4.2. Here the encoder path consists of three sets of convolution and maxpooling, while the decoder path constitutes two sets of upsampling and convolution. Final outcome of the autoencoder is the decoded *fused* image extracted from $(20+20=)$ 40 components.

Principal Component Analysis (PCA)

1. Flatten the images by converting each image matrix I into a column vector x with the stacking of its pixels.
2. Stack the flattened image vectors x from both images to create a combined image data matrix

$$X_{\text{combined}} = [x_1, x_2]. \quad (4.1)$$

3. Calculate the mean vector $\bar{x}$ of the combined data, and center the data matrix by subtracting the mean, as

$$X_{\text{centered}} = X_{\text{combined}} - \bar{x}. \quad (4.2)$$

4. Compute covariance matrix C of the centered data, as

$$C = \frac{1}{N} X_{\text{centered}} X_{\text{centered}}^T. \quad (4.3)$$

5. Perform eigenvalue decomposition on C , to obtain the eigenvectors V and eigenvalues Λ .

$$C = V \Lambda V^T. \quad (4.4)$$

6. Select the Principal Components as the top k eigenvectors, corresponding to the largest eigenvalues. These eigenvectors form the matrix V_{top} .

For this purpose a PCA graph, showing the cumulative proportion of variance in the reconstructed image (ordinate) w.r.t. the number of principal components (abscissa). It was observed that for both the CT and PET images, the set of top 20 principal components enabled the cumulative

proportion of variance to reaches nearly 100%. Therefore, we selected these (top 20) principal components in our experiments.

7. Transform the centered data, using the selected principal components

$$Y_{\text{PCA}} = V_{\text{top}} X_{\text{centered}}. \quad (4.5)$$

8. Apply the inverse PCA transformation to reconstruct the image in the 128×128 feature space as

$$Y_{\text{PCAInverted}} = \frac{1}{4} Y_{\text{PCA}} V_{\text{top}}^T. \quad (4.6)$$

Here we essentially perform dimensionality reduction to the top 20 principal components, while retaining as much relevant information as possible.

Autoencoder

The pair of 128×128 CT and PET images, reconstructed from the principal components, are supplied at the input of the autoencoder. The architecture of the autoencoder is illustrated in Fig. 4.2. Here the encoder path consists of three sets of convolution and maxpooling, while the decoder path constitutes two sets of upsampling and convolution. The final outcome of the autoencoder is the decoded *fused* image extracted from the (20+20=) 40 components. The procedure is outlined as follows.

1. Train an autoencoder using $Y_{\text{PCAInverted}}$ as input, to learn a compressed representation Z of the PCA-transformed data. Here
Encoder function: $Z = E(Y_{\text{PCAInverted}})$.
2. Decode the compressed representation Z to obtain the reconstructed PCA-transformed data. We have
Decoder function: $Y_{\text{reconstructed}} = D(Z)$.
3. Apply fusion of PCA inverted data to obtain the fused data in the original feature space as

$$X_{\text{fused}} = Y_{\text{reconstructed}}^{CT} + Y_{\text{reconstructed}}^{PET}. \quad (4.7)$$

4. Reshape and Visualize: Reshape X_{fused} back into the 128×128 image format, to visualize I_{fused} as

$$I_{\text{fused}} = \text{Reshape}(X_{\text{fused}}). \quad (4.8)$$

4.2.2 Ensemble classification

Ensemble deep models combine multiple deep learning algorithms to enhance predictive accuracy, improve generalization, and reduce overfitting. Their diverse perspectives foster robust predictions and mitigate individual model biases, resulting in reliable outcomes across complex tasks. The DEMF builds an

ensemble of the classifiers VGG16, ResNet, Inception-Net, Xception-Net and Efficient-Net, pre-trained over the ImageNet, with the final outcome being decided by majority voting. Each of these classifiers takes the augmented PET-CT fused image set as input, to classify between the cancerous and healthy lung slices. The entire procedure is summarized in Algorithm 1.

Algorithm 1 Ensemble classification in DEMF

```

1:  $X_i = \text{Image } i$ 
2:  $Y_i = \text{corresponding label of Image } i$ 
3:  $N = \text{Number of images}$ 
4:  $\text{Data} = \{(X_i, Y_i)\}_{i=1}^N$ 
   #Training
5: for each model in Ensemble do
6:   for each epoch do
7:     for each batch in Train_Data do
8:       Batch of images  $X_b$  and labels  $Y_b$ 
9:        $\text{out} = \text{model}(X_b)$ 
10:      loss:  $\ell = \text{Loss}(Y_b, \text{out})$ 
11:      gradients:  $\nabla_{\theta} \ell$ 
12:      weights:  $\theta \leftarrow \text{Optimizer}(\theta, \nabla_{\theta} \ell)$ 
13:     end for
14:   end for
15: end for
16: Save the trained ensemble model weights
   #Testing
17: for each model in Ensemble do
18:   Test_images  $X_v$  and labels  $Y_v$ 
19:    $\text{test\_out} = \text{model}(X_v)$ 
20:   test_loss:  $\ell_{\text{test}} = \text{Loss}(Y_v, \text{test\_out})$ 
21: end for
22:  $\text{ensemble\_votes} = \text{MajorityVoting}(\text{test\_out})$ 
23:  $\text{acc}_{\text{test}} = \text{Metric}(Y_v, \text{ensemble\_votes})$ 

```



4.3 Implementation and Experimental Results

While the Figshare dataset was used for training, Y-PET/CT and Lung-PET-CT-Dx were kept for testing the model. Data details are provided in Table 1.1. The axial slices of the lung were extracted from the whole body scan, as depicted in Fig. 4.3.

All axial slices, within the lung region, were first extracted from the Figshare data. Next the multimodal fusion methodology of Section 4.2.1 was employed. A subset of such combined images, used for training, is shown in Fig. 4.4. The PET scan identifies both tumor and certain non-tumor regions in the lung, with equal intensity values, based on their metabolic activity. This makes it challenging to manually distinguish between cancer-affected and healthy tissues. When the effect of both CT and PET are combined, then the green bounding boxes (in Fig. 4.4) become indicative of the actual tumor regions; while the red bounding boxes correspond to the incorrectly highlighted regions which are not essentially tumors. Our goal is to locate the tumor region(s), in each of the fused images, through our ensemble DEMF.

Table 4.1 depicts the data distribution for training and testing, with the distribution of cancerous and healthy slices used. We split the data in 4:1

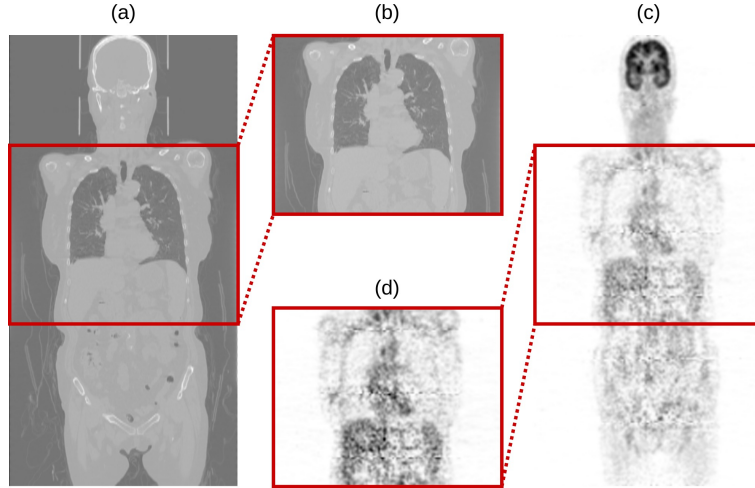


FIGURE 4.3: Extracting lung region from full body scans. CT of (a) full body, (b) lung area, and PET of (c) full body, (d) lung, from coronal view

proportion for training and testing. A 17-fold augmentation was employed for the cancerous slices, as shown in Fig. 4.5, with a 3-fold augmentation being used for the healthy ones. Since the dataset contains fewer cancerous slices than healthy ones, a higher augmentation factor was employed to balance the classes. The learning rate was set to 1×10^{-5} with a decay rate of 10^{-8} . The augmentation procedure involved rotation ± 10 , width shift ± 0.1 , height shift ± 0.1 , zoom ± 0.2 , shear ± 0.2 .

TABLE 4.1: Train-test split of the cancerous & healthy slices of the lung

Dataset	Purpose		Cancerous slices	Healthy slices
Figshare Data	Training data	Total $\Rightarrow$	80	453
		Before Augmentation $\Rightarrow$	64	362
		After Augmentation $\Rightarrow$	1088	1086
	Testing data	$\Rightarrow$	16	91
Y-PET/CT	Testing data	$\Rightarrow$	43	737
Lung-PET-CT-Dx	Testing data	$\Rightarrow$	1659	6893

4.3.1 Classification

The binary cross-entropy was used, involving batch size of 16 with Adam optimizer. Each individual pre-trained model was further trained upto 100 epochs. The metrics used to evaluate the performance of the classifier were Accuracy, F1 score, Precision and Recal. from eqns. (1.9-1.10), (1.13-1.14).

Table 4.2 demonstrates a comparative study of these metrics for evaluating the effectiveness of the ensemble classifier w.r.t. the individual classifier models. It is evident that the ensemble classifier outperformed all individual classifiers.

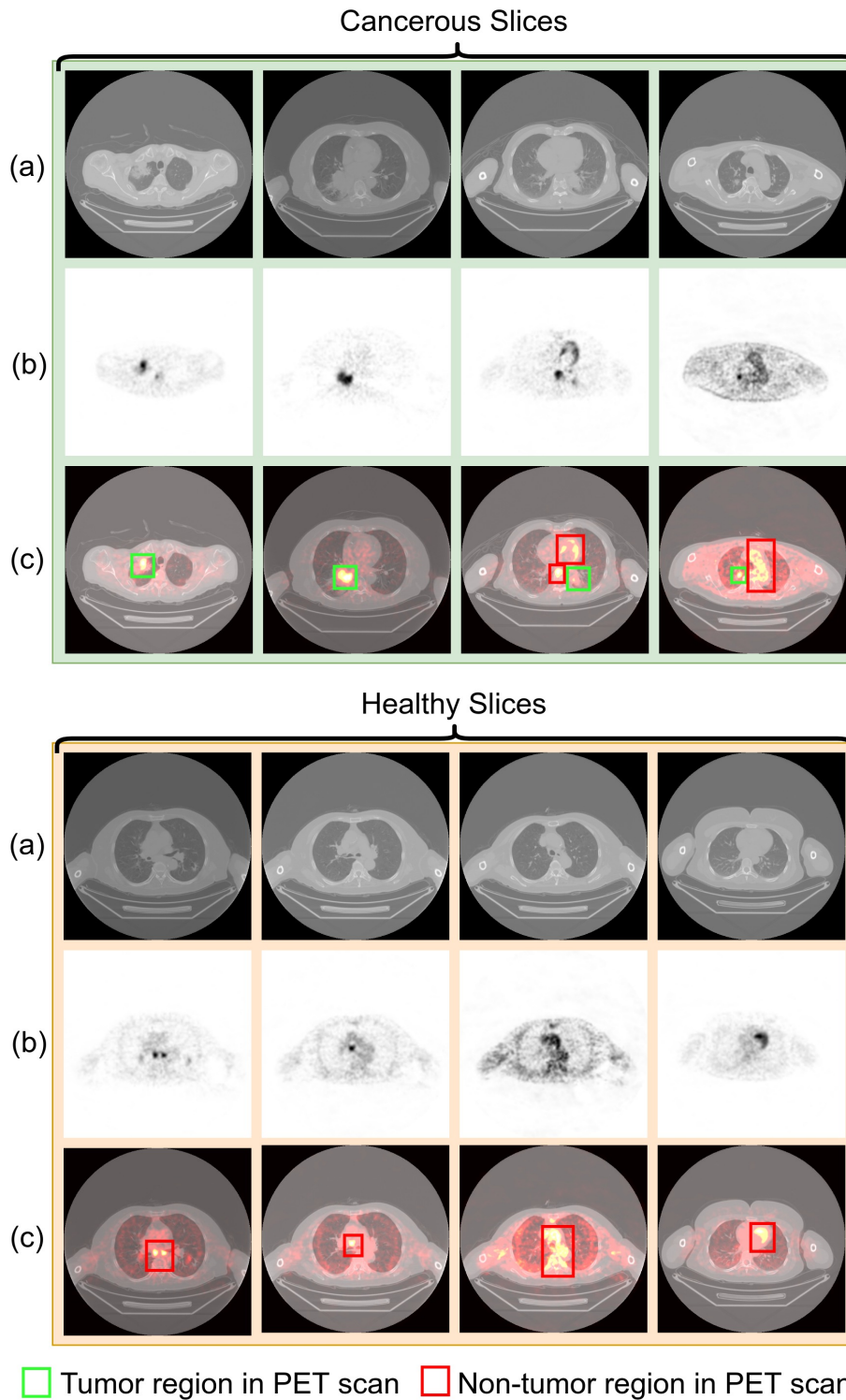


FIGURE 4.4: Training samples from (a) CT, (b) PET, and (c) the corresponding fused images, encompassing cancer-affected sample slices along with healthy ones

4.3.2 Qualitative analysis

Fig 4.6 depicts sample visual attention, as generated through Grad-CAM, from the ensemble classifier output over the test dataset. It demonstrates the robustness of our model performance, in terms of accurately detecting the cancerous

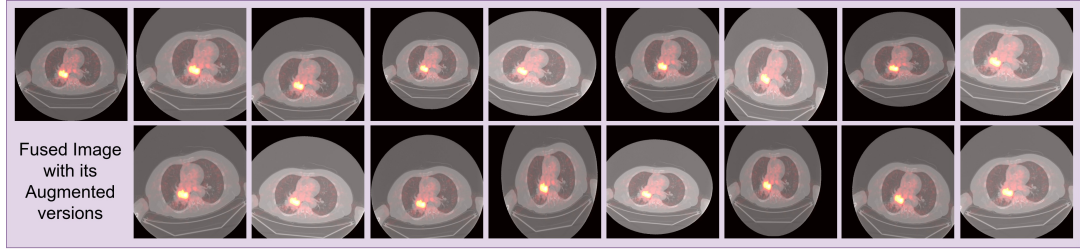


FIGURE 4.5: Sample augmentation of fused training data

TABLE 4.2: Comparative study of the individual classifiers and their ensemble, for the fused multimodal image data

Dataset	Classifier	Metrics			
		Accuracy	F1-Score	Precision	Recall
Figshare Data	VGG 16	0.4766	0.0968	0.0652	0.1875
	ResNet	0.5234	0.1905	0.1277	0.3750
	Inception-Net V3	0.7850	0.3784	0.3334	0.4375
	Xception-Net	0.6262	0.2857	0.2012	0.5003
	Efficient-Net	0.8505	0.5789	0.5070	0.6875
	Ensemble DEMF	0.9813	0.9412	0.8889	1.0000
Y-PET/CT	VGG 16	0.8269	0.3216	0.2051	0.7442
	ResNet	0.7163	0.1905	0.1126	0.6190
	Inception-Net V3	0.9436	0.6000	0.4925	0.7674
	Xception-Net	0.9154	0.5417	0.3861	0.9070
	Efficient-Net	0.9244	0.4957	0.3919	0.6744
	Ensemble DEMF	0.9923	0.9333	0.8936	0.9767
Lung-PET-CT-Dx	VGG 16	0.8239	0.5786	0.5399	0.6233
	ResNet	0.8313	0.6190	0.5508	0.7064
	Inception-Net V3	0.8214	0.6114	0.5291	0.7239
	Xception-Net	0.8736	0.6954	0.6529	0.7438
	Efficient-Net	0.9107	0.7877	0.7308	0.8541
	Ensemble DEMF	0.9760	0.9410	0.9003	0.9855

slices while visualizing the exact location of tumor region. The exact location of tumors are depicted by the green bounding box in the fused image, and by the sky-blue bounding box over the correct prediction in the attention map. The red bounding box, in the fused image, is indicative of incorrect detection by PET. This is appropriately corrected in the CT-PET fused image. The importance of the attention map is evident from its precise identification of tumor locations.

