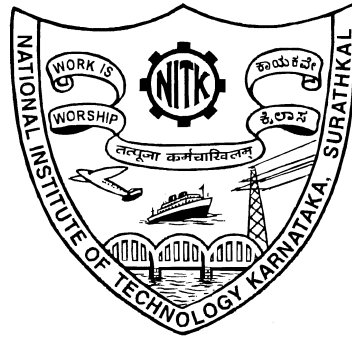


Deep Learning For Nuclei Segmentation and Classification of Histopathology Images

Thesis

Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by
AMIT KUMAR CHANCHAL



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING,
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,
SURATHKAL, MANGALORE -575025

APRIL, 2023

DECLARATION

I hereby *declare* that the Research Thesis entitled **Deep Learning For Nuclei Segmentation and Classification of Histopathology Images** which is being submitted to the *National Institute of Technology Karnataka, Surathkal* in partial fulfillment of the requirements for the award of the Degree of *Doctor of Philosophy* in **Department of Electronics and Communication Engineering** is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.


Amit Kumar Chanchal,

Reg.No: 187075EC001

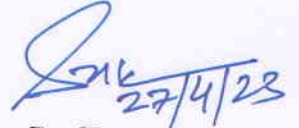
Department of Electronics and Communication Engineering.

Place: NITK-Surathkal.

Date: 27-04-2023

CERTIFICATE

This is to *certify* that the Research Thesis entitled **Deep Learning For Nuclei Segmentation and Classification of Histopathology Images** submitted by **AMIT KUMAR CHANCHAL** (Register Number: 187075EC001) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of degree of **Doctor of Philosophy**.



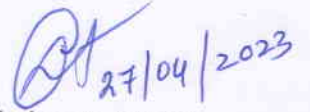
Dr. Shyam Lal

Research Guide

Assistant Professor

Dept. of Electronics and Communication Engg.

NITK Surathkal-575025



Chairman-DRPC

(Signature with Date and Seal)

प्राध्यापक एवं विभागाध्यक्ष / PROF & HEAD
ई एवं सी विभाग / E & C Department
एन आई टी के, सुरतकल/NITK, Surathkal
मंगलूर / MANGALORE - 575 025

Acknowledgements

First and foremost, I would like to thank God almighty for his blessings, giving me the opportunity, knowledge, and strength to undertake this research study and to persevere and complete it satisfactorily.

I offer my sincerest gratitude to my research supervisor Dr. Shyam Lal, who brilliantly guided me throughout, patiently sharing his knowledge and experience, insightful discussions while carrying out research, and having confidence in me. I am very much impressed with your positive energy and understanding attitude. I am deeply expressing my appreciation towards you for everything, and I promise that I will always be responsible for you.

A special word of gratitude to Prof. T. Laxminidhi, Prof. Ashvini Chaturvedi the former Heads, and Prof. Neelawar Shekar Vittal Shet, the present Head, Department of Electronics and Communication Engineering for their invaluable advice and administrative support. My heartfelt thanks to Dr. Raghavendra B.S. and Dr. Ajay Kumar Yadav, RPAC members, for their assessment and useful guidance. Both have contributed significantly by providing valuable suggestions to uplift the content and quality of the research. I am grateful to Dr. Ramesh Kini M. for extending his helping hand, in need.

I thank all my fellow research scholars at NITK for their friendship, motivation, and support. I am thankful to the technicians of the laboratory of EC department, NITK, for providing all computing resources. I am particularly grateful for the assistance given by all teaching and non-teaching faculty of EC department, NITK.

I greatly acknowledge the Science Engineering and Research Board, Department of Science and Technology, Govt. of India funded project entitled "Design and Development of Automated Kidney Cancer Detection System from H&E Stained Kidney Histopathological Images." for the financial assistance under special grants, (Grant No. EEG/2018/000323, 2019) which enabled me to pursue the research work.

Nobody has been more important to me in the pursuit of this research work than the members of my family. I express my warm gratitude to my wife, Preeti Kumari, who is so understanding, and encouraging, and kept my spirits always high. I recognize your presence and faithful support throughout this journey.

My son, Aditya, made this journey joyful for me. I appreciate his understanding and pleasurable disturbances which always relieve my stress and pressure. I will never forget that most of the time you were seated overnight with me and waiting for my work to be done. I thank my whole family for providing me with unfailing support and continuous encouragement throughout my entire tenure of study and research.

Dedicated to
MY
FAMILY AND TEACHERS

Abstract

To improve the process of diagnosis and treatment of cancer disease, automatic segmentation and classification of haematoxylin and eosin (H & E) stained histopathology images are important steps in digital pathology. The advent of new computation systems like GPU, fast digital scanners, and the availability of lots of data, Deep Learning (DL) techniques have shown superior performance in different applications of medical image analysis. The potential and applicability of deep learning models for the analysis of histopathology images have been demonstrated by many researchers. Due to variations in the appearance and complex clinical structure of histopathology slides, reported results still needed to be improved for accurate diagnosis of disease. An accurate and efficient classification algorithm that exactly resembles the clinical feature of cancer disease is still open-ended research. This thesis investigates a detailed methodology for the design and implementation of deep learning architectures which includes nuclei detection and segmentation, characterization of subtypes of cancer, and grading of histopathological tissues.

In the first part of the thesis, the analysis of histopathology images by using efficient segmentation algorithms is presented. In this study, an effective encoder-decoder architecture with a separable convolution pyramid pooling network (SCPP-Net) is designed and implemented for automatically segmenting complex nuclei present in digital histopathology images. The SCPP unit focuses on two aspects: first, it increases the receptive field by varying four different dilation rates, keeping the kernel size fixed, and second, it reduces the trainable parameter by using depth-wise separable convolution. For multi-organ histopathology analysis, a new deep learning framework is proposed, that consists of a high-resolution encoder path, an atrous spatial pyramid pooling (ASPP) bottleneck module, and a powerful decoder. The proposed network is wide and deep that effectively leverages the strength of residual learning as well as encoder-decoder architecture. The problem of the vanished boundary of detected nuclei is addressed by proposing an efficient loss function that better trains the proposed deep structured residual encoder-decoder network (DSREDN) and reduces the false prediction. The obtained score of nuclei segmentation indicated that the proposed architectures achieved a considerable margin over state-of-the-art deep learning models on three different publicly

available histopathology image datasets.

Next, in the thesis, a novel dataset and an efficient deep-learning framework for the classification of subtypes of renal cell carcinoma (RCC) from kidney histopathological images are proposed. The proposed RenalNet is intended to capture cross-channel and inter-spatial features at three different scales parallelly and held them together. The proposed model contains a new convolutional neural network (CNN) block called multiple channel residual transformation (MCRT), to focus on the most relevant morphological features of RCC by fusing the information of multiple paths. Further, to improve the network's representation power, a novel block called group convolution deep localization (GCDL) is introduced that effectively integrates three different feature descriptors. A new benchmark dataset for the classification of subtypes of RCC from kidney histopathology images is also introduced as a part of this study. The results of the proposed model are compared with the existing DL models trained from scratch as well as networks leveraged by transfer learning of pre-trained weights. During the experimentation, the proposed network achieved an *accuracy* 91.67%, and *F1-Score* 91.65% on the proposed kidney dataset that is the highest among all competitive models. The experimental results show that the proposed RenalNet architecture is best in terms of training and prediction time, classification accuracy, F1 score, and computational complexity.

A pathologist report affirmed that the stage and grade of diagnosis is the most important prognostic factor. In these cases, continuous staging and grading evaluation is extremely important for the clinical management of patients. This study proposed a robust and computationally efficient fully automated Renal Cell Carcinoma Grading Network (RCCGNet) from kidney histopathology images. The proposed shared channel residual (SCR) block shares the information between two different layers and operates the shared data separately by providing beneficial supplements to each other. As a part of this study, a new dataset also has been introduced for the grading of RCC with five different grades. The simulation results include deep learning models trained from scratch as well as transfer learning techniques using pre-trained weights of the ImageNet. The performance of the proposed RCCGNet is evaluated by the most preferred quality metrics and achieved 90.14% of *accuracy*, and 89.06% *F1-score* on the introduced kidney dataset.

Another proposed architecture is called Robust CNN (RoCNN) for grading (Normal, Grade-1, Grade-2, Grade-3, and Grade-4) and classification (Normal, KIRC, KIRP, KICH) in kidney cancer tissue. To demonstrate that the proposed model is generalized and independent of the dataset, it has experimented on two well-known datasets, the KMC kidney dataset of five different grades and the TCGA dataset of four classes. The RoCNN is capable of learning features at varying convolutional filter sizes because of the inception modules employed in it. Squeeze and Extract (SE) blocks are used to remove unnecessary contributions from noisy channels and improve model accuracy. Regarding the computational complexities the proposed RoCNN is extremely efficient compared to the reference models. Due to a substantial reduction in the computational complexities, incorporation of the proposed method into FPGA board processing for next-generation histopathological image analysis is a significant step in the right direction. Compared to the best-performing state-of-the-art model the accuracy of RoCNN shows significant improvement of about 4.22% and 3.01% for two different datasets.

All proposed deep learning algorithms evolved to be the most promising, stable, and computationally efficient for the analysis of histopathological images.

Keywords: Deep learning, Kidney cancer, Breast cancer, Nuclei segmentation, Histopathology images, Renal cell carcinoma, Cancer detection, RCC grading, Cancer classification, Residual Learning, Dilated convolution

Contents

Acknowledgements	i
Abstract	i
List of Figures	viii
List of Tables	xii
Nomenclature	xv
Abbreviations	xv
1 INTRODUCTION	1
1.1 Motivations	3
1.2 Histopathology Image Analysis	3
1.2.1 Preparing histopathology slides	7
1.3 Deep Learning	8
1.4 Deep Learning Tools	10
1.4.1 Activation function	10
1.4.2 Loss function	11
1.4.3 Gradient descent optimization methods	14
1.4.4 Handling overfitting problems	15
1.4.5 Batch normalization	16
1.5 Functional Diagram of a Deep Learning Architecture for Histopathology Image Segmentation	17
1.6 Functional Diagram of a Deep Learning Architecture for Histopathology Image Classification	19
1.7 Problem Statement	21
1.8 Research Objectives	21
1.9 Main Contributions of the Thesis	21
1.10 Organization of the Thesis	23

2	LITERATURE REVIEW	25
2.1	Introduction	25
2.2	Literature Survey on Segmentation Techniques	25
2.2.1	Literature survey of conventional segmentation techniques	26
2.2.2	Literature survey on CNN based segmentation approaches	29
2.3	Literature Survey on Classification and Grading	35
2.3.1	Deep learning solutions for classification of H&E stained images	35
2.3.2	Techniques to extract relevant information	38
2.3.3	Techniques to retain extracted features	38
2.3.4	Deep learning models which utilize channel shuffle	39
2.4	Research Gap Analysis	40
2.4.1	Gap analysis for nuclei segmentation from H&E stained histopathology images	40
2.4.2	Gap analysis for classification of subtypes of histopathology images	40
2.4.3	Gap analysis for grade prediction of histopathology images	41
2.5	Quality Metrics for Nuclei Segmentation, Classification, and Grading of Histopathology Images	41
3	Deep Learning Architectures for Nuclei Segmentation of Histopathology Images	45
3.1	Introduction	45
3.2	Proposed Architecture-1	46
3.2.1	Description of SCPP-Net architecture	46
3.3	Implementation and Training	51
3.3.1	Datasets used for nuclei segmentation	52
3.3.2	Training details	52
3.4	Results and Discussions	53
3.4.1	Ablation study	53
3.4.2	Comparison with state-of-the-art segmentation methods	55
3.4.3	Computational complexity analysis	61
3.5	Proposed Architecture-2	62
3.5.1	High resolution layer	63
3.6	Implementation and Training	65
3.7	Results and Discussions	65

3.7.1	Ablation study	66
3.7.2	Comparison with state-of-the-art segmentation methods	67
3.7.3	Computational complexity analysis	70
3.8	Proposed Architecture-3 and a Novel Loss Function	70
3.8.1	Proposed loss function	72
3.9	Implementation and Training	74
3.10	Results and Discussions	75
3.10.1	Ablation study	75
3.10.2	Comparison with state-of-the-art segmentation methods	76
3.10.3	Computational complexity analysis	81
3.11	Summary	82
4	Deep Learning Architecture for Classification of Histopathology Images	85
4.1	Introduction	85
4.2	Proposed Method	86
4.2.1	Multiple channel residual transformation	86
4.2.2	Group convolution deep localization network	90
4.3	Implementation and Training	92
4.3.1	Novel kidney histopathology dataset	92
4.3.2	Training details	94
4.4	Results and Discussions	95
4.4.1	Ablation study	95
4.4.2	Comparison with state-of-the-art classification methods	98
4.4.3	Cross-validation	103
4.4.4	Statistical analysis	106
4.4.5	Computational complexity analysis	107
4.5	Summary	107
5	Deep Learning Architectures for Grading of Histopathology Images	109
5.1	Introduction	109
5.2	Proposed Architecture-1	109
5.2.1	Shared channel residual block	111
5.3	Implementation and Training	114
5.3.1	Novel RCC grading dataset	115
5.3.2	Training details	116

5.4	Results and Discussions	118
5.4.1	Ablation study	118
5.4.2	Comparison with state-of-the-art models	122
5.4.3	Cross-validation	126
5.4.4	Statistical analysis	126
5.4.5	Computational complexity analysis	127
5.5	Proposed Architecture-2	128
5.5.1	Inception-ResNet block	129
5.5.2	Squeeze and excitation block	130
5.5.3	Hyper-column	132
5.6	Implementation and Training	132
5.6.1	Dataset	132
5.6.2	Training details	133
5.7	Results and Discussions	134
5.7.1	Ablation study	135
5.7.2	Comparison with state-of-the-art models	136
5.7.3	Cross-validation	142
5.7.4	Computational complexity	144
5.8	Summary	144
6	Conclusion and Future Work	147
6.1	Conclusion	147
6.2	Future Work	149
	Bibliography	150
	References	163
	Publications based on the Thesis	165

List of Figures

1.1	Nuclear characteristics of normal and cancerous breast tissues	4
1.2	Visualization of histology and clinical features of RCC	5
1.3	Visualization of different grades of renal tumor (Normal to Grade-4 from left to right)	8
1.4	Visualization of different classes of Hepatocellular carcinoma (HCC) (Normal, Low sub-type liver HCC tumor, medium sub-type liver HCC tumor, high sub-type liver HCC tumor- from left to right)	8
1.5	A systematic view of CAD system for cancerous/non-cancerous classification	8
1.6	Block diagram of preparation of histopathology slides	9
1.7	Comparison of traditional approach versus deep learning method	9
1.8	Internal covariate shift	17
1.9	Functional diagram of image segmentation from histopathology images . .	18
1.10	Functional diagram of image classification from histopathology images . .	20
2.1	Confusion matrix (a) Binary class problem (b) Multi-class problem	44
3.1	Detailed diagram of proposed separable convolution pyramid pooling network (SCPP-Net)	46
3.2	Max-pooling with kernel $K=2$ and stride $S=2$	48
3.3	Visualization of the receptive field (global view) of atrous convolution with different dilation rates	49
3.4	Visualization of depth-wise separable convolution	49
3.5	Separable convolution pyramid pooling block	50
3.6	Binary and overlay images of proposed SCPP-Net architecture (with and without separable convolution) on kidney and TNBC datasets	54
3.7	Visual segmentation results comparison of different models on kidney dataset	56
3.8	Visual segmentation results comparison of different models on TNBC dataset	57

3.9	Visual segmentation results comparison of different models on MoNuSeg dataset	58
3.10	(a), (c), (e)-Training and validation accuracy for Kidney, TNBC, and MoNuSeg datasets. (b), (d), (f) Training and validation loss for Kidney, TNBC, and MoNuSeg datasets.	59
3.11	Computational complexity comparison of different CNN models	62
3.12	Architecture of the ASPPU model	64
3.13	Proposed high resolution deep transferred ASPPU-Net	64
3.14	High resolution encoder path, ASPP bottleneck path and decoder path . . .	65
3.15	Comparison of visual segmentation of different models of two sample images	68
3.16	(a), (c), (e)-Training and validation accuracy for Kidney, TNBC, and MoNuSeg datasets. (b), (d), (f) Training and validation loss for Kidney, TNBC, and MoNuSeg datasets.	69
3.17	Proposed deep structured residual encoder-decoder network	71
3.18	Accuracy and loss plot (Kidney Dataset), (a), (c) - Learning curve using proposed loss, (b), (d) - Learning curve using BCE loss	76
3.19	Accuracy and loss plot (TNBC Dataset), (a), (c) - Learning curve using proposed loss, (b), (d) - Learning curve using BCE loss	77
3.20	Accuracy and loss plot (MoNuSeg Dataset), (a), (c) - Learning curve using proposed loss, (b), (d) - Learning curve using BCE loss	78
3.21	Comparison of predicted nuclear regions of different models	79
4.1	Structure of proposed RenalNet model for classification of RCC subtypes and normal renal tumor from H&E stained histopathology data	87
4.2	Multiple channel residual transformation network	88
4.3	Visualization of depth-wise separable convolution	92
4.4	Visual performance comparison of different CNN variations using activation maps. (Red: Very high probability score regions, Orange: Medium probability score regions, Light blue: Low probability score regions.) . . .	97
4.5	Learning curve of RenalNet on introduced TCGA kidney dataset (a) Comparison of validation accuracy of state-of-the-art models (b) Training and validation loss of three trials (c) Confusion matrix of best trial (d) ROC-AUC of test data for four class classification class-0 (KICH) class-1 (KIRC), class-2 (KIRP), class-3 (Normal)	100

4.6	Learning curve of RenalNet on KMC liver dataset (a) Comparison of validation accuracy of state-of-the-art models (b) Training and validation loss of three trials (c) Confusion matrix of best trial (d) ROC-AUC of test data for four class classification class-0 (HCC-1) class-1 (HCC-2), class-2 (HCC-3), class-3 (HCC-4)	103
4.7	Learning curve of RenalNet on TCGA liver dataset (a) Comparison of validation accuracy of state-of-the-art models (b) Training and validation loss of three trials (c) Confusion matrix of best trial (d) ROC-AUC of test data for three class classification class-0 (Normal) class-1 (Primary), class-2 (Secondary)	104
4.8	5-Fold cross-validation learning curve of RenalNet on newly introduced TCGA kidney dataset. (a) to (e) Comparison of validation accuracy of state-of-the-art models of five different fold	105
4.9	Box Plot: (a), (b), (c) Accuracy of different models for new TCGA kidney dataset, KMC Liver dataset, TCGA Liver dataset.	106
5.1	Structure of the proposed RCCGNet for grading of renal cell carcinoma from kidney histopathology images	110
5.2	Shared channel residual block	112
5.3	Comparison of intermediate features with and without channel sharing between layers	119
5.4	Visual comparison of different CNN variations using activation maps. (Red: Very high probability score regions, Orange: Medium probability score regions, Light blue: Low probability score regions.)	120
5.5	Learning curve of RCCGNet on introduced KMC kidney dataset (a) Training and validation accuracy of the proposed model (b) Training and validation loss of proposed model (c) Confusion matrix of proposed model (d)-(h) Receiver operating characteristic (ROC) curve of five best-performing state-of-the-art models using one versus rest approach.	123
5.6	Box Plot: (a), (b) Accuracy, and F1 Score of different models for new KMC kidney dataset. (c), and (d) Accuracy and F1 Score of different models for BreakHis dataset.	125
5.7	Complete overview of the proposed architecture	130
5.8	Inception-ResNet block	131
5.9	Inception-SE-ResNet block	131

5.10 Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on KMC kidney dataset (<i>cont.</i>)	138
5.10 Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on KMC kidney dataset	139
5.11 Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on TCGA dataset (<i>cont.</i>)	140
5.11 Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on TCGA dataset	141

List of Tables

1.1	Histology and clinical features of RCC subtypes and normal tissue regions at 40X resolution.	6
1.2	Microscopic features of different grades of renal tumor (Normal to Grade-4 from left to right)	7
2.1	Summary of conventional segmentation techniques	27
2.2	Summary of state-of-the-art DL techniques used for segmentation of medical images	32
2.3	A summary of existing deep learning solutions for classification of H&E stained images	36
3.1	Performance measure of proposed model with kidney dataset	54
3.2	Performance measure of proposed model with TNBC dataset	55
3.3	Performance measure of proposed model with MoNuSeg dataset	55
3.4	Performance comparison of architectures on kidney dataset	55
3.5	Performance comparison of architectures on TNBC dataset	60
3.6	Performance comparison of architectures on MoNuSeg dataset	60
3.7	Comparison of different variants of proposed model on Kidney dataset . . .	66
3.8	Comparison of different variants of proposed model on TNBC dataset . . .	66
3.9	Comparison of different variants of proposed model on MoNuSeg dataset .	66
3.11	Experimental performance comparison of different architectures with three datasets	67
3.12	Network computational complexity in millions	70
3.13	F1/AJI metric performance comparison of Unet Model and Proposed model on different datasets	75
3.14	Performance comparison of different models with three datasets	80
3.15	Computational complexity comparison of different models	82

4.1	Datasets class distribution at 40X	92
4.2	Details of hyper-parameters	94
4.3	Overall quality comparison of eight variants of proposed RenalNet MCRT: Multiple channel residual transformation, GCDL: Group convolution deep localization, ADSC:Aggregated depthwise separable convolution, AM: Attention mechanism, SE with excitation ratio (R)= 2, 4, 8.	96
4.4	Performance metrics comparison of proposed RenalNet with other competitive model on introduced TCGA kidney dataset	99
4.5	Performance metrics comparison of proposed RenalNet architecture with other competitive model on KMC liver dataset	101
4.6	Performance metrics comparison of proposed RenalNet architecture with other competitive model on TCGA liver dataset	101
4.7	Performance metrics comparison of proposed RenalNet trained from scratch with other models using pre-trained weights from the ImageNet weight on introduced kidney dataset	102
4.8	5-Fold cross validation average quality metrics comparison of proposed RenalNet with other competitive models on introduced kidney dataset	102
4.9	Computational complexity comparison of architectures	107
5.1	Introduced dataset grades distribution	116
5.2	BreaKHis dataset at 40X magnification	117
5.3	Details of hyper parameters	117
5.4	Overall quality comparison of three variants of proposed RCCGNet on new KMC Kidney dataset	119
5.5	Overall quality comparison of three variants of proposed RCCGNet on BreakHis dataset	120
5.6	Performance metrics comparison of proposed model with other competitive models on KMC kidney dataset	122
5.7	Overall quality comparison of the proposed model with other competitive models (BreakHis dataset-Eight Class)	124
5.8	3-Fold cross validation average quality metrics comparison of proposed RCCGNet with other competitive models (KMC kidney dataset)	127
5.9	Computational complexity comparison of architectures	127
5.10	TCGA dataset class distribution	133
5.11	Details of hyper parameters	134

5.12	Summary of the ablation study	136
5.13	Comparative study of detection of grades using KMC kidney dataset	137
5.14	Comparative study of prediction of classes on TCGA dataset	142
5.15	Computational complexity comparison of the suggested architecture against various comparison architecture	142
5.16	Average quality comparison of three-fold KMC kidney dataset	143
5.17	Average quality comparison of five-fold TCGA dataset	144

Nomenclature

Symbol	Meaning	Symbol	Meaning
I	Image	K	Kernel
f	Filter size	W, H	Width and height of the image
p	Amount of padding	L	Number of layers in network
N	Number of pixels in the image	Y_i	Pixel values in the ground
Y_i	Predicted probabilities of the class-wise pixels	$\log(p)$	Logarithmic of probability
L_{BCE}	Binary Cross Entropy loss	T	Transformation
$L_{Composite}$	Composite loss function	C	Channel depth
$*$	Convolution	W_1^i	Weight vector
r	Dilation rate	\odot	Concatenation operation
p_i	Probabilities of outcomes	$f_{sig}(x)$	Sigmoid activation
$f'_{sig}(x)$	Derivative of sigmoid	$f_{anh}(x)$	Tanh activation
$f'_{tanh}(x)$	Derivative of Tanh	$f_{ReLU}(x)$	ReLU activation
$f'_{ReLU}(x)$	Derivative of ReLU	y	Actual output
\hat{y}	Predicted output	β	Used to assign weight
η	Learning rate	λ	Regularization parameter
ϕ	Parameter of softmax classifier	L	Loss
W_{New}	Updated weight	W_{Old}	Old weight
FP	False positive	FN	False negative
TP	True positive	TN	True negative
(i, j)	pixel indices	$\frac{\delta L}{\delta w}$	Derivative of loss with respect to weight
$\ W\ _1$	L1 Norm	$\ W\ _2^2$	L2 Norm
min	Minimum value	max	Maximum value

Abbreviations

Abbreviation	Expansion
DL	Deep Learning
AI	Artificial Intelligence
DNNs	Deep Learning Neural Networks
GCNN	Group Equivalent Convolutional Neural Networks
WHO	World Health Organization
ISUP	International Society of Urologic Pathology
H&E	Hematoxylin & Eosin
TNBC	Triple Negative Breast Cancer
FLOPs	Floating Point Operations
ReLU	Rectified Linear Unit
CNN	Convolutional Neural Network
MoNuSeg	Multi-Organ Nuclei Segmentation
BCE	Binary Cross Entropy
WBCE	Weighted Binary Cross-Entropy
WSI	Whole Slide Imaging
CAD	Computer Assisted Diagnosis
IoU	Intersection Over Union
ASPP	Atrous Spatial Pyramid Pooling
SCPP-Net	Separable Convolution Pyramid Pooling Network
BHCNet	Breast Cancer Histopathological Image Classification
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
GD	Gradient Descent
SGD	Stochastic Gradient Descent
NAG	Nesterov Accelerated Gradient
Adam	Adaptive Moment Estimation
GPU	Graphics Processing Unit
RCC	Renal Cell Carcinoma
KIRC	Renal Clear Cell Carcinoma
KIRP	Renal Papillary Cell Carcinoma
KICH	Kidney Chromophobe
VHL	Von Hippel Lindau
MCRT	Multiple Channel Residual Transformation
GCDL	Group Convolution Deep Localization
DCIS	Ductal Carcinoma Insitu
SE	Squeeze and Excitation
TCGA	The Cancer Genome Atlas
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

Abbreviation	Expansion
AJI	Aggregated Jaccard Index
RCCGNet	Renal Cell Carcinoma Grading Network
SCR Block	Shared Channel Residual Block
ROC Curve	Receiver Operating Characteristic Curve
AUC	Area Under the Curve
BN	Batch Normalization
ROI	Region of Interest
GP	Global Pooling
FCN	Fully Connected Network
PP	Pyramid Pooling
ELU	Exponential Linear Unit
RNN	Recurrent Neural Network
BreaKHis	Breast Histopathology Image
HMEDN	High Resolution Multi-Scale Encoder-Decoder Network

Chapter 1

INTRODUCTION

Good health is a precious asset for us. Our ability to work efficiently, learn new things, serve the nation, and provide support to ourselves and our families depends on our physical and mental fitness. If we are not fit, our families and communities lag in all respects. World Health Organization (WHO) annual world health statistics reports present the most recent health statistics for the WHO member states, where the report also summarizes the impact of increasing cancer cases worldwide. Cancer is currently considered to be the leading cause of death globally. Breast, kidney, liver, prostate, colon, and stomach are the affected organs and the most common causes of death due to cancer. Genetic factors, tobacco exposure, obesity, and hypertension are the most significant and generally accepted risk factor for cancer disease.