4.3.3 Ablations

Fig. 4.7 illustrates a comparative graphical study of the multimodal fusion strategy PCAE, of our ensemble classifier DEMF, with that of a few other multimodal image fusion techniques over the three test datasets involving CT and PET scans. The fusion strategies compared are GSF [50], DCNN [188], and HF [191]. Fusion GSF employs structural similarity-based gradient filtering to emphasize features in source images, for subsequent reconstruction with fused gradients. The DCNN mechanism uses stacked autoencoders to extract the

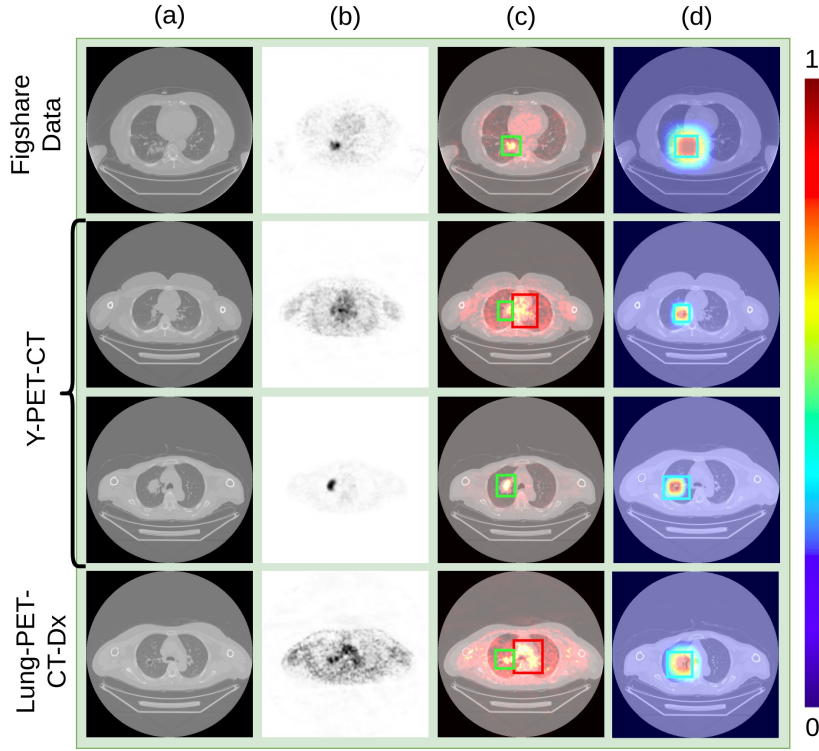


FIGURE 4.6: Visualization with Grad-CAM, over sample test set images from (a) CT, (b) PET, and (c) Fused versions, with the corresponding (d) attention map generated by DEMF

high- and low-frequency components from images, and reintegrates them into the final image. The HF technique uses wavelets to identify optimal parameters of frequencies, for achieving an optimal combination of low- and high-pass filters. Our PCAE fusion strategy, combining PCA and Autoencoder, effectively captures input image patterns and features by reducing dimensionality. It minimizes redundancy, while retaining essential information towards enabling the model to learn hierarchical and abstract features. It helps preserve relevant details and offers flexibility across diverse datasets. The comparative performance of these fusion methodologies is evaluated in terms of output *Accuracy* of individual classifiers like VGG16, ResNet, Inception-Net, Xception-Net, Efficient-Net, as well as the ensemble version DEMF. The fusion result underscores the efficacy of PCAE in improving both image quality and interoperability, thereby leading to enhanced accuracy in classification.

Next the effectiveness of our multimodal image fusion technique was evaluated with respect to each of the individual methodologies, *viz.* PCA and Autoencoder. This is shown in Fig. 4.8, in terms of *Accuracy* by different classifiers, over all three datasets. It is observed that our fusion methodology always resulted in uniformly good performance, over all the classifiers used.

The visualization effect on sample test images is depicted in Fig. 4.9, over the individual imaging modalities *viz.* CT and PET, along with that of the fused image (incorporating both modalities). Results from the classifier DEMF, trained with only CT images, is shown in row (b). It is observed that the prediction is often incorrect, as marked by the red bounding boxes. The classifier,

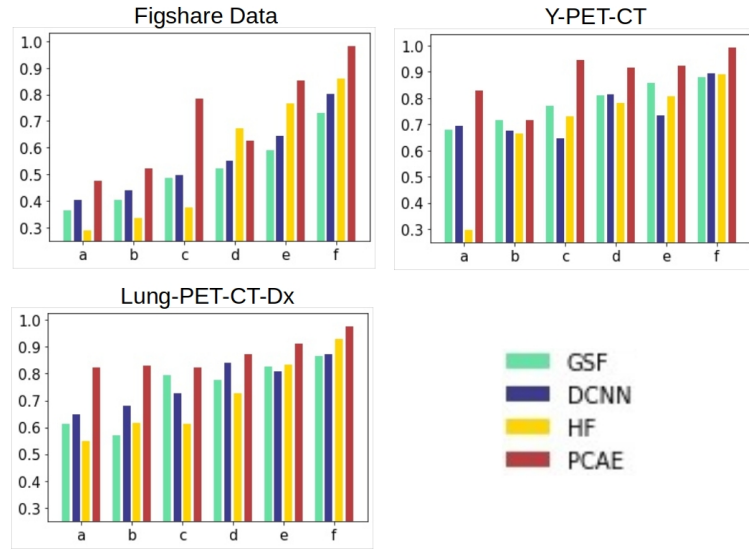


FIGURE 4.7: Comparative study of image fusion strategies, in terms of *Accuracy* over each test set, on the performance of (f) ensemble DEMF and the (a) VGG16, (b) ResNet, (c) Inception-Net, (d) Xception-Net, (e) Efficient-Net.

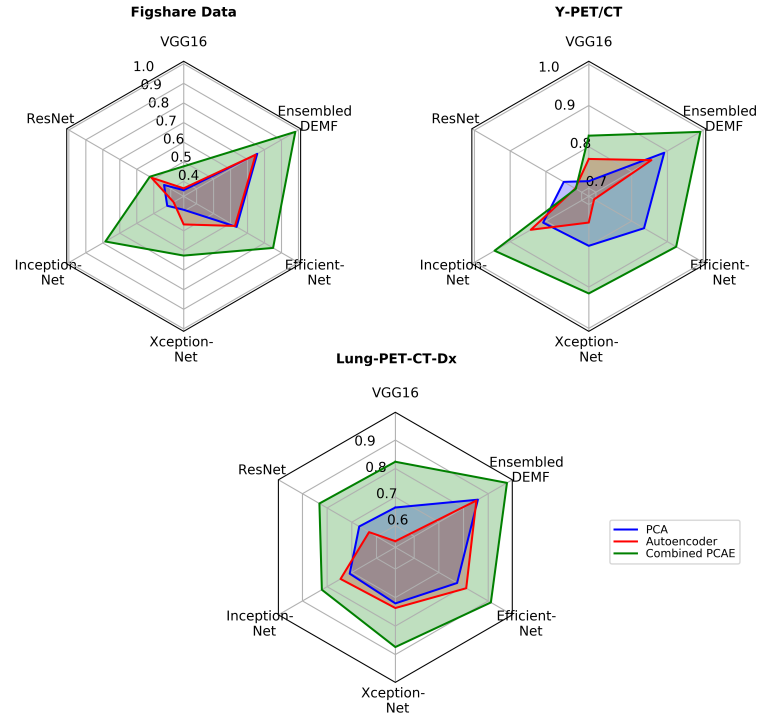


FIGURE 4.8: Ablation of multimodal fusion strategies, in terms of *Accuracy* by different classifiers

trained only over the corresponding PET images, is sometimes found to locate the correct tumor position in row (d) [green bounding box] and otherwise falsely point to incorrect locations [red bounding box]. On the other hand, using the fused images, the classifier is able to correctly locate the actual tumor region [green box] in row (f).

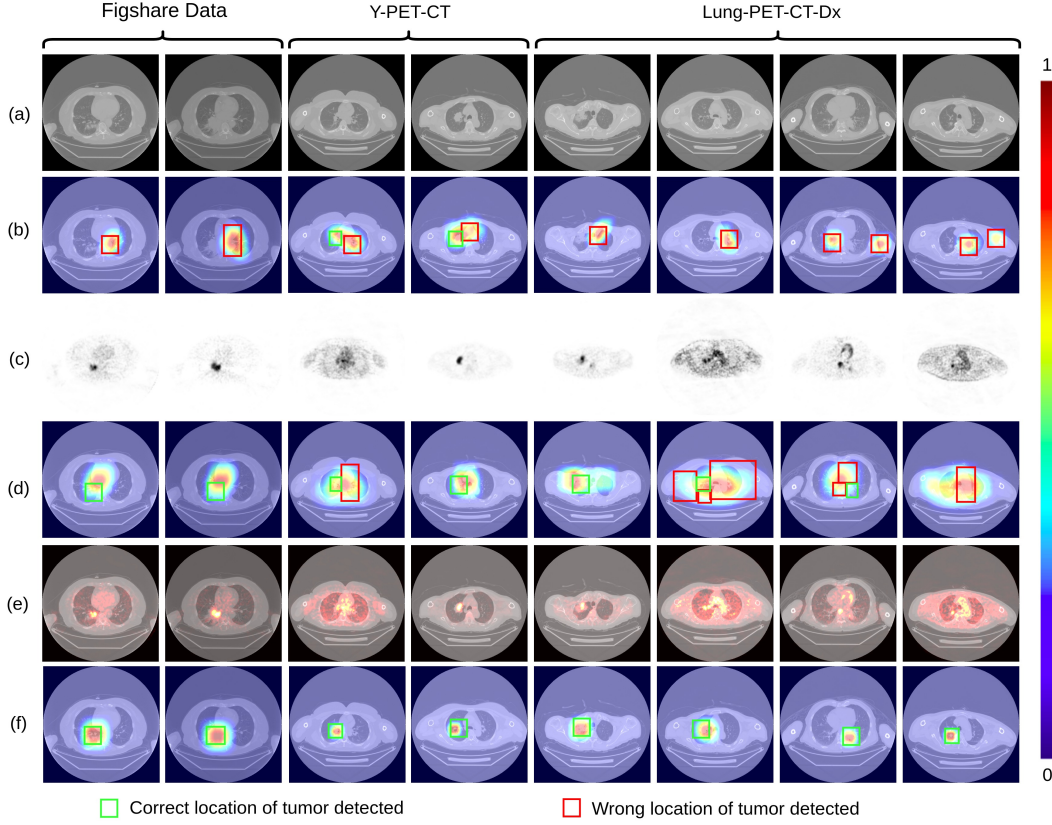


FIGURE 4.9: Visualization over sample test cases, using our DEMF with Grad-CAM, in terms of the (a) CT image, with (b) its visualization; the (c) PET image, with (d) its visualization; and the (e) Fused image, with (f) its visualization

An analogous comparative study of our DEMF, with other related deep classifiers, is included in Fig. 4.10. It explores the effect of the individual imaging modalities, along with that of the multimodal fused image, when provided as input to these classifiers. It is evident that the ensemble classification strategy in DEMF, incorporating the multimodal fusion PCAE, performed the best over all three datasets.



4.4 Conclusions and Discussion

A new deep ensemble multimodal fusion network DEMF was developed, for effectively classifying lung CT slices as cancerous or healthy. The ensemble model combined a number of distinct classifiers using majority voting. The approach was robust and less prone to overfitting, as it relied on a collective decision from diverse models. Majority voting was computationally efficient and simple to implement. It enhanced generalization by reducing the risk of memorizing noise or outliers in the limited training data. Ablation studies and visualization revealed that the ensemble deep multimodal classifier helped emphasize cancerous regions within the images.

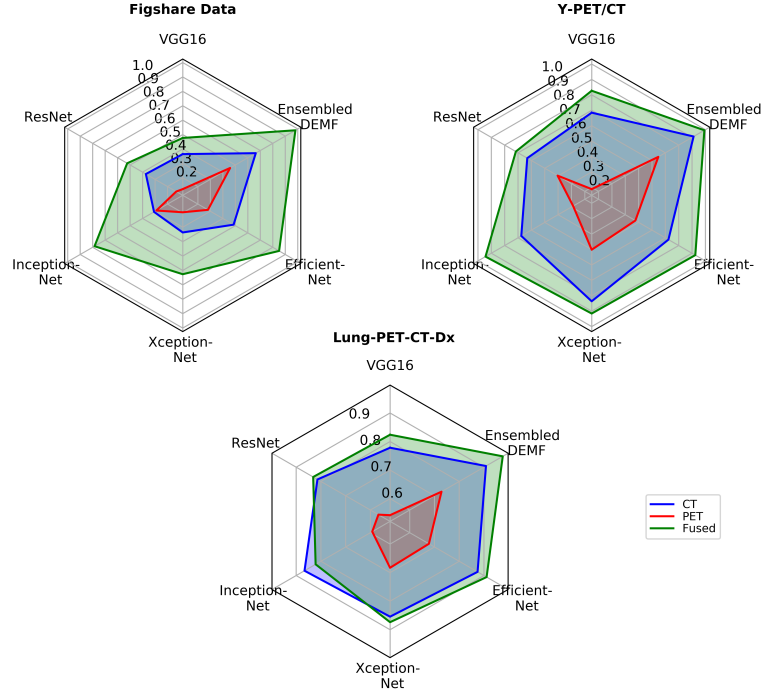


FIGURE 4.10: Effect of the imaging modalities, in terms of *Accuracy* by the classifiers

The fusion strategy PCAE helped reduce dimensionality, involving the PCA and autoencoder components. Leveraging multimodal data with reduced feature cardinality, and ensemble fusion, generated more resilient results. The methodology can be extended to other diseases across different organs, including different imaging modalities. In the following chapter we investigate synthetic generation of images, in the EIT modality, for classification.



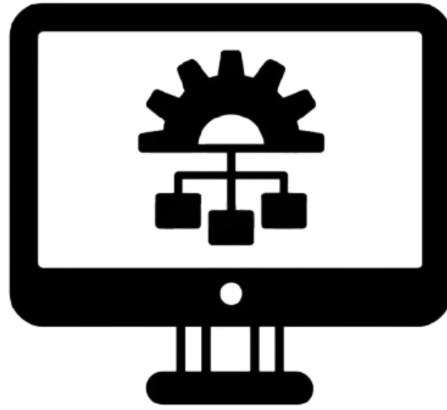


Chapter 5



Multihead Deep Classification of Lung Diseases





5.1 Introduction

In this chapter we move our focus to the multi-class framework of classification using deep networks. The CNN has been employed in the classification of malignant and benign lung tumors from CT scans [149]. A lightweight parallel depth-wise separable CNN, with a ridge regression extreme learning machine, was formulated for the classification of adenocarcinoma, large cell carcinoma, and squamous cell carcinoma [120]. An Attribute-Driven Generative Adversarial Network was developed for the synthesis and multi-class classification of Pulmonary Nodules in CT [145]. A deep learning model integrating convolutional neural networks with Fish Swarm Optimization (CFSO) was employed for the classification of malignant lung tumors from CT images, aiming to reduce misclassification and computational overhead through metaheuristic optimization [162]. The UNet3+ architecture was employed for CT lung image segmentation, and its integration with SE-ResNeXt50 and ResNet50 was proposed to enhance COVID-19-infected region identification from medical data [121]. A 13-layer convolutional network, comprising 10 encoder layers and 3 decoder dense layers, was proposed for binary classification of COVID-19 and normal cases from CT images [85]. A hybrid classification optimization model combining recurrent learning and fuzzy logic (HCO-RLF) was developed to identify infected lung regions in CT scans by capturing pixel-wise feature variations and compensating for missing information through fuzzy-based derivative substitution [15]. A privacy-preserving self-supervised learning framework was proposed using federated learning with Paillier homomorphic encryption, where BYOL-based contrastive learning with a VGG19 encoder and attention CNN blocks was applied for COVID-19 classification from lung CT scans [30].

A way of overcoming the scarcity and unavailability of data in the medical domain is the synthetic generation of images in other modalities, in order to complement the hard-to-obtain annotated authentic medical data. The benefits lie in circumventing the time-consuming task of manual labelling and eliminating the long process of obtaining consent while preserving privacy. GANs are employed for data generation and augmentation, in data-scarce scenarios. This

is very typical in medical decision-making [169]. The generator produces realistic synthetic data, while the discriminator distinguishes between real and generated data, resulting in high-quality samples. The GAN framework involves simultaneous minimization (maximization) of generative (discriminative) loss to reach equilibrium. While adversarial techniques typically rely on conventional augmentation operations like rotation, flipping, etc., these cannot capture the nuances of different imaging modalities.

Here we present the novel Deep Generative Multihead Classifier (*DGMC*) [111], for efficient detection and distinction between certain diseases of the lung from CT scans. Augmentation by the CTGAN overcomes the issue of scarcity of data. Subsequently, EIT-like visual images are generated from the CT slices to introduce an alternative perspective and generate a processed version of CT. Standard CT scans often present both healthy and diseased tissues with similar intensities, making it difficult for models to clearly distinguish abnormal regions. To address this limitation, we generate the EIT-processed version of the CT that emphasizes the anomalous regions of the lung while suppressing the healthy areas. The primary motivation behind creating this enhanced view is to improve the visibility of disease-specific patterns, thereby enabling the model to focus more precisely on clinically relevant regions and make more accurate classification decisions.