Currently, the clinical examination steps to diagnose cancer disease usually begins with basic imaging tests like CT scans, magnetic resonance imaging (MRI), nuclear scan, and others. If the doctor suspects any unusual growth through these tests then the next step is a biopsy. A biopsy is a procedure in which a doctor removes a sample of diseased tissue and a pathologist looks at this tissue under a microscope and decides whether the tissue is cancerous or not. Digital histopathological images are obtained from the microscopic examination of the stained biopsy tissues. Once it is confirmed that the tissue is cancerous then they need to do some more tests to figure out the stage and grade of cancer. Identification of stage and grade is very important for planning further treatment. A pathologist report affirmed that the stage of diagnosis is the most important prognostic factor. In these cases, continuous staging evaluation is extremely important for the clinical management of patients. The above process is time-consuming, prone to human error, and highly depends on the expertise of a pathologist. Early detection and classification of cancer tissues enable doctors and prac-

titioners to decide the further course of treatment. Therefore, fast and precise analysis of cancerous tissue images is extremely important for proper diagnosis. With the exponential growth of cancer cases, the computer-assisted diagnosis (CAD) of histopathology images has been used in preference and replaced the manual diagnosis for fast and precise assessment.

To improve the process of diagnosis and treatment of cancer disease, automatic segmentation of haematoxylin and eosin (H & E) stained cell nuclei from histopathology images is the first step in digital pathology. Preparation of histopathology slides for automatic segmentation involves tissue collection, fixation, embedding, sectioning, staining, and visualization. Development in sophisticated microscopes has enabled the collection of high-quality histopathology images in large quantities. Nuclei segmentation can be categorized into conventional and CNN-based deep learning segmentation approaches. The traditional method of image segmentation includes the discontinuity-based approach, similarity-based approach, global clustering, superpixel segmentation, watershed segmentation, active contour segmentation techniques, etc. These methods performed better in the field of biomedical application but the main challenge is to deal with complex histology images where conventional techniques result in under-segmentation or over-segmentation. Deep learning-based segmentation methods has gained significant popularity and attention in digital pathology due to its promising results and high prediction rate.

Classification and grading of different subtypes of cancerous tissue is an important prognostic parameter for the diagnosis of cancer disease. Due to the advent of new computation systems like GPU, fast digital scanners, and the availability of lots of data, Deep Learning (DL) techniques have shown superior performance in different applications of medical image analysis. To get optimal speed and accuracy, many efforts have been made to design deep learning architectures computationally efficient and lightweight. Many works retain spatial information by incorporating skip connections to avoid degradation of information between output and input and utilized all preceding layers as input to the next layer throughout the network. The concept of attention in deep learning has been widely explored in recent decades. To focus on the most informative component of the input images, the attention or gating mechanism is incorporated in diverse application domains. The application of deep learning to analyze the histopathological images of kidney, breast, liver, prostate, colon, and other organs includes a number of tasks such as nuclei detection and segmentation, characterization of subtypes of cancer, and grading.

1.1 Motivations

Cancer is a major threat to life. Early detection and classification of cancer tissues enable doctors and practitioners to decide the further course of treatment. A fully automated and precise method of cancer detection from histopathology images is in high demand for identifying malignant tumors and their prognosis. Grading of the cancerous cell describes how much cancer cell is abnormal compared to normal tissue. With the knowledge of proper grade, pathologists get an idea that how quickly it will grow and how much spread to other parts of the body. Therefore continuous staging evaluation is extremely important for the clinical management of patients and early detection of cancer stage and grade is today's necessity. Deep learning algorithms like CNNs and the easy availability of high-end GPUs to run fully automated algorithms have greatly improved the results in all image classification applications, including histopathology image-based cancer classification and grading.

The deep learning-based feature extraction method has gained significant popularity and attention in digital pathology due to its promising results and high prediction rate. Convolutional neural networks (CNNs) have shown superior performance in the field of digital pathology which involves tasks such as pixel-wise segmentation, instance-segmentation, classification of different classes, and grading of cancers. The deep learning framework is driven by large amounts of data, an activation function, an effective optimization algorithm, and a powerful loss function. The deep neural network has achieved a dominant breakthrough in digital pathology and computer vision tasks. For automatic diagnosis and prognosis of cancer disease from H&E-stained histology data, the design of an efficient and robust computer-aided system is in high demand. Recent trends of cancer incidence worldwide are expected to increase persistently and this inspires the modification in the traditional diagnosis system to respond to future challenges.

1.2 Histopathology Image Analysis

A histopathology image consists of hematoxylin stains cell cores called nuclei, and eosin stain cells with other connected tissue called cytoplasm which is of pink shading. As per (Elston and Ellis, 1991), it has been discovered that the state of cancer cells may be bigger or smaller than ordinary cells, and that the nuclei of cancerous cells are larger than normal cells and deviate from the center of mass. The nucleus of a cancer cell is darker than that of a normal cell. Histopathology refers to the microscopic examination of tissue or surgical

specimen by pathologists. A pathologist report by (Washington *et al.*, 2009), affirmed that the grade of a cancer can be determined by the size, shape, and other morphology of the associated tissue. The stage or extent of cancer in the body is determined by tumor size, location, and spread. Fig. 1.1 is the histopathological image of breast tissue. In normal breast tissue, the cells look well arranged and have the same morphology. One can see irregularities associated with cancerous breast tissue in Fig. 1.1. Breast tissues are broadly classified into two categories. In situ breast cancer (ductal carcinoma in situ or DCIS) and other is invasive. Insitu's are limited to a certain area but invasive is infiltrating cancer and penetrates through the basement membrane through the stroma. Renal cell carcinoma

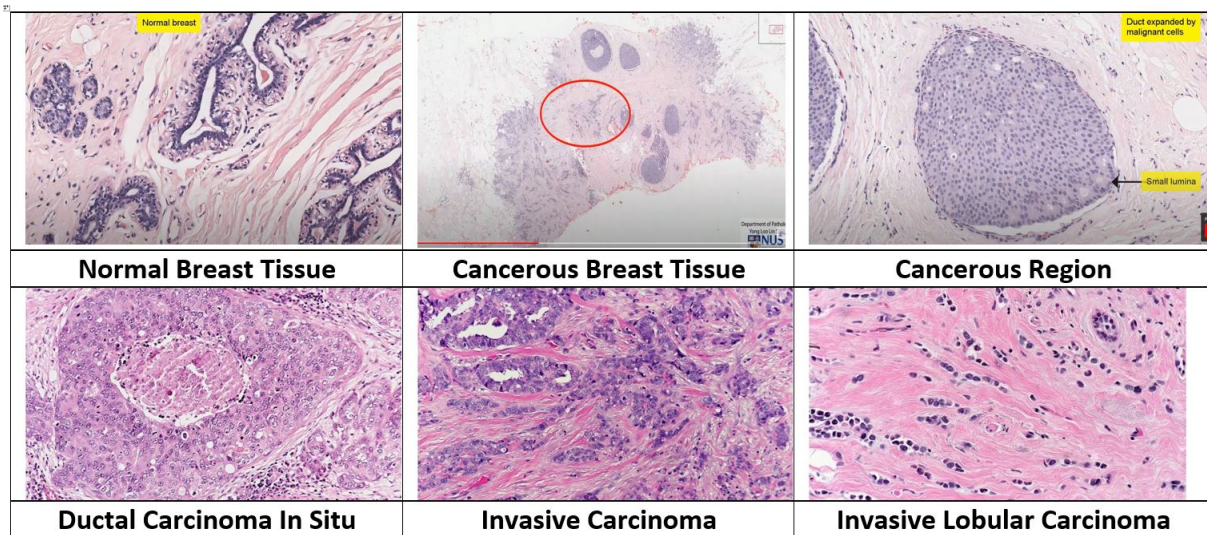


Figure 1.1: Nuclear characteristics of normal and cancerous breast tissues

(RCC) is the another most common kidney cancer (Sung *et al.*, 2021), (Du *et al.*, 2020) that accounts for 2-3% of adult cancers and constitutes approximately 85% of all primary renal malignant tumors.

The major histologic subtypes of RCC (Ricketts *et al.*, 2018) include clear cell RCC (KIRC), papillary cell RCC (KIRP), and chromophobe RCC (KICH) accounts for (70-80%), (14-17%), and (4-8%) of RCC respectively. The classification of the subtypes of RCC is based on morphology and cytogenetic characteristics. Each of these subtypes has different prognoses and differs in cytogenetic and histopathologic profiles. Clear cell RCC is the most common subtype of cancer in adults. Cells of clear cell RCC have clear cytoplasm and have spherical masses that distort the renal outline. Tumors of clear cell RCC are composed of large uniform cells with abundant cytoplasm that is rich in glycogen. Hered-

itary syndromes related to clear cell RCC is von hippel lindau (VHL) and mutations in the VHL gene leads to the development of tumors. Papillary RCC is the second most common renal cancer and is characterized by prominent papillary structures lined by monomorphic cuboidal epithelial cells and cells with abundant eosinophilic cytoplasm. Tumor cells in KIRC have a rounded or polygonal shape while KIRP is of cuboidal structure and low columnar cells are arranged like papillary formations.

Chromophobe RCC characterized by pale cells and smaller eosinophilic tumor cells in variable proportions with wrinkled nuclei and perinuclear halos. Cell membranes of KIRP appear very distinct with prominent margins. Fig. 1.2 provides the detailed view of tissue pattern and overall organization of RCC subtypes and Normal tissue regions. Histology and clinical features of RCC subtypes and normal tissue are presented in Table 1.1. Generally, the cancer grading is organ specific and is critically important for a better prognosis. It is noteworthy that the cancer grading is different from cancer staging. Cancer staging is more about the location, size, number of tumors and how far the cancer has been spread to the nearby lymph nodes; whereas the cancer grade describes how different the cancer cells and tissues appear from normal healthy cells under the microscope.

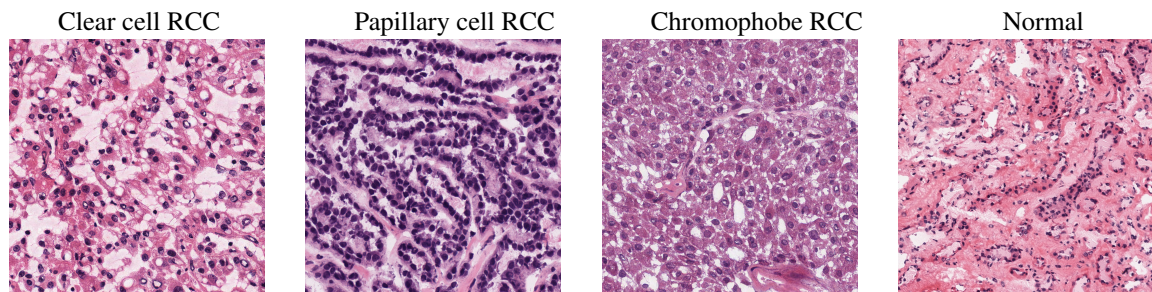


Figure 1.2: Visualization of histology and clinical features of RCC

According to Fohrman’s 4-tier grading system used by (Hong *et al.*, 2011), the Grade-1 nuclei are of tubular structure and uniform and it is very similar to normal nuclei. Grade-2 nuclei have slightly irregular contour compared to normal nuclei. Grade-3 nuclei have more irregular contour. Grade-4 nuclei have pleomorphic cells, mitoses, multilobate with the Grade-3 feature. Another method of grading system by WHO /international society of urologic pathology (ISUP) (Samaratunga *et al.*, 2014), where Grade-1 to Grade-3 tumor is decided based on nuclear prominence whereas Grade-4 is recognized by the presence of those cells showing extreme nuclear pleomorphism. The ISUP nucleolar grading system has been shown superior to the traditional Fuhrman grading system by (Delahunt *et al.*,

Table 1.1: Histology and clinical features of RCC subtypes and normal tissue regions at 40X resolution.

Clear cell RCC	Papillary cell RCC	Chromophobe RCC	Normal
<ul style="list-style-type: none"> • Clear cells with thin-walled. • Few cells are eosinophilic and granular. • Tumor shows large uniform cells with abundant cytoplasm that is rich in glycogen. • Cells are organized in acinar, alveolar, and tubular form. • Well defined margins and confined to the renal capsule. • Tumors in KIRC are composed of cells that have clear cytoplasm and have spherical masses. • Staghorn-shaped vasculature. • Nest cells are seen. • Tumors have delicate branching fibrovascular septae. 	<ul style="list-style-type: none"> • Prominent papillary structures lined by monomorphic cuboidal epithelial cells. • Bilateral and multiple masses. • Finger like projections with fibro-vascular cores. • Papillary carcinoma are multi-focal in origin. • Surrounded by fibrous capsule. • Type-1 Cells form single layer on papillae • Type-1 Composed of small cells having scanty basophilic or pale cytoplasm while Type-2 is of large cells with abundant acidophilic cytoplasm. 	<ul style="list-style-type: none"> • Large eosinophilic vegetal-like cells, in finely granular cytoplasm. • Smaller eosinophilic tumor cells are in variable proportions with wrinkled nuclei • Cell membranes of KIRC appear very distinct with prominent margins. • Nuclear grade does not apply. • Typically solid with parenchymal extension. • Irregular nuclei. • Perinuclear clearing • Pale to eosinophilic cytoplasm. • Arranged in solid sheets. 	<ul style="list-style-type: none"> • The cells of the proximal tubules have central nuclei and very acidophilic cytoplasm. • Cells show prominent vacuoles. • Each renal corpuscle has a vascular pole and a tubular pole. • The renal cortex on microscopy reveals glomeruli and tubeless. • A normal glomerulus by light microscopy shows glomerular capillary loops which are thin and delicate. • Endothelial and mesangial cells are normal in number. • The surrounding tubules are lined by cuboidal epithelium.

2019) and recommended as the main basis for grading by pathologists. In this study, the slide preparation phase was carried out under the supervision of experienced pathologists who performs routine diagnoses and carefully follow the WHO/ISUP grading criteria. Visualization of nuclear morphology of five different grades is shown in Fig. 1.3. Nucleolar prominence, and nuclear membrane irregularities of different grades of renal tumor is presented in Table 1.2. Visualization of nuclear morphology of four different classes of liver histopathology images is shown in Fig. 1.4.

The traditional method to diagnose cancer disease is to examine the morphological structure and distribution of nuclei manually by pathologists. After taking the image of the sample tissue pathologists decides whether tissue regions are cancerous or not and how

Table 1.2: Microscopic features of different grades of renal tumor (Normal to Grade-4 from left to right)

Normal (Grade-0)	Grade-1	Grade-2	Grade-3	Grade-4
<ul style="list-style-type: none"> • The cells of the proximal tubules have central nuclei and very acidophilic cytoplasm. • Cells are well arranged and are normal in number • A normal glomerulus structure. 	<ul style="list-style-type: none"> • Nucleoli are basophilic. • Nucleoli are not visible even 400X magnification. • Morphology is very similar to normal nuclei 	<ul style="list-style-type: none"> • Nucleoli are seen as eosinophilic at 400X magnification but not very prominent at 100X magnification. • Slightly irregular contour compared to normal nuclei. 	<ul style="list-style-type: none"> • Nucleoli conspicuous and eosinophilic at 100X magnification. • Clearly visible tumors were graded as grade-3. • Grade-3 nuclei have a more irregular contour compared to normal nuclei. 	<ul style="list-style-type: none"> • Pronounced nuclear pleomorphism. • Rhabdoid or sarcomatoid differentiation. • Contains tumor giant cells.

much is the malignancy level. A lot of time and effort is required in the manual analysis of histopathology images due to the complexity involved in such images. In recent trends, Computer Assisted Diagnosis (CAD) has become one of the most reliable digital techniques for the diagnosis and prognosis of cancer patients. CAD has outperformed manual cancer detection, done by pathologists, as CAD is more capable for early cancer detection than manual process and it is not dependent on human errors. The important component of CAD system for Cancerous/Noncancerous classification is shown in Fig. 1.5. Generally, histopathology images are having good contrast, due to the use of advanced digital slide scanners (Olympus, Aperio, Hamamatsu, etc.). Color is an important feature that is significant for cancer detection. For example, more colorful nuclei with respect to standard color indicate the sign of a malignant tumor. With the advancement of image processing and computer vision technology, the manual diagnosis of histopathology images has replaced automated analysis due to the inability to assess large amounts of nuclei present in histopathology images.

1.2.1 Preparing histopathology slides

With the help of solidifying material doctors make it in the form of a block called embedding. To make the slice of an embedded block is called sectioning. Generally, a thin slice is useful because nuclear portions are more expressive in thin slices. The thicker slice looks

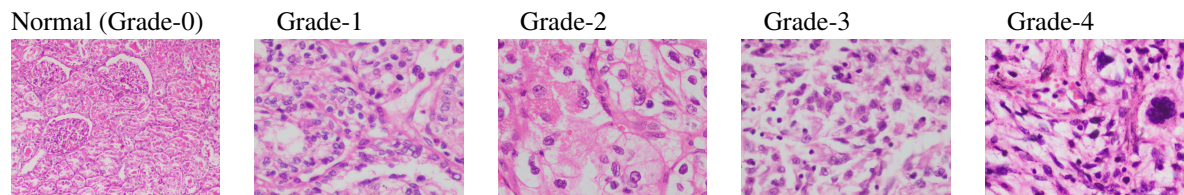


Figure 1.3: Visualization of different grades of renal tumor (Normal to Grade-4 from left to right)

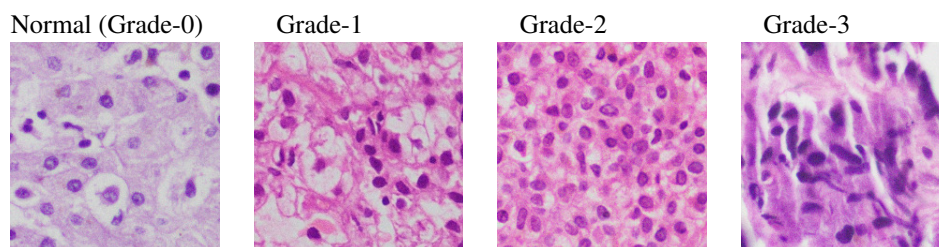


Figure 1.4: Visualization of different classes of Hepatocellular carcinoma (HCC) (Normal, Low sub-type liver HCC tumor, medium sub-type liver HCC tumor, high sub-type liver HCC tumor- from left to right)

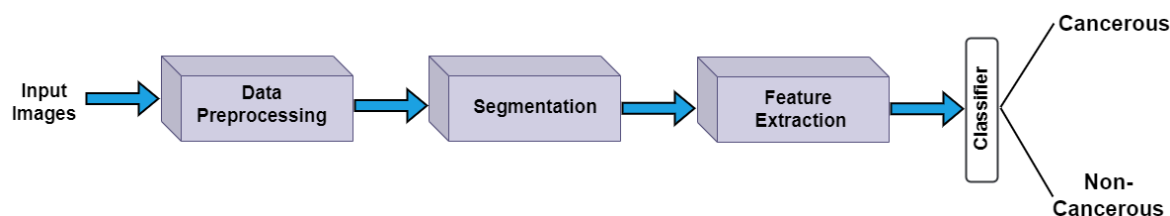


Figure 1.5: A systematic view of CAD system for cancerous/non-cancerous classification

darker and contains less information. Next, staining is done to make nuclei more expressive, and finally, with the help of a whole slide scanner, these slides are diagnosed. The schematic block diagram of histopathology slide preparation is shown in Fig. 1.6.

1.3 Deep Learning

A lot of success happened from 2012 onwards and currently, a lot of state-of-the-art systems are based on deep learning. Numerous amounts of data, a new type of computation system like GPU, an effective method to handle overfitting issues, and new optimization strategies make deep learning more powerful than traditional feature extraction algorithms. Deep

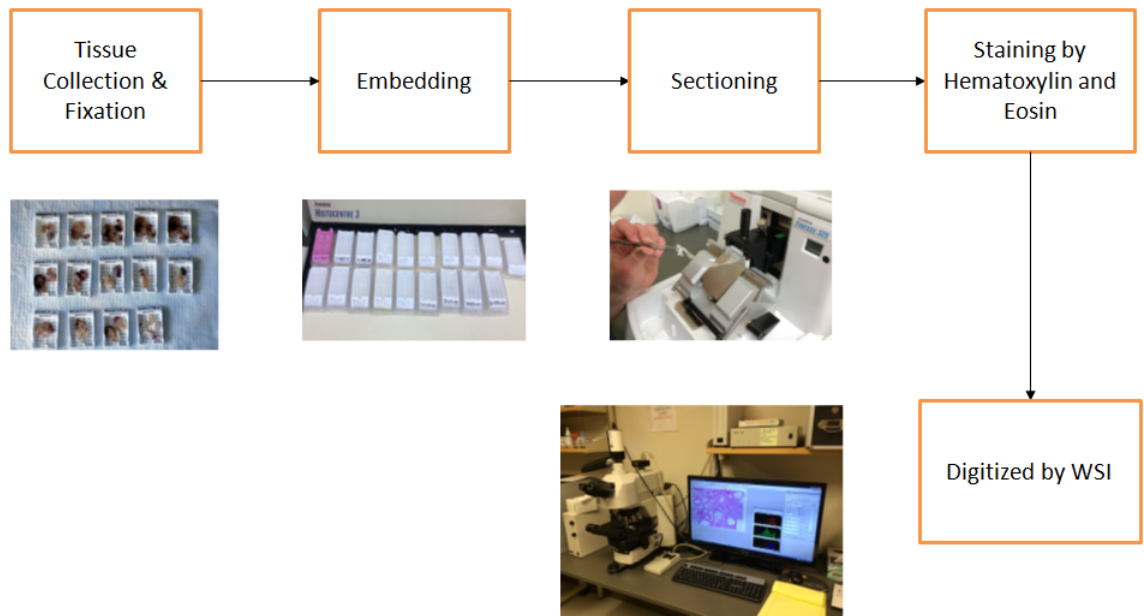


Figure 1.6: Block diagram of preparation of histopathology slides

learning is driven by the fastest-growing language called python where most of the codes are prepackaged in the form of a library. To perform the advanced operation we need to just write code in a very straightforward manner and don't need to write a whole bunch of code. Deep learning (DL) and artificial intelligence (AI) user communities are very easily reachable. Deep learning takes the data as the input and feeds it directly to the algorithms

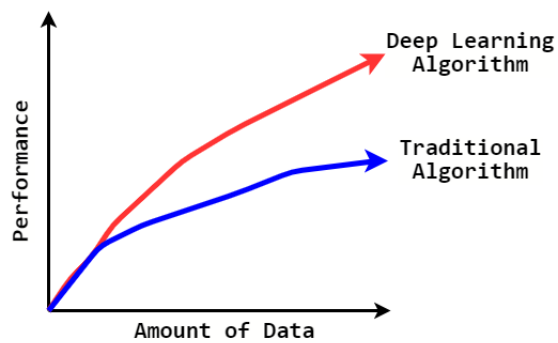


Figure 1.7: Comparison of traditional approach versus deep learning method

without requiring any manual feature extraction step. The data pass to the different layers of the artificial neural network and predict the final output. Machine learning needs an additional step of feature extraction however deep learning works technically in the same way as machine learning but with different capabilities and approaches. Deep learning is

inspired by the functionality of the human brain and it learns high-level features from the data on its own. Although machine learning depends on a huge amount of data, it can work with a smaller amount of data. Deep learning highly depends on a large amount of data.

Nowadays data is everywhere, it is in the form of images, it can be in the form of analytical data, it can be even in the form of text. No matter what field we are from, it is important to analyze data and extract information from those data. In general, when we work with a huge amount of data, the performance of traditional segmentation methods (Discontinuity-based approach, Similarity-based approach, Global clustering, Super-pixel segmentation, Watershed segmentation, Active contour segmentation techniques, etc.) and traditional classification methods of feature extraction (Linear regression, Logistic regression, Decision tree algorithm, Random Forest Algorithm, Naive Bayes classifier, Support Vector Machine, etc.) degrades. Based on the literature survey in this study a simple comparison of the traditional approach to feature extraction and the deep learning-based feature extraction method is shown in Fig. 1.7.

1.4 Deep Learning Tools

1.4.1 Activation function

Any linear function has less capability to learn any complex input. Since the outcome of the convolution layer is based on linear operation, a nonlinear operation is needed to map complex input to a meaningful output. The nonlinear activation function allows the network to learn complex data, compute and learn almost any function, and provide accurate predictions. It also helps the model to generalize or adapt to a variety of data and to differentiate the output. The activation function must be differentiable and in particular easily differentiable. Since we need millions and millions of gradients to update the weights. Easy differentiation makes the weight update process fast. The most popular activation function in the pre-deep learning era is Sigmoid and Tanh. The wonderful properties of Sigmoid and Tanh are that its derivative can be expressed in terms of their function themselves. equations (1.1, 1.2, 1.3, 1.4) represent the Sigmoid and Tanh activations and their derivatives.

$$f_{sig}(x) = \frac{1}{1 + e^{-x}} \quad (1.1)$$

$$f'_{sig}(x) = f_{sig}(x)(1 - f_{sig}(x)) \quad (1.2)$$

$$f_{\tanh}(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (1.3)$$

$$f'_{\tanh}(x) = 1 - f_{\tanh}(x)^2 \quad (1.4)$$

In a deeper network, the above activation functions resulted in a vanishing gradient problem. Rectified Linear Unit (ReLU) activation function rectified the vanishing gradient problem. ReLU is the most popular activation function in deep learning models. ReLU takes care of several problems faced by the sigmoid and the tanh activation function, and it also avoids and rectifies the vanishing gradient problems. The ReLU activation function simply has a value of zero, if it receives any negative input, but for any positive value, it returns that value like a linear function. It is computationally economical compared to sigmoid and tanh. Mathematically ReLU activation function and its derivative is expressed by equations (1.5, 1.6).

$$f_{ReLU}(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (1.5)$$

$$f'_{ReLU}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \quad (1.6)$$

1.4.2 Loss function

The deep learning model learns by means of loss function which is driven by a calculation of error. This error is the difference of actual value and predicted value as in the regression problem and the difference of actual distribution and predicted distribution in the case of a classification problem. With the help of suitable optimization algorithm, the loss function learns to reduce the error. Following are the most preferred loss function in the case of binary segmentation. Following are the most frequently used loss function in deep learning. (a) Binary Cross-Entropy: Binary Cross-Entropy is a combination of sigmoid activation and Cross-Entropy, which is discussed by (Naylor *et al.*, 2019), and (Chanchal *et al.*, 2021a). Let consider a countable set of symbols $X = \{x_1, x_2, \dots, x_i, x_n\}$. Y is the discrete probability distribution which represents the probability of occurrence of those symbols. $Y = \{y_1, y_2, \dots, y_i, y_n\}$. Let $y_i = p(x_i)$ where y_i is the probability of occurrence of symbol x_i . According to the concept of entropy the minimum number of bits required to represent the i^{th} symbol is $x_i = \log(1/y_i)$. If I consider entire distribution Y to achieve the optimal number of bits per transmission through some channel, then the optimal number of bits is known

as entropy. Mathematically it is just the expected number of bits per encoding and can be shown in equation (1.7).

$$H(Y) = \sum_{i=1}^n y_i \cdot \log \frac{1}{y_i} = - \sum_{i=1}^n y_i \cdot \log y_i \quad (1.7)$$

Now let $X' = \{x'_1, x'_2, \dots, x'_i, \dots, x'_n\}$ with probability of occurrence $Y' = \{y'_1, y'_2, \dots, y'_i, \dots, y'_n\}$. If encode the symbol X using different symbol X' then encoding will require $x'_i = \log \frac{1}{y'_i}$ instead of $\log \frac{1}{y_i}$ then it can be define Cross-Entropy $H(Y, Y') = \sum_{i=1}^n y_i \log \left(\frac{1}{y'_i} \right)$. If distributions of Y and Y' are equal then entropy and cross-entropy are equal. For binary cross-entropy there will be two classes, class C_1 and class C_2 . (a) $Label_1 \{0, 1\}$ represents the ground truth label for class C_1 , S_1 represents the sigmoid score for class C_1 (b) $Label_2 = \{1 - Label_1\}$ represents the ground truth label for class C_2 , $S_2 = (1 - S_1)$ represents the sigmoid score for class C_2 . Binary cross-entropy (BCE) used by (Ronneberger *et al.*, 2015) and is defined by equations (1.8), and (1.9).

$$S_i = \frac{1}{1 + e^{-(Prediction)_i}} \quad (1.8)$$

$$L_{BCE}(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (1.9)$$

(b) **Weighted Binary Cross-Entropy:** For skewed data sometimes weighted binary cross entropy (WBCE) loss function performs better. Normally for the minority class, it assigns higher weights. The purpose of using class weights is to change the loss function so that the training loss cannot be minimized within a limit. It is a way of passing weights to the binary cross entropy loss function used by (Jadon, 2020) and (Sugino *et al.*, 2021). equation (1.10) describes the weighted binary cross entropy loss function. Here β is used to assign weights to the more relevant objects. β value can be used to tune false negatives and false positives. If we want to reduce the number of false negatives then set $\beta > 1$, similarly to decrease the number of false positives, set $\beta < 1$. y , and \hat{y} represents ground truth and predicted results respectively.

$$L_{Weighted\ BCE}(y, \hat{y}) = -(\beta * y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (1.10)$$

(c) **Dice Loss:** In case of binary segmentation and having class imbalance dice loss is

a most preferred overlap measure used by (Milletari *et al.*, 2016), and (Sudre *et al.*, 2017). There is little difference in the calculation of loss using intersection over union (IOU) and dice. equations (1.11), (1.12), and (1.13) describes the dice loss function.

$$IOU = \frac{Intersection}{Union} = \frac{TP}{FP + TP + FN} \quad (1.11)$$

While the calculation of dice is by calculating the harmonic mean of precision and recall.

$$Dice \ Coefficient = \frac{2 * Intersection}{Union + Intersection} = \frac{2 * TP}{FP + 2 * TP + FN} \quad (1.12)$$

$$Loss = 1 - Dice \ Coefficient \quad (1.13)$$

(d) Categorical Cross-Entropy Loss: Cross-entropy loss followed by softmax activation is called softmax Cross-entropy or categorical cross-entropy loss. It is useful for multiclass classification problems. Regularized softmax loss function calculation in the training process can be expressed as in equation 1.14. λL_2 is regularization term, where λ is regularization parameter and L_2 regularization penalizes the sum of square of weights.

$$L_{Regularized} = Cross \ Entropy \ Loss + \lambda L_2 \quad (1.14)$$

$$Cross \ Entropy \ Loss = - \sum_{i=1}^C y_i \log(f(x)_i) \quad (1.15)$$

Where $f(x)_i = \frac{e^{x_i}}{\sum_{j=1}^C e^{x_j}}$ is softmax activation for multi-class classification problem and C is the number of classes.

For an input vector x_i with length equal to the number of classes k , $x = (x_1, x_1, \dots, x_k) \in \mathbb{R}^k$, The output function of the softmax $f_\phi(x_i)$ can be expressed in equation 1.16.

$$f_\phi(x_i) = \begin{bmatrix} p(y_i = 1) | x_i, \phi \\ p(y_i = 2) | x_i, \phi \\ \vdots \\ p(y_i = c) | x_i, \phi \end{bmatrix} = \frac{1}{\sum_{j=1}^c e^{\phi_j^T x_i}} \begin{bmatrix} e^{\phi_1^T x_i} \\ e^{\phi_2^T x_i} \\ \vdots \\ e^{\phi_c^T x_i} \end{bmatrix} \quad (1.16)$$

The term $\frac{1}{\sum_{j=1}^c e^{\phi_j^T x_i}}$ is used to normalized the output to 1 and Φ is the parameter of softmax classifier Wei *et al.* (2017).

1.4.3 Gradient descent optimization methods

The most important part of any optimization method in deep learning is about gradient which is used to update the weights. equation (1.17) is a general weight update equation.

$$W_{New} = W_{Old} - \eta \left[\frac{\delta L}{\delta w} \right] \quad (1.17)$$

Where W_{New} and W_{Old} represents updated weight and old weight respectively. η is learning rate and $\frac{\delta L}{\delta w}$ is derivative of loss with respect to weight. Following are the frequently used optimization strategy in deep learning to update the weights.

(1) Gradient Descent (GD) (2) Stochastic Gradient Descent (SGD) (3) Minibatch SGD (4) SGD with Momentum (5) Nesterov Accelerated Gradient (NAG) (6) Adagrad (7) AdaDelta (8) RMSProp (9) Adam (Adaptive Moment Estimation)

GD updates the weights using all the data points in the data set at a time. The cost of computing GD is very large and time expensive. SGD uses only one point at a time to update the weights and Minibatch SGD uses a random subset of k points in the data set to update the weights. In today's trend, the Minibatch SGD strategy is much used. The path of convergence of GD is slightly different from SGD and Minibatch SGD. Minibatch SGD is an approximation of GD. The major problem of minibatch SGD and SGD is that it gets stuck at the saddle point and gives a noisy update to the gradient. (Minibatch SGD, SGD update noisy as compared to GD updates). Denoising leads the concept of SGD with a momentum that gives more weightage to the most recent point and lesser and lesser weight to previous data points. SGD plus Momentum speeds up convergence and is a very important concept in the whole of optimization. One another strategy is called NAG which is similar to SGD plus Momentum, one difference is that SGD calculates gradient and momentum at the same point but NAG first calculates momentum and then calculates gradient so the path of convergence is different. Both of them are works well in most cases.

All methods discussed above have the same learning for each weight and the key idea of Adagrad is for each parameter/weight has a different learning rate. Since the sparse feature has different behavior than the dense feature, AdaGrad performs larger updates for infrequent parameters and smaller updates for frequent parameters. One of the major limitations of Adagrad is it has monotonically decreasing learning rates. The number of iterations

increases the learning rate of that weight adaptively reduces, this could lead to slower convergence. AdaDelta tries to address this monotonically decreasing tuning of learning rate by restricting the accumulation of past gradients within some fixed window by calculating a running average. Adadelta has some advantage over Adagrad as it avoids large denominators which cause slow convergence. RMSProp tries to address the issue of Adagrad's rapidly diminishing learning rate by using a moving average of the squared gradients.

Adam (Adaptive Moment Estimation) optimization: For efficient learning by making a faster convergence of an algorithm towards the minimum loss with less memory requirement, Adam (Kingma and Ba, 2014) is used as an optimizer. It gives the advantage of both Gradient Descent with momentum and RMSProp algorithms. It uses two exponential decays, β_1 (usually 0.9), the rate for the first moment estimates and β_2 (usually 0.999) for the second. The moving averages are calculated for a given iteration t with equations 1.18 and 1.19.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (1.18)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (1.19)$$

where g_t is the gradient at t^{th} . Since the $\beta_1, \beta_2 \approx 1$, the tendency of the moving averages to be biased towards zero can be corrected by an additional step, called the bias correction step, given by the equations 1.20 and 1.21.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (1.20)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (1.21)$$

Weights updation can be done by using the equation 1.22.

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (1.22)$$

In equation 1.22, η is the learning rate, w_{t-1} and w_t are weights at $(t - 1)^{\text{th}}$ and t^{th} iterations and ϵ is a small constant to avoid the error due to division by zero.

1.4.4 Handling overfitting problems

As the number of layers increases, our model learns too well but in parallel, the complexity of the model increases such that training error goes on decreasing but poor performance

on unseen data. If the number of layers increases we will end up with a case of high variance and more weights and parameters, resulting in a higher chance of overfitting. If the number of layers is too less, there will be a higher chance of underfitting. A network having fewer weights ends with a case of a high bias problem. Regularization is a technique that makes the algorithm such that the model generalizes better and this, in turn, improves the performance of the model on the unseen data as well. To avoid overfitting regularization is used. L1 and L2 are the most common regularization method. equation (1.23) is L1 regularization, where L1 norm is calculated as the sum of the absolute values of the vector.

$$W^* = \text{Logistic Loss for Training Data} + \lambda \|W\|_1 \quad (1.23)$$

L2 norm is calculated as the square root of the sum of the squared vector values. equation (1.24) represents L2 regularization. L1 regularization create sparsity in W where all the unimportant or less important features becomes zero.

$$W^* = \text{Logistic Loss for Training Data} + \lambda \|W\|_2^2 \quad (1.24)$$

1.4.5 Batch normalization

Due to lots of operations happening between layers, if the input changes slightly in the deep network, this can lead to large changes at later layers. The distribution of each layer's input changes during training. During training process, the network weight keeps changing. If the weight changes randomly in successive iterations, it becomes difficult for the network to adjust each layer's input. Batch normalization makes sure the distribution does not change too much. During the training procedure gradient of the loss, plays an important role in the convergence. Visualization of internal covariate shift in the deeper network can be seen in Fig. 1.8. Batch normalization used by (Ioffe and Szegedy, 2015), and (Santurkar *et al.*, 2018), it can be thought of as an additional layer and it works well in a very deep network. It has the following advantages (a) Faster convergence (b) Works as a regularizer. (c) Avoids internal covariate shift. (d) It is possible to train a deeper network.

Let T_1, T_2 are two transformations at a particular layer L shown in equation (1.25). Two layers are characterized by weights W_1 and W_2 . If W_1 changes, T_1 will change, and input to T_2 will change. If W_2 changes L will change. If these changes are random and large then

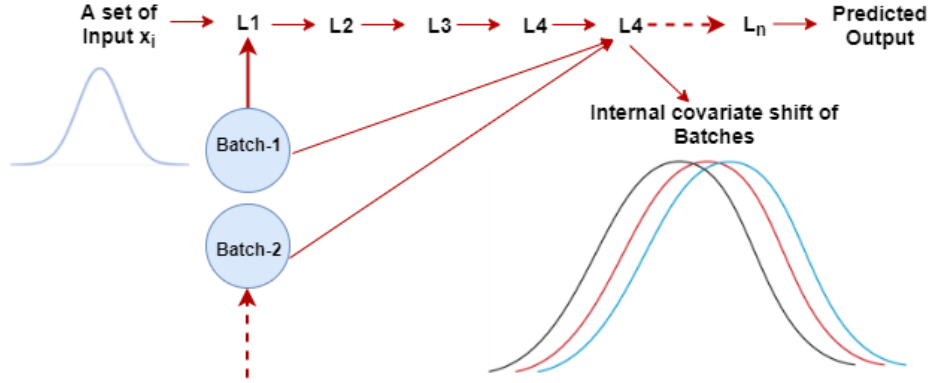


Figure 1.8: Internal covariate shift

there is the problem of convergence in deep neural networks.

$$L = T_2(T_1(U, W_1), W_2) \quad (1.25)$$

$$\text{Batch Norm}(y_o)^{(m)} = \beta_1 \left(\frac{(y_o)^{(m)} - \mu(y_o)}{\sigma(y_o)} \right) + \beta_2 \quad (1.26)$$

Batch normalization normalizes each activation independently by controlling the mean and standard deviation of the layer's output. The process expressed in equation (1.26), where y_o^m is the value of the output y_o on the m^{th} input of a batch and β_1, β_2 are trainable parameters.

1.5 Functional Diagram of a Deep Learning Architecture for Histopathology Image Segmentation

Here deep learning framework has been used for the segmentation process where features are learned automatically from the model, and based on that they predict the segmented image. Deep learning architectures learn the model parameters to separate the nuclei in histology images. The deep learning framework is driven by large amounts of data, activation function, an effective optimization algorithm, and powerful loss function. An activation function can map the input to a meaningful number and an optimization function is responsible for the fast convergence of model parameters. A loss function effectively calculates the difference between the true value and predicted value in the case of regression problems while calculates true probability distribution and predicted probability distribution in the case of classification problems. The functional diagram for automatic image segmentation

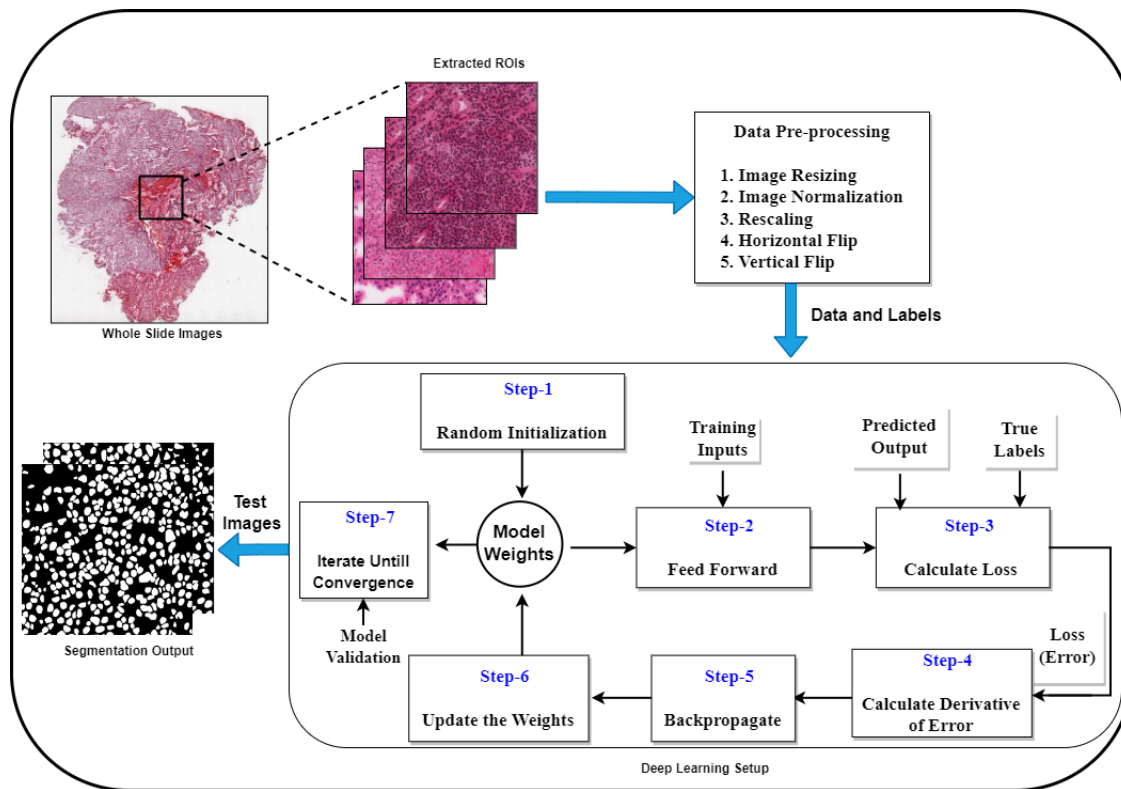


Figure 1.9: Functional diagram of image segmentation from histopathology images

from H&E stained histopathology images is shown in Fig. 1.9. There is a whole slide image (WSI), and we have extracted the meaningful region of interest (ROI) from those WSI. Depending on the characteristics of these images, sometimes there is a need for data pre-processing and data augmentation. To increase the diversity of data for training, the strategy of data augmentation by the creation of artificial variations, such as rotation, flipping, padding, and zooming, dataset were expanded. The purpose of augmentation is to prevent CNN from undergoing overfitting, and the augmentation approach depends on the type of architecture used.

In this case, we have used different sets of data augmentation for the different histology datasets, so that it is possible to obtain more images for the training set as well as the validation set. We have used flipping, rotation, dihedral, and random lighting (the threshold is 0.5). We have also performed color normalization for the breast and multi-disease histopathology datasets. Color normalization is needed to match the color characteristics of an image to those of the target image. In this process, we matched the mean and standard deviation of each color channel in the two images. If ROIs are of good resolution and

have a good number then data augmentation step may not be required. Now, this data and labels are fed to the deep learning setup, where first of all there is random initialization of all weight and bias or may set all weight and bias to zero. A set of data and labels are forward pass-through hidden layers and non-linear activation. This feed-forward network gives actual output. In supervised learning, the actual output is also present, so the loss is calculated. This loss is the difference between true distribution and predicted distribution. Now, this is time to back-propagate and for that gradient of the loss with respect to weight is required. Here model calculates the derivative of errors then back-propagates. Using the optimization algorithm, the model update the weight and this optimization method set the weight such that it minimizes the loss. This forward and backward process of complete data once through a network is called one epoch. During each epoch, the model validates with validation data and iterates until convergence. Based on the final optimized weight and bias, the network gives segmented results.

1.6 Functional Diagram of a Deep Learning Architecture for Histopathology Image Classification

The application of CNN for histopathological image classification is booming. The implementation of these networks on edge devices has been gaining attention due to the fast and precise assessment of histopathological tissues. The CNNs have been proven to be best suited for image classification tasks. Different architectures of CNNs with broader perspectives and novel ideas have been discussed in literature which provides a deep insight into different types of deep learning networks and their development. The overall implementation flow is as depicted in Fig. 1.10. There are four phases in implementing a CNN for classification described in below subsections. The data preparation phase is carried out under the supervision of experienced pathologists and it has been cropped only those parts of WSI which have histologically defined clinical features. To improve the model generalization capability and to avoid overfitting problems data augmentation such as horizontal flip, vertical flip, and rotation has been performed. Total obtained patches after data augmentation were slidewise distributed into training (80%), and test (10%).

All the images are resized to a size of 224x224 and normalized to zero mean and unit variance before passing them through deep learning architecture for training. Training phase in which a neural network's weights are continuously updated with back-propagation.

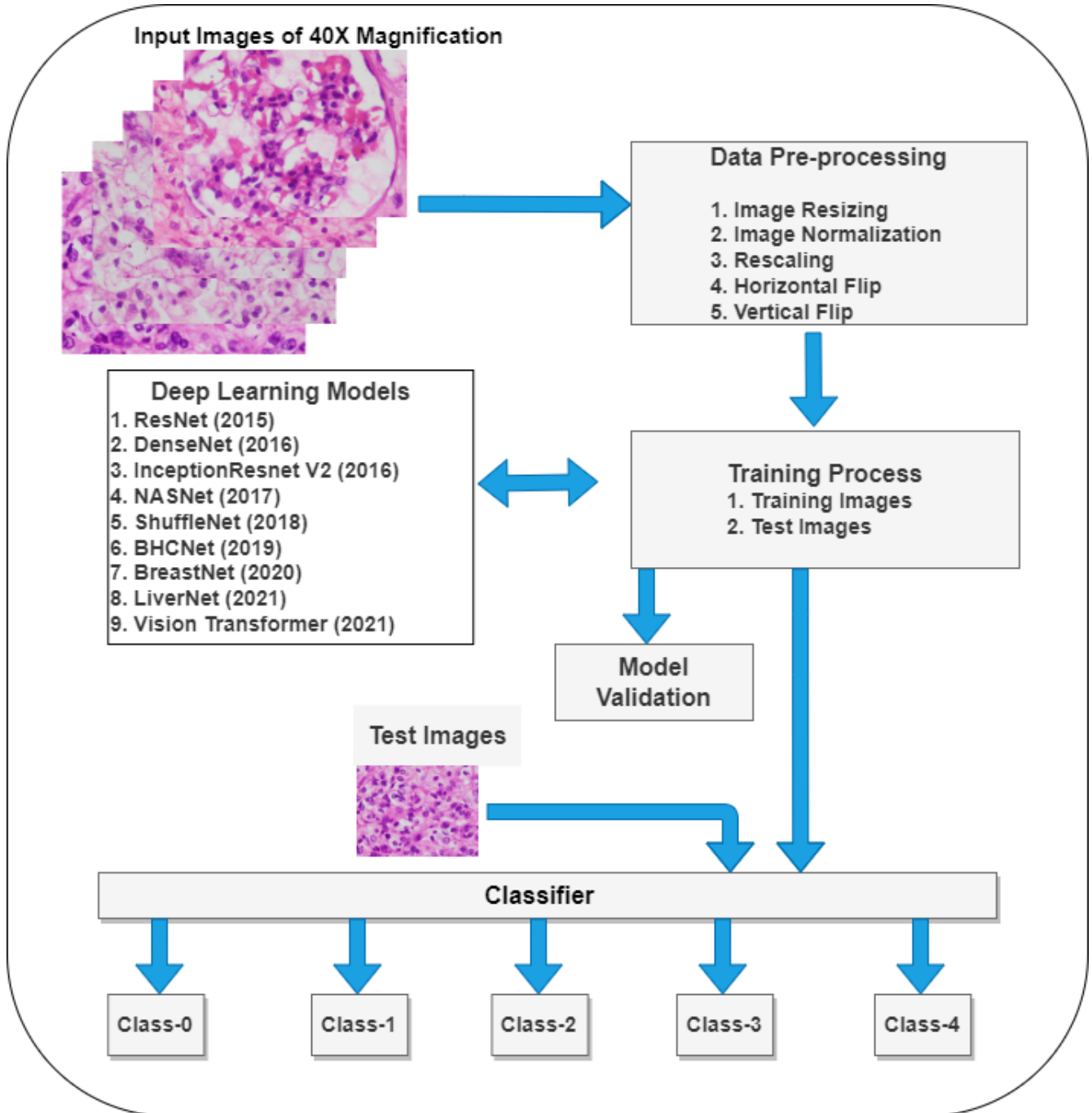


Figure 1.10: Functional diagram of image classification from histopathology images

Training a neural network is the most computationally intensive phase, it is preferred to train the CNNs using GPUs, and inference is run on edge devices that have lower latency and higher computational power. Thus, in this research, a set of deep neural network models are trained on GPU, and then testing is done based on the best weight saved in the entire training process. In this study, the validation dataset works as a calibration set. The validation set was created by randomly selecting patches from the training set. Validation data is used to continuously monitor the loss and act accordingly. Deep learning models have been trained on the training set for a given range of hyperparameters and evaluates the efficiency of the model on the test datasets. Test phase in which only forward propagation is performed to predict the output. The testing phase was performed by the patches of those WSIs which were completely unseen for the model.