Despite the widespread use of CT imaging for lung disease diagnosis, existing automated methods often face challenges in accurately distinguishing abnormal regions due to visual similarity between healthy and diseased tissues. Prior multimodal approaches have attempted to improve performance by combining data from different imaging modalities; however, these often rely on weakly aligned data, generic pre-trained models, or simplistic fusion strategies that may not effectively capture complementary features. To address these limitations, this study aims to develop a robust deep learning framework that enhances CT data using a processed view inspired by EIT, and integrates both original and enhanced representations through a novel multihead classifier. The main focus is to improve disease localization and classification accuracy while offering interpretable visual cues, thereby bridging the gap between technical innovation and clinical applicability.

Experimental results from lung cancer and COVID-19 infections, involving CT slices from publicly available datasets, are used to demonstrate the effectiveness of our model in terms of ablation studies and visualization. The contribution of this research is summarized as follows [111].

- (i) Generation of CT images through CTGAN, followed by transformation to corresponding synthetic EIT images in an augmented framework.
- (ii) Design of a novel multihead classifier for efficient classification of COVID-19 infected, lung tumour affected, and healthy slices from input images.
- (iii) Learning from a combination of CT and EIT-processed CT perspectives, to capture granular details of the infection.
- (iv) Efficient visualization of activation regions (i.e., the regions that contribute to the model's decision-making) in enhanced image framework.

- (v) Ablation and comparative studies to demonstrate the effectiveness of our *DGMC*, over seven different publicly available CT datasets.

The rest of the chapter is organized as follows. Section 5.2 presents the framework of the Deep Generative Multihead Classifier (*DGMC*), along with its component modules – encompassing generation of synthetic CT images (*CTGAN*), creation of corresponding cross-modal EIT images, and multihead classification. Qualitative and quantitative experimental results on seven publicly available datasets, in Section 5.3, help establish the effectiveness of the implementation along with comparison with related state-of-the-art models in literature. Finally, Section 5.4 concludes the chapter.

5.2 Methodology

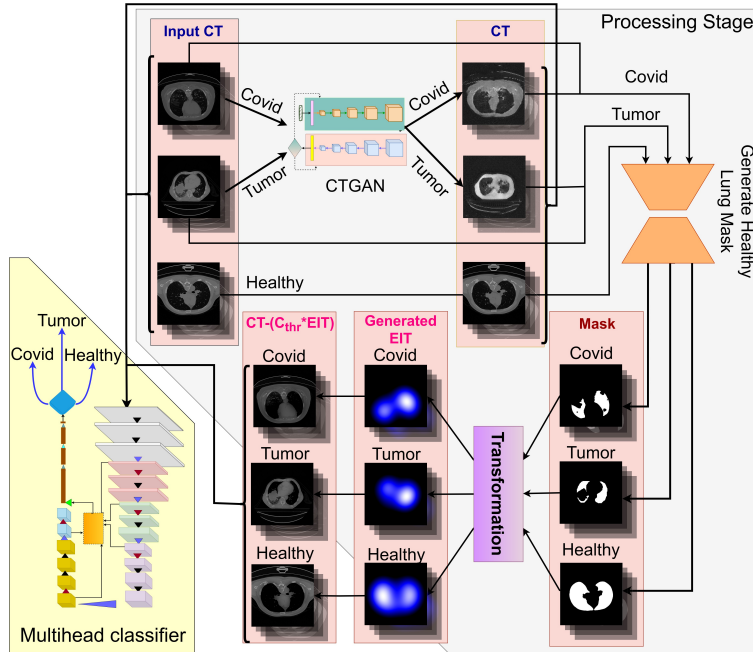


FIGURE 5.1: The workflow of *DGMC*

The overview of the Deep Generative Multihead Classifier (*DGMC*) model is depicted in Fig. 5.1. The input consists of extracted CT scan slices from the lung, containing COVID-19 infection and/or tumor-affected portions, as well as the unaffected ones (termed Healthy or control). The COVID-19 and Tumor slices are augmented by *CTGAN*, and (together with the control slices) get passed through an encoder-decoder network to segment out the non-infected part of the lung therein. The delineation is necessary for preparing the masks, such that the infected regions in the CT slices may be subsequently reconstructed in the corresponding EIT modality (with higher impedance, as indicated by black). This transformation aspect is novel to the *DGMC*, and helps to generate cross-modal information from a different perspective to circumvent the issue of scarcity in annotated data. Finally the entire set of augmented CT

slices, along with the corresponding synthetic EIT slices, and the processed CT scans are partitioned into the three categories [COVID-19, Tumor, and Healthy or control] using the new multihead classifier module of *DGMC*.

5.2.1 Augmentation of CT samples

The *CTGAN* of Fig. 5.2 is designed to generate high resolution synthetic CT images, involving spatial dimensions of 512×512 . Strided transpose convolution [55] layers in the Generator upscale input noise, while strided convolution [55] layers in the Discriminator effectively distinguish real and generated images at higher resolution. This seamless synthesis of high-resolution CT images faithfully generates intricate details, elevating the quality with remarkable realism and accuracy.

Convolution, key to deep learning, processes images and structured data by sliding a filter over input, multiplying elements, and summing, for extracting features. Utilizing strided transpose convolution (instead of conventional fully connected layers) improves the retention of spatial structure, crucial for image generation. The stride parameter determines filter shifts across the input, reducing spatial dimensions during convolution while preserving fine-grained spatial features. Transpose convolution is flexible across varying input sizes, making it adaptable to different data dimensions. Shared weights in convolution filters reduce overfitting and computational demands. Local feature detection gets enhanced for capturing the fine-grained details of image data. Additionally, it offers interpretability through hierarchical feature representation and implicit regularization.

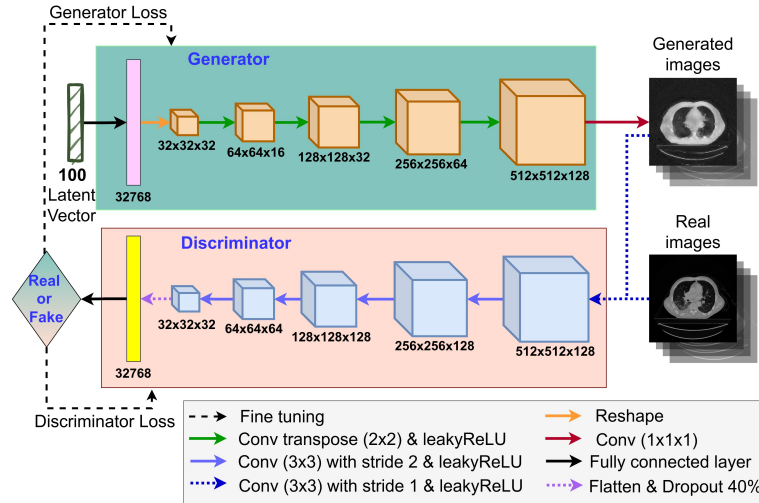


FIGURE 5.2: The *CTGAN* architecture

Handling higher spatial dimensions in images captures fine-grained details, crucial for detecting smaller features in tumor or COVID-19 affected CT slices. Using larger images enhances model accuracy by providing greater visual information, mitigating information loss during downsampling. This facilitates involvement of deeper layers, potentially learning more abstract and hierarchical features.

Given an input tensor X of dimension (H, W, C) , with H, W, C , representing the height, width, and number of input channels, respectively, let us assume a convolution kernel K with dimension (k, k) , padding size (b, b) to control the spatial dimensions of the output feature map, for stride (s, s) . The output tensor Z , for each output location (i, j, c_{out}) , is expressed as

$$Z(i, j, c_{out}) = \sum_{c_{in}=1}^C \sum_{p=0}^{k-1} \sum_{q=0}^{k-1} X(\overline{i \cdot s - b + p}, \overline{j \cdot s - b + q}, c_{in}) \cdot K(p, q), \quad (5.1)$$

in strided transpose convolution. The triple summation ranges over the spatial dimensions (p, q) of K and the input channels $c_{in} = \{1, \dots, C\}$ that the filter operates on. The $c_{out} = \{1, \dots, C'\}$ refer to the output channels. The stride (s, s) is applied as part of the indexing of the input tensor X , for effectively expanding the dimensions of the output tensor. Here $X(\overline{i \cdot s - b + p}, \overline{j \cdot s - b + q}, c_{in})$ represents the values in the input tensor X at specific locations based on the current output location (i, j, c_{out}) , the kernel dimensions, and the stride. The operation computes a weighted sum of values from the input tensor X , using K at each output location, and results in the transposed convolutional output tensor Z of dimension (H', W', C') .

Analogously, for strided convolution (involving no padding), we have the output tensor Y for each output location (i, j, c_{out}) as

$$Y(i, j, c_{out}) = \sum_{c_{in}=1}^C \sum_{p=0}^{k-1} \sum_{q=0}^{k-1} X(\overline{i \cdot s + p}, \overline{j \cdot s + q}, c_{in}) \cdot K(p, q). \quad (5.2)$$

Generator architecture

The generator takes random noise as input and transforms it into an image. The CTGAN consists of several layers of strided transposed convolution layers, each followed by a LeakyReLU activation function. The architecture can be summarized as follows.

1. Input latent vector: A vector of random noise $z \in \mathbb{R}^{100}$.
2. Fully connected layer: $FC(z) \rightarrow \mathbb{R}^{32 \times 32 \times 32}$ maps the noise vector to a higher-dimensional representation.
3. Reshape the output of the fully connected layer: $Reshape(FC(z)) \rightarrow \mathbb{R}^{32 \times 32 \times 32}$.
4. Strided transpose convolution: Perform $G_{out} = Z$ by eqn. (5.1), with stride $s = 2$, padding $b = 2$ and kernel size $k = 4$; followed by LeakyReLU activation, four times.

Discriminator architecture

The discriminator takes an image as input and tries to classify it as real (original) or fake (generated). In CTGAN this comprises several strided convolution

layers, each followed by a LeakyReLU activation function. The architecture is summarized as follows.

1. Input image: $x \in \mathbb{R}^{512 \times 512 \times 1}$.
2. Strided convolution: Perform $G_{out} = Y$ by eqn. (5.2), with stride $s = 2$ and kernel size $k = 4$; followed by LeakyReLU activation, four times.
3. Flattened layer: D_{out} is a feature map from the last convolution layer with dimension $H \times W \times \mathcal{C}$, where H is the height, W is the width, and $\mathcal{C}$ is the number of channels.
4. Fully connected layer: $FC = FC(Flattened_layer)$ maps the output of the convolutional layers to a single scalar value, representing the likelihood that the input is real.

Loss functions

The generative loss L_G and discriminative loss L_D [161] are used to train the generator and discriminator, respectively. While the generative loss measures how well the generator can fool the discriminator, the discriminative loss evaluates how well the discriminator can distinguish between the real and generated samples.

The generative loss aims to encourage the generator to produce realistic samples that can deceive the discriminator. It is formulated as a minimax between the generator and the discriminator, expressed as

$$L_G = -\mathbb{E}[\log(D\{G(z)\})], \quad (5.3)$$

where z represents the random noise vector input to the generator, $\{G(z)\}$ is the generated sample, $D\{G(z)\}$ is the discriminator output for the generated sample, and $\mathbb{E}$ denotes the expectation over the random noise distribution. The generator tries to minimize this loss by generating samples that the discriminator cannot distinguish with respect to the real samples.

The discriminative loss, on the other hand, measures how well the discriminator can differentiate between the real and generated samples. It aims to maximize the discrimination capability of the discriminator, and is typically defined as

$$L_D = -\mathbb{E}[\log\{D(x)\}] - \mathbb{E}[\log(1 - D\{G(z)\})]. \quad (5.4)$$

Here x represents a real sample, $D(x)$ is the discriminator output for the real sample, and $1 - D\{G(z)\}$ is the discriminator output corresponding to the generated sample. The discriminator tries to maximize this loss by correctly classifying between the real and generated samples.

5.2.2 EIT for processed CT

It is known that lung function changes could be visualized through electrical impedance differences within the lungs [22]. EIT primarily images regions of

the lung with normal ventilation to assess lung function [200]. It helps to visualize ventilation abnormalities caused by diseases, thereby aiding in diagnosis and monitoring of respiratory conditions; especially when combined with other modalities. Given the scarcity of real EIT images, it was observed that most deep learning studies typically generated phantom data for evaluating their algorithms and/or models [83]. We, on the other hand, generate EIT-like visual slices from the corresponding available (augmented) CT slices towards gathering cross-modal information.

First the CT lung scans are segmented to eliminate the diseased region(s), such that the remaining healthy portion (in each slice) serves as the mask for the transformation. This is done with the help of the attention *U*-Net [124]. Note that the impedance in EIT is high in the anomalous region (here the COVID-19 infected or tumor affected), and low in the healthy regions [90]. This serves to generate masks corresponding to healthy parts of each slice.

The segmented healthy regions of the CT scans are convolved to generate the corresponding EIT maps, as depicted in Fig. 5.3. Here the *black* $\rightarrow$ *blue* $\rightarrow$ *white* shades are indicative of *high* $\rightarrow$ *mid* $\rightarrow$ *low* impedance of the output EIT images. While a high impedance implies a low ventilated (i.e. anomalous/infected) region, a low impedance is indicative of appropriately ventilated (i.e. healthy) region. The convolution kernel K , defined in Eq. 5.5, is a combination of a Gaussian-like decay and a local averaging operation. Parameters $\alpha = 4$, $n = 3$, and $m = 5$ were chosen based on empirical tuning. However, to validate their robustness, a sensitivity analysis was conducted by varying each parameter within a plausible range and evaluating the model's classification accuracy and EIT quality. The model showed consistent performance for $3 \leq \alpha \leq 5$, $2 \leq n \leq 4$, and $4 \leq m \leq 6$, confirming the stability of the chosen values.

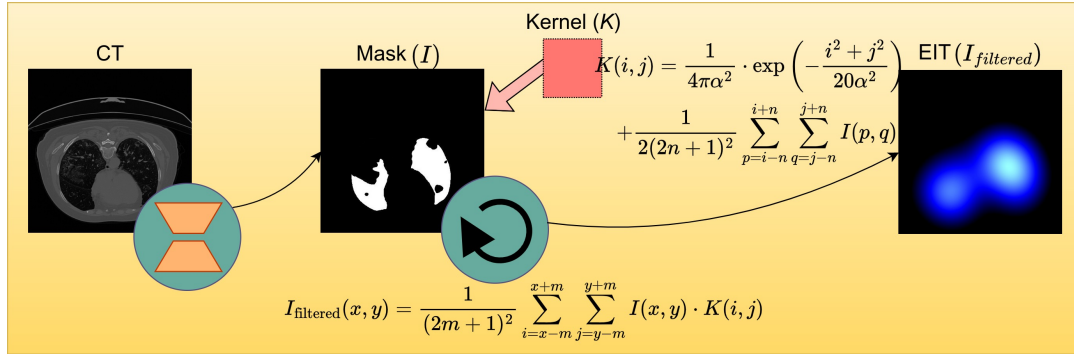


FIGURE 5.3: Mapping between modalities, from extracted mask I in CT to I_{filtered} in EIT

Let the extracted mask from the input CT image be denoted as I , and the output transformed EIT image be I_{filtered} . The processing filter can be represented as the convolution of the extracted mask I with a kernel K , where K is defined as

$$K(i, j) = \frac{1}{4\pi\alpha^2} \cdot \exp\left(-\frac{i^2 + j^2}{20\alpha^2}\right) + \frac{1}{2(2n+1)^2} \sum_{p=i-n}^{i+n} \sum_{q=j-n}^{j+n} I(p, q). \quad (5.5)$$

Here i and j represent the spatial coordinates, $I(p, q)$ is the value of pixel in extracted mask I at location p, q , with α and n being the hyper-parameters. We selected $\alpha = 4$ and $n = 3$, after several experiments on train data.

The convolution operation between the extracted mask I and the kernel K at a specific pixel (x, y) in the output filtered image I_{filtered} is expressed as

$$I_{\text{filtered}}(x, y) = \frac{1}{(2m+1)^2} \sum_{i=x-m}^{x+m} \sum_{j=y-m}^{y+m} I(x, y) \cdot K(i, j), \quad (5.6)$$

where i and j represent the spatial coordinates of the kernel center, and m is a hyper-parameter with value 5.

Next the generated EIT is used to obtain corresponding EIT-processed CT scans. Subtracting the EIT pixel values, with a threshold value of 0.5 throughout, helped to reduce the intensity of the healthy regions in the processed CT scans. This enabled focus on the anomalous or infectious regions, which got characterized by their higher impedance. Fig. 5.4 illustrates relevant details.

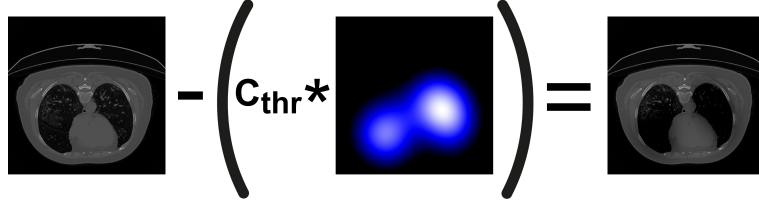


FIGURE 5.4: Formation of processed CT, with $C_{thr} = 0.5$

5.2.3 Multihead classification

The set of 512×512 CT and corresponding EIT-processed CT scans are supplied at the input of the classifier, along two channels. The classifier architecture is illustrated in Fig. 5.5. The first block from each of the layers 2 to 5, and last block from layer 6 are passed through the Multihead gradient computing module to generate a flattened tensor. Finally a ‘softmax’ activation determines the class of the input image.