1.7 Problem Statement

Design and development of automatic methods for nuclei segmentation and classification of histopathology images.

1.8 Research Objectives

Two research objectives are formulated after analyzing the research gaps.

1. To propose an efficient and robust segmentation method and evaluate its performance on histopathology images.
2. To design and develop a deep learning framework for the classification and grading of kidney histopathology images.

1.9 Main Contributions of the Thesis

1. An efficient and robust deep learning architecture (SCPP-Net) for the segmentation of nuclei in three datasets of kidney, breast, and multi-organ histopathology images has been proposed. For fast assessment, the proposed algorithm used depth-wise separable convolution that effectively reduced the total number of parameters and floating point operations (FLOPs) in the SCPP-Net.

2. A segmentation framework named high-resolution deep transferred ASPPU-Net has been proposed for multi-organ nuclei segmentation. To enrich the features at five distinct stages, the proposed model effectively utilized residual connection throughout the network and atrous spatial pyramid pooling (ASPP) layer at the bottleneck.
3. A deep structured residual encoder-decoder network with a novel loss function for nuclei segmentation in three datasets of kidney, breast, and multi-organ histopathology images has been proposed.
4. A computationally efficient deep-learning framework for the classification of subtypes of renal cell carcinoma from kidney histopathological images has been proposed. This study proposes a CNN module called multiple channel residual transformation (MCRT) to focus on the most relevant morphological features of RCC. Another novel module called group convolution deep localization (GCDL) that effectively integrates three different feature descriptors. These blocks contribute maximally to the proposed RenalNet model and complete the classification process.
5. A new benchmark kidney histopathology dataset for the classification of subtypes of renal cell carcinoma is developed from the WSIs of TCGA data. The introduced dataset is class balanced compared to any other RCC dataset.
6. An efficient and end-to-end fully automated deep learning architecture called Renal Cell Carcinoma Grading Network (RCCGNet) is proposed for grading renal tumors from H&E stained kidney histopathology images. This work proposes a novel CNN block called the shared channel residual (SCR) block which shares the information between different layers and strengthens the local semantic features at multiple stages within the network. The SCR block contributes maximally to the proposed method.
7. A new benchmark RCC dataset of H&E stained kidney histopathology images is created for grading of renal tumors, obtained from the Department of Pathology, KMC Mangalore, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India. The newly introduced dataset is available for public use.
8. Comparable experiments include transfer learning methods, as well as deep learning networks trained from scratch, performed on multiple organ histopathology datasets. The proposed architecture requires reduced computational resources and outperforms

the recent benchmark models in terms of prediction accuracy and computational complexity.

9. An efficient and robust CNN (RoCNN) architecture is proposed to automate the grading and classification task for kidney cancer tissue. The RoCNN is capable of learning features at varying convolutional filter sizes because of the inception modules employed in it. To demonstrate that the proposed model is generalized and independent of the dataset, it has experimented on two well-known datasets, the KMC kidney dataset of five different grades and the TCGA dataset of four classes. In addition, the proposed method attains a substantial reduction in the number of parameters and FLOPs. Minimum computational complexity among state-of-the-art models, our method represents a step toward possible incorporation in FPGA board processing on next-generation histopathological image analysis.

1.10 Organization of the Thesis

The thesis is arranged into six chapters, as follows.

Chapter 1 briefly overviews the nuclear characteristics and tissue patterns of multiple organ histopathological images, including the digital histopathology slide preparation method. The motivation behind the development of automatic methods for disease detection is also presented. This chapter provides details of functional diagrams of a deep learning architecture for segmentation and classification of the nuclear region from histopathology image. The problem statement along with the research objectives put forward for fulfilling those are discussed in this chapter. The chapter concludes by highlighting the major contributions towards this research work.

Chapter 2 provides an exhaustive literature survey of non-deep learning and deep learning-based segmentation and classification of histopathology images which are conducted to formulate the research gaps and problem statement.

Chapter 3 discusses the need for nuclei segmentation from histopathology images. A detailed explanation of the newly proposed segmentation algorithms is included in this chapter, followed by a summarization of the work done to fulfil the first objective.

Chapter 4 includes the details about proposed classification algorithm RenalNet for the classification of subtypes of renal cell carcinoma from kidney histopathology images. The implementation details, mathematical description of the proposed model, computational

complexity analysis, and simulation results of the proposed RenalNet are specified in this chapter. The summary of the work done for the classification of subtypes of RCC from histopathology images is also presented.

Chapter 5 provides details about two computationally efficient fully automated renal cell carcinoma grade prediction models from kidney histopathology images. This chapter provides information about the mathematical representation of proposed networks, the training and implementation details, the ablation study, computational complexity, simulation results, and a summary of the proposed deep learning architectures.

Chapter 6 reports the conclusion and summaries the overall work done to accomplish the research objectives. The chapter also indicates the future scope of this research work.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

Convolutional neural networks (CNN's) have shown superior performance in the field of digital pathology which involves tasks such as pixel-wise segmentation, instance segmentation, classification of different sub-types of cancer, and grading of cancers. The development of automatic methods for nuclei segmentation, classification, and grading from histopathology images is considered as a major challenging problem in the field of biomedical image processing. The major contributions achieved so far in the field of segmentation, classification, and grading of nuclear region from histopathology images are provided in the below Sections 2.2, and 2.3, respectively. The research gap analysis in the related field of nuclei segmentation and classification of histopathology images is discussed in Section 2.4. Finally, the quality metrics used to evaluate segmentation, classification, and grading approaches are presented in the Section 2.5.

2.2 Literature Survey on Segmentation Techniques

Segmentation tasks can be categorized into traditional or handcrafted feature extraction techniques and CNN-based deep learning approaches. As a segmentation task, our focus is to separate hematoxylin stains cells from eosin stain cells and other connected tissue. A brief review of conventional image segmentation is presented in Section 2.2.1. The traditional feature extraction techniques performed better in the field of biomedical application but the main challenge is to deal with complex histology images where these techniques

result in under- segmentation or over-segmentation. In recent research work, most of the authors reported that, the segmentation technique based on a deep convolutional neural network performs far better than the conventional segmentation approach. A concise review of CNN based approach is presented in Section 2.2.2.

2.2.1 Literature survey of conventional segmentation techniques

Traditional segmentation methods are mostly based on similarity-based approach, discontinuity based approach, watershed techniques, active contour methods and their variants, superpixel, and clustering-based methods, etc. The similarity based approach discussed by (Gonzalez and Woods, 2008), is based on local thresholding, global thresholding, adaptive thresholding, Otsu's thresholding, region growing, region splitting, and merging, where these methods try to group and segment similar pixels. For image histogram having flat valleys, the similarity-based approach does not work well and the wrong selection of threshold value may result in over-segmentation and under- segmentation in this case. The discontinuity based approach tries to segment those pixels which are isolated in some manner like point, line, edges, and it is a mask processing-based approach. This method requires different operators at different stages. (Cousty *et al.*, 2010) proposed watershed segmentation method, based on the split, merge, and marker controlled watershed. Detected boundaries in the watershed method depend on cell complexity. (Song *et al.*, 2017) proposed active contour segmentation, where they consider intensity information and local edge information for the detection of object boundaries.

Superpixel segmentation method used by (Albayrak and Bilgin, 2019), is based on the cluster of connected pixels having identical features. It considers the color and coordinate information of neighbor pixels. This technique provides better regional information but is not very effective in the case of cell segmentation. Clustering based segmentation proposed by (Win *et al.*, 2017), performs grouping based on their similarity. A summary of conventional image segmentation is presented in Table 2.1.

Table 2.1: Summary of conventional segmentation techniques

Approach	Description	Observations
Discontinuity based approach (Gonzalez and Woods, 2008) (Boundary approach) <ul style="list-style-type: none"> • Point detection • Line detection • Edge detection 	<ul style="list-style-type: none"> • This method is based on the principle of intensity variations among the pixels. • Extract regions that differ in properties like intensity, color, texture etc. 	<ul style="list-style-type: none"> • Works well for images having good contrast between regions. • Second-order derivative gives a reliable result. • Boundaries determined by this method is generally discontinuous. • The single operator does not suit for all types of images.
Similarity based approach (Gonzalez and Woods, 2008) (Region approach) <ul style="list-style-type: none"> • Region growing • Region splitting and merging • Thresholding method 	<ul style="list-style-type: none"> • Group those pixels which are similar in some sense. • Common properties of the pixels are utilized. 	<ul style="list-style-type: none"> • Gives better result compared to other conventional segmentation methods. • Flow from inner point to outer region generates clear object boundaries. • Expensive in terms of computation time and memory. • Selection of noisy seed leads to wrong segmentation result.
Global clustering based approach <ul style="list-style-type: none"> • K-means (Win <i>et al.</i>, 2017) clustering • Fuzzy-C means (Dhanachandra and Chanu, 2020) clustering 	<ul style="list-style-type: none"> • In clustering, all pixels in the plane have equal weights. • K-means performs grouping based on their similarity by comparing the data with each other. • In fuzzy C-means samples to be clustered belongs to a cluster center in a certain plane. 	<ul style="list-style-type: none"> • Relatively simple to implement. • Time taken to cluster K-means rises linearly with the number of data points. • Sensitive to initial values. • Falling easily into local minima. • Sensitive to noise.

Continued on next page

Table 2.1 – continued from previous page

Approach	Description	Observations
<p>Superpixel (A cluster of connected pixels with similar features like color, brightness, texture, etc.) segmentation methods (Stutz <i>et al.</i>, 2017), (Albayrak and Bilgin, 2019)</p> <ul style="list-style-type: none"> • Simple linear iterative clustering (SLIC). • Density-based spatial clustering of application with noise (DBSCAN). • Topological preserved regular superpixel (TPRS). • Entropy rate superpixel (ERS). • Mean Shift. • Quick Shift. • Normalized Cut (Ncut). 	<ul style="list-style-type: none"> • SLIC superpixel segmentation is used as pre-segmentation to the clustering-based segmentation method. • SLIC, clusters neighbor pixels by considering their color and coordinate information. • DBSCAN is a clustering algorithm based on the measurement of density in regions close to the given object. • DBSCAN improves the performance of segmentation by adding local geometric boundaries. • In TPRS arranged seeds relocated to the pixel with locally maximal edge magnitude. • ERS finds compact and homogeneous superpixels by using a graph-based approach. 	<ul style="list-style-type: none"> • Applicable in MRI image segmentation, Glaucoma screening, Optic disc segmentation. • Superpixel methods extract the meaningful regions in the image and improve the computation based on the pixel. • SLIC method controls the tradeoff between superpixel compactness and boundary adherence. • Provides regional information. • High computational efficiency. • Good adherence to object boundaries. • Superpixel methods used in histopathological analysis are about the segmentation of regions rather than cell segmentation. • The superpixel method with large parameters leads to slow running time and poor segmentation performance. • Different superpixel methods have different advantages and drawbacks. We should choose the proper method according to the problem.
Continued on next page		

Table 2.1 – continued from previous page

Approach	Description	Observations
Watershed segmentation methods (Cousty <i>et al.</i> , 2010), (Gamarra <i>et al.</i> , 2019) <ul style="list-style-type: none"> • Marker Controlled (MC) watershed. • Split and Merge (SM) watershed. 	<ul style="list-style-type: none"> • MC watershed identifies the cells from the background. • SM watershed separate the overlapped cell. 	<ul style="list-style-type: none"> • Popular in cell image analysis. • This method does not achieve great accuracy but extracted features supports cell classification. • Issue of under segmentation and over-segmentation depending on cell complexity.
Active contour segmentation methods <ul style="list-style-type: none"> • Dual channel active contour (Song <i>et al.</i>, 2017) (DCAC). • Edge based active contour. (Ciecholewski, 2016) 	<ul style="list-style-type: none"> • DCAC employs initial contour to the previously delineated nuclear boundary by considering intensity information. • The edge-based model uses local edge information to fit the boundaries of the approximated shape. 	<ul style="list-style-type: none"> • Works well for nuclei having a clear nuclear boundary. • Autonomous and self-adapting method of finding object boundary. • Need of initial contour, proper parameters, and long runtime. • Difficult to segment the nearest objects. • When image size is too large, this method works slowly.

2.2.2 Literature survey on CNN based segmentation approaches

Most of the CNN architecture for the cell segmentation task consists of an encoder-decoder path for feature extraction. Many of the recent research utilizes lots of potential opportunities like improving training strategies, handling overfitting problems, better optimization methods, and many other strategies to obtain better prediction accuracy. However, many authors reported their result which is very efficient but an accurate and efficient segmentation algorithm is still open-ended research due to complexity in histopathology images.

One of the significant contributions by (Ronneberger *et al.*, 2015), called UNet, provides a very good direction and a dramatic breakthrough in the field of biomedical image segmentation. UNet is a symmetric encoder-decoder convolutional network and has a large number of feature channels that allow to extract feature to the higher layer in a deep network. Repeated application of (3 x 3) convolution kernel followed by ReLU activation, (2 x 2) maxpooling and (2 x 2) up-sampling with stride size of 2 and (1 x 1) convolution

followed by sigmoid activation at final layer, total of 23 layers in the network. The original U-Net does not use batch normalization, residual connections, and many other potential opportunities for further enhancements in segmentation applications. (Veit *et al.*, 2016), realized through their experiment that if a network has a collection of paths, then a shorter path is enough during training or a very deep path is not required during training. These multiple paths do not strongly depend on each other and their smooth co-relation with multiple valid paths increases the performance of the network. (Milletari *et al.*, 2016) proposed an encoder-decoder convolutional network for three-dimensional data by utilizing dice loss as a loss function. Their empirical evaluation achieves better performance on the strong imbalance dataset. (Nogues *et al.*, 2016) proposed an architecture for detection of lymph nodes by two fully nested supervised convolutional networks and a structured conditional random field optimization strategy.

Degradation of information in deeper network addressed by (He, 2016), by introducing deep residual network, which is easier to train and optimize. The residual connection is realized by skipping one or more layers to restore the information in a deep network. With FCN and atrous spatial pyramid pooling, (Graham *et al.*, 2018) re-introduced the input image within the network at several points to minimize the degradation of context in an image. For instant nucleus segmentation, (Graham and Rajpoot, 2018) proposed a network that has an encoder, decoder, and a series of residual blocks. (Xing *et al.*, 2016)'s CNN model used a bottom-up and top-down shape deformation approach to precisely segment the nuclei, where they combined the region iteratively and merged it for better initialization of tissue morphology. For segmentation and detection of histological objects, (Chen *et al.*, 2016), introduced a contour-aware model that extracts multi-level information under auxiliary supervision. (Huang *et al.*, 2017) proposed a convolutional network, which strengthens the input feature map by feeding preceding layer input as well as original input both. Their experiment also indicates that due to integration of identity mapping, model learns more compact features and reduces the vanishing gradient problem. In the case of an imbalanced dataset, predictions are biased towards high precision and low recall which is not tolerable especially in the medical field. This problem is addressed by (Salehi *et al.*, 2017), which trained the deep network, even with a highly imbalanced dataset, and handled effectively where false negative prediction is much dangerous than false positive.

The behavior of loss function such as weighted cross-entropy and dice loss with different learning rates examined by (Sudre *et al.*, 2017), on medical images and house dataset. Their experiment found that as the level of imbalance increases, overlap measure-based

loss function is more effective. A very efficient in terms of memory and time for semantic segmentation of road and indoor scene, an encoder-decoder architecture called SegNet by (Badrinarayanan *et al.*, 2017). SegNet network has an encoder path to produce a sparse feature map, a corresponding decoder network followed by a classifier (Softmax), which is used for pixel-wise classification. SegNet generates a sparse feature decoder that up-samples with the transferred pool and its lower resolution input from its encoder. For the automated segmentation of breast tissue, (Pan *et al.*, 2017) proposed sparse reproduction, multi-layer convolution systems, and morphological operations. To accurately segment near boundary regions, (Zhou *et al.*, 2018) used a residual network with a dilated convolution block. They utilize many hierarchical blocks in parallel to retrieve meaningful semantic information. To handle class imbalance problems or reducing false-negative predictions in healthcare, (Hashemi *et al.*, 2019) proposes a 3D-dense CNN with Tversky index-based asymmetric similarity loss that trains the network with the lowest surface distance. Complex boundary-related segmentation problem addressed by (Naylor *et al.*, 2019), by formulating a loss function based on intra- nuclear distance. They solved the segmentation problem as a regression task by calculating the distance between nuclei and formulating a global loss function that precisely segments the touching nuclei. Their encoder-decoder model outperform on FCN, FCN+PP, Mask R-CNN, U-Net, U-Net+PP experimented with TNBC and MoNuSeg datasets.

For the separation of overlapped nuclei, (Graham *et al.*, 2019a) used the synchronous segmentation method to calculate the vertical and horizontal separations of nuclear pixels to their centers of mass. For gland and lumen segmentation in colon histology, (Graham *et al.*, 2019b) introduced a rotation equivalent network that uses group-equivalent convolutional neural networks (G-CNN) having an additionally rotated kernel. For automatic segmentation of ductal carcinoma in situ in whole slide images, (Seth *et al.*, 2019) used a standard segmentation model UNet with ELU activation and batch normalization. Meaningful extensions in standard encoder-decoder by incorporating an additional module called attention gate by (Schlemper *et al.*, 2019), and attention as well as residual mechanism by (Lal *et al.*, 2021), where network is trained in such a way that it suppress irrelevant features while highlights the meaningful feature. By assimilating residual block, attention block, and bottleneck layer, (Lal *et al.*, 2021) incorporated three efficient blocks in encoder-decoder architecture to regain the meaningful information lost during max-pooling operation and correct localization of objects in the decoder section. For road scene segmentation (Malekijoo and Amirhossein, 2019), utilized the autoencoder-based model where convolu-

tion, deconvolution, and pyramid pooling were applied for reinforcing the local feature. For the segmentation of microscopic, MR, and CT images an encoder-decoder architecture by (Zhou *et al.*, 2020), linked meaningful connections to precisely locate the complex boundaries. For the segmentation of nuclei in pathology images, (Lal *et al.*, 2020) model, consists of adaptive color deconvolution, multiscale thresholding followed by morphological operations, and other post-processing steps. For the segmentation of medical images, a novel loss function by (Karimi and Davood, 2020), estimated Hausdorff distance using the morphological operation method, distance transformation method, and by circularly convoluted kernels of different radius. Utilizing methods of reduction of the Hausdorff distance, they train CNN for various microscopy images and compare their results with a commonly used loss function.

(Mehta *et al.*, 2019) achieved significant improvements compared to other existing models by utilizing a new convolutional unit called dimension-wise convolution to learn the spatial representation of channels in a better manner. In (Ding *et al.*, 2020), a multi-scale structure precisely segments the glands in histological images where the network has a fully convolutional residual structure with a dilated convolution. (Hanif and Muhammad Shehzad, 2020), proposed a competitive residual network by stacking multiple residual units called wide network. Their study concluded that the performance of such a wide network is better than the deep and thin network. (Chanchal *et al.*, 2021a) and (Aatresh *et al.*, 2021b) used separable convolution pyramid pooling and dimension-wise pyramid pooling for nuclei segmentation tasks. To capture the multilayer features of complex nuclei, Wan *et al.* (2020)'s encoder-decoder network used atrous spatial pyramid pooling, a concave point detection method, a combination of dice loss and binary cross-entropy loss as a joint loss function. A summary of DL based segmentation approaches is presented in Table 2.2.

Table 2.2: Summary of state-of-the-art DL techniques used for segmentation of medical images

Author	Application	Dataset	Method	Performance
(Ronneberger <i>et al.</i> , 2015)	Cell segmentation	ISBI-2012, Microscopic images	2D-CNN, Repeated application of Max-pooling and Up-convolution	IOU=92% for PhC-U373 and 77% for DIC-HeLa

Continued on next page

Table 2.2 – continued from previous page

Author	Application	Dataset	Method	Performance
(Milletari <i>et al.</i> , 2016)	Clinical	PROMISE 2012, MRI	Volumetric convolution, Feature channels doubles at each stage	Avg. Dice=86.9%
(Nogues <i>et al.</i> , 2016)	Lymph node cluster segmentation	MICCAI 2015, CT	Holistically-nested neural network, Structured optimization	Mean Dice coefficient=82.1% ± 9.6%
(He, 2016)	Image recognition	ImageNet, CIFAR-10 (Colored images)	Deep residual learning, Identity mapping by shortcuts	Error= 3.57%, 6.43%
(Chen <i>et al.</i> , 2016)	Gland segmentation	Warwick-QU, Histopathology	2D-CNN, Auxiliary supervision, Transfer learning	F_1 Score=91% and 77%
(Huang <i>et al.</i> , 2017)	Classification of natural images	CIFAR-10, 100, SVHN, and ImageNet	Feature-maps of all preceding layers are used as inputs	Error=5.19%, 19.64%, 1.59%
(Sudre <i>et al.</i> , 2017)	Training with unbalanced data	BRATS, MRI	2D and 3D deep learning framework	ReLU and Sigmoid / Dice / SGD / Dice loss was found to be more robust than the other loss functions
(Badrinarayanan <i>et al.</i> , 2017)	Indoor and road segmentation	CamVid, SUN RGB-D	Fully convolutional encoder decoder	Mean IoU=60.10, 31.84
(Hashemi <i>et al.</i> , 2019)	Medical image segmentation	MSSEG-2016, ISBI, MRI	3D fully convolutional, 3D densely connected architecture	Average Dice=69.9%, 65.74%
(Naylor <i>et al.</i> , 2019)	Segmentation of Nuclei	TNBC, TCGA, Histopathology	Fully convolutional network	AJI=55.98%, F_1 =78.63%

Continued on next page

Table 2.2 – continued from previous page

Author	Application	Dataset	Method	Performance
(Schlemper <i>et al.</i> , 2019)	Medical image analysis	TCIA Pancreas, Multi-class abdominal, CT	Focuses on target structures by employing attention gate	Dice score (TCIA)=82%, (Multi-class abdominal)=84%
(Lal <i>et al.</i> , 2021)	Liver cancer analysis	KMC liver, Histopathology	Employed residual block, attention mechanism, and joint loss function	$F_1=83.59\%$, and $JI=72.06\%$
(Malekijoo and Amirhossein, 2019)	Indoor and road segmentation	CamVid, RGB	Convolution deconvolution, Pyramid pooling	Mean IoU =48.90%, Accuracy= 88.49%
(Zhou <i>et al.</i> , 2020)	Low contrast medical image segmentation	Pelvic CT, Brain tumor, Nuclei segmentation	Multiscale dense connections, high resolution pathways	Dice ratio=95% , 90%, and F_1 score= 91.1 ± 10.2
(Lal <i>et al.</i> , 2020)	Nuclei segmentation	Gold-standard, KMC, Histopathology	Adaptive colour de-convolution, Multilevel thresholding	Pr, Re, $F_1 = 94.6\%$, 89.6% 91.2%, Pr, Re, $F_1 = 78.7\%$, 76.6% 77.2%
(Karimi and Davood, 2020)	Medical image segmentation	TRUS, PROMISE12, 3D CT-Liver, and 3D CT-Pancreas	Distance transform, Morphological erosion, Spherical convolution kernels	DSC= 95%, 87%, 94%, 78%
(Kumar <i>et al.</i> , 2017)	Clinical	Multiple organs, Histopathology	A CNN that produces a ternary map	AJI=50.83%, F_1 Score=82.67%
(Chanchal <i>et al.</i> , 2021a)	Clinical	TNBC, Kidney, MoNuSeg	Separable convolution pyramid pooling in encoder-decoder network	AJI=70%, 86%, 67%, F_1 Score=82%, 92%, 80%

Continued on next page

Table 2.2 – continued from previous page

Author	Application	Dataset	Method	Performance
(Aatresh <i>et al.</i> , 2021b)	Clinical	Kidney, TNBC	Attention based encoder-decoder, ASPP, Dimension-wise convolutions	AJI=87%, 70%, F_1 Score=93%, 82%
(Chanchal <i>et al.</i> , 2021b)	Clinical	TNBC, Kidney, MoNuSeg	High-resolution encoder-decoder path, An ASPP at bottleneck	AJI=73%, 94%, 72%, F_1 Score=84%, 97%, 83%

2.3 Literature Survey on Classification and Grading

In this section, a comprehensive review of the state-of-the-art deep learning solution for the classification and grading of histopathological images has been performed, techniques to extract relevant information from the digital histopathologic patches, and techniques to retain the extracted information. Deep learning solution for classification of H&E stained images is discussed in Section 2.3.1. Techniques to extract relevant information and techniques to retain extracted features is presented in 2.3.2 and 2.3.3 respectively. Deep learning models which utilize channel shuffle are briefly discussed in Section 2.3.4.

2.3.1 Deep learning solutions for classification of H&E stained images

Histological structure of clear cell RCC primarily has eosinophilic cells with thin-walled, staghorn-shaped vasculature. Papillary RCC has prominent papillary structures and foamy macrophages, while Chromophobe RCC consists of the prominent cell membrane, irregular nuclei, perinuclear halo, and pale to eosinophilic cytoplasm. With the use of the CNN model (Elazab *et al.*, 2020), the inherent features of renal tumors can be identified by capturing their unique characteristics. The automatic classification framework (Tabibu *et al.*, 2019), (Zhu *et al.*, 2021) of RCC subtypes employed CNN, pre-trained on ImageNet, and for grading (Tian *et al.*, 2019) used the Lasso model which includes a certain number of features to predict the grade.

Table 2.3: A summary of existing deep learning solutions for classification of H&E stained images

Author	Organ/Dataset	Method	Observations
(Araujo <i>et al.</i> , 2017)	Breast (Bioimaging-2015)	Processed patchwise on CNN + SVM	Captures nuclear feature as well as overall tissue organization
(Cruz-Roa <i>et al.</i> , 2018)	Breast (HUP, UHCMC, CWRU)	A 3-layer ConvNet architecture constituted	Trained ConvNet classifier module can be easily integrated into other computational frameworks
(Coudray <i>et al.</i> , 2018)	Lung (TCGA)	Inception-V3	Both deep-tuning as well as fine-tuning of network has been compared
(Xie <i>et al.</i> , 2019)	Breast (BreakeHis)	Inception-V3 and Inception-Resnet-V2 with transfer learning	Applied post-processing is helpful
(Zhou <i>et al.</i> , 2019)	Colorectal (CRC)	Cell graph convolutional neural network	Captures both the local and global features
(Tian <i>et al.</i> , 2019)	RCC (TCGA)	Lasso, a machine learning based method	Utilises nuclear feature for grading
(Tabibu <i>et al.</i> , 2019)	RCC (TCGA)	CNN and directed cyclic graph-support vector machine is used	Performed multi-class classification task into multiple binary classification tasks
(Togacar <i>et al.</i> , 2020)	Breast (BreakeHis)	A residual architecture built on attention module	Performance is better than transfer learning used in all presented methods
Jiang <i>et al.</i> (2019)	Breast (BreakeHis)	Convolutional layers with small SE-ResNet module	Most accurate on BreakeHis data and computationally efficient
(Iizuka <i>et al.</i> , 2020)	Stomach and Colon (Hiroshima and Hardoi hospital, Japan)	Inception-V3 network with depth multiplier of 0.35	Both CNN and RNN is used for final classification

Continued on next page

Table 2.3 – continued from previous page

Author	Organ/Dataset	Method	Observations
(Sun <i>et al.</i> , 2020)	Liver (TCGA)	ResNet-50 pretrained on ImageNet	Simple to implement using global label
(Wei <i>et al.</i> , 2020)	Colorectal (Multi-institutional data)	ResNet with pre-trained weights of ImageNet	A comprehensive annotation process was performed
(Shahidi <i>et al.</i> , 2020)	Breast (BreaKHis, TMA, Camelyon, BACH)	Applies the various pre-trained model	Analyzed the effect of pre-processing, data augmentation, and transfer learning
(Chen <i>et al.</i> , 2020)	Liver (TCGA)	Pre-trained Inception-V3	Genomic analysis of Hepatocellular carcinoma has been performed
(Zhu <i>et al.</i> , 2021)	RCC (DHMC-USA, TCGA)	Sliding-window method for patches creation and ResNet-18 is used for patch-prediction	Analyzed on both surgical resection and biopsy whole-slide images of RCC
(Hirra <i>et al.</i> , 2021)	Breast (HUP, CWRU, CINJ, TCGA)	A deep belief network is formed by stacking the layers of restricted boltzmann machine	Proposed model needs quite minimal hardware resources
Aatresh <i>et al.</i> (2021a)	Liver (TCGA, KMC Hospital)	A CNN using attention mechanism and atrous spatial pyramid pooling	Effective in extracting HCC features, Parameter is less

The works, (Araujo *et al.*, 2017), (Cruz-Roa *et al.*, 2018), (Xie *et al.*, 2019), (Togacar *et al.*, 2020), (Shahidi *et al.*, 2020), (Hirra *et al.*, 2021), where deep learning is successfully applied for the detection of malignant tumor cells in breast tissues. The study Jiang *et al.* (2019) used a highly optimized CNN module and not using any pre-trained weights is very effective in tumor detection of BreaKHis data. The study of hepatocellular carcinoma (HCC) in liver tissue Chen *et al.* (2020), Sun *et al.* (2020) used transfer learning, while Aatresh *et al.* (2021a) CNN model, trained from scratch on TCGA liver dataset. A summary

of existing deep learning solution for classification of H&E stained images shown in Table 2.3.

2.3.2 Techniques to extract relevant information

Attention mechanisms is widely employed in the field of segmentation and classification of histopathology images to focus on relevant features. Attention-based feature descriptor (Woo *et al.*, 2018) (Schlemper *et al.*, 2019), (Chen *et al.*, 2015), (Li *et al.*, 2020), where two different attention modules called channel attention and spatial attention are simultaneously incorporated to assign different weight on object of different scales. The work (Zhang and Chen, 2020) generates high resolution spatial maps by attention pooling to predict class from X-ray images. A naive stacking of the attention mechanism leads to a performance drop in the network. The new residual attention mechanisms (Wang *et al.*, 2017) solved the problem by incorporating methods like stacked network structures, and a mixed attention mechanism.

The representative works (Zhao *et al.*, 2018), (Zhang *et al.*, 2019) deployed attention module in the network and focused to avoid local constraints and aggregates global context. The behavior of channel-wise discriminative features adaptively calibrated in squeeze-and-excitation (Hu *et al.*, 2020), (Fu *et al.*, 2020) to model the interdependences between channels. BHCNet (Jiang *et al.*, 2019) demonstrated the use of attention mechanism in histopathology image classification by using a small ResNet architecture along with squeeze and excitation block for binary and eight malignant subtypes classification of BreakHis dataset with less computational complexity. BreastNet (Togacar *et al.*, 2020) contains channel attention, spatial attention, residual unit and hyper column unit to achieve better classification accuracy. LiverNet (Aatresh *et al.*, 2021a) for four-class classification of liver hepatocellular carcinoma from liver histopathology images, incorporated attention modules and atrous spatial pyramid pooling block to capture the representative feature. This work adopted a transformation method based on squeeze and excitation (SE), a lightweight gating mechanism that helps to obtain improved performance.

2.3.3 Techniques to retain extracted features

Many works retain spatial information by incorporating skip connections (He, 2016), (Huang *et al.*, 2017) to avoid degradation of information between output and input and utilized all preceding layers as input to the next layer throughout the network. This idea is effective in

feature reuse and thus reduces the vanishing gradient problem and therefore training of a large number of deep neural network layers became simpler. CNN models (Simonyan and Zisserman, 2014), (Szegedy *et al.*, 2014) were proved to capture more complex features as the model depth increases, and by increasing width, models can capture fine-grained features. Two layer shallow model and a twenty two layer deep network (Khoshdeli *et al.*, 2018) tested on TCGA kidney dataset and concluded that deep model classified low grade granular tumor and high grade clear cell carcinoma better than shallow model.

Resolution is another factor that improves accuracy but was observed to saturate and adding more layers further leads to higher training errors. Parallel layers of convolution modules (Szegedy *et al.*, 2016), (Xie *et al.*, 2016), (Alom *et al.*, 2018) showed that increased cardinality works better than going deeper or wider in the network and adopts residual connection to concatenate multiple paths of a set of transformations. Although these methods are effective, the above techniques used the traditional transformation method to retain the extracted features. This work proposed a CNN model that effectively enriches nuclear features by fusing the information of multiple parallel branches and attains a considerable margin on the existing network.

2.3.4 Deep learning models which utilize channel shuffle

To get optimal speed and accuracy, many efforts have been made to design deep learning architectures computationally efficient and lightweight. Recently, the shuffling of channels within a layer was exploited for efficient model design in convolutional neural network. The shuffle operation proposed (Zhang *et al.*, 2017), (Ma *et al.*, 2018), is a stack of channel shuffle units and group convolution, and works better under smaller computational budgets. To meet different target complexities, these techniques effectively utilize many convolution groups and the advantages of adjustment of channels. However, these networks are specially designed for small models but provide a promising research direction to adjust the trade-off between accuracy and computational complexity.

Some works (Yang, 2021), (Yang *et al.*, 2022), adopts shuffle unit and applied various attention mechanism to the shuffled version of multiple sub-features. The idea is effective since it enables sufficient communication across channels and helps in object detection by information communication between sub-features. The operation of channel shuffle and patch shuffle (Kang *et al.*, 2017), (Kumawat *et al.*, 2022) studied as a regularization technique in deep learning. Here small patches from the same spatial location of multiple channels in the feature map are extracted and to produce a better activation map, these small

patches are randomly shuffled between selected channels.

2.4 Research Gap Analysis

2.4.1 Gap analysis for nuclei segmentation from H&E stained histopathology images

- Many of the authors reported their results which are very efficient, but due to the processing of large-scale histopathology image datasets, the feature extraction model needs to be both time and memory efficient. Reducing computational complexity and training time is still a challenge.
- One of the major challenges in the segmentation of nuclei in histopathology images is the separation of clumped nuclei. An accurate and efficient segmentation algorithm is still open-ended research due to the complexity of histopathology images.
- There are many potential opportunities for further improvement in the segmentation process by incorporating new loss functions to better train the deep learning model, extracting relevant features using efficient CNN block, and retaining those extracted features at the latter layer.
- Most of the deep learning algorithms for nuclei segmentation in literature are organ specific. An efficient deep learning framework for multi-organ nuclei segmentation needs to be explored further.

2.4.2 Gap analysis for classification of subtypes of histopathology images

- The ultimate aim of using AI in a clinical workflow is speed and accuracy. The relevance and potential of automatic framework for classification of histopathology image has demonstrated by many authors, but the reported result are still sub-optimal for clinical use.
- In most of the CNN models, classification accuracy depends on the patch-level feature. Patch level accuracy needs to be improved such that essential local detail makes the architecture more robust.

- Many recent classification algorithms, networks consists of a large number of layers or fuse several deep architectures to get optimal accuracy. The number of parameters and training time need to be optimized.
- In most of the deep learning architecture used in literature so far, one major concern is to classify subtypes of cancerous tissues. Subtle distinctions between different cancer tissues require extracted features should be highly expressive.
- Most of the datasets used in literature for histopathological image analysis have an issue of high-class imbalance.
- There is no publicly accessible processed data for Renal cell carcinoma (RCC) classification for the reproducibility of the result.

2.4.3 Gap analysis for grade prediction of histopathology images

- The previous study on deep learning applications on kidney cancer histopathology images is mostly focused on the classification of RCC subtypes. There is no publicly available dataset for grading of clear cell RCC.
- Most of the previous work focused on transfer learning techniques and using pre-trained weights of the ImageNet dataset. Of course, pre-trained, ImageNet weights transfer powerful texture features, even then the development of an end-to-end deep learning model that is intended to capture morphological and clinical features from the overall tissue organization of histopathological images needs to explore more.

2.5 Quality Metrics for Nuclei Segmentation, Classification, and Grading of Histopathology Images

For evaluating the model, different performance metrics used are known as *Accuracy*, *Precision*, *Recall*, F_1 , ($\beta = 1$) (Aatresh *et al.*, 2021a), (Naylor *et al.*, 2019), *Intersection over union (IoU)* Lal *et al.* (2021), *Aggregated Jaccard index (AJI)* (Kumar *et al.*, 2017). The model predictions are finally grouped into four, namely True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

1. Precision

Precision is the ratio of total number of correctly classified samples to the number of samples classified as positive. This Metric gives us an estimate of how well the model identifies negative samples. Given TP is True Positive (correctly classified positive sample) and FP is False Positive (incorrectly classified negative samples), The equation 2.1 defines Precision score.

$$Precision = \frac{TP}{TP + FP} \quad (2.1)$$

2. Recall

Recall is the ratio of total number of correctly classified samples to the total number of an actual positive sample. This Metric gives us an estimate of how well the model identifies positive samples. Given TP is True Positive (correctly classified positive sample) and FN is False Negative (incorrectly classified positive samples), the equation 2.2 defines Recall.

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

3. F1 score

F1 score is the harmonic mean of Precision and Recall score. Being a combination of both, it gives an excellent estimate of how well the model performs in estimating both positive and negative samples and is the state-of-the-art metric in the study of image classification. Mathematically, equation 2.3 defines F1 score.

$$F1 = \frac{Recall * Precision}{Recall + Precision} \quad (2.3)$$

4. Jaccard Coefficient

It is a commonly used method for the measurement of overlap of two sets. It is a measure of similarity or dissimilarity between binary data. Jaccard Coefficient measurement is the best method to evaluate the performance in the case of nuclei segmentation from histopathology images in deep learning. equation 2.4 defines $Jaccard(A, B)$.

$$Jaccard(A, B) = \frac{Number\ of\ items\ in\ A \cap B}{Number\ of\ items\ in\ A \cup B} \quad (2.4)$$

If consider the Jaccard Coefficient of a set with itself, the ratio will be one and the Jaccard Coefficient will be one, $Jaccard(A, B) = 1$. If two sets are disjoint and have no members in common then Jaccard Coefficient will be zero, $Jaccard(A, B) = 0$. If two sets are not the same size, then the Jaccard Coefficient will always assign a number between zero and one, $0 \leq Jaccard(A, B) \leq 1$. The Jaccard distance loss is defined in equation 2.5.

$$Jaccard\ distance/Jaccard\ loss = (1 - Jaccard\ Coefficient) \quad (2.5)$$

5. Aggregated Jaccard index (AJI)

The concept of AJI was proposed by (Kumar *et al.*, 2017). AJI compute and measure the performance of segmentation better than the global Jaccard index by incorporating the concept of the connected components. AJI compute and measure the performance of segmentation better than the global Jaccard index by incorporating the concept of the connected components.

$$AJI = \frac{\sum_{k=1}^M |G_k \cap P_S^k|}{\sum_{k=1}^M |G_k \cup P_S^k| + \sum_{R \in U} |P_R|} \quad (2.6)$$

AJI is defined in equation 2.6, Where G_k is k^{th} nuclei of ground truth having M nuclei. P_S^k is S^{th} Connected component in prediction result having the highest Jaccard index with ground truth. Each index of S cannot be utilized more than once. U represents a set of the connected components in the prediction result without corresponding ground truth. P_R represents the ground truth and intersection that are not in the associated space. AJI is a connected component-based method that is the improved version of the pixel-based global Jaccard index. A higher AJI value indicates a better segmentation method.

6. **IoU Score** To measure classification performance, Intersection over union (IoU) is one of the most preferred metrics. IoU of two non-empty sets X and Y can be calculated to measure the similarity between the sets. The IoU score is given in equation 2.7.

$$IoU\ Score = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{(TP + FP + FN)} \quad (2.7)$$

7. **Accuracy** It is the ratio of the correctly classified samples to the total number of samples. Accuracy gives a good estimate of model performance in case of balanced test data. In terms of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The equation 2.8 defines accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

8. **Confusion Matrix** A confusion matrix is a tabular tool for evaluating the performance of the model. The target variable has two classes in binary classification, whereas in multi-class classification, the target variable has more than two classes. The TP, TN, FP, and FN measures of the model's performance are summarized in the confusion matrix of a classifier. Fig. 2.1 describes the confusion matrix for binary class and multi-class problems.

	Predicted		
		P	N
Actual		TP	FN
	P	TP	FN
	N	FP	TN

(a)

	Predicted				
		A	B	C	D
Actual	A	TP_A	E_{AB}	E_{AC}	E_{AD}
	B	E_{BA}	TP_B	E_{BC}	E_{BD}
	C	E_{CA}	E_{CB}	TP_C	E_{CD}
	D	E_{DA}	E_{DB}	E_{DC}	TP_D

(b)

Figure 2.1: Confusion matrix (a) Binary class problem (b) Multi-class problem

Chapter 3

Deep Learning Architectures for Nuclei Segmentation of Histopathology Images

3.1 Introduction

This Chapter discusses about the need for nuclei segmentation prior to tissue subtypes classification. Further, the details of the newly proposed method for nuclei segmentation are presented. The first proposed algorithm automatically segments the complex nuclei present in histology images by implementing an effective encoder–decoder architecture with a separable convolution pyramid pooling network (SCPP-Net). The detailed information about the proposed SCPP-Net technique is explained in the Section 3.2. The next proposed algorithm focuses on the segmentation of the nuclear region of multi-organ histopathology images. Compared to previous architecture this architecture is wide and deep that effectively leverages the strength of residual learning as well as encoder-decoder architecture. Proposed architecture addressed the problem of degradation of information in deeper network by introducing deep and wide residual network which is easier to train and optimize. Detailed mathematical analysis, implementation, and visual segmentation result is presented in Section 3.5. In the literature review, it has been observed that to deal the vanished boundary of detected nuclei is a challenging because of the lack of a robust approach. This problem is addressed by proposing an efficient loss function that better trains the proposed deep structured residual encoder-decoder network. The experimental results of proposed technique are discussed in the Section 3.8.

3.2 Proposed Architecture-1

An efficient and robust deep learning architecture called separable convolution pyramid pooling network (SCPP-Net) for segmentation of kidney and breast histopathology images has been proposed and also evaluated its effectiveness. The working method of proposed SCPP-Net model was based on the principle of extracting more relevant features at a higher level. The SCCP layer increased the receptive field by varying four different dilation rates and keeping a fixed kernel size. Dilation rate was an additional parameter that varied to the resultant feature maps to visualize larger areas. This is advantageous in the sense that we have added little computational complexity and were able to obtain a larger receptive field. This CNN-based architecture is helpful for overcoming the problem of closeness and overlapping nuclei by increasing the receptive field at a higher level.

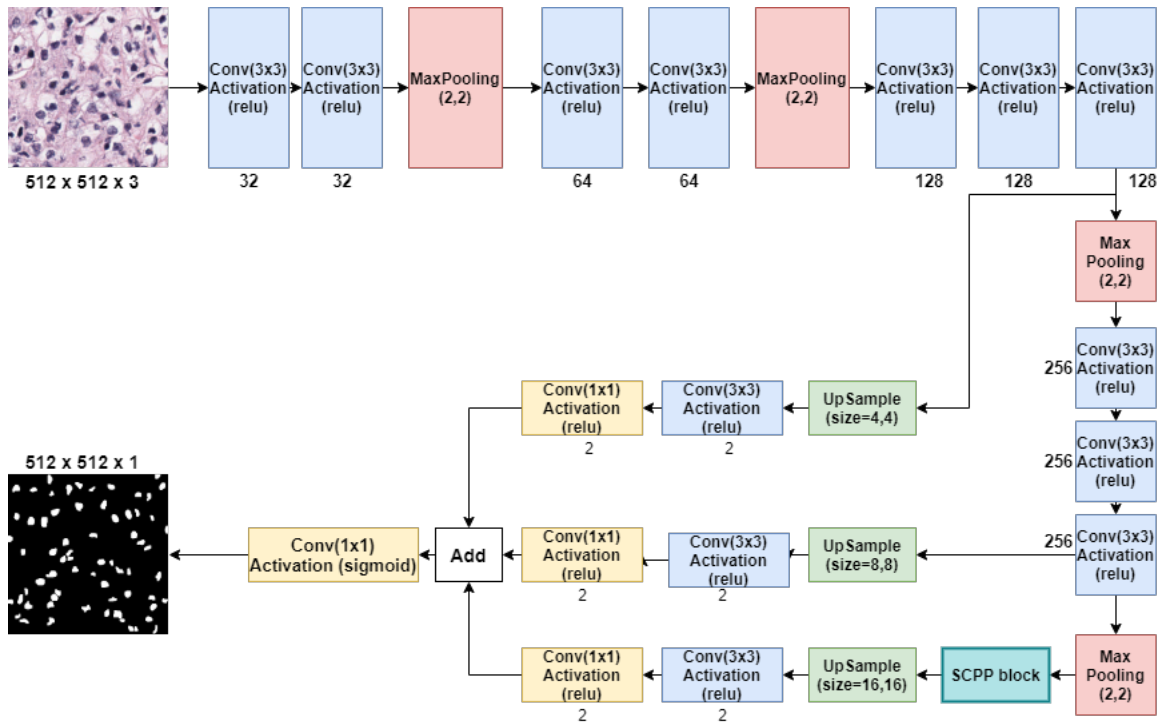


Figure 3.1: Detailed diagram of proposed separable convolution pyramid pooling network (SCPP-Net)

3.2.1 Description of SCPP-Net architecture

Based on (Ronneberger *et al.*, 2015), We conceived of the idea of an encoder and decoder. In the down-sampling path, performed convolution and max-pooling operations for the input

images. In this path, the focus was on capturing the context in the image. In the contraction path, the image size gradually reduced, while the depth of the image gradually increased. In the deep layer of the convolution operation, SCPP block was used to extract more relevant features to increase the receptive field by varying four different dilation rates and keeping a fixed kernel size. In the expanding path, the size of the image gradually increased and the depth gradually diminished. Decoder path enabled precise localization by gradually applying up-sampling. For each step in the expanding path, utilized a skip connection by concatenating the output of the decoder convolution layers with feature maps from the contraction path at a similar level. This is an end-to-end completely convolutional system, and it did not contain any dense layer. This architecture is invariant to image size, which means it can accept any size of input image. The detailed diagram of the proposed SCPP-Net model is presented in Fig. 3.1.

Standard convolution layer: Convolution layers consist of a set of learnable parameters. The input image has three color channels (RGB) and has a dimension of $512 \times 512 \times 3$, on which a filter of size (3×3) is applied. For a standard 2D convolution and if one have $x(p,q)$ is the input feature map and $h(p,q)$ is the filter kernel, then $y(p,q)$ is the output of 2D convolution expressed in equation 3.1.

$$y(p, q) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} x(j, k)h(p - j, q - k) \quad (3.1)$$

When apply a kernel of $K \times K$ on the image of $N \times N$ with padding P and stride S , will get the image size as in equation 3.2.

$$NXN \rightarrow \left(\frac{N - K + 2P}{S} + 1 \right) \left(\frac{N - K + 2P}{S} + 1 \right) \quad (3.2)$$

Without padding and single-step, stride, the relation between the input image and the output image is in equation 3.3.

$$NXN \rightarrow (N - K + 1)(N - K + 1) \quad (3.3)$$

Pooling layer: Pooling is a concept that makes a deep learning model location invariant, scale-invariant, and rotation invariant. It acts as an additional layer. Whatever an image has a higher value within the kernel dimension, where higher value corresponds to object region, only that area will reflect in the higher layer which makes detection task easy. By

applying a kernel of size $K=(2 \times 2)$ and stride size $S=2$ on (4×4) two-dimensional image the resultant feature size is shown in Fig. 3.2.