The Multihead classifier offers several distinctive advantages. Firstly, it exhibits the ability to simultaneously process a pair of image scans within a single architecture. Such innovative handling of input allows the model to leverage information from both images, potentially enhancing classification accuracy. Including the Multihead gradient module sets the classifier apart. It enhances gradient information flow, by selectively processing feature maps from specific layers, and capturing intricate data patterns at various scales. Additionally, employing softmax activation emphasizes its role in categorizing input scans into specific classes, especially in medical imaging for tasks like disease diagnosis. Overall, the novelty of this classifier extends to its domain of application; *viz.* medical imaging, where it augments the data from different perspectives, for improved diagnostic accuracy.

Let the input feature map X be of size $H \times W \times C$, where H represents height, W corresponds to width, and C is the number of input channels. The

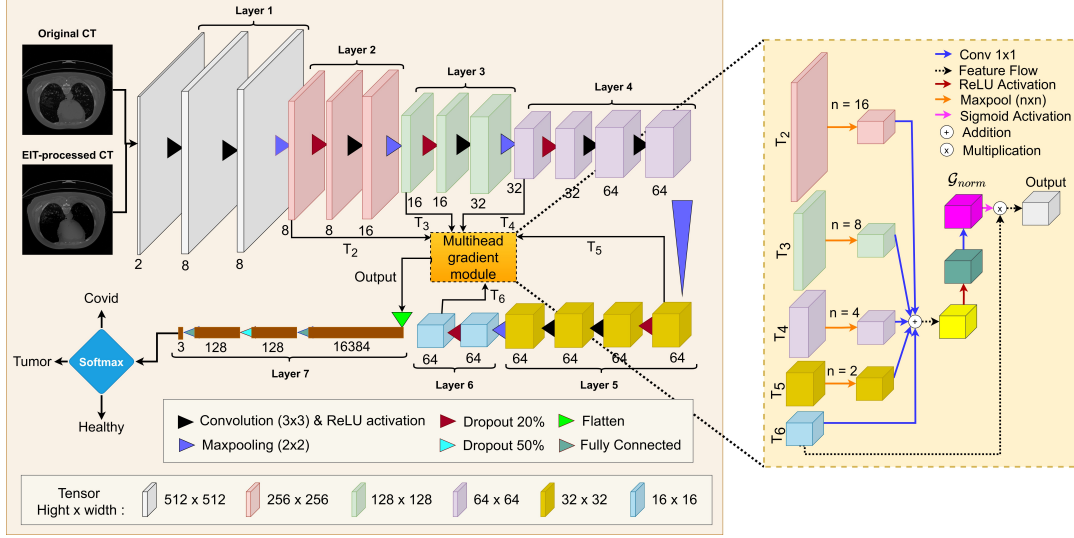


FIGURE 5.5: Multihead classifier

convolution operation (kernel size $K \times K$) produces an output feature map $\mathcal{F}$ of size $H \times W \times M$, with M indicating the number of output channels. Let $\omega = \{\omega^i \mid 1 \leq i \leq M; \omega^i \in (K \times K)\}$ be the set of learned kernel weights of size $K \times K \times M$. The output feature $\mathcal{F}$, corresponding to pixel location (i, j) and output channel m , is computed as

$$\mathcal{F}_{i,j}^m = \sum_{c=1}^C \sum_{p=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \sum_{q=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \omega_{p,q}^m X_{\{i+p\}, \{j+q\}}^c, \quad (5.7)$$

where $c = \{1, \dots, C\}$ refers to the input channels, $m = \{1, \dots, M\}$ corresponds to the output channels, and K is assumed to be odd.

Multihead gradient module

This module plays a crucial role in enhancing deep neural network for training, combining information across layers for a holistic data view. It introduces non-linearity with ReLU activation, stabilizes training with sigmoid normalization, and focuses on relevant features, for enhancing discrimination. Efficient operations minimize computational overhead while customizing gradients to tasks, improving gradient flow and feature learning; thereby, making the deep model effective across diverse tasks.

The module is introduced for combining gradients from different layers of the classifier. Let $\mathbf{T} = \{T_i \mid 2 \leq i \leq 6\}$, where T_2, T_3, T_4, T_5 and T_6 correspond to the first feature map blocks from layers 2, 3, 4, 5 and 6, respectively. The multihead gradient module initially performs $M^{T_i} = \text{MaxPool}_{n \times n}(T_i)$, where $n = \{2^{(6-i)} \mid 2 \leq i \leq 5\}$ and $\text{MaxPool}_{n \times n}$ denotes $n \times n$ maxpooling operations. Here maxpooling achieves same spatial resolution for all feature maps from different levels.

Let $\mathbf{M}^T = \{M^{T_i} \mid 2 \leq i \leq 5\}$. Now $\mathbf{M}^T$ and T_6 are passed through the convolution operation of eqn. (5.7). We have $\mathcal{C}^{\mathbf{M}^T} = \text{Conv}_{1 \times 1}(\mathbf{M}^T)$ and

$\mathcal{C}^{T_6} = \text{Conv}_{1 \times 1}(T_6)$, where $\text{Conv}_{1 \times 1}$ denotes the 1×1 convolution. Thereafter, an elementwise addition of each feature is performed. The gradient matrix is computed as

$$\begin{aligned}\mathcal{G} &= \text{Conv}_{1 \times 1 \times 1}\{\text{ReLU}(\mathcal{C}^{M^T} \oplus \mathcal{C}^{T_6})\} \\ &= \text{Conv}_{1 \times 1 \times 1}\{\text{ReLU}(\mathcal{C}^{M^{T_2}} \oplus \mathcal{C}^{M^{T_3}} \oplus \mathcal{C}^{M^{T_4}} \oplus \mathcal{C}^{M^{T_5}} \oplus \mathcal{C}^{T_6})\},\end{aligned}\quad (5.8)$$

where $\oplus$ is element-wise addition, as passed through *ReLU* activation, and $\text{Conv}_{1 \times 1 \times 1}$ denotes a $1 \times 1 \times 1$ convolution. The normalized gradient matrix becomes

$$\mathcal{G}_{\text{norm}} = \sigma(\mathcal{G}), \quad (5.9)$$

with a sigmoid operation (σ) being applied along the spatial dimensions. The final gradient $\mathcal{G}_{\text{final}}$ is computed by element-wise multiplication $\otimes$ of the normalized gradient matrix with the feature map T_6 . Therefore

$$\mathcal{G}_{\text{final}} = \mathcal{G}_{\text{norm}} \otimes T_6. \quad (5.10)$$

The module combines and normalizes feature maps from various layers, generating a final gradient for training or optimizing the neural network. This enhances gradient information flow across layers in the deep neural network.

***n*-ary classification**

Here we consider three classes of lung scan images, *viz.* COVID-19 infected, Tumor affected and Healthy controls. Given the output vector $h = [h_1, h_2, \dots, h_n]$ from the final fully connected layer (output of layer 7), the softmax activation calculates the predicted probabilities $\hat{y}_k$ for each class $k \in n$ as

$$\hat{y}_k = \frac{e^{h_k}}{\sum_{j=1}^n e^{h_j}}. \quad (5.11)$$

Here the denominator $\sum_{j=1}^n e^{h_j}$ ensures that predicted probabilities sum up to 1, making them valid probabilities for multiclass classification. Generally, e^{h_k} is a component of the softmax function that helps convert raw scores (logits) from the final layer into probabilities for each class.

The categorical cross-entropy loss is expressed as [198]

$$L = - \sum_{k=1}^n y_k \cdot \log(\hat{y}_k), \quad (5.12)$$

where y_k is the class value (ground truth) of the input belonging to class k .



5.3 Implementation and Results

The Deep Generative Multihead Classifier (*DGMC*) was modeled on publicly available annotated CT datasets, comprised of three from COVID-19 and four

from lung tumors, as elaborated in Table 1.1. In this section we describe the dataset preparation, followed by sample results from CT augmentation, synthetic EIT image translation to generate processed CT, and multihead classification. Ablation studies help to establish the effectiveness of our model, both qualitatively and quantitatively. Comparative studies, with related state-of-the-art models in literature, demonstrate the superiority of *DGMC*.

5.3.1 Data preparation

For **COVID-19** the Kaggle-COVID-19:Part-1 dataset was used for training. The Kaggle-COVID-19:Part-2, CT-Seg and Seg-nr.2 were kept for testing. In case of **Lung tumor**, the NSCLC-radiomics was used for training. The NSCLC-radiogenomics, Decathlon and Moffitt data were selected for testing. The unseen test datasets were used only for evaluating the generalization performance of model *DGMC*.

The separate training datasets, for COVID-19 and lung tumor, were kept together for training. Initially, all axial slices were extracted from the combined dataset. Based on the presence of COVID-19 or tumor infliction, the slices were categorized; while the remaining slices were labeled as healthy. These were supplied as input to the processing stage of the model in Fig. 5.1. The objective of the model *DGMC* is to classify between the three categories of COVID-19 infected, lung tumor affected and healthy CT scan images of lung.

5.3.2 Augmentation by CTGAN

It is observed from Table 1.1 that the training set comprises of 1351 COVID-19-infected slices, 4649 tumor-affected slices, and $(1230 + 26240 =)$ 27470 Healthy slices. The huge class imbalance between the classes COVID-19, Tumor and Healthy is clearly evident, with a distribution ratio $\approx 1 : 3 : 20$. Therefore it becomes imperative to augment the COVID-19-infected and Tumor-affected slices using the CTGAN developed in Fig. 5.2. A total of 3649 COVID-19 images, and 351 tumor images, were thus generated. A random selection of 5000 Healthy scans were chosen from the existing total of 27470 images. Thereby, the augmented set of training scans now consisted of COVID-19: $(1351 + 3649 =)$ 5000, Tumor: $(4649 + 351 =)$ 5000, and Healthy: 5000 images. The class imbalance was thus reduced. The CTGAN was trained separately for the three classes, using the Adam optimizer with a learning rate 0.0002, and momentum of 0.5. The size of the mini batches were 64, and trained for 120 (90) epochs for the classes COVID-19 (tumor), respectively.

Fig. 5.6 depicts sample CT images (obtained from the databases) belonging to these three classes, along with a few of the augmented images generated by CTGAN. The high resolution of the generated images add to their authenticity. The infected regions are marked by a rectangle in each case.

5.3.3 Generating EIT-processed CT scans

Fig. 5.7 illustrates the transformation of CT images to their corresponding EIT-like versions (as in Fig. 5.3), along with the corresponding processed CT

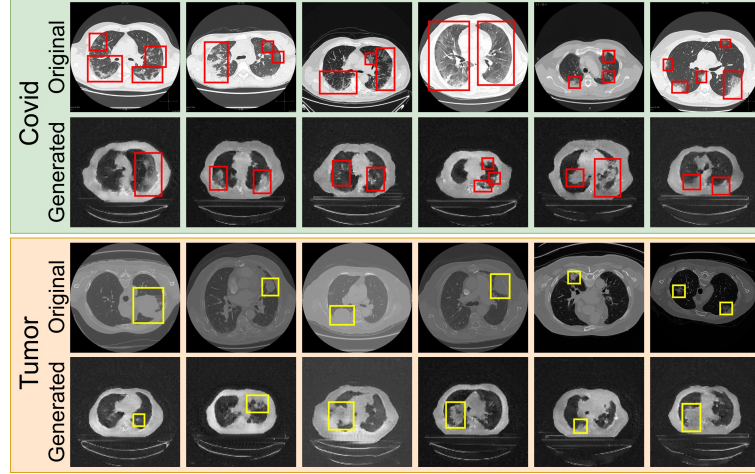


FIGURE 5.6: Sample augmented COVID-19-infected and tumor-affected CT slices, as generated by CTGAN

scans, for all the three classes involved. The infected regions in the CT scans are manually marked by bounding boxes (for better understandability of readers). The legend at the bottom of the figure depicts the color bar, demonstrating the impedance of the respective regions in the transformed EIT image space.

Subtracting the synthetic EIT images from their corresponding CT versions provides a remarkable advantage in output interpretation, potentially enhancing the feature space and offering an additional perspective. This allows the end user to visualize the effect of EIT impedance on CT images, especially in the infected regions of the lung. For example, the third row of Fig. 5.7(c) depicts higher impedance of EIT in the tumor-affected region of the lung. Analogously, the second row of Fig. 5.7(c) illustrates high to medium impedance of EIT on the right side of the lung. This is because most of the right lung is affected by COVID-19. The left lung, on the other hand, shows infection only in the upper part. In general, it becomes easier to distinguish between the healthy (lower impedance) and the diseased (medium to higher impedance) regions of the lung in the EIT version.

5.3.4 Multihead classification

The classifier is trained using the categorical cross entropy of eqn. (5.12), over 100 epochs, with the rmsprop optimizer. The evaluation metrics include *AUC*, *Accuracy*, *Macro – F1* score, and *mAP* of eqns. (1.11), (1.13) - (1.16).

Table 5.1 summarizes a comparative study of the classification generated by our model *DGMC* with respect to other related state-of-the-art deep models in literature (like VGG [158], RA-MDIN [175], HyEx-Net [60], ResNet-50 [64], SE-ResNeXt-50 [140], DenseNet-121 [70], MobileNet-V3 [68]). The input space is extended to uniformly include CT scans, CTGAN-generated augmented images, along with the corresponding processed CT scan generated from the reconstructed synthetic EIT versions, over the six different publicly available COVID-19 and Tumor test data (as outlined in Table 1.1). The generalization performance of model *DGMC* is found to be always the best in terms

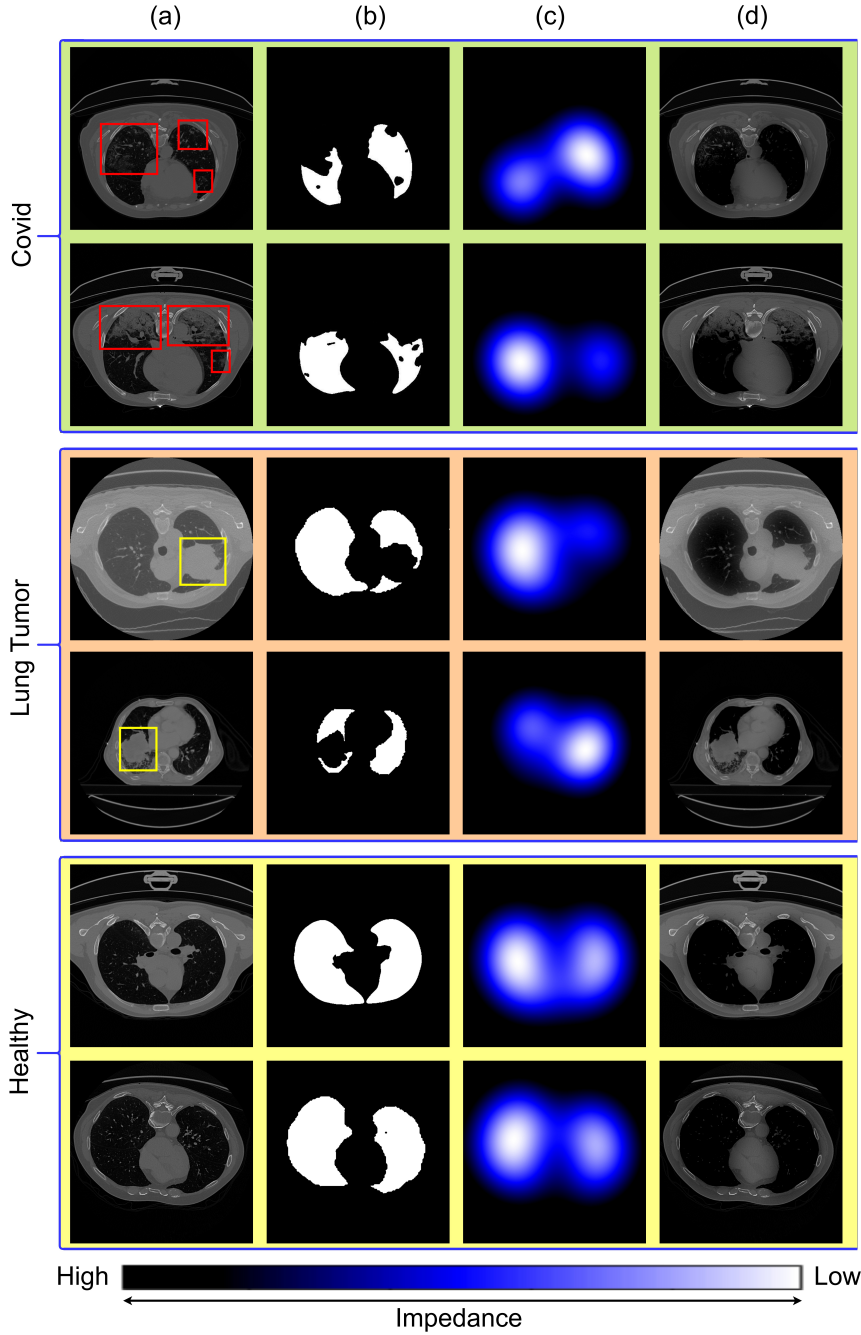


FIGURE 5.7: (a) Sample CT images, with (b) portion of Healthy Lung (regions without infection), (c) the corresponding synthetic EIT version, and (d) the EIT-processed CT image, for all the three classes

of the evaluation metrics defined in eqns. (1.13)-(1.16). The best values are marked in bold in the table. The comparative analysis in Table 5.1 highlights the trade-off between model complexity and classification performance across various deep learning classifiers. VGG16, with the highest parameter count of 568.38M, achieves moderate performance, indicating inefficiency in parameter utilization. In contrast, the proposed *DGMC* network attains the best performance across all datasets with only 2.23M parameters, showcasing superior parameter efficiency. RA-MDIN (35.45M) and HyEx-Net (24.14M) also deliver