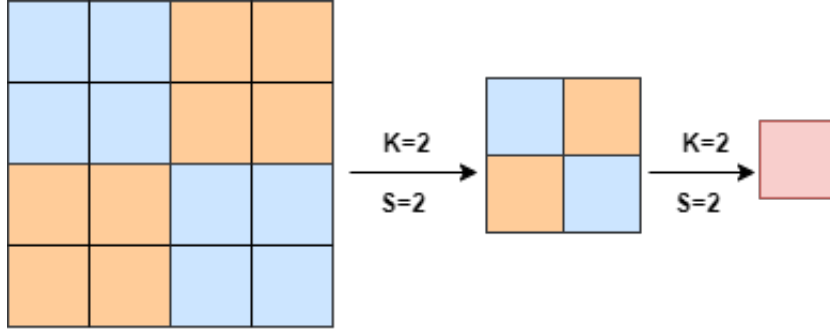


Figure 3.2: Max-pooling with kernel $K=2$ and stride $S=2$

Atrous spatial pyramid pooling: In atrous spatial pyramid pooling (ASPP), atrous convolution is used with different dilation rates for a better receptive field of view. Atrous convolution is also known as dilation convolution. As described by (Chen *et al.*, 2017), assuming a one-dimensional signal, the output $y[n]$ of the atrous convolution of a one-dimensional input signal $x[n]$ with a filter $w[n]$ of length M is defined as:

$$y[n] = \sum_{i=1}^M x[n + r \cdot i]w[i] \quad (3.4)$$

For each spatial position n on the output y and kernel w , atrous convolution is applied to input x . Convolution of input x with up-sampled filters produced by inserting $(r - 1)$ zero between two consecutive filters produced one-dimensional output of length M . Each location of (m, n) of the output y , atrous convolutions applied to input x results in a two-dimensional image of size $M \times N$ given in equation (3.5). In equations (3.4) and (3.5), r corresponds to the dilation rate with which the input signal goes into the sample. When $r=1$, a normal convolution operation occurs and it is normal CNN networks. When $r>1$, then it allows us to expand the filter strides with different rates, so this convolution becomes atrous or dilation convolution, which increases receptive field of view.

$$y(m, n) = \sum_{i=1}^M \sum_{j=1}^N x(m + r \cdot i, n + r \cdot j)w(i, j) \quad (3.5)$$

The receptive field (global view) visualization of atrous convolution with dilated rates, $r = 2, 4, 6,$ and 8 , is shown in Fig. 3.3. The selection of dilation rates was based on obtaining

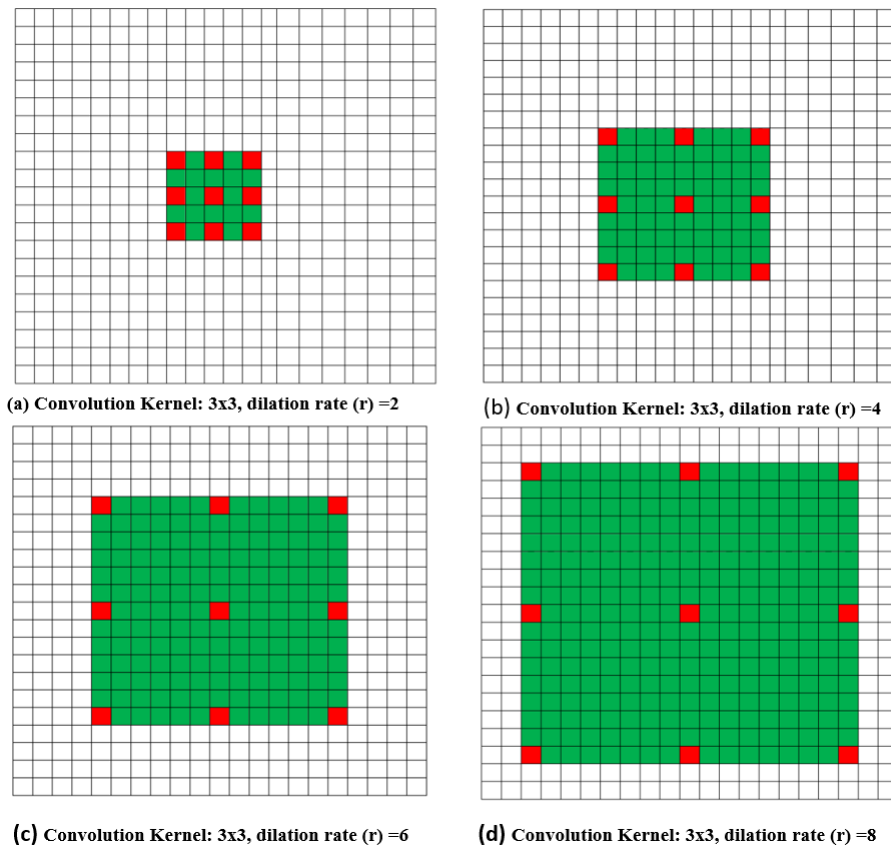


Figure 3.3: Visualization of the receptive field (global view) of atrous convolution with different dilation rates

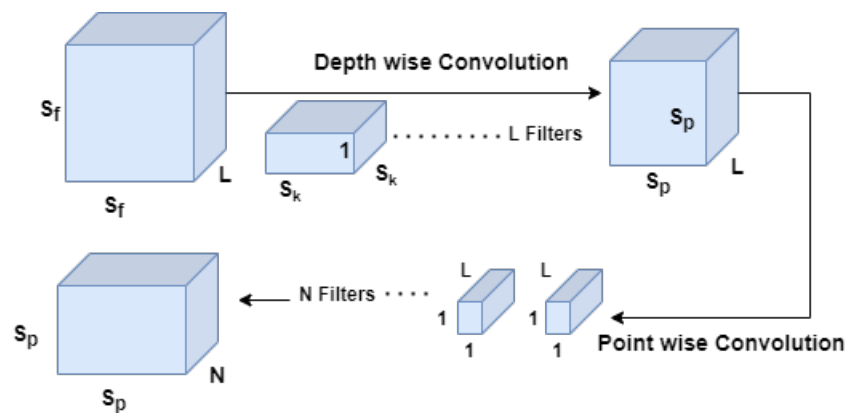


Figure 3.4: Visualization of depth-wise separable convolution

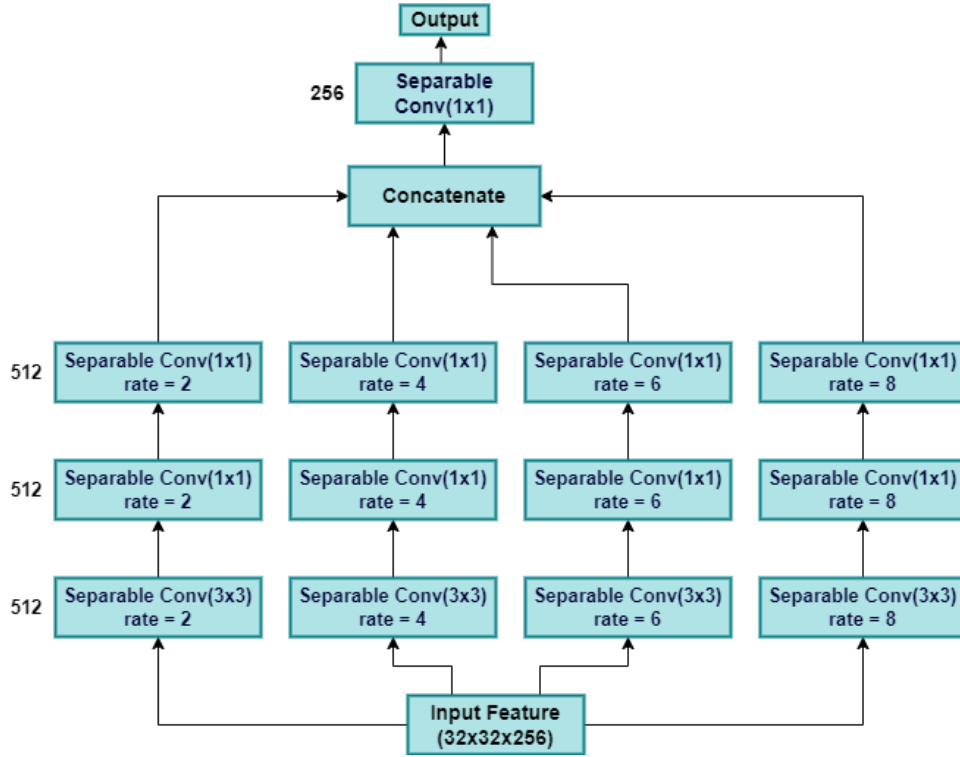


Figure 3.5: Separable convolution pyramid pooling block

optimum accuracy while keeping the number of parameters minimum. In Fig. 3.3, the green color indicates how the receptive field increases with different dilation rates. ASPP is a version of spatial pyramid pooling, which is used by (He *et al.*, 2015). It has been convolved the input feature map parallelly with the different dilation rates (r) and fused them. ASPP helped us identify the object of the same class with different scales in the image. So, with the identification of the different object scales, accuracy also improved. With this motivation, the proposed design used ASSP, but instead of a normal convolution, separable convolution (He *et al.*, 2019) is used.

Visualization of detailed computational complexity in standard convolution and depth-wise separable convolution is shown in Fig. 3.4. Convolution of $S_k \times S_k \times 1$ with feature size $S_f \times S_f \times L$ and L filters yields feature size of $S_p \times S_p \times L$. Convolution of N filters of $1 \times 1 \times L$ with $S_p \times S_p \times L$ results $S_p \times S_p \times N$. By this way total number of multiplications in depth-wise separable convolution is $(S_k \times S_k \times 1)(S_p \times S_p \times L) + (1 \times 1 \times L)(S_p \times S_p \times N)$, which is significantly less than standard convolution. ASPP used separable convolution with 3×3 kernel sizes and dilation rates of 2, 4, 6, 8. It has been concatenated all of the above with the same depth and convoluted (separable) with 1×1 kernel size so that the model was able to

extract maximum features. The decoder path was pipelined in such a way that it up-sampled all the feature maps parallelly. The final feature map for the model using the upsampling technique (Hariharan *et al.*, 2014) has been obtained. Feature maps from different parts of the network have been taken, and to scale into similar dimensions Up-Sampled to the same height and width. The block diagram shown in Fig. 3.5 describes the detailed architecture of the proposed SCPP block.

3.3 Implementation and Training

Training and implementation have been conducted in Google Colab GPU with the latest version of Keras and TensorFlow Python-3 framework to speed up the development process. Before performing the training, it has been reviewed the dataset to test its validity with the proper data and labels. In the process of training, some data augmentation as well as pre-processing has been performed. To increase the diversity of data for training, the strategy of data augmentation by the creation of artificial variations, such as rotation, flipping, padding, and zooming, dataset were expanded. The purpose of augmentation is to prevent CNN from undergoing overfitting, and the augmentation approach depends on the type of architecture used. This case used different types of data augmentation for the Triple Negative Breast Cancer (TNBC) and Multiple organs nuclei segmentation (MoNuSeg) histopathology datasets, so that obtained more images for the training set as well as the validation set. Flipping, rotation, dihedral, and random lighting (the threshold is 0.5) has been used. It is also performed color normalization for the TNBC and multi-disease histopathology datasets.

Color normalization is needed to match the color characteristics of an image to those of the target image. In this process, it has been matched the mean and standard deviation of each color channel in the two images. We have trained the proposed model on the training set for a given range of hyperparameters and evaluated the efficiency of the model on the test and validation datasets. Description of histopathology datasets used for this purpose is presented in Section 3.3.1. Training details in Section 3.3.2 includes details of hyperparameters, optimizer, and loss function used in this work.

3.3.1 Datasets used for nuclei segmentation

1. Dataset-1

Hematoxylin and Eosin (H&E) stained Triple Negative Breast Cancer (TNBC) dataset: This dataset was created by (Naylor *et al.*, 2019). This dataset was a corpus of 50 H&E-stained histopathology images of human breast tissue and consisted of 4022 cells that were accurately annotated by a pathologist.

2. Dataset-2

H&E stained kidney dataset: This dataset was prepared by (Irshad *et al.*, 2015). This corpus of kidney renal clear cell carcinoma images was obtained from the National Cancer Institute. A subset of 730 H&E-stained histopathology images for the task of nuclei segmentation has been obtained.

3. Dataset-3

Multiple organs nuclei segmentation (MoNuSeg) histopathology dataset: Multi-organ nucleus segmentation dataset prepared by (Kumar *et al.*, 2017). This dataset consists of 44 images and included seven different organs of kidney, liver, colon, breast, prostate, bladder, and stomach, with a total of 26747 annotated nuclei. These images of size 1000x1000 pixels are acquired from TCGA and annotated by expert pathologists. For my purpose, it has been created image patches of 512x512 pixels and divided into the train, test, and validation sets. This has 288 images for the training of deep neural networks. The test set consists of a total of 56 images, 8 images of each organ for evaluation of the model.

3.3.2 Training details

Adam described in (Kingma and Ba, 2014) used as an optimizer, and batch size was four (for kidney dataset) and one (for breast dataset and multiple organs multi-disease dataset) for the best possible result. These are the hyperparameters that varied: (1) learning rate α , (2) weight decay constant λ , (3) the size of the convolution filter n_f , (4) dilation rate, and (5) batch size.

According to (Kandel and Castelli, 2020), batch size and learning rates have a significant impact on the model performance. The data folder consists of 654 training images in the kidney dataset without any data augmentation. After data augmentation, the TNBC and multi-organ datasets had 200 and 288 training images, respectively. The learning rate

of 0.0001 for training all the models with three datasets has been selected. Selection of learning rate and weight decay constant was based on trial and error, as it cannot be calculated analytically. As there was a strong correlation between batch size and learning rate, proposed model yielded satisfactory results for batch size four for the kidney dataset and one for the other two cases. The selection of those two hyperparameters, such as batch size and learning rates, sped up the training process. The proposed model also utilized flexible size convolution filters based on available resources. To measure the performance of the deep learning model for nuclei segmentation, four different performance metrics are mostly preferred namely *Precision*, *Recall*, F_1 , ($\beta = 1$) Lal *et al.* (2021), (Naylor *et al.*, 2019), and *Aggregated Jaccard index (AJI)* (Kumar *et al.*, 2017) were used. Binary cross-entropy (BCE) loss function used in (Ronneberger *et al.*, 2015) is opted in this work.

3.4 Results and Discussions

After the simulation, we were able to compute and analyze the experimental results of nuclei segmentation on three publicly available H&E-stained histopathology image datasets. It has been used the image quality metrics F1 score used by (Naylor *et al.*, 2019), (Lal *et al.*, 2021) and Aggregated Jaccard Index (AJI) scores used by Naylor *et al.* (2019) to quantify the effectiveness of the proposed SCPP-Net architecture and other existing deep learning segmentation methods. The number of parameters and FLOPs quantified the GPU computing power of the proposed SCPP-Net architecture and other existing deep-learning segmentation methods.

3.4.1 Ablation study

To demonstrate the effectiveness of the proposed SCPP-Net architecture, an ablation study was performed on two publicly available kidney and TNBC histopathology datasets. In this study, the effectiveness of each module was demonstrated quantitatively and qualitatively. The results of the proposed SCPP-Net architecture, with and without separable convolution, are reported in Tables 3.2, 3.1, 3.3 and Fig. 3.6.

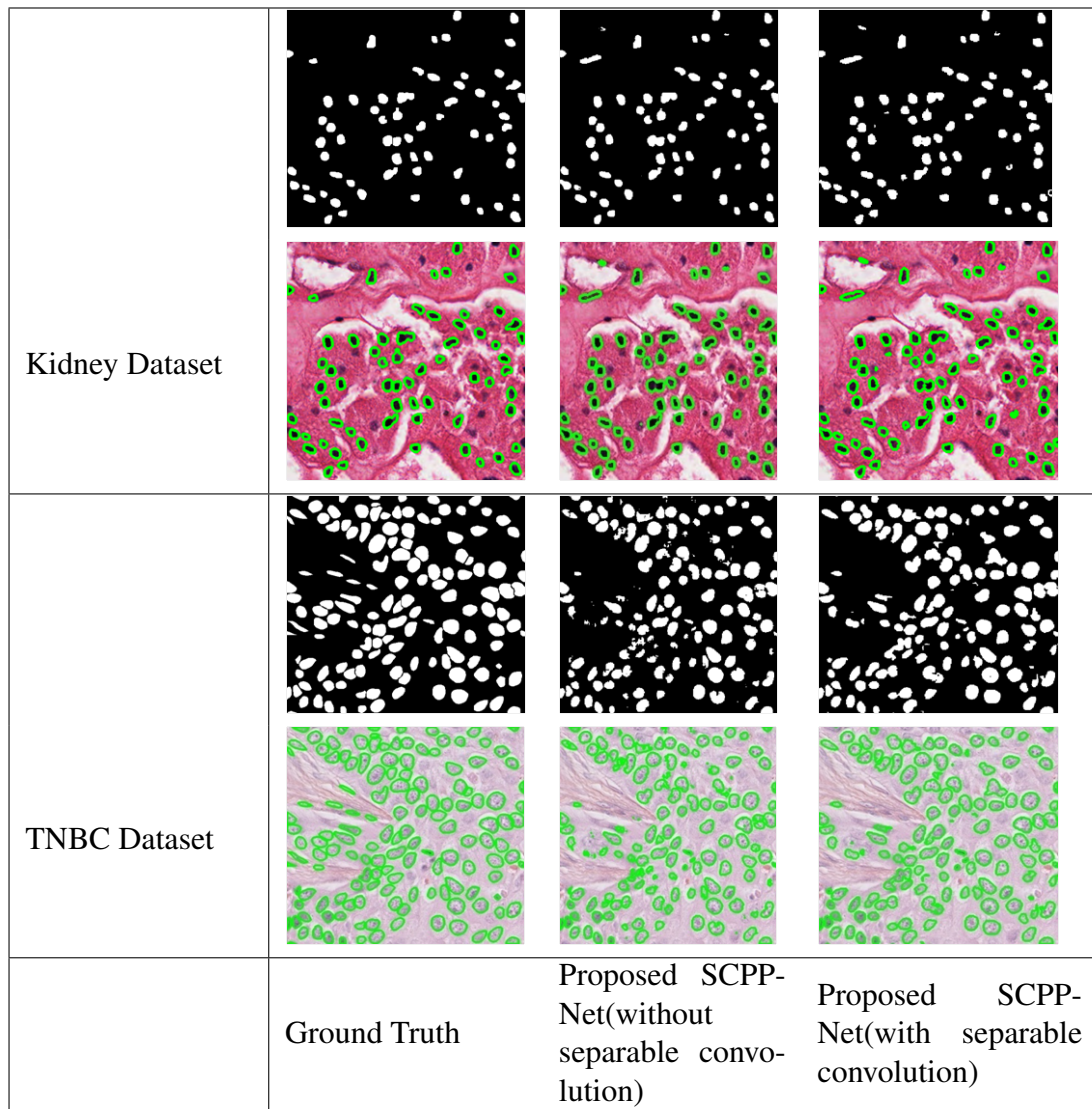


Figure 3.6: Binary and overlay images of proposed SCPP-Net architecture (with and without separable convolution) on kidney and TNBC datasets

Table 3.1: Performance measure of proposed model with kidney dataset

Model	F1 Score	AJI Score	Parameters	FLOPs
Proposed SCPP-Net (without Separable)	0.8975	0.8290	9,267,899	18,520,320
Proposed SCPP-Net (with Separable)	0.9203	0.8592	5,088,955	10,162,445

Table 3.2: Performance measure of proposed model with TNBC dataset

Model	F1 Score	AJI Score	Parameters	FLOPs
Proposed SCPP-Net (without Separable)	0.7661	0.6551	9,267,899	18,520,320
Proposed SCPP-Net (with Separable)	0.8168	0.6998	5,088,955	10,162,445

Table 3.3: Performance measure of proposed model with MoNuSeg dataset

Model	F1 Score	AJI Score	Parameters	FLOPs
Proposed SCPP-Net (without Separable)	0.7694	0.6341	9,267,899	18,520,320
Proposed SCPP-Net (with Separable)	0.8010	0.6710	5,088,955	10,162,445

3.4.2 Comparison with state-of-the-art segmentation methods

To demonstrate the effectiveness of the proposed SCPP-Net architecture with other existing deep learning segmentation methods (Ronneberger *et al.*, 2015), (Badrinarayanan *et al.*, 2017), (Schlemper *et al.*, 2019), (Naylor *et al.*, 2019), (Wan *et al.*, 2020), nuclei segmentation results were evaluated and compared on three publicly available kidney, TNBC, and MoNuSeg histopathology image datasets.

Table 3.4: Performance comparison of architectures on kidney dataset

Model	F1 Score	AJI Score	Parameters	FLOPs
U Net(2015)	0.8537	0.7489	31,378,945	62,746,139
Segnet(2017)	0.8972	0.8304	18,826,753	47,065,663
Attention UNet(2019)	0.9135	0.8590	31,902,629	63,791,573
DIST(2019)	0.8992	0.8272	7,771,873	15,525,181
ASPPU-Net(2020)	0.9052	0.8293	4,307,713	8,610,587
Proposed SCPP-Net	0.9203	0.8592	5,088,955	10,162,445

Table 3.4 shows a quantitative segmentation results comparison of the proposed SCPP-Net architecture with five other existing deep-learning segmentation architectures on the kid-

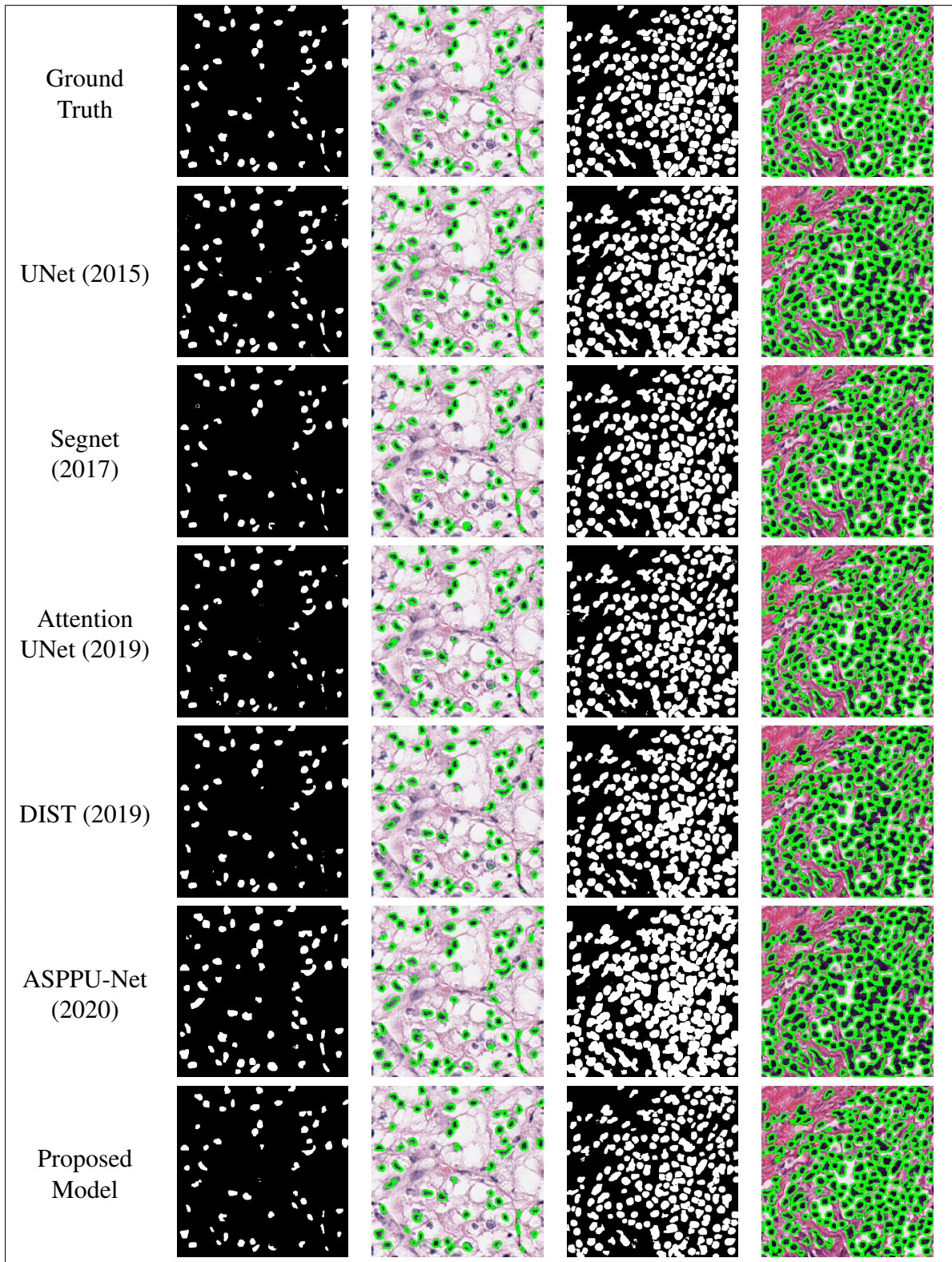


Figure 3.7: Visual segmentation results comparison of different models on kidney dataset

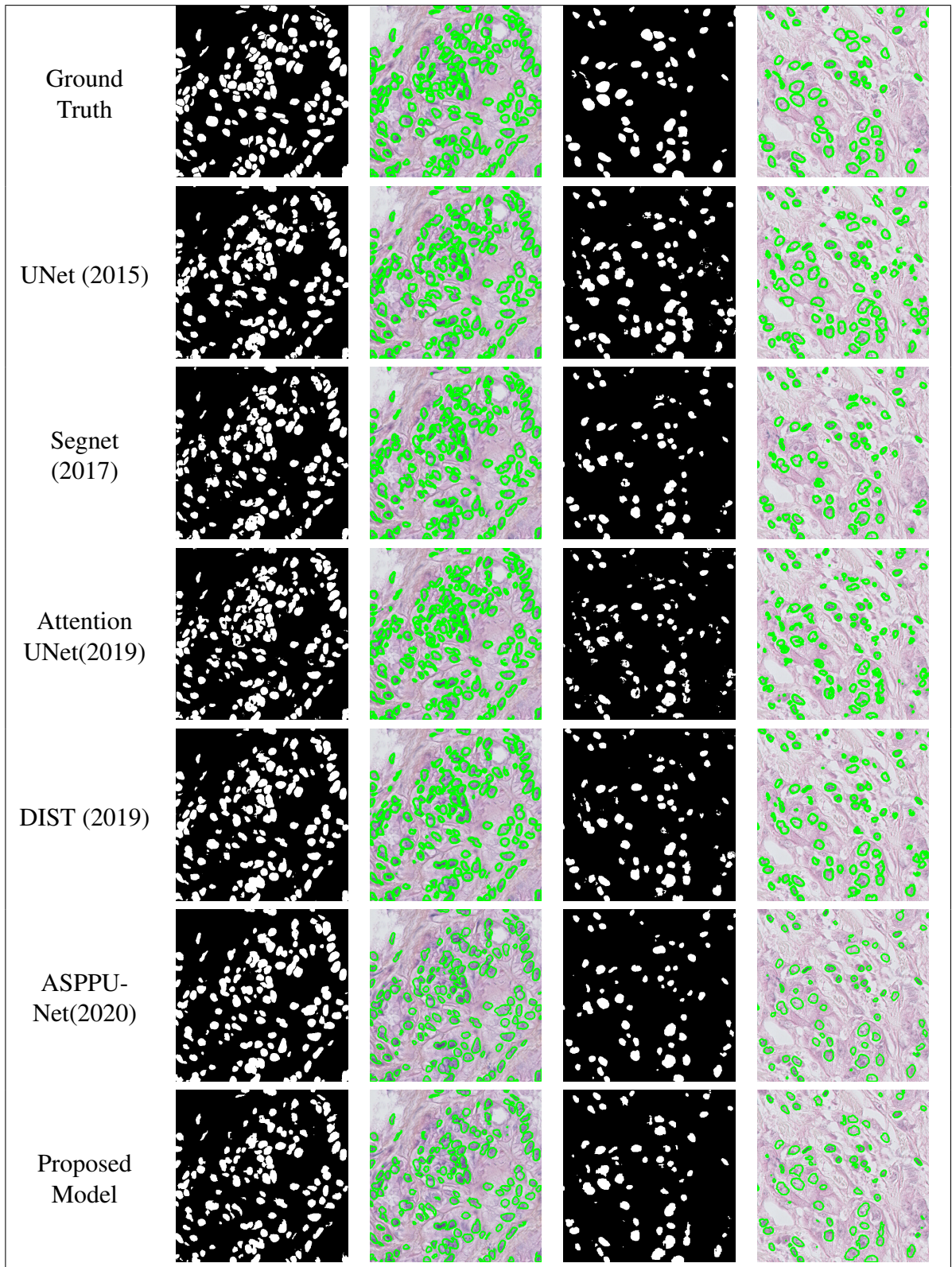


Figure 3.8: Visual segmentation results comparison of different models on TNBC dataset