TABLE 5.1: Comparative performance of different classifiers, in terms of the evaluation metrics and network size

		Dataset	Metrics			
			Accuracy	AUC	Macro F1	mAP
Classifier (# Parameters)	VGG 16 [158] (568.38 M)	Kaggle-Covid-19: Part 2	0.7721	0.7727	0.7719	0.6891
		CT-Seg	0.9200	0.9600	0.5792	0.1111
		Seg-nr-2	0.8552	0.8477	0.8514	0.8114
		NSCLC radiogenomics	0.9006	0.9087	0.7944	0.9811
		Decathlon	0.8586	0.8270	0.7143	0.9660
		Moffitt	0.9276	0.9329	0.9516	0.9515
	RA-MDIN [175] (35.45 M)	Kaggle-Covid-19: Part 2	0.8062	0.8070	0.8061	0.7267
		CT-Seg	0.9300	0.9650	0.5930	0.1250
		Seg-nr-2	0.9023	0.8954	0.8998	0.8610
		NSCLC radiogenomics	0.9102	0.9185	0.8099	0.9832
		Decathlon	0.8637	0.8269	0.7195	0.9660
		Moffitt	0.9285	0.9715	0.9800	0.9788
	HyEx-Net [60] (24.14 M)	Kaggle-Covid-19: Part 2	0.8946	0.8979	0.8945	0.8252
		CT-Seg	0.9900	0.9950	0.8308	0.5000
		Seg-nr-2	0.8613	0.8479	0.8544	0.8041
		NSCLC radiogenomics	0.9059	0.9044	0.8003	0.9801
		Decathlon	0.8681	0.8317	0.7258	0.9669
		Moffitt	0.9342	0.9400	0.9570	0.9565
	Proposed <i>DGMC</i> (2.23 M)	Kaggle-Covid-19: Part 2	0.9617	0.9618	0.9616	0.9383
		CT-Seg	1.0000	1.0000	1.0000	1.0000
		Seg-nr-2	0.9795	0.9822	0.9806	0.9831
		NSCLC radiogenomics	0.9939	0.9789	0.9827	0.9954
		Decathlon	0.9943	0.9809	0.9831	0.9961
		Moffitt	0.9971	0.9994	0.9996	0.9996
	ResNet-50 [64] (25.5 M)	Kaggle-Covid-19: Part 2	0.8235	0.8257	0.7214	0.6932
		CT-Seg	0.8780	0.8800	0.5549	0.1010
		Seg-nr-2	0.8437	0.8459	0.7012	0.6834
		NSCLC radiogenomics	0.8941	0.9015	0.8132	0.8765
		Decathlon	0.8517	0.8535	0.7277	0.8280
		Moffitt	0.9092	0.9130	0.8547	0.8103
	SE-ResNeXt-50 [140] (27.6 M)	Kaggle-Covid-19: Part 2	0.8397	0.8432	0.8039	0.7213
		CT-Seg	0.9200	0.9600	0.5824	0.1250
		Seg-nr-2	0.8542	0.8575	0.7712	0.7239
		NSCLC radiogenomics	0.9092	0.9135	0.8247	0.7385
		Decathlon	0.8772	0.8803	0.7292	0.8300
		Moffitt	0.9259	0.9273	0.8639	0.8533
	DenseNet-121 [70] (7.2 M)	Kaggle-Covid-19: Part 2	0.8439	0.8475	0.8215	0.7539
		CT-Seg	0.8800	0.8950	0.5916	0.1111
		Seg-nr-2	0.8724	0.8770	0.8159	0.7074
		NSCLC radiogenomics	0.9217	0.9239	0.8034	0.7715
		Decathlon	0.8549	0.8572	0.7725	0.8430
		Moffitt	0.9532	0.9557	0.8722	0.8617
	MobileNet-V3 [23] (5.4 M)	Kaggle-Covid-19: Part 2	0.8932	0.8987	0.8539	0.7824
		CT-Seg	1.0000	1.0000	1.0000	1.0000
		Seg-nr-2	0.8737	0.8759	0.7372	0.8017
		NSCLC radiogenomics	0.9093	0.9134	0.8872	0.8013
		Decathlon	0.8952	0.8990	0.8127	0.8532
		Moffitt	0.9731	0.9782	0.8951	0.9192

competitive results but at significantly higher model sizes. Lightweight models such as MobileNet-V3 (5.4M) and DenseNet-121 (7.2M) show reasonable performance, yet they are outperformed by *DGMC* both in terms of accuracy and

model compactness. Overall, *DGMC* offers the best balance between performance and computational complexity, making it highly suitable for resource-constrained deployment scenarios.

5.3.5 Visualization and explainability

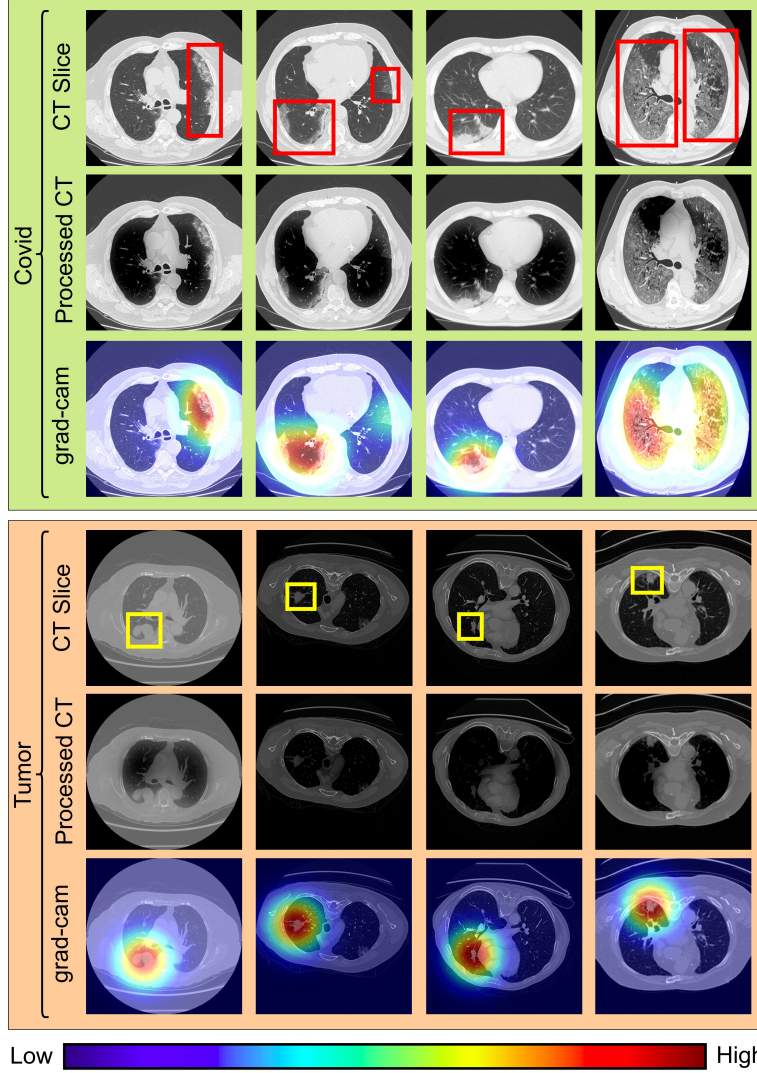


FIGURE 5.8: Visual attention of classifier *DGMC*, as generated through Grad-Cam, highlighting the COVID-19 infected and tumour affected regions in the eight sample CT scans.

Fig. 5.8 displays the GradCAM-based visual attention produced by our classifier *DGMC*. Manual bounding boxes are drawn on the CT image to pinpoint the infectious locations. Performance over four random test cases, from each of COVID-19 infected and tumour affected slices, demonstrate the robustness of *DGMC*. It is observed that the diseased regions of the CT slices get highlighted in all the cases. The model makes decisions based on the regions which are highlighted by GradCAM. For example, it focuses on COVID infected areas to classify an image as COVID, and similarly for tumors. This focused attention on the infected areas is achieved through the EIT-processed CT scans.

5.3.6 Ablations

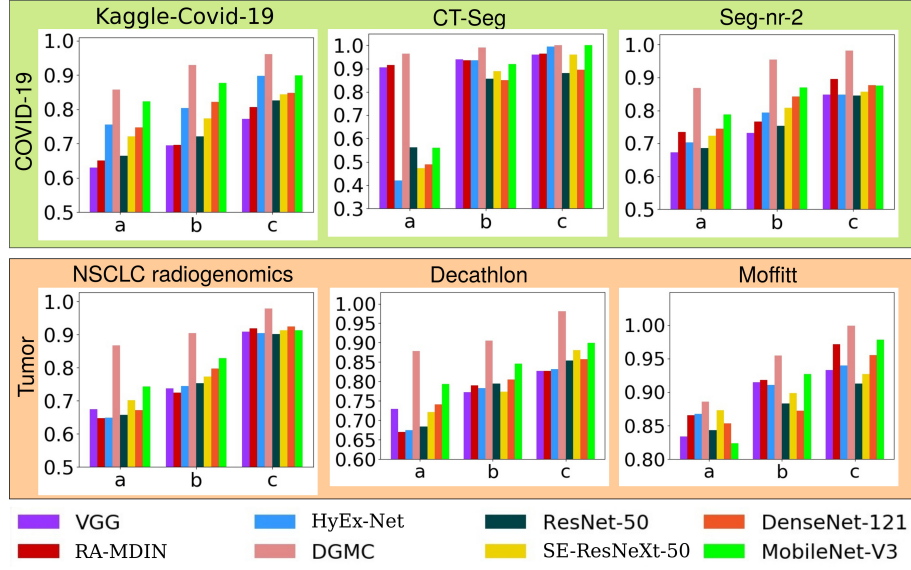


FIGURE 5.9: Comparative performance of different classifiers, in terms of AUC , on the six test datasets (of COVID-19 and Tumor slices) with input as (a) only CT images, (b) CT and CTGAN-generated images, and (c) CT and CTGAN-generated images with corresponding EIT-processed CT.

Fig. 5.9 demonstrates the effect of input images with and without augmentation (of CT and EIT-processed CT) on the performance of classifier *DGMC*, as evaluated on the six publicly available COVID-19 and Tumor test data (outlined in Table 1.1) in terms of metric AUC . We explore the three cases of using (a) only CT images, (b) CT images along with the CTGAN-augmented images, and (c) corresponding EIT-processed CT, along with the original CT images plus the CTGAN-generated versions. A comparative study with related state-of-the-art models in literature (like VGG [158], RA-MDIN [175], HyEx-Net [60], ResNet-50 [64], SE-ResNeXt-50 [140], DenseNet-121 [70], MobileNet-V3 [68]) is also included in the figure, to validate the superiority of *DGMC* in each scenario. It is evident that our *DGMC* classifier is found to perform the best in all cases. Overall, the inclusion of EIT-processed CT images along with the original CT and CTGAN-generated CT, is seen to provide the best results for all classifiers considered in this study. This is because the EIT-processed CT helps the model maintain greater focus on the anomalous regions compared to the original (or augmented) CT scans. As a result, the model is able to capture complementary details from both CT and EIT-processed CT images.

The confusion matrices of Fig. 5.10 depict the three-class performance of our *DGMC*, for the Healthy, COVID-19 infected, and Tumor-affected slices, in comparison with those of the related state-of-the-art classifiers considered. The three input cases refer to using (a) only CT images, (b) CT images along with the CTGAN-augmented images, and (c) EIT-processed CT scans, along with the original CT images plus the CTGAN-augmented versions. We combined images from all test datasets to evaluate differentiation into the three classes. Again, *DGMC* was found to produce the best overall performance in all cases.



FIGURE 5.10: Confusion matrices generated by different classifiers, with input as (a) only CT images, (b) CT and CTGAN-generated images, and (c) CT and CTGAN-generated images plus EIT-processed CT version.

Finally the role of the multihead classifier module was explored, in Table 5.2, on the dual-mode input images encompassing the CT and corresponding EIT-processed CT scans. As before, the six publicly available COVID-19 and Tumor test data (outlined in Table 1.1) were used in the experiments. It is evident that with the incorporation of the multihead classifier module the output performance improved over all cases considered in the study.

TABLE 5.2: Impact of multihead classifier module, as evaluated in terms of *Accuracy*, on the input images

		Dataset	Metrics			
			Accuracy	AUC	Macro F1	mAP
Classifier	Proposed Classifier (Without multihead module)	Kaggle-Covid-19: Part 2	0.9010	0.8999	0.9005	0.8569
		CT-Seg	0.9901	0.9950	0.8308	0.5000
		Seg-nr-2	0.9626	0.9636	0.9623	0.9580
		NSCLC radiogenomics	0.9868	0.9695	0.9635	0.9934
		Decathlon	0.9855	0.9690	0.9583	0.9937
		Moffitt	0.9861	0.9987	0.9991	0.9990
		Kaggle-Covid-19: Part 2	0.9617	0.9618	0.9616	0.9383
		CT-Seg	1.0000	1.0000	1.0000	1.0000
	Proposed Classifier (With multihead module)	Seg-nr-2	0.9795	0.9822	0.9806	0.9831
		NSCLC radiogenomics	0.9939	0.9789	0.9827	0.9954
		Decathlon	0.9943	0.9809	0.9831	0.9961
		Moffitt	0.9971	0.9994	0.9996	0.9996

5.4 Conclusions and Discussion

A new Deep Generative Multihead Classifier (*DGMC*) network was described, in this chapter, for the detection and classification of lung diseases from CT images. While CTGAN was used to augment the scans for addressing data scarcity, the EIT-like slices provided a separate perspective of the CT scans for better decision-making. The *DGMC* model outperformed in lung cancer and COVID-19 detection across seven datasets, showcasing its superiority through ablation studies and visualization. Learning from multiple perspectives enhanced the overall capabilities of the model, by leveraging strengths from both perspectives.

CT scans offered detailed structural information with high precision, while EIT helped to monitor dynamic physiological changes. Subtracting EIT-like images from the corresponding CT scans provided a holistic understanding, linking structural anomalies with real-time functional insights. It enhanced visibility of the infected regions within EIT-processed CT, aiding in pinpointing abnormalities and understanding their impact on tissue functionality.

Existing multimodal fusion techniques in medical imaging typically integrate information from diverse sources such as CT, X-ray, ultrasound, or MRI through transfer learning or feature-level fusion [66], [205]. In contrast, DGMC introduces a unique two-stage enhancement and classification pipeline. First, CTGAN is employed to augment the CT dataset and mitigate data scarcity. Separately, EIT-like images are generated from CT slices to provide an alternative processed view that emphasizes abnormal regions. This dual representation

of original and EIT-processed CT is then leveraged by a novel multihead classifier, enabling the model to learn complementary features and improve decision-making. Unlike conventional approaches that rely on direct modality fusion or pretrained networks [3], [72], [143], DGMC effectively separates augmentation, enhancement, and classification, resulting in improved disease localization, interpretability through activation mapping, and classification accuracy, as demonstrated through extensive experiments.

Our proposed method can be integrated into existing diagnostic workflows as a decision-support tool for radiologists. By processing standard CT scans to generate EIT-like images that emphasize anomalous regions, our approach can help clinicians quickly localize and interpret disease-relevant areas. This enhanced visualization, combined with automated classification, can serve as a second opinion, potentially improving diagnostic confidence and reducing interpretation time, especially in high-throughput or resource-limited clinical settings. Integration into PACS (Picture Archiving and Communication Systems) or as a plugin to existing radiology software could facilitate its real-world deployment.

The primary difficulty encountered in this study revolved around distinguishing between tumor-affected and COVID-infected regions, as these two often exhibited similar patterns. The validation relied on Grad-CAM visualization. As there is a general scarcity of EIT data from the domain, we resorted to transformation from the CT to the EIT domain. The utility of the approach is two-fold. It uses the information from different viewpoints, to arrive at a more efficient classification and diagnosis of the disease. It also indicates an effective strategy to generate mapping projections from one imaging modality to another, as and when necessary. This aspect holds promise in overcoming the inherent problem of scarcity of data in automated medical imaging applications. The possible role of EIT, combined with CT in lung disease management, can thus be explored and validated; in spite of relatively lower availability of EIT.

Real-time EIT dataset generation is ongoing; and we intend to extend the results. Although the model has been applied to the lung domain, encompassing the CT and EIT-processed CT scans, the general principles can be easily extended to include other imaging modalities depending on their applicability. Moreover the design can be extended to focus on diseases occurring in other organs of the body, as well as explore diseases beyond the set considered here.





Chapter 6



Conclusions and Future Scope





This chapter critically analyzes the results obtained from the algorithms developed in the thesis. The discussion includes highlights of our contributions towards designing an intelligent computer-aided system, for improved classification, detection, and segmentation of affected regions in lung diseases (COVID-19 and lung cancer) from CT images. It also discusses on the applicability of the solutions in wider industrial and real-life scenarios. We examine how these solutions can assist clinical decision-making, while suggesting cost-effective treatment schedules for improved prognosis. Finally, the challenges that remain unaddressed or have emerged in the course of our research are presented to provide insights into future work.

6.1 Conclusions

The thesis introduced several innovative noninvasive approaches for improved segmentation and classification of diseased regions caused by lung diseases, such as COVID-19 and lung cancer, using publicly-available multi-institutional lung images. The new frameworks were validated against ground truth data provided by doctors and radiologists. Cross-dataset evaluation was made, during model generalization and robustness checking, by training on one dataset and testing on multiple others. The computer-aided segmentation and classification systems facilitated automated detection and delineation of regions of interest, essential for surgical and treatment planning. Developing these automated algorithms presented challenges due to their inherent spatial and structural variability. Additionally, analyzing different types of anomalies with the same algorithm was often difficult. Comparisons with existing methods were challenging, because they were typically designed and validated on different private datasets. Critical factors influencing the results included (i) imaging modalities used, (ii) the disease stage at the time of imaging, and (iii) the grade of tumor or infection severity. For example, in case of COVID-19 the severity grading depends on the percentage of lung volume covered by the infection. In the case of lung tumors there can be different grades of the cancer. This research detailed the development of efficient algorithms for automated and deep learning-based strategies in segmenting COVID-19 infection and lung tumor affected regions from CT scans, followed by a generalized classification of the diseases in multimodal frameworks like CT-PET or CT-EIT.

- The Ensemble of selective Focus-based Multi-resolution Convolution network (*EFMC*) model of Chapter 2 developed a deep convolutional network to efficiently segment COVID-19 infected regions in the lung. Two new modules selective Focusing Module (*FM*) and Multi-Resolution block (*MR*) were introduced for obtaining selective focus based multi-resolution contextual information. This helped reduce the error along the segmentation boundary, for improved accuracy. Class imbalance was addressed with Focal Loss (*FL*), to balance between precision and recall with proper hyper-parameter settings. The *EFMC* achieved high accuracy and minimal misclassification. Patch-based training was employed to handle overfitting, and was found to be robust across diverse CT image sources – often surpassing state-of-the-art methods in performance. The novel Leave-One-Patient-Out (LOPO) ensemble enhanced precise delineation of anomalous regions, while categorizing disease severity based on limited training data. Ablation studies corroborated its potential in CAD systems for assisting radiologists.
- The novel Weighted Deformable *U*-Net (*WDU*-Net) of Chapter 3 was designed for efficient segmentation of the regions-of-interest possessing no discernible known geometric shape. It was applied for delineating COVID-19-infected and lung cancer-affected regions of the lung, and was found to outperform state-of-the-art models in multiple aspects. It could capture irregularly-shaped regions using Deformable Convolution (*DC*) module, while the Weight Generation (*WG*) module highlighted the relevant features. The Focal Asymmetric Similarity (*FAS*) loss function, combining Focal Loss (*FL*) and Asymmetric Similarity Loss (*ASL*), effectively addressed class imbalance. Patch-based training ensured balanced learning, while enhancing overall performance. Combining outputs from ten classifiers, through majority voting, refined the segmentation decisions. Bins of separate patient samples were created, with one of these being kept for validation in each fold. The *WDU*-Net demonstrated superior performance in lung anomaly segmentation, with quick learning, good generalization, and precise accuracy on the region of interest. It was observed that in datasets where the ROI had a specific geometrical shape, both *WDU*-Net and *EFMC*-Net performed equally well. However, when the ROI lacked any distinctive geometrical shape, the *WDU*-Net was significantly better. This was mainly due to its adaptability to the varying patterns and shapes of such ROI. Reduced training and testing times, combined with ensemble learning, contributed to its higher accuracy and robustness while minimizing over-and-under segmentation.
- The new *DEMF* model of Chapter 4 effectively classifies lung images as cancerous or healthy, and operates on multimodal images PET and CT. Combining diverse classifiers through majority voting, the *DEMF* improves robustness and minimizes overfitting. Ensembles help in leveraging varied model decisions to mitigate biases. The integration of five base classifiers, in parallel, ensures efficiency and strong generalization – even with limited data. Ablation studies and visualization underscore

its efficacy in identifying cancerous regions. The PCAE fusion strategy reduces dimensionality, by incorporating PCA and autoencoder components, while producing robust outcomes.

- The Deep Generative Multihead Classifier (*DGMC*) of Chapter 5 could detect and classify lung diseases from CT images, utilizing CTGAN for data augmentation along with the generation of EIT-like slices for producing corresponding EIT-processed CT scans. It could successfully distinguish between lung cancer and COVID-19 infections, across seven datasets, as validated through ablation studies and visualization. Learning the structural details from CT images, in conjunction with the dynamic insights obtained from the corresponding EIT-processed CT scans, allowed enhancement of overall performance. Challenges in distinguishing between tumor-affected and COVID-infected regions were validated through Grad-CAM visualization. The principle of the *DGMC* can be extended to diverse imaging modalities and diseases; thereby, broadening its applicability in multimodal medical image analysis.

These solutions are well-suited for industrial and real-life scenarios due to their ability to automate complex image analysis tasks, thus enhancing their diagnostic accuracy and efficiency. By leveraging deep learning and multimodal imaging, these methods provided precise detection and delineation of lung infections and diseases. This is crucial for any timely intervention. The integration of multimodal scans, along with novel network architectures, ensured comprehensive analysis; thereby, aiding in accurate diagnosis and effective monitoring. These advancements hold potential to assist medical practitioners, by reducing their manual workload and bias while allowing for better resource allocation and patient management. Cost-effective treatment schedules can be derived from these accurate diagnostic tools by facilitating early detection and personalized treatment plans. All these would ultimately lead to improved patient outcomes with optimized use of healthcare resources.

6.2 Limitations

The thesis focused on supervised learning. Although several challenges have been addressed, there exist certain limitations.

- **Dependence on labeled data:** The models developed in this research required labeled data. This is often expensive and time-consuming to obtain, and might require expert knowledge.
- **Limited to the scope of training data:** The models were inherently limited by the diversity and scope of the training data.
- **Scalability:** Training large supervised models can be computationally intensive and time-consuming. As the dataset grows, the amount of computational resources and time required for training simultaneously increase.

- **Sensitivity to noise and outliers:** Appropriate preprocessing and cleaning of data are essential to minimize the effect of noise and outliers.
- **Need for retraining:** Sometimes the models may need to be retrained with new data, to preserve accuracy over time. This can be resource-intensive.
- **Potential for human error in labeling:** Human-labeled data can be subject to errors, biases, and inconsistencies; which can negatively impact the performance of a model.

These limitations highlight the importance of careful data preparation, regular model evaluation, and the need for complementary approaches such as unsupervised or semi-supervised learning in certain scenarios.

6.3 Future Scope

Deep learning models in clinical settings face challenges due to data variability and patient demographics, which can affect their performance. Variations in data quality, imaging protocols, and patient characteristics may hinder model generalization across different populations or healthcare institutions. Additionally, the ability of models to handle rare conditions or unseen manifestations is crucial but remains a concern. Overcoming these challenges necessitates robust model development strategies, including diverse training data and rigorous evaluation methodologies. Collaboration among healthcare institutions, data scientists, and clinicians is essential for developing and deploying deep learning models that demonstrate reliable performance across various patient populations and clinical scenarios.

The thesis presented advancements in developing a pipeline for non-invasive lung disease detection and diagnosis within the rapidly evolving field of medical image analysis. While the algorithms and methodologies demonstrated efficacy and efficiency in experimental settings, their real-world applicability requires further investigation. Future investigations should involve testing the CAD system with larger multi-institutional datasets encompassing diverse scanners, modalities, and scan settings. Additionally, theoretical performance analysis of the developed algorithms is warranted. The following outlines potential future extensions of the research presented in this thesis.

- **Enhanced algorithmic validation:** Further validation of the segmentation algorithms (*EFMC* and *WDU-Net*) can be conducted on larger and more diverse datasets to ensure their robustness across different patient demographics and imaging conditions.
- **Clinical integration and validation:** Integration of the developed algorithms into clinical practice settings for real-time validation and feedback from healthcare professionals can provide insights into their effectiveness in real-world scenarios.

- **Exploration of advanced strategies:** Investigation of advanced deep learning techniques, such as transformers and diffusion networks, could lead to the development of more accurate and efficient models for disease diagnosis and segmentation.
- **Multimodal data fusion:** Exploration of multimodal imaging data fusion, including CT, PET, EIT, and others, can provide comprehensive insights into lung pathology and improve the accuracy of disease detection and characterization. Conditional Variational Autoencoders (CVAEs) can be used for fusion purposes, due to their inherent ability for generating data which is conditioned on additional information or labels. This can enable controlled and specific data generation, allowing effective modeling of complex relationships between input data and conditioning variables.
- **Advanced machine learning algorithms:** Incorporation of advanced machine learning strategies, such as reinforcement learning and transfer learning, can optimize model performance and adaptability to evolving medical datasets.
- **Clinical decision support systems:** Development of AI-driven clinical decision support systems can assist healthcare professionals in making informed treatment decisions, by providing accurate and timely insights into lung disease diagnosis and management.
- **Collaboration with medical experts:** Collaboration with medical experts and healthcare providers, for clinical validation and feedback, can ensure the relevance and applicability of the developed algorithms in real-world healthcare scenario.
- **Iterative development process:** The iterative process of development, validation, and refinement, based on interactive feedback from medical experts along with clinical validation studies, can lead to continuous improvement and optimization of the developed algorithms for improved patient outcome and quality of care.



List of Publications

1. Published/Accepted:

Journals:

- J1. **Surochita Pal**, Sushmita Mitra and B. Uma Shankar. Collective intelligent strategy for improved segmentation of COVID-19 from CT. *Expert Systems with Applications*, volume: 235, pages: 121099, 2024. DOI: [10.1016/j.eswa.2023.121099](https://doi.org/10.1016/j.eswa.2023.121099).
- J2. **Surochita Pal**, Sushmita Mitra and B. Uma Shankar. Weighted Deformable Network for Efficient Segmentation of Lung Tumors in CT. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, volume: 55, pages: 898-909, 2024. DOI: [10.1109/TSMC.2024.3489029](https://doi.org/10.1109/TSMC.2024.3489029)

Conferences:

- C1. **Surochita Pal** and Sushmita Mitra. Deep Ensembling with Multimodal Image Fusion for Efficient Classification of Lung Cancer. in *Proceedings of the 15th International IEEE Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages: 1–7, 2024. DOI: [10.1109/ICCCNT61001.2024.10726043](https://doi.org/10.1109/ICCCNT61001.2024.10726043)
- C2. **Surochita Pal** and Sushmita Mitra. A Comparative Analysis of Deep Learning Architectures for Segmentation in Lung. *IEEE Region 10 Symposium (TENSYP)*, pages: 1–6, 2024. DOI: [10.1109/TENSYP61132.2024.10751813](https://doi.org/10.1109/TENSYP61132.2024.10751813)

Patent Granted:

- P1. Sushmita Mitra, **Surochita Pal Das**, and B. Uma Shankar. *Efficient segmentation of tumours from lung CT*, Indian Patent No. 558671, Jan. 24, 2025.

2. Under Processing:

Journals:

- J3. **Surochita Pal** and Sushmita Mitra and and Tingwen Huang. Deep Multihead Classification of Some Lung Diseases. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024 (Manuscript Id: TETCI-2024-1160, Submission Date: 30.09.2024) (submitted).

Patents Filed:

- P2. Sushmita Mitra, **Surochita Pal Das**, and B. Uma Shankar. *Efficient segmentation of tumours from lung CT*, US Patent Application # 18/409,137, Jan. 10, 2024.
- P3. Sushmita Mitra and **Surochita Pal Das**. *Detection of lung diseases with multimodal imaging*, Indian Patent Application # 202331070180, Oct. 16, 2023.

Bibliography

- [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*, vol. ISBI, 2019, pp. 683–687.
- [2] H. J. Aerts, E. R. Velazquez, *et al.*, “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach,” *Nature Communications*, vol. 5, pp. 1–9, 2014.
- [3] M. Ahmad, U. I. Bajwa, *et al.*, “Lightweight ResGRU: A deep learning-based prediction of SARS-CoV-2 (COVID-19) and its severity classification using multimodal chest radiography images,” *Neural Computing and Applications*, vol. 35, pp. 9637–9655, 2023.
- [4] I. Ali and O. M. Alharbi, “COVID-19: Disease, management, treatment, and social impact,” *Science of the Total Environment*, vol. 728, p. 138 861, 2020.
- [5] M. Z. Alom, M. Hasan, *et al.*, “Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation,” *CoRR*, vol. abs/1802.06955, 2018.
- [6] V. Ambrosini, S. Nicolini, *et al.*, “PET/CT imaging in different types of lung cancer: An overview,” *European Journal of Radiology*, vol. 81, pp. 988–1001, 2012.
- [7] M. Antonelli, A. Reinke, *et al.*, “The medical segmentation Decathlon,” *Nature Communications*, vol. 13, pp. 1–13, 2022.
- [8] S. M. Anwar, M. Majid, *et al.*, “Medical image analysis using convolutional neural networks: A review,” *Journal of Medical Systems*, vol. 42, pp. 1–13, 2018.
- [9] D. Ardila, A. P. Kiraly, *et al.*, “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” *Nature Medicine*, vol. 25, pp. 954–961, 2019.
- [10] S. G. Armato III, G. McLennan, *et al.*, “The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans,” *Medical Physics*, vol. 38, pp. 915–931, 2011.
- [11] R. Azad, L. Niggemeier, *et al.*, “Beyond self-attention: Deformable large kernel attention for medical image segmentation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 1287–1297.
- [12] X. Bai, H. Wang, *et al.*, “Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence,” *Nature Machine Intelligence*, vol. 3, pp. 1081–1089, 2021.
- [13] S. Bakr, O. Gevaert, *et al.*, “A radiogenomic dataset of non-small cell lung cancer,” *Scientific Data*, vol. 5, pp. 1–9, 2018.

- [14] H. M. Balaha, M. H. Balaha, and H. A. Ali, "Hybrid COVID-19 segmentation and recognition framework (HMB-HCF) using deep learning and genetic algorithms," *Artificial Intelligence in Medicine*, vol. 119, p. 102 156, 2021.
- [15] K. Balasamy and V. Seethalakshmi, "HCO-RLF: Hybrid classification optimization using recurrent learning and fuzzy for COVID-19 detection on CT images," *Biomedical Signal Processing and Control*, vol. 100, p. 106 951, 2025.
- [16] S. Banerjee, S. Mitra, and B. Uma Shankar, "Automated 3D segmentation of brain tumor using visual saliency," *Information Sciences*, vol. 424, pp. 337–353, 2018.
- [17] I. Bankman, *Handbook of Medical Image Processing and Analysis*. San Diego: Elsevier, 2008.
- [18] R. Bashir, R. Junejo, *et al.*, "SWT and PCA image fusion methods for multi-modal imagery," *Multimedia Tools and Applications*, vol. 78, pp. 1235–1263, 2019.
- [19] 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)*, IEEE, 2020, pp. 2521–2527.
- [20] A. El-Baz, X. Jiang, and J. S. Suri, *Biomedical Image Segmentation: Advances and Trends*. Natick: CRC Press, 2016.
- [21] 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)*, IEEE, 2022, pp. 151–155.
- [22] T. K. Bera, "Applications of electrical impedance tomography (EIT): A short review," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 331, 2018, p. 012 004.
- [23] A. Bhagwat, S. Dutta, *et al.*, "An online 11 kv distribution system insulator defect detection approach with modified YOLOv11 and MobileNetV3," *Scientific Reports*, vol. 15, p. 15 691, 2025.
- [24] D. Bhattacharyya, N. Thirupathi Rao, *et al.*, "A bi-directional deep learning architecture for lung nodule semantic segmentation," *The Visual Computer*, vol. 39, pp. 5245–5261, 2023.
- [25] P. Bühlmann, "Bagging, boosting and ensemble methods," *Handbook of Computational Statistics: Concepts and Methods*, pp. 985–1022, 2012.
- [26] H. Cao, H. Liu, *et al.*, "Dual-branch residual network for lung nodule segmentation," *Applied Soft Computing*, vol. 86, p. 105 934, 2020.
- [27] X. Chen, X. Wang, *et al.*, "Recent advances and clinical applications of deep learning in medical image analysis," *Medical Image Analysis*, vol. 79, p. 102 444, 2022.
- [28] S. Chheang and K. Brown, "Lung cancer staging: Clinical and radiologic perspectives," *Semin. Intervent. Radiol.*, vol. 30, pp. 99–113, 2013.