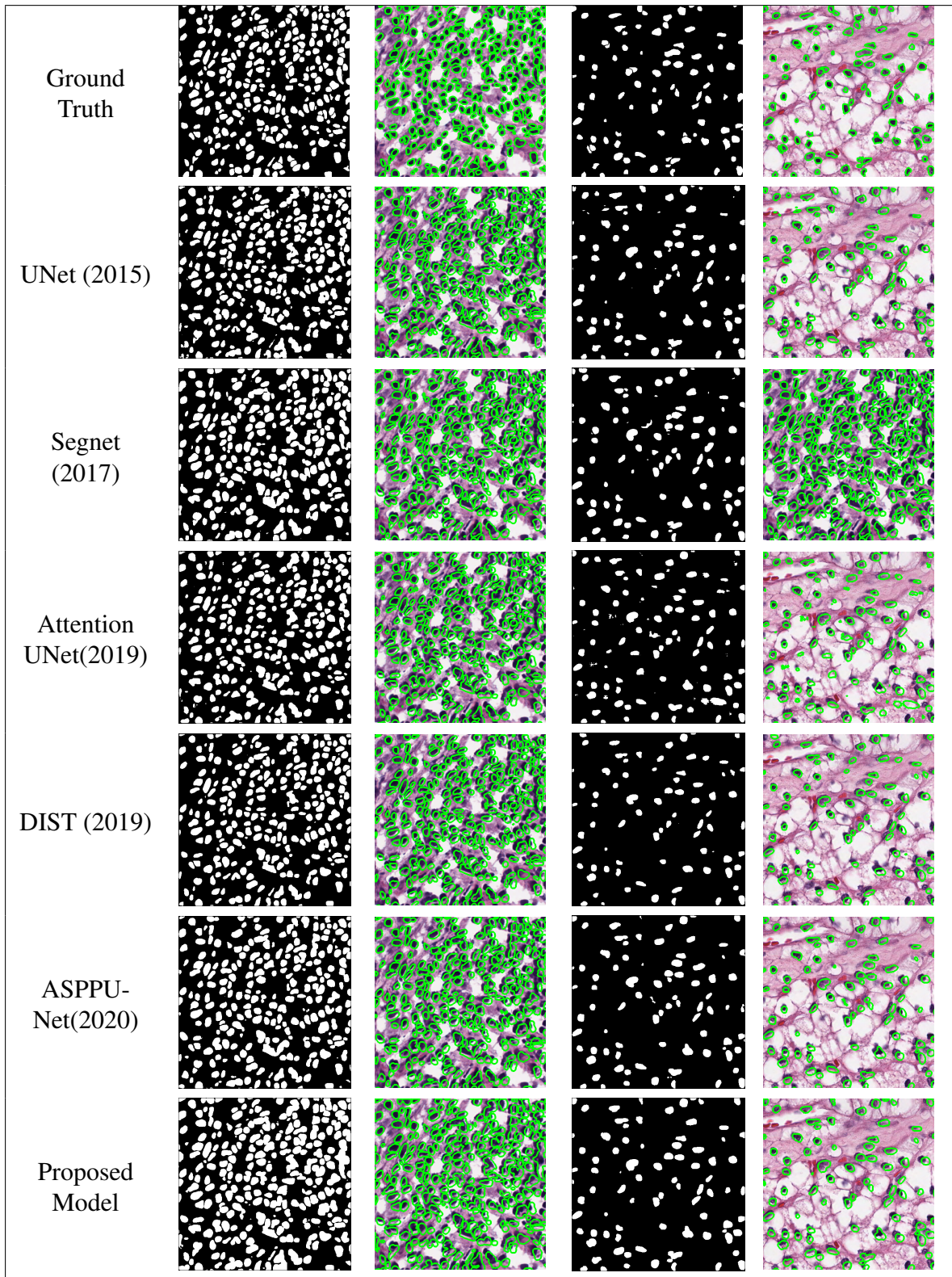
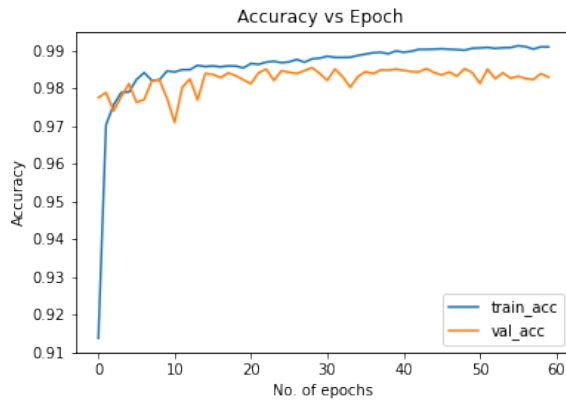
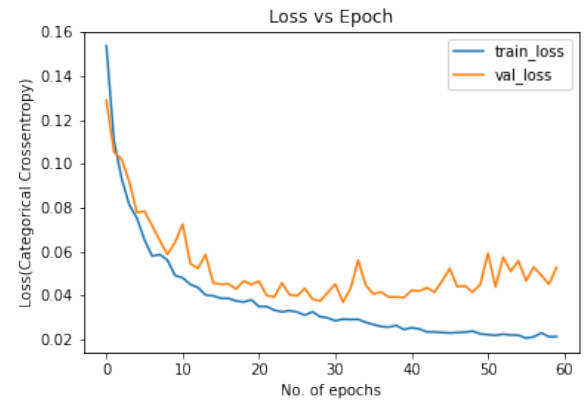


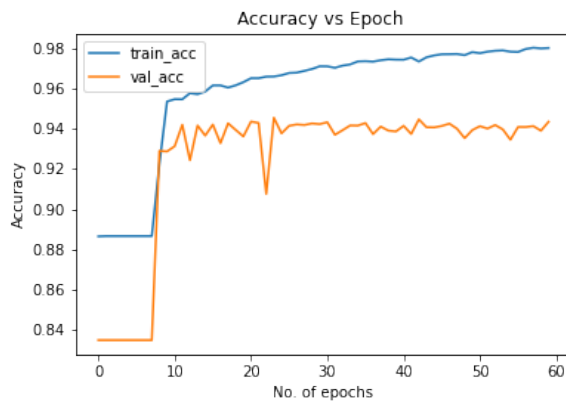
Figure 3.9: Visual segmentation results comparison of different models on MoNuSeg dataset



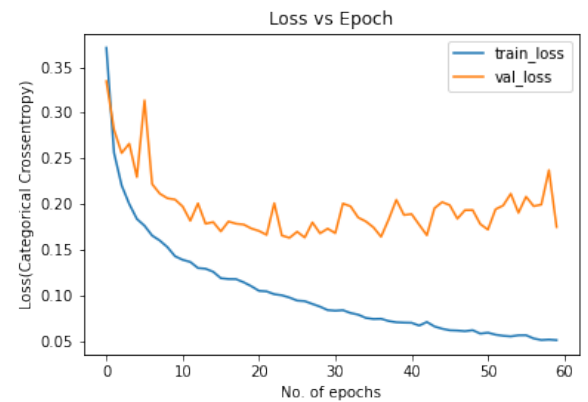
(a)



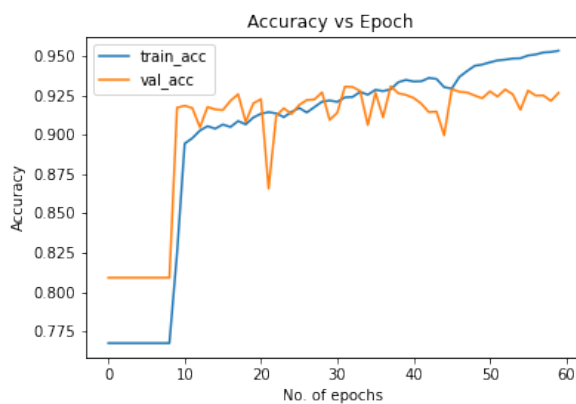
(b)



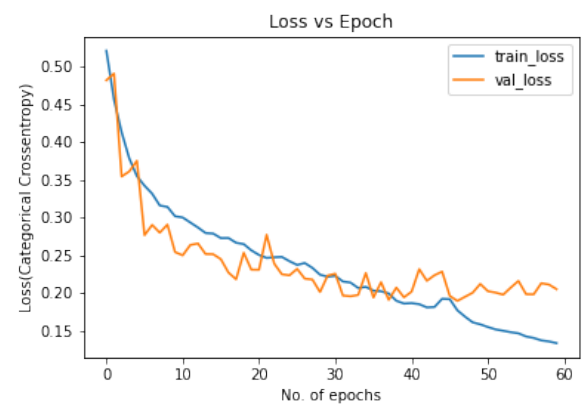
(c)



(d)



(e)



(f)

Figure 3.10: (a), (c), (e)-Training and validation accuracy for Kidney, TNBC, and MoNuSeg datasets. (b), (d), (f) Training and validation loss for Kidney, TNBC, and MoNuSeg datasets.

Table 3.5: Performance comparison of architectures on TNBC dataset

Model	F1 Score	AJI Score	Parameters	FLOPs
U-Net(2015)	0.7324	0.6559	31,378,945	62,746,139
Segnet(2017)	0.7685	0.6434	18,826,753	47,065,663
Attention UNet (2019)	0.7216	0.6194	31,902,629	63,791,573
DIST(2019)	0.7516	0.6727	7,771,873	15,525,181
ASPPU-Net(2020)	0.7781	0.6378	4,307,713	8,610,587
Proposed SCPP-Net	0.8168	0.6998	5,088,955	10,162,445

Table 3.6: Performance comparison of architectures on MoNuSeg dataset

Model	Avg. F1 Score of 7 Different Organs	Avg. AJI Score of 7 Different Organs	Parameters	FLOPs
U-Net(2015)	0.7681	0.6295	31,378,945	62,746,139
Segnet(2017)	0.7846	0.6488	18,826,753	47,065,663
Attention UNet (2019)	0.7538	0.6098	31,902,629	63,791,573
DIST(2019)	0.7831	0.6480	7,771,873	15,525,181
ASPPU-Net(2020)	0.7839	0.6483	4,307,713	8,610,587
Proposed SCPP-Net	0.8010	0.6710	5,088,955	10,162,445

ney histopathology dataset. Performance measurement in terms of F1 Score, AJI, the total number of trainable parameters, and FLOPs described the training time and computational complexity. Table 3.5 shows a quantitative segmentation results comparison of the proposed SCPP-Net architecture with five other existing deep-learning segmentation architectures on the TNBC histopathology dataset. Table 3.6 shows a quantitative segmentation results comparison of the proposed SCPP-Net architecture with five other existing deep-learning segmentation architectures on the multiple organ multi-disease histopathology dataset. The results indicated that the proposed SCPP-Net architecture was able to retrieve more information compared to other existing deep learning segmentation architectures. Training and validation accuracy, training and validation loss curve on three datasets are shown in Fig.3.10. These plots indicated that the proposed model better learns the parameters for the

prediction on each of mentioned datasets. Visual nuclei segmentation results comparison of proposed and different existing deep learning segmentation architectures on kidney dataset, TNBC dataset, and multiple organs multi-disease dataset are shown with two sample images and their overlay images in Fig. 3.7, Fig. 3.8, and Fig. 3.9 respectively. Number of clustered predictions was less in the case of SegNet as compared to U-Net and Attention U-Net. False detection was very low in the case of SegNet. The number of partially detected nuclei was maximum in the case of Attention U-Net compared to any other model. Partially detected nuclei in Dist were fewer compared to SegNet, U-Net, and Attention U-Net. Prediction accuracy in terms of partially detected nuclei and clearly identified nuclei slightly improved in ASPPU-Net compared to U-net and SegNet.

The problem of separating overlapped nuclei was better handled by Dist. In the proposed model, there were very few false-detected nuclei and partially detected nuclei. The problem of separating clustered nuclei was not completely solved. Visual results indicated that the proposed SCPP model yielded very good results in terms of the number of clearly identified nuclei, the number of absent nuclei, partially detected nuclei, the number of overlapped nuclei, and false detection.

3.4.3 Computational complexity analysis

The computational complexity of the proposed SCPP-Net architecture and existing deep learning segmentation methods (i.e., U-Net (Ronneberger *et al.*, 2015), SegNet(Badrinarayanan *et al.*, 2017), Attention UNet(Schlemper *et al.*, 2019), Dist(Naylor *et al.*, 2019), and ASPPU-Net(Wan *et al.*, 2020)) were analyzed in terms of the number of trainable parameters and FLOPs. The value of FLOPs describes the computing power of given hardware, such as the GPU; the smaller the value, the faster the computing ability. The trainable parameters and FLOPs of the proposed SCPP-Net architecture and existing deep learning segmentation methods are presented in the fourth and fifth columns of Tables 3.4, 3.5, and 3.6.

From the tables, it is clear that the proposed SCPP-Net required fewer trainable parameters and FLOPs compared to the most recent deep learning segmentation methods because the use of the ASPP with separable convolution. Training time was proportional to the number of trainable parameters and the size of the training dataset. In this process, the average training time of the proposed model was 27 s, 17 s, and 14 s per epoch for the kidney, TNBC, and multi-organ datasets, respectively. It had been stopped the training process when validation accuracy was not improved in 10 successive iterations. Based on the improvement in validation accuracy, training the model an average of 45 to 65 epochs was possible. A

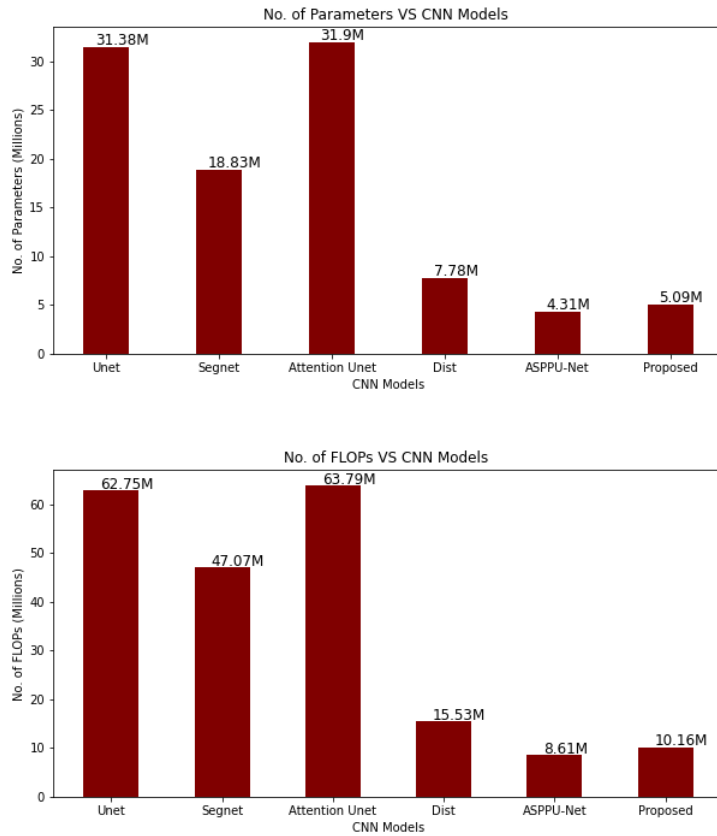


Figure 3.11: Computational complexity comparison of different CNN models

graphical illustration of the computational complexity comparison of nuclei segmentation results of the different CNN models (Model-1 (UNet (2015)), Model-2 (SegNet (2017)), Model-3 (Attention U-Net (2019)), Model-4 (Dist (2019)), Model-5 (ASPPU-Net(2016)) and the proposed SSCP-Net model) is shown in Fig. 3.11.

3.5 Proposed Architecture-2

A segmentation framework called high resolution deep transferred ASPPU-Net for nuclei segmentation of histopathology images has been proposed. In deep learning framework, segmentation strategy involves encoder path where model learns the ‘what’ information or context in the image. In decoder path, the model learns the ‘where’ information or location in the image. In proposed architecture 2 encoder path consists of repeated application of

standard convolution layer and an additional parallel path to the main network through a residual connection to minimize the loss in pooling operation. Each stage of the downsampling path has a (2x2) max-pooling layer which effectively reduces the spatial size of the image. Similar level features are concatenated with upsampling layers. In the up-sampling path, the features having different shapes and size has cropped and merged into the next layer. The proposed wide network used a powerful decoder by aggregating the similar level spatial features and collecting maximum possible residual information. At the final stage, a (1x1) convolution is used to map the size (512x512x16) to (512x512x1).

ASPPUNet (Wan *et al.*, 2020) shown in Fig.3.12 proved to be the best-performing model because it performs better on the datasets used in this study and other multi-class datasets. Therefore we chose that to be our base model and made changes to improve the performance further. To improve the performance of the network the proposed model introduce an additional information retrieval module to extract a more relevant feature used by (Wan *et al.*, 2020) called Atrous spatial pyramid pooling (ASPP) bottleneck path where a CNN uses multiple dilation rate. Dilation rate is an additional parameter that differed from the resultant feature map to visualize a larger area. Different dilation rates are applied to the same layer and results are then concatenated, which allows the model to take advantage of multi scale feature extraction. This yields very good results because hierarchical information with varying sizes can be identified in the same layer. The detailed diagram of the proposed model is shown in Fig.3.13.

3.5.1 High resolution layer

Proposed architecture addressed the problem of degradation of information in deeper network by introducing deep and wide residual network which is easier to train and optimize. The residual connection is realized by creating an additional path parallel to the main encoder-decoder path of the network to restore the flow of information in deep network. Instead of deep and thin encoder and the decoder path the proposed network has wide and deep path that effectively leverage the strength of residual learning as well as encoder decoder architecture. Proposed architecture has a high resolution encoder in Fig. 3.14(a), ASPP bottleneck path in Fig. 3.14(b) for multilevel feature extraction and an effective decoder shown in Fig. 3.14(c). In the encoder-decoder of the model, two paths are not substantially dependent on one another, and their smooth co-relation with valid paths improves the network's performance and it reduces the vanishing gradient problem.

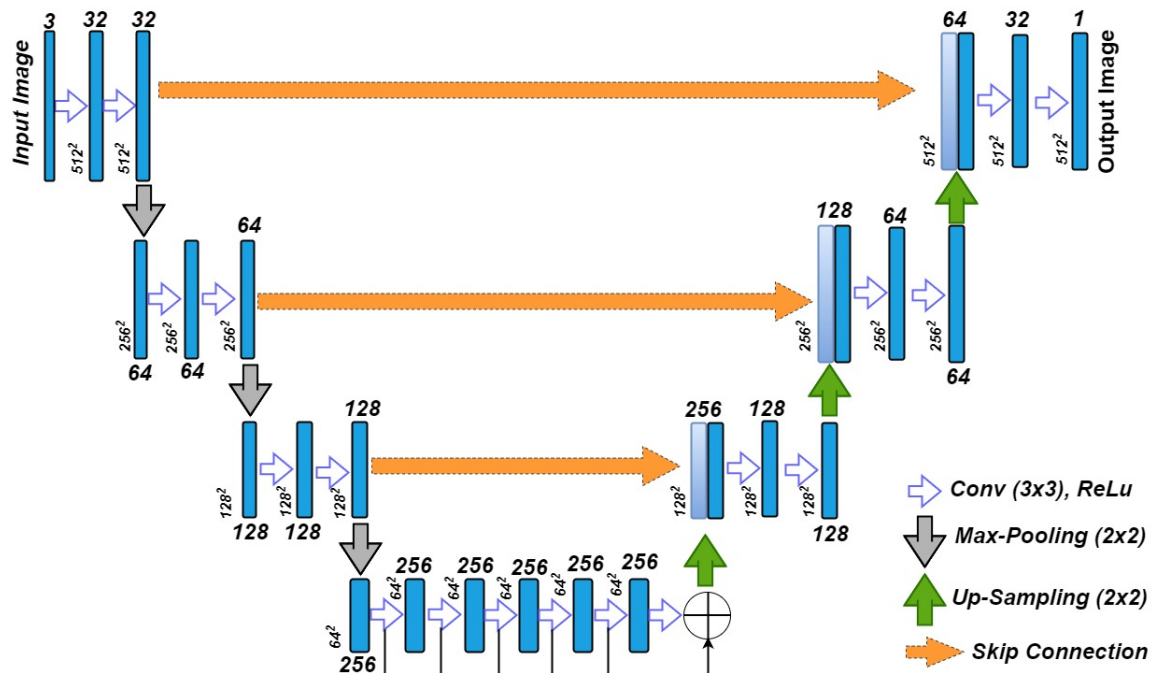


Figure 3.12: Architecture of the ASPPU model

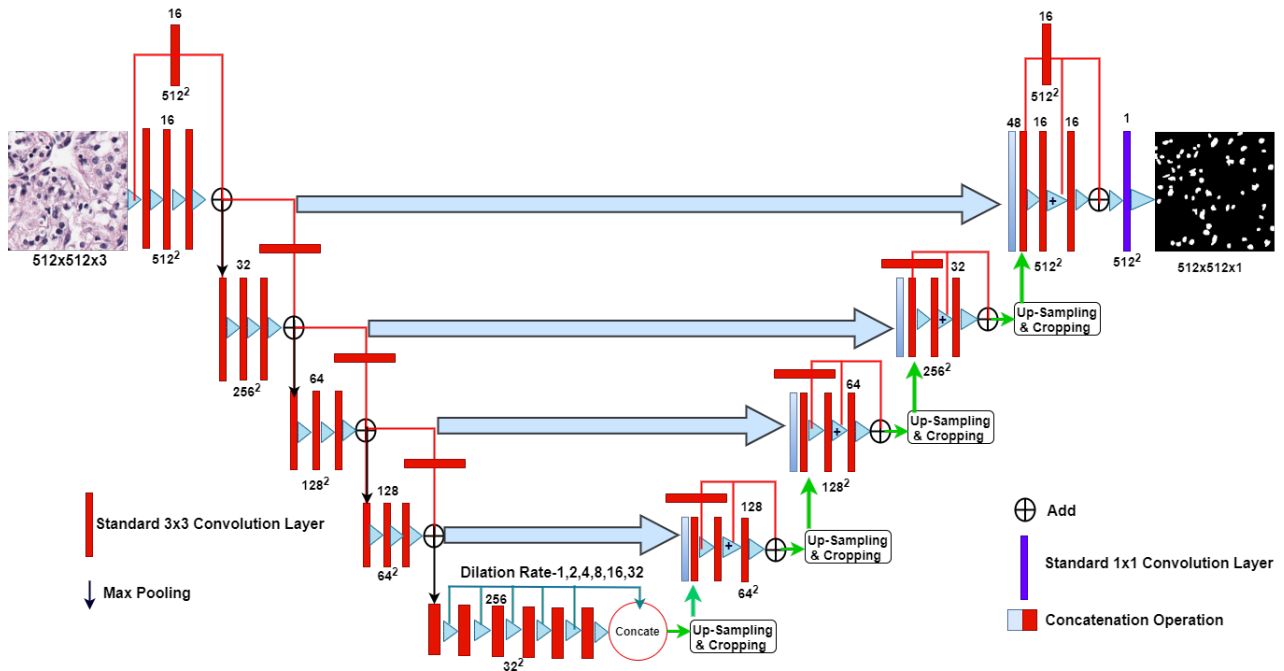


Figure 3.13: Proposed high resolution deep transferred ASPPU-Net

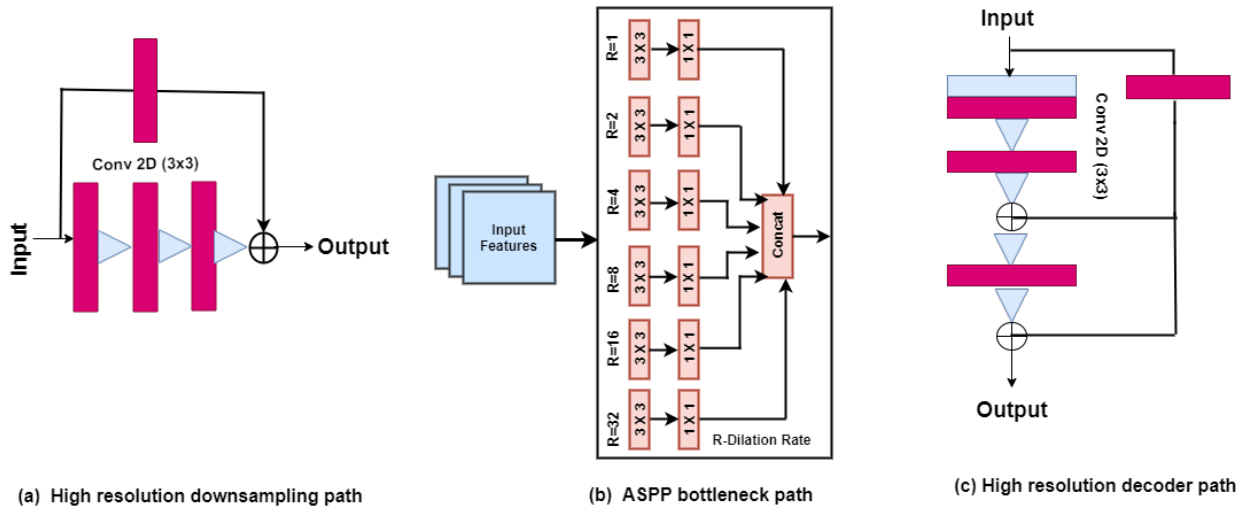


Figure 3.14: High resolution encoder path, ASPP bottleneck path and decoder path

3.6 Implementation and Training

This experiment is implemented on the latest version of python with Keras and TensorFlow framework. Google GPU and Colab notebook were the resources for conducting the experiments. Adam (Kingma and Ba, 2014), is the optimization method and binary cross-entropy used in Naylor *et al.* (2019) is the loss function used in this study. The reported results of all deep learning models are the averages of three trials conducted independently. Final result is calculated based on an average of three trials by initializing random weights in each trial. The final quality metrics were measured as the average quality metrics of all images in the test dataset.