- [29] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
- [30] S. S. Chowa, M. R. I. Bhuiyan, *et al.*, “An automated privacy-preserving self-supervised classification of COVID-19 from lung CT scan images minimizing the requirements of large data annotation,” *Scientific Reports*, vol. 15, p. 226, 2025.
- [31] C. Collazo, I. Vargas, B. Cara, C. J. Weinheimer, R. P. Grabau, D. Goldgof, L. Hall, *et al.*, “Synergizing deep learning-enabled preprocessing and human–AI integration for efficient automatic ground truth generation,” *Bioengineering*, vol. 11, p. 434, 2024.
- [32] R. Cong, Y. Zhang, *et al.*, “Boundary guided semantic learning for real-time COVID-19 lung infection segmentation system,” *IEEE Transactions on Consumer Electronics*, vol. 68, pp. 376–386, 2022.
- [33] J. Dai, J. K. Udupa, *et al.*, “A deep-learning-based geographic attention model for body composition tissue segmentation,” in *Proceedings of SPIE Medical Imaging 2023: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 12468, 2023, pp. 12–18.
- [34] J. Dai, H. Qi, *et al.*, “Deformable convolutional networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 764–773.
- [35] D.-P. Dao, H.-J. Yang, *et al.*, “Survival analysis based on lung tumor segmentation using global context-aware transformer in multimodality,” in *Proceedings of the 26th International Conference on Pattern Recognition (ICPR)*, IEEE, 2022, pp. 5162–5169.
- [36] S. Das, R. Sasidharan, S. Ghosh, S. Chatterjee, S. Ray, M. Arunsingh, T. Bhattacharyya, and J. Mukherjee, “Predicting the radiotherapeutic treatment response of non-small cell lung cancer,” *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100 065, 2022.
- [37] J. Deng, W. Dong, *et al.*, “ImageNet: A large-scale hierarchical image database,” in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, pp. 248–255.
- [38] F. C. Detterbeck, D. J. Boffa, *et al.*, “The 8th Edition of Lung Cancer Stage Classification,” *Chest*, vol. 151, pp. 193–203, 2017.
- [39] T. G. Dietterich, “Ensemble methods in machine learning,” in *Proceedings of the International Workshop on Multiple Classifier Systems*, Springer, 2000, pp. 1–15.
- [40] X. Ding, J. Xu, *et al.*, “Chest CT findings of COVID-19 pneumonia by duration of symptoms,” *European Journal of Radiology*, vol. 127, p. 109 009, 2020.
- [41] Y. Dong, W. Yang, *et al.*, “An improved supervoxel 3D region growing method based on PET/CT multimodal data for segmentation and reconstruction of GGNs,” *Multimedia Tools and Applications*, vol. 79, pp. 2309–2338, 2020.

- [42] R. O. Duda, P. E. Hart, *et al.*, *Pattern Classification*. New York: John Wiley & Sons, 2006.
- [43] P. Dutande, U. Baid, *et al.*, “Deep residual separable convolutional neural network for lung tumor segmentation,” *Computers in Biology and Medicine*, vol. 141, p. 105 161, 2022.
- [44] 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)*, vol. ICAS 2021, 2021, pp. 1–6.
- [45] Y. Fang, H. Zhang, *et al.*, “Sensitivity of chest CT for COVID-19: Comparison to RT-PCR,” *Radiology*, vol. 296, E115–E117, 2020.
- [46] D. Ferrario, B. Grychtol, *et al.*, “Toward morphological thoracic EIT: Major signal sources correspond to respective organ locations in CT,” *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 3000–3008, 2012.
- [47] R. Fogarty, D. Goldgof, L. Hall, *et al.*, “Classifying malignancy in prostate glandular structures from biopsy scans with deep learning,” *Cancers*, vol. 15, p. 2335, 2023.
- [48] A. Frangi, J. Prince, and M. Sonka, *Medical Image Analysis*. Cambridge: Academic Press, 2023.
- [49] V. Fredriksen, S. O. M. Sevre, *et al.*, “Teacher-student approach for lung tumor segmentation from mixed-supervised datasets,” *PLoS ONE*, vol. 17, e0266147, 2022.
- [50] Z. Fu, Y. Zhao, *et al.*, “Gradient structural similarity based gradient filtering for multi-modal image fusion,” *Information Fusion*, vol. 53, pp. 251–268, 2020.
- [51] L. Geng, S. Zhang, *et al.*, “Lung segmentation method with dilated convolution based on VGG-16 network,” *Computer Assisted Surgery*, vol. 24, pp. 27–33, 2019.
- [52] S. Ghosh, N. Das, *et al.*, “Understanding deep learning techniques for image segmentation,” *ACM Computing Surveys*, vol. 52, pp. 1–35, 2019.
- [53] M. Goncharov, M. Pisov, *et al.*, “CT-Based COVID-19 triage: Deep multitask learning improves joint identification and severity quantification,” *Medical Image Analysis*, vol. 71, p. 102 054, 2021.
- [54] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. New York: Pearson Education India, 2009.
- [55] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge: MIT Press, 2016.
- [56] C. Gridelli, A. Rossi, *et al.*, “Non-small-cell lung cancer,” *Nature Reviews Disease Primers*, vol. 1, pp. 1–16, 2015.

- [57] Y. Gu, X. Lu, *et al.*, “Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs,” *Computers in Biology and Medicine*, vol. 103, pp. 220–231, 2018.
- [58] Y. Gu, V. Kumar, *et al.*, “Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach,” *Pattern Recognition*, vol. 46, pp. 692–702, 2013.
- [59] H. Guan and M. Liu, “Domain adaptation for medical image analysis: A survey,” *IEEE Transactions on Biomedical Engineering*, vol. 69, pp. 1173–1185, 2022.
- [60] S. Gupta and R. P. Mahapatra, “HyEx-Net: Hybrid efficient xception network for human activity recognition using multimodalities,” *Cluster Computing*, vol. 28, p. 240, 2025.
- [61] R. Harkness, A. F. Frangi, *et al.*, “Multi-centre benchmarking of deep learning models for COVID-19 detection in chest X-rays,” *Frontiers in Radiology*, vol. 4, p. 1386906, 2024.
- [62] S. R. Hashemi, S. S. M. Salehi, *et al.*, “Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection,” *IEEE Access*, vol. 7, pp. 1721–1735, 2018.
- [63] S. Hayes, M. Pietanza, *et al.*, “Comparison of CT volumetric measurement with RECIST response in patients with lung cancer,” *European Journal of Radiology*, vol. 85, pp. 524–533, 2016.
- [64] K. He, X. Zhang, *et al.*, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [65] K. He, X. Zhang, *et al.*, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1904–1916, 2015.
- [66] M. J. Horry, S. Chakraborty, *et al.*, “COVID-19 detection through transfer learning using multimodal imaging data,” *IEEE Access*, vol. 8, pp. 149808–149824, 2020.
- [67] S. Hossain, S. Najeeb, *et al.*, “A pipeline for lung tumor detection and segmentation from CT scans using dilated convolutional neural networks,” in *Proceedings of the ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2019, pp. 1348–1352.
- [68] A. Howard, M. Sandler, *et al.*, “Searching for mobilenetv3,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.
- [69] K.-L. Hua, C.-H. Hsu, *et al.*, “Computer-aided classification of lung nodules on computed tomography images via deep learning technique,” *OncoTargets and Therapy*, vol. 8, pp. 2015–2022, 2015.

- [70] G. Huang, Z. Liu, *et al.*, “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [71] D. T. Huff, A. J. Weisman, *et al.*, “Interpretation and visualization techniques for deep learning models in medical imaging,” *Physics in Medicine & Biology*, vol. 66, 04TR01, 2021.
- [72] D. M. Ibrahim, N. M. Elshennawy, *et al.*, “Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases,” *Computers in Biology and Medicine*, vol. 132, p. 104348, 2021.
- [73] N. Ibtehaz and M. S. Rahman, “MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation,” *Neural Networks*, vol. 121, pp. 74–87, 2020.
- [74] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the International Conference on Machine Learning*, PMLR, 2015, pp. 448–456.
- [75] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewoods Cliff: Prentice-Hall, Inc., 1989.
- [76] Z. Ji, Z. Zhao, *et al.*, “ResDSda_U-Net: A novel U-Net-Based residual network for segmentation of pulmonary nodules in lung CT images,” *IEEE Access*, vol. 11, pp. 87775–87789, 2023.
- [77] J. Jiang, Y.-C. Hu, *et al.*, “Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images,” *IEEE Transactions on Medical Imaging*, vol. 38, pp. 134–144, 2018.
- [78] C. Jin, J. K. Udupa, *et al.*, “Object recognition in medical images via anatomy-guided deep learning,” *Medical Image Analysis*, vol. 81, p. 102527, 2022.
- [79] Q. Jin, H. Cui, *et al.*, “Domain adaptation based self-correction model for COVID-19 infection segmentation in CT images,” *Expert Systems with Applications*, vol. 176, p. 114848, 2021.
- [80] S. Kamada and T. Ichimura, “A segmentation method of lung tumor by using adaptive structural deep belief network,” in *Proceedings of the 62nd Annual Conference of the Society of Instrument and Control Engineers (SICE)*, IEEE, 2023, pp. 1529–1534.
- [81] U. Kamal, A. M. Rafi, *et al.*, “Lung cancer tumor region segmentation using recurrent 3D-denseUNet,” in *Proceedings of the Thoracic Image Analysis: Second International Workshop, TIA 2020, MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2*, Springer, 2020, pp. 36–47.
- [82] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [83] Y.-F. Ko and K.-S. Cheng, “Semi-Siamese U-Net for separation of lung and heart bioimpedance images: A simulation study of thorax EIT,” *PLoS ONE*, vol. 16, e0246071, 2021.

- [84] B. Koonce, “EfficientNet,” in *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*. Apress, 2021, pp. 109–123.
- [85] S. Kordnoori, M. Sabeti, *et al.*, “A deep learning framework for accurate COVID-19 classification in CT-Scan images,” *Machine Learning with Applications*, p. 100 628, 2025.
- [86] A. Krizhevsky, I. Sutskever, *et al.*, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, pp. 84–90, 2017.
- [87] A. Kumar, M. Fulham, *et al.*, “Co-learning feature fusion maps from PET-CT images of lung cancer,” *IEEE Transactions on Medical Imaging*, vol. 39, pp. 204–217, 2019.
- [88] D. Kumar, A. Wong, *et al.*, “Lung nodule classification using deep features in CT images,” in *Proceeding of the 12th Conference on Computer and Robot Vision*, IEEE, 2015, pp. 133–138.
- [89] S. Kumar, S. K. Gupta, *et al.*, “Ensemble multimodal deep learning for early diagnosis and accurate classification of COVID-19,” *Computers and Electrical Engineering*, vol. 103, p. 108 396, 2022.
- [90] O. Kwon, J. R. Yoon, *et al.*, “Estimation of anomaly location and size using electrical impedance tomography,” *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 89–96, 2003.
- [91] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.
- [92] G. Lee and H. Fujita, *Deep Learning in Medical Image Analysis: Challenges and Applications*. Cham: Springer, 2020, vol. 1213.
- [93] T. Lei, R. Wang, *et al.*, “DefED-Net: Deformable encoder-decoder network for liver and liver tumor segmentation,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 6, pp. 68–78, 2021.
- [94] P. Li, S. Wang, *et al.*, “A large-scale CT and PET/CT dataset for lung cancer diagnosis [dataset],” *The Cancer Imaging Archive*, 2020.
- [95] Y. Li and L. Xia, “Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management,” *American Journal of Roentgenology*, vol. 214, pp. 1280–1286, 2020.
- [96] F. Liao, M. Liang, *et al.*, “Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pp. 3484–3495, 2019.
- [97] M. Lin, Q. Chen, *et al.*, “Network in network,” in *Proceedings of the 2nd International Conference on Learning Representations*, ICLR, 2014, pp. 1–10.
- [98] T.-Y. Lin, P. Goyal, *et al.*, “Focal loss for dense object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 318–327, 2017.

- [99] J. Liu, H. Yang, *et al.*, “Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining,” *arXiv preprint arXiv:2402.03302*, 2024.
- [100] J. Liu, J. Zhao, *et al.*, “Automated segmentation of COVID-19 lung infections from chest CT images via 2.5d dual-path encoder with multi-scale dynamic weight deep supervision mechanism,” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–1, 2024.
- [101] Z. Liu, Y. Lin, *et al.*, “Swin Transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 012–10 022.
- [102] J. Long, E. Shelhamer, *et al.*, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [103] L. Lu, D. Wang, *et al.*, “A quantitative imaging biomarker for predicting disease-free-survival-associated histologic subgroups in lung adenocarcinoma,” *European Radiology*, vol. 30, pp. 3614–3623, 2020.
- [104] J. Ma, F. Li, and B. Wang, “U-Mamba: Enhancing long-range dependency for biomedical image segmentation,” *arXiv preprint arXiv:2401.04722*, 2024.
- [105] J. Ma, Y. Wang, *et al.*, “Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation,” *Medical Physics*, vol. 48, pp. 1197–1210, 2021.
- [106] R. Mahum and A. S. Al-Salman, “Lung-RetinaNet: Lung cancer detection using a retinanet with multi-scale feature fusion and context module,” *IEEE Access*, pp. 53 850–53 861, 2023.
- [107] A. Malhotra, S. Mittal, P. Majumdar, S. Chhabra, K. Thakral, M. Vatsa, R. Singh, S. Chaudhury, *et al.*, “Multi-task driven explainable diagnosis of COVID-19 using chest X-ray images,” *Pattern Recognition*, vol. 122, p. 108 243, 2022.
- [108] S. P. K. Malladi, J. Mukherjee, M.-C. Larabi, and S. Chaudhury, “Towards explainable deep visual saliency models,” *Computer Vision and Image Understanding*, vol. 235, p. 103 782, 2023.
- [109] S. Mercieca, J. Belderbos, *et al.*, “Comparison of SUV_{max} and SUV_{peak} based segmentation to determine primary lung tumour volume on FDG PET-CT correlated with pathology data,” *Radiotherapy and Oncology*, vol. 129, pp. 227–233, 2018.
- [110] S. Minaee, Y. Boykov, *et al.*, “Image segmentation using deep learning: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, pp. 3523–3542, 2021.
- [111] S. Mitra and S. P. Das, “Detection of lung diseases with multimodal imaging,” Indian Patent Application # 202331070180, Oct. 16, 2023.
- [112] S. Mitra, S. P. Das, and B. U. Shankar, “Efficient segmentation of tumours from lung CT,” Indian Patent No. 558671, Jan. 24, 2025.

- [113] S. Mitra, S. P. Das, and B. U. Shankar, "Efficient segmentation of tumours from lung CT," US Patent Application # 18/409,137, Jan. 10, 2024.
- [114] S. Mittal and Y. Hasija, "Applications of deep learning in healthcare and biomedicine," *Deep Learning Techniques for Biomedical and Health Informatics*, vol. 68, pp. 57–77, 2020.
- [115] S. Modi, Y. Lin, *et al.*, "A socially inspired framework for human state inference using expert opinion integration," *IEEE/ASME Transactions on Mechatronics*, vol. 16, pp. 874–878, 2011.
- [116] S. P. Morozov, Andreychenko, *et al.*, "MosMedData: Data set of 1110 chest CT scans performed during the COVID-19 epidemic," *Digital Diagnostics*, vol. 1, pp. 49–59, 2020.
- [117] M. Muzammil, I. Ali, *et al.*, "Pulmonary nodule classification using feature and ensemble learning-based fusion techniques," *IEEE Access*, vol. 9, pp. 113 415–113 427, 2021.
- [118] S. Nabavi, A. Ejmalian, M. E. Moghaddam, A. A. Abin, A. F. Frangi, *et al.*, "Medical imaging and computational image analysis in COVID-19 diagnosis: A review," *Computers in Biology and Medicine*, vol. 135, p. 104 605, 2021.
- [119] S. A. Nadeem, A. P. Comellas, E. A. Hoffman, and P. K. Saha, "Airway detection in COPD at low-dose CT using deep learning and multi-parametric freeze and grow," *Radiology: Cardiothoracic Imaging*, vol. 4, e210311, 2022.
- [120] M. Nahiduzzaman, L. Faisal Abdulrazak, *et al.*, "A novel framework for lung cancer classification using lightweight convolutional neural networks and ridge extreme learning machine model with SHapley Additive exPlanations (SHAP)," *Expert Systems with Applications*, vol. 248, p. 123 392, 2024.
- [121] H. T. Nguyen, N. M. Nguyen, *et al.*, "An enhanced UNet3+ model for accurate identification of COVID-19 in CT images," *Digital Signal Processing*, vol. 163, p. 105 205, 2025.
- [122] Y. Ni, Z. Xie, *et al.*, "Two-stage multitask U-Net construction for pulmonary nodule segmentation and malignancy risk prediction," *Quantitative Imaging in Medicine and Surgery*, vol. 12, p. 292, 2022.
- [123] M. Nishio, K. Fujimoto, *et al.*, "Lung cancer segmentation with transfer learning: Usefulness of a pretrained model constructed from an artificial dataset generated using a generative adversarial network," *Frontiers in Artificial Intelligence*, vol. 4, p. 694 815, 2021.
- [124] O. Oktay, J. Schlemper, *et al.*, "Attention U-Net: Learning where to look for the pancreas," *CoRR*, vol. abs/1804.03999, 2018. arXiv: [1804.03999](https://arxiv.org/abs/1804.03999).
- [125] M. Owais, N. R. Baek, *et al.*, "DMDF-Net: Dual multiscale dilated fusion network for accurate segmentation of lesions related to COVID-19 in lung radiographic scans," *Expert Systems with Applications*, vol. 202, p. 117 360, 2022.

- [126] S. Pal and S. Mitra, "A comparative analysis of deep learning architectures for segmentation in lung," in *Proceedings of the IEEE Region 10 Symposium (TENSYP)*, Netaji Subhas University of Technology (NSUT), New Delhi, 2024, pp. 1–6.
- [127] S. Pal and S. Mitra, "Deep ensembling with multimodal image fusion for efficient classification of lung cancer," in *Proceedings of the 15th International IEEE Conference on Computing Communication and Networking Technologies (ICCCNT)*, IIT Mandi, 2024.
- [128] S. Pal, S. Mitra, and B. U. Shankar, "Collective intelligent strategy for improved segmentation of COVID-19 from CT," *Expert Systems with Applications*, vol. 235, p. 121 099, 2024.
- [129] S. Pal, S. Mitra, and B. U. Shankar, "Weighted deformable network for efficient segmentation of lung tumors in ct," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, pp. 898–909, 2024.
- [130] C. Parmar, E. Rios Velazquez, *et al.*, "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PLoS ONE*, vol. 9, e102107, 2014.
- [131] R. Paul, S. H. Hawkins, *et al.*, "Combining deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from diagnostic CT," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, 2016, pp. 002 570–002 575.
- [132] R. Paul, S. H. Hawkins, Y. Balagurunathan, M. Schabath, R. J. Gillies, L. O. Hall, *et al.*, "Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma," *Tomography*, vol. 2, pp. 388–395, 2016.
- [133] R. Paul, S. Kariev, D. Cherezov, M. B. Schabath, R. J. Gillies, L. O. Hall, *et al.*, "Deep radiomics: Deep learning on radiomics texture images," in *Medical Imaging 2021: Computer-Aided Diagnosis*, SPIE, vol. 11597, 2021, pp. 8–17.
- [134] M. Pennisi, I. Kavasidis, *et al.*, "An explainable AI system for automated COVID-19 assessment and lesion categorization from CT-scans," *Artificial Intelligence in Medicine*, vol. 118, p. 102 114, 2021.
- [135] D. Pessoa, B. M. Rocha, *et al.*, "BRACETS: Bimodal repository of auscultation coupled with electrical impedance thoracic signals," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107 720, 2023.
- [136] J. Peter, "Medical imaging modalities—An introduction," in *Advanced Imaging in Biology and Medicine: Technology, Software Environments, Applications*, Springer, 2009, pp. 225–254.
- [137] G. Pezzano, V. R. Ripoll, *et al.*, "CoLe-CNN: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation," *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105 792, 2021.

- [138] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, pp. 21–45, 2006.
- [139] S. P. Primakov, A. Ibrahim, *et al.*, "Automated detection and segmentation of non-small cell lung cancer computed tomography images," *Nature Communications*, vol. 13, p. 3423, 2022.
- [140] A. Priya and P. S. Bharathi, "SE-ResNeXt-50-CNN: A deep learning model for lung cancer classification," *Applied Soft Computing*, p. 112 696, 2025.
- [141] N. C. Purandare and V. Rangarajan, "Imaging of lung cancer: Implications on staging and management," *Indian Journal of Radiology and Imaging*, vol. 25, pp. 109–118, 2015.
- [142] N. Ravikumar, A. Zakeri, Y. Xia, and A. F. Frangi, "Chapter 16 - Deep learning fundamentals," in *Medical Image Analysis*, ser. MICCAI Society, Academic Press, 2024, pp. 415–450.
- [143] H. C. Reis and V. Turk, "COVID-DSNet: A novel deep convolutional neural network for detection of coronavirus (SARS-CoV-2) cases from CT and chest X-Ray images," *Artificial Intelligence in Medicine*, vol. 134, p. 102 427, 2022.
- [144] O. Ronneberger, P. Fischer, *et al.*, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference*, N. Navab, J. Hornegger, *et al.*, Eds., ser. LNCS, vol. 9351, Springer, 2015, pp. 234–241.
- [145] R. Roy, S. Mazumdar, *et al.*, "ADGAN: Attribute-driven generative adversarial network for synthesis and multiclass classification of pulmonary nodules," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, pp. 2484–2495, 2024.
- [146] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.
- [147] P. K. Saha, S. A. Nadeem, *et al.*, "A survey on artificial intelligence in pulmonary imaging," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 13, e1510, 2023.
- [148] K. Santosh and S. Ghosh, "COVID-19 versus lung cancer: Analyzing chest CT images using deep ensemble neural network," *International Journal on Artificial Intelligence Tools*, vol. 31, p. 2 250 049, 2022.
- [149] P. Saravanaprasad and S. Anbu Karuppusamy, "Advanced lung tumor diagnosis using a 3D deep neural network based CAD system," *Biomedical Signal Processing and Control*, vol. 89, p. 105 650, 2024.
- [150] S. Schulz, A.-C. Woerl, *et al.*, "Multimodal deep learning for prognosis prediction in renal cancer," *Frontiers in Oncology*, vol. 11, p. 788 740, 2021.

- [151] R. R. Selvaraju, M. Cogswell, *et al.*, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618–626.
- [152] 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*, vol. 149, p. 106 033, 2022.
- [153] P. M. Shakeel, M. Burhanuddin, *et al.*, “Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier,” *Neural Computing and Applications*, vol. 34, pp. 9579–9592, 2022.
- [154] G. A. Shanbhag, K. A. Prabhu, *et al.*, “Prediction of lung cancer using ensemble classifiers,” in *Journal of Physics: Conference Series*, IOP Publishing, vol. 2161, 2022, p. 012 007.
- [155] S. Sharif, R. A. Naqvi, *et al.*, “Deep perceptual enhancement for medical image analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, pp. 4826–4836, 2022.
- [156] D. Shen, G. Wu, *et al.*, “Deep learning in medical image analysis,” *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.
- [157] F. Shi, B. Chen, *et al.*, “Semi-supervised deep transfer learning for benign-malignant diagnosis of pulmonary nodules in chest CT images,” *IEEE Transactions on Medical Imaging*, vol. 41, pp. 771–781, 2022.
- [158] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, 2015, pp. 1–14.
- [159] D. Singh, V. Kumar, *et al.*, “Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks,” *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 39, pp. 1379–1389, 2020.
- [160] V. K. Singh, E. Y. Kalafi, *et al.*, “Prior wavelet knowledge for multi-modal medical image segmentation using a lightweight neural network with attention guided features,” *Expert Systems with Applications*, vol. 209, p. 118 166, 2022.
- [161] P. L. Suárez, A. D. Sappa, *et al.*, “Learning to colorize infrared images,” in *Trends in Cyber-Physical Multi-Agent Systems. Proceedings of the PAAMS Collection-15th International Conference, PAAMS 2017 15*, Springer, 2018, pp. 164–172.
- [162] R. Sudha and K. U. Maheswari, “Analysis on grading of lung nodule images with segmentation using U-Net and classification with convolutional neural network fish swarm optimization,” *Biocybernetics and Biomedical Engineering*, vol. 45, pp. 90–104, 2025.
- [163] P. Suetens, *Fundamentals of Medical Imaging*. Cambridge: Cambridge University Press, 2017.

- [164] C. Szegedy, S. Ioffe, *et al.*, “Inception-v4, Inception-ResNet and the impact of residual connections on learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, 2017, pp. 1–12.
- [165] C. Szegedy, W. Liu, *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [166] C. Szegedy, V. Vanhoucke, *et al.*, “Rethinking the Inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
- [167] S. Taylor, J. M. Brown, *et al.*, “Monitoring disease progression with a quantitative severity scale for retinopathy of prematurity using deep learning,” *JAMA Ophthalmology*, vol. 137, pp. 1022–1028, 2019.
- [168] R. Thirunavukarasu, R. Gnanasambandan, *et al.*, “Towards computational solutions for precision medicine based big data healthcare system using deep learning models: A review,” *Computers in Biology and Medicine*, vol. 149, p. 106 020, 2022.
- [169] R. Toda, A. Teramoto, *et al.*, “Lung cancer CT image generation from a free-form sketch using style-based pix2pix for data augmentation,” *Scientific Reports*, vol. 12, p. 12 867, 2022.
- [170] C. Tong, B. Liang, *et al.*, “Pulmonary nodule classification based on heterogeneous features learning,” *IEEE Journal on Selected Areas in Communications*, vol. 39, pp. 574–581, 2020.
- [171] Y. Tong, J. K. Udupa, *et al.*, “Prediction of lymphoma response to CAR T cells by deep learning-based image analysis,” *PLoS ONE*, vol. 18, e0282573, 2023.
- [172] M. Usama, B. Ahmad, *et al.*, “Deep feature learning for disease risk assessment based on convolutional neural network with intra-layer recurrent connection by using hospital big data,” *IEEE Access*, vol. 6, pp. 67 927–67 939, 2018.
- [173] M. Usman, B.-D. Lee, *et al.*, “Volumetric lung nodule segmentation using adaptive ROI with multi-view residual learning,” *Scientific Reports*, vol. 10, p. 12 839, 2020.
- [174] B. J. Van, Y. K. Dewaraja, *et al.*, “Y-90 SIRT: evaluation of TCP variation across dosimetric models,” *EJNMMI physics*, vol. 8, p. 45, 2021.
- [175] A. Vinisha and R. Boda, “DeepBrainTumorNet: An effective framework of heuristic-aided brain tumour detection and classification system using residual attention-multiscale dilated inception network,” *Biomedical Signal Processing and Control*, vol. 100, p. 107 180, 2025.
- [176] G. Wang, W. Li, *et al.*, “Interactive medical image segmentation using deep learning with image-specific fine tuning,” *IEEE Transactions on Medical Imaging*, vol. 37, pp. 1562–1573, 2018.
- [177] L. Wang, H. Wang, *et al.*, “Trends in the application of deep learning networks in medical image analysis: Evolution between 2012 and 2020,” *European Journal of Radiology*, vol. 146, p. 110 069, 2022.

- [178] 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*, vol. 67, p. 175 010, 2022.
- [179] S. Wang, Y. Zha, *et al.*, “A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis,” *European Respiratory Journal*, vol. 56, 2000775: Pages–9, 2020.
- [180] W. Wang and G. Chakraborty, “Automatic prognosis of lung cancer using heterogeneous deep learning models for nodule detection and eliciting its morphological features,” *Applied Intelligence*, vol. 51, pp. 2471–2484, 2021.
- [181] W. Wang and G. Chakraborty, “Evaluation of malignancy of lung nodules from CT image using recurrent neural network,” in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics*, IEEE, 2019, pp. 2992–2997.
- [182] X. Wang, L. Jiang, *et al.*, “Joint learning of 3D lesion segmentation and classification for explainable COVID-19 diagnosis,” *IEEE Transactions on Medical Imaging*, vol. 40, pp. 2463–2476, 2021.
- [183] Z. Wang, J. Men, and F. Zhang, “Improved V-Net lung nodule segmentation method based on selective kernel,” *Signal, Image and Video Processing*, vol. 17, pp. 1763–1774, 2023.
- [184] 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*, vol. 24, pp. 3529–3538, 2020.
- [185] W. Wu, L. Gao, *et al.*, “Segmentation of pulmonary nodules in CT images based on 3D-UNET combined with three-dimensional conditional random field optimization,” *Medical Physics*, vol. 47, pp. 4054–4063, 2020.
- [186] X. Wu, H. Hui, *et al.*, “Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study,” *European Journal of Radiology*, vol. 128, p. 109 041, 2020.
- [187] Y. H. Wu, S. H. Gao, *et al.*, “JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation,” *IEEE Transactions on Image Processing*, vol. 30, pp. 3113–3126, 2021.
- [188] K. Xia, H. Yin, *et al.*, “A novel improved deep convolutional neural network model for medical image fusion,” *Cluster Computing*, vol. 22, pp. 1515–1527, 2019.
- [189] J. Xiao, L. Xie, C. He, and X. Jiang, “Dynamic classifier ensemble model for customer classification with imbalanced class distribution,” *Expert Systems with Applications*, vol. 39, pp. 3668–3675, 2012.
- [190] H. Xie, D. Yang, *et al.*, “Automated pulmonary nodule detection in CT images using deep convolutional neural networks,” *Pattern Recognition*, vol. 85, pp. 109–119, 2019.

- [191] L. Xu, Y. Si, *et al.*, “Medical image fusion using a modified shark smell optimization algorithm and hybrid wavelet-homomorphic filter,” *Biomedical Signal Processing and Control*, vol. 59, p. 101 885, 2020.
- [192] P. Yadav, N. Menon, *et al.*, “Lung-GANs: Unsupervised representation learning for lung disease classification using chest CT and X-ray images,” *IEEE Transactions on Engineering Management*, vol. 70, pp. 2774–2786, 2023.
- [193] C. Yang and Z. Zhang, “PFD-Net: Pyramid fourier deformable network for medical image segmentation,” *Computers in Biology and Medicine*, vol. 172, p. 108 302, 2024.
- [194] A. Zakeri, Y. Xia, N. Ravikumar, and A. F. Frangi, “Chapter 17 - Deep learning for vision and representation learning,” in *Medical Image Analysis*, ser. MICCAI Society, Academic Press, 2024, pp. 451–474.
- [195] B. Zhang, S. Qi, *et al.*, “Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images,” *Computer Methods and Programs in Biomedicine*, vol. 222, p. 106 946, 2022.
- [196] 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*, vol. 181, 1423–1433.e11, 2020.
- [197] Y. Zhang, S. Liu, *et al.*, “Rethinking the Dice loss for deep learning lesion segmentation in medical images,” *Journal of Shanghai Jiaotong University (Science)*, vol. 26, pp. 93–102, 2021.
- [198] Z. Zhang and M. Sabuncu, “Generalized cross entropy loss for training deep neural networks with noisy labels,” *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [199] B. Zhao, L. P. James, *et al.*, “Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer,” *Radiology*, vol. 252, pp. 263–272, 2009.
- [200] Z. Zhao, K. Moller, *et al.*, “Determination of lung area in EIT images,” in *Proceedings of the 3rd International Conference on Bioinformatics and Biomedical Engineering*, IEEE, 2009, pp. 1–4.
- [201] C. Zhou, C. Ding, *et al.*, “One-pass multi-task networks with cross-task guided attention for brain tumor segmentation,” *IEEE Transactions on Image Processing*, vol. 29, pp. 4516–4529, 2020.
- [202] D. Zhou, J. Fang, *et al.*, “IoU loss for 2D/3D object detection,” in *Proceedings of the International Conference on 3D Vision*, vol. 3DV, 2019, pp. 85–94.
- [203] L. Zhou and C. D. Hansen, “A survey of colormaps in visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, pp. 2051–2069, 2015.
- [204] 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*, vol. 4, pp. 494–503, 2022.

- [205] P. Zhou, Y. Li, *et al.*, “Coco-attention for tumor segmentation in weakly paired multimodal MRI images,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, pp. 2944–2955, 2023.
- [206] Z. Zhou, M. M. R. Siddiquee, *et al.*, “UNet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1856–1867, 2019.