3.7 Results and Discussions

The proposed model is evaluated on three publicly available H&E-stained histopathology image datasets and the experimental result is compared with five benchmark segmentation models. It has been used the image quality metrics F1 score used by (Naylor *et al.*, 2019), (Lal *et al.*, 2021) and Aggregated Jaccard Index (AJI) scores used by Naylor *et al.* (2019) to quantify the effectiveness of the proposed high resolution deep transferred ASPPUNet architecture and other existing deep learning segmentation methods. The number of parameters and FLOPs are discussed in Section

3.7.1 Ablation study

The proposed model is composed of three main modules at different resolutions of the network. (I) An encoder-decoder path which consists of a regular convolution layer and max-pooling layer (II) Atrous spatial pyramid pooling (ASPP) block at bottleneck (III) Residual path. To demonstrate the effectiveness of the proposed model, it has been measured the effectiveness of the individual module as well as some combination of its by detaching its main module from the proposed model. Different versions of the network were trained in the same environment and performance was compared with the proposed model where all the important modules jointly worked. The F1-score, and IoU value of different versions are listed in Table 3.7, Table 3.8, and Table 3.9 for kidney, TNBC, and MoNuSeg dataset respectively.

Table 3.7: Comparison of different variants of proposed model on Kidney dataset

Model Variations	Residual path	ASPP	Regular Conv.	F1 Score	AJI
Variation-1	✓	×	✓	0.9515	0.9079
Variation-2	×	✓	✓	0.9399	0.8873
Variation-3	×	×	✓	0.9369	0.8826
Proposed Model	✓	✓	✓	0.9684	0.9394

Table 3.8: Comparison of different variants of proposed model on TNBC dataset

Model Variations	Residual path	ASPP	Regular Conv.	F1 Score	AJI
Variation-1	✓	×	✓	0.8234	0.7008
Variation-2	×	✓	✓	0.7991	0.6675
Variation-3	×	×	✓	0.7719	0.6313
Proposed Model	✓	✓	✓	0.8419	0.7282

Table 3.9: Comparison of different variants of proposed model on MoNuSeg dataset

Model Variations	Residual path	ASPP	Regular Conv.	F1 Score	AJI
Variation-1	✓	×	✓	0.7999	0.6721
Variation-2	×	✓	✓	0.7871	0.6536
Variation-3	×	×	✓	0.7780	0.6425
Proposed Model	✓	✓	✓	0.8344	0.7169

3.7.2 Comparison with state-of-the-art segmentation methods

Table 3.11 shows a comparison of the proposed architecture with five other segmentation architectures on three datasets. Performance measurement in terms of F1 Score, AJI score, the total number of trainable parameters that describe the training time and complexity and floating point operations per second (FLOPs). Visual segmentation comparisons of predicted images of proposed model and reference models are shown in Fig.3.15. The visual segmentation results show that all the deep learning models separate the nuclear region to a certain extent. Some common region in all predicted image has been picked to demonstrate that at what extent individual model performs. The red box and green box show how many nuclei are correctly identified, how many are partially identified, how many false detections in each model are present, and how many nuclei are not detected.

Visual results are the comparison of predicted images of five state-of-the-art models with the proposed model of two images of the mentioned datasets. In row-1 of Fig.3.15, corresponding ground truth image has 57 annotated nuclei. U-Net (Ronneberger *et al.*, 2015), detected 53 nuclei clearly, two nuclei are not detected and two are in clustered form. Four additional ducts are detected in U-Net which are not desirable and not present in the ground truth. SegNet (Badrinarayanan *et al.*, 2017), also not detected any additional things, but out of 57 nuclei only 47 nuclei detected accurately, three are detected partially and seven nuclei are not detected. The best part of SegNet architecture is that it separates the overlapped nuclei. Attention U-Net (Schlemper *et al.*, 2019), is an extended version of the original U-Net that detected 49 nuclei clearly, two nuclei are in overlapped form, five nuclei are partially detected, one nucleus not detected and five additional things found that are not required. Dist (Naylor *et al.*, 2019), architecture detected 49 nuclei almost similar to

Table 3.11: Experimental performance comparison of different architectures with three datasets

Model	Kidney		TNBC		MoNuSeg	
	F1	AJI	F1	AJI	F1	AJI
U Net(2015)	0.8537	0.7489	0.7324	0.6559	0.7248	0.6029
Segnet(2017)	0.8972	0.8304	0.7685	0.6434	0.7879	0.6514
Attention UNet(2019)	0.9135	0.8590	0.7216	0.6194	0.7735	0.6315
DIST(2019)	0.8992	0.8272	0.7516	0.6727	0.7795	0.6467
ASPPU-Net(2020)	0.9052	0.8293	0.7781	0.6378	0.8087	0.6803
Proposed Model	0.9684	0.9394	0.8419	0.7282	0.8344	0.7169

ground truth, three nuclei not able to detect, three partially detected, and two are in clustered

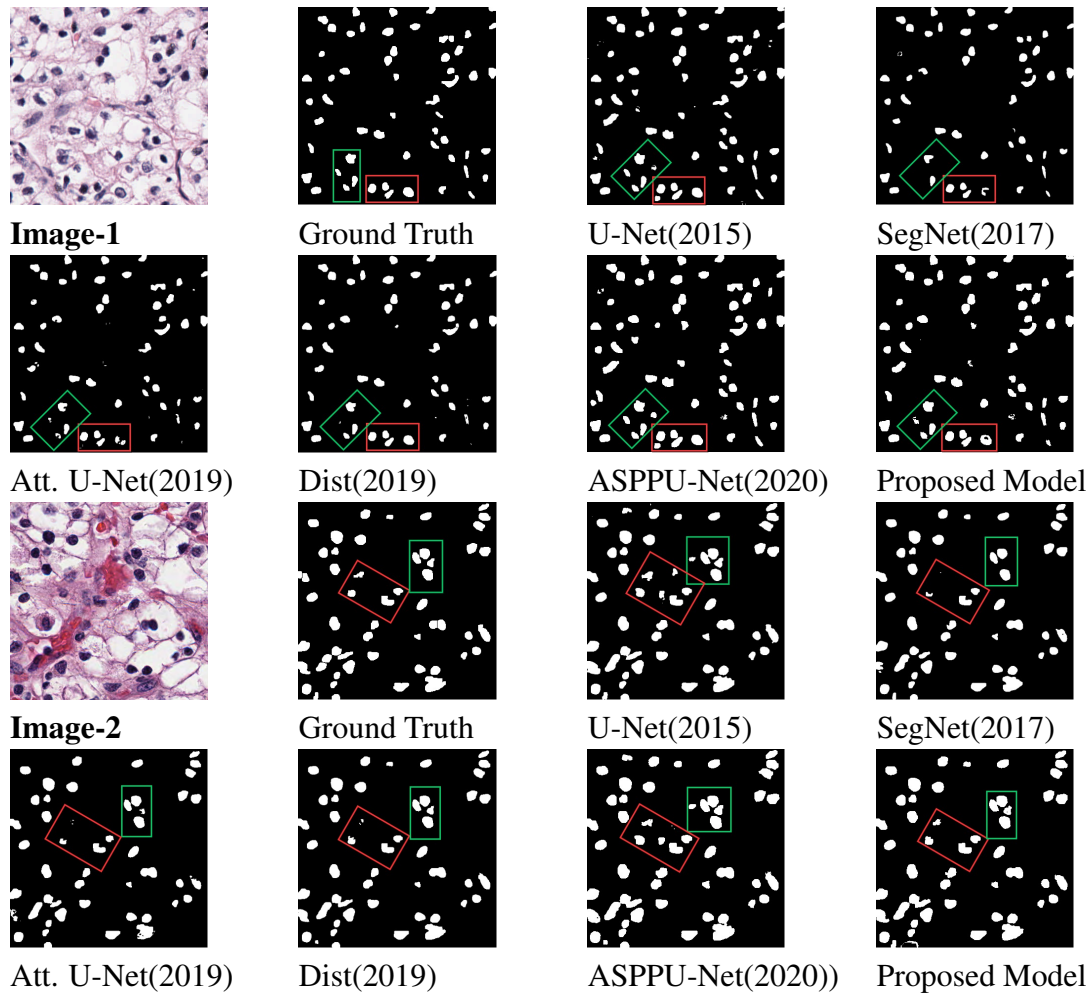
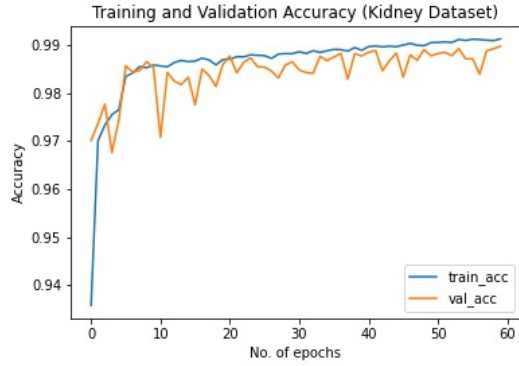
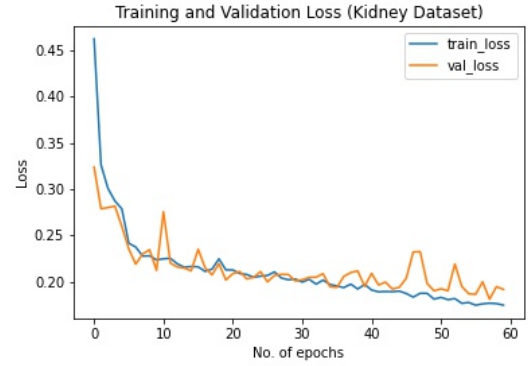


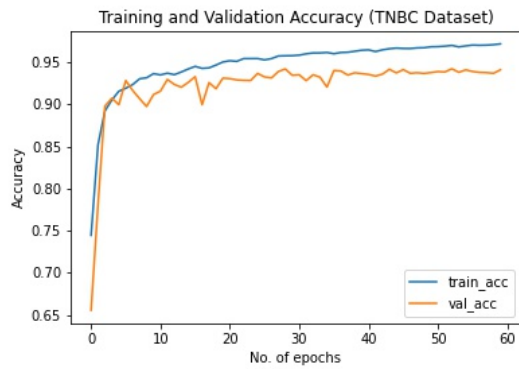
Figure 3.15: Comparison of visual segmentation of different models of two sample images



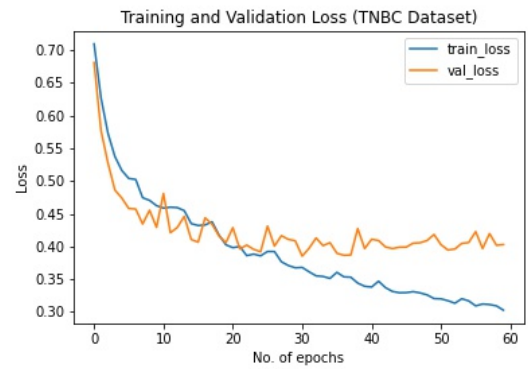
(a)



(b)



(c)



(d)



(e)



(f)

Figure 3.16: (a), (c), (e)-Training and validation accuracy for Kidney, TNBC, and MoNuSeg datasets. (b), (d), (f) Training and validation loss for Kidney, TNBC, and MoNuSeg datasets.

Table 3.12: Network computational complexity in millions

Model	Parameters(millions)	FLOPs(millions)
UNet	31.3	62.7
SegNet	18.8	47
Attention UNet	31.9	63.7
Dist	7.7	15.5
ASPPU-Net	4.3	8.6
Proposed	8.7	21.6

form. ASPPU-Net (Wan *et al.*, 2020) identifies 7 additional shapes, 2 nuclei are in clustered form out of 57 nuclei. The best part of the proposed model is out of 57 nuclei, 55 nuclei are clearly identified, no undesirable things are detected, the morphology of the detected nuclei are exactly similar to the ground truth image and it also reflects from performance metrics shown in tabular form. Results indicated that the proposed architecture is able to retrieve more information compared to others. Training and validation accuracy, training and validation loss curve on three datasets are shown in Fig.3.16. These plots indicated that the proposed model better learns the parameters for the prediction on each of mentioned datasets.

3.7.3 Computational complexity analysis

Computational complexity is expressed in terms of the total trainable parameters and floating-point operations (FLOPs) shown in Table 3.12. The proposed model uses 8.7 million parameters which is the least among all reference models except ASPPU-Net and Dist architecture. Dist architecture has marginal difference in trainable parameter compared to proposed model and it has 7.7 million parameters. ASPPU-Net has least number of FLOPs. Attention UNet has highest number of trainable parameters and FLOPs.

3.8 Proposed Architecture-3 and a Novel Loss Function

A deep structured residual encoder-decoder network (DSREDN) and a novel loss function for nuclei segmentation of multiple organ histopathological images has been proposed. For the segmentation of microscopy images, an encoder-decoder architecture is best suited due to the fact that if an encoder has regular convolution layers, max-pooling layers, then it

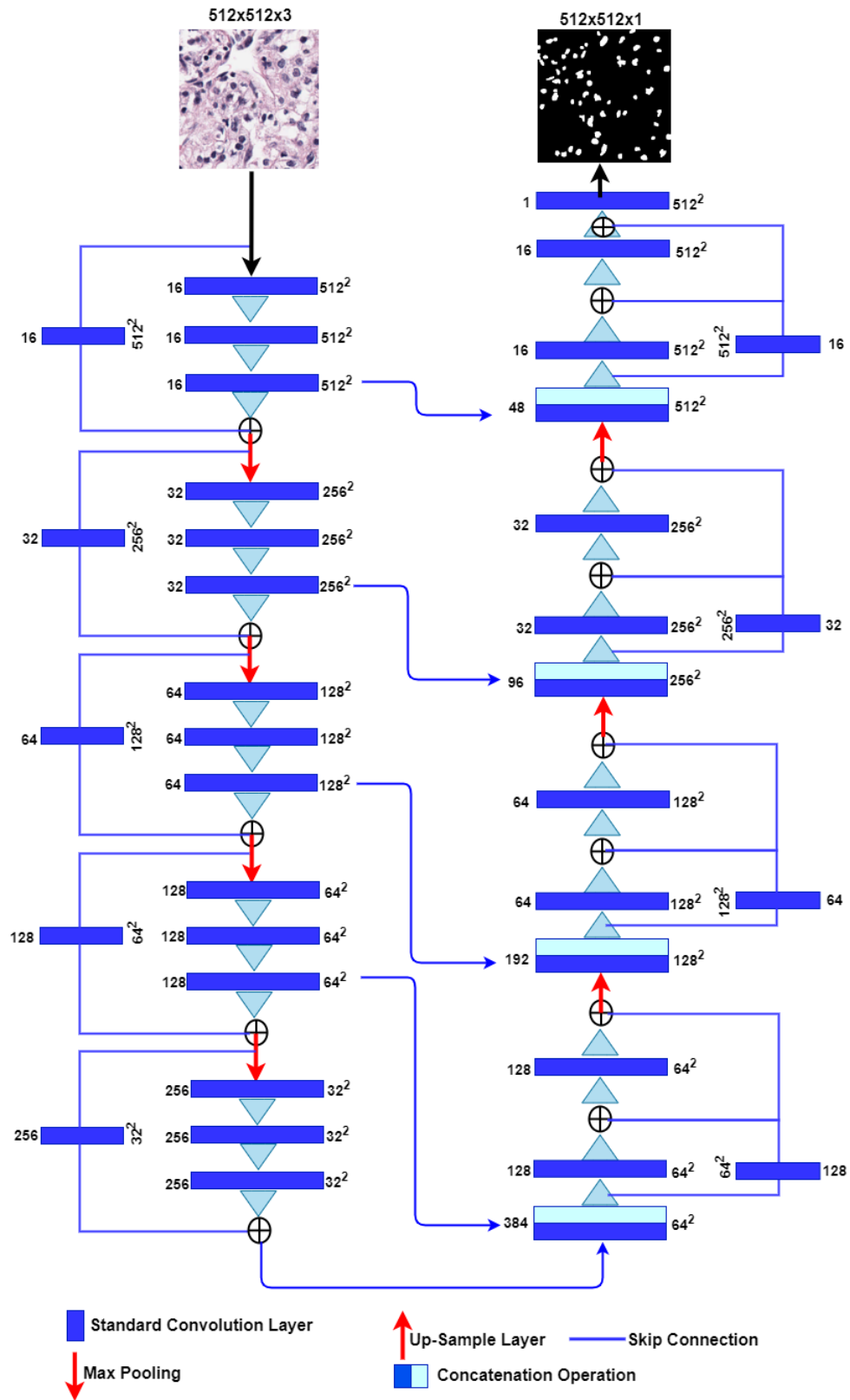


Figure 3.17: Proposed deep structured residual encoder-decoder network

captures the context in the image very effectively. Decoder path presents the output by gradually applying up-sampling, collecting relevant features from the encoder, and enable precise localization. For each of the filters in the encoder side of the DSREDN network shown in Fig. 3.17, accepts input of flexible size. A regular (3 x 3) 2D standard convolution, batch normalization, and max-pooling layer are applied. To avoid the saturation problems and loss of information while going deeper into the network, the proposed model restored the lower layered information by creating an additional path parallel to the main path of the network. These two paths are not strongly correlated to each other and it avoids vanishing gradient problems.

For each of the filter sizes, the entire encoder side of the DSREDN network consists of three convolution layers in parallel with a single convoluted path that focused to flow the more contextual feature in the network. As the collection of contextual data from the encoder side determines how well the decoder path performs in producing the final output, the network uses a slightly different path on the decoder side to ensure that the features are processed as effectively as possible. By this procedure, the DSREDN network becomes wide and deep instead of thin and deep. DSREDN network trained with RGB images of size (512 x 512 x 3). Feature maps from different parts of the network have been taken, and to scale into similar dimensions Up-Sampled to the same height and width. (e) At the final stage a (1 x 1) convolution is used to map the size (512 x 512 x 16) to (512 x 512 x 1) with sigmoid activation.

3.8.1 Proposed loss function

The proposed loss function has been conceived from the idea of focal loss. It is also called as dynamically scaled cross-entropy loss. Focal loss is a pixel-based loss that automatically assigns more weight to the portion of our interest or object portion and down-weight to those which are of less interest with the help of hyper parameter α_t and γ coefficient.

Focal Loss: For highly imbalanced datasets dynamically scaled cross-entropy works better. From BCE loss, Focal Loss (Jadon, 2020), (Sugino *et al.*, 2021) can be expressed in equations (3.6), (3.7), (3.8), (3.9), and (3.10).

$$BCE = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (3.6)$$

$$BCE Loss = \begin{cases} -\log(\hat{y}) & \text{for } y = 1 \\ -\log(1 - \hat{y}) & \text{otherwise} \end{cases} \quad (3.7)$$

To make the notation more convenient it can be written as:

$$p_t = \begin{cases} \hat{y} & \text{for } y = 1 \\ 1 - \hat{y} & \text{otherwise} \end{cases} \quad (3.8)$$

Now BCE can be written in a very precise manner as follows:

$$BCE(y, \hat{y}) = BCE(p_t) = -\log(p_t) \quad (3.9)$$

A modulation factor $(1 - p_t)^\gamma$, $\gamma > 0$ to down-weight the background and assign more weight to object region and a hyper-parameter α_t , $0 < \alpha < 1$ make the above loss function dynamically scaled cross-entropy loss. This is also known as a focal loss described in (Jadon, 2020).

$$Focal Loss (p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3.10)$$

Dice Loss: This loss (Sudre *et al.*, 2017) is a widely used metric in computer vision community to calculate the similarity between two images. Equation (3.11) represents Dice Loss. Here, 1 is added in the numerator and denominator to ensure that the function is not undefined in edge-case scenarios.

$$DL(y, \hat{y}) = 1 - \frac{2y\hat{y} + 1}{y + \hat{y} + 1} \quad (3.11)$$

Tversky Loss: This loss (Salehi *et al.*, 2017) can also be seen as a generalization of the Dice Loss. It adds weight to FP (false positives) and FN (false negatives) with the help of the β coefficient. Here, when $\beta = 0.5$, It can be solved into regular Dice coefficient. Similar to Dice Loss, Tversky loss can also be defined in equation (3.12).

$$Tversky Loss = \frac{\hat{y} \cdot y}{\hat{y} \cdot y + \beta(1 - y)\hat{y} + (1 - \beta)y(1 - \hat{y})} \quad (3.12)$$

Approximation of Dice Loss: This loss derived from dice coefficient by setting different weight to false positive and false negative with the help of β coefficient. Mathematically

this loss can be expressed in equation (3.13).

$$\text{Approximated Dice Loss} = (1 - DI) \text{ where } DI = \frac{\sum \hat{y} \cdot y}{\sum \hat{y} \cdot y + \beta(1 - y)\hat{y} + (1 - \beta)y(1 - \hat{y})} \quad (3.13)$$

Proposed loss function calculated based on the concept of dynamically scaled cross entropy loss in equation (3.10). Predicted results and learning curves show that using proposed loss in equation (3.14), training is better with three different histopathology datasets.

$$\text{Dynamically Scaled Dice Loss} = \alpha_t(1 - DI)^\gamma \quad (3.14)$$

DI indicates Dice index and for our purpose, we have used $\alpha_t = 0.5$ and $\gamma = 2$ for the best possible results.

3.9 Implementation and Training

Google Colab notebook (with 12 GB Google GPU), Tensorflow 2.0, and Keras API for simulation of all models have been used. The most important part of any optimization method in deep learning is about gradient which is used to update the weights. equation (3.15) is a general weight update equation.

$$w_t = w_{t-1} - \eta \left[\frac{\delta L}{\delta w} \right] \quad (3.15)$$

Adam (Adaptive Moment Estimation) optimizer is used to get the best weight for the model. SGD with Momentum and RMSProp are mostly preferred algorithms. Adam (Kingma and Ba, 2014) takes advantage its two predisessor and used in this experiment. Followings are the significant controlling parameters that we have fine-tuned: (a) Learning rate= 0.001 for training of all the model with three datasets (b) Weight decay constant, selected learning rate and weight decay constant based on trial and error (c) Batch size=4, Larger batch size allow us to parallelize computations to a greater degree but lead to poor generalization (d) The size of the convolution filter used as per available resources. F1-score used by (Lal *et al.*, 2020), (Aatresh *et al.*, 2021b), Aggregated Jaccard Index (AJI) used by (Naylor *et al.*, 2019) has been used to measure the performance. To evaluate the results on three different datasets, we have used the loss function described in equation (3.14).

3.10 Results and Discussions

The proposed model has been trained using a new loss function and evaluated on three publicly available H&E-stained histopathology image datasets. The impact on network training using the new loss function has been separately studied in Section 3.10.1. and the experimental result is compared with five benchmark segmentation models. It has been used the image quality metrics F1 score used by (Naylor *et al.*, 2019), (Lal *et al.*, 2021) and Aggregated Jaccard Index (AJI) scores used by Naylor *et al.* (2019) to quantify the effectiveness of the proposed DSREDN architecture and other existing deep learning segmentation methods. The number of parameters and FLOPs are discussed in Section 3.10.3.

3.10.1 Ablation study

A comprehensive experiment on different loss functions has been performed. Binary cross-entropy (BCE), Dice loss, Weighted binary cross-entropy (WBCE), Focal loss, Tvesky loss, combined loss of Binary cross-entropy, and Dice loss has experimented on three publicly available H&E-stained histopathology image datasets. These different loss function were trained in the same environment. We have evaluated the reference models and the proposed model based on the proposed loss function described in equation (3.14). Superiority of proposed loss function with F1-score/IoU score of two benchmark model is shown in Table 3.13. The proposed loss function appears to better describe the validation data, according

Table 3.13: F1/AJI metric performance comparison of Unet Model and Proposed model on different datasets

Loss Functions	(Kidney) UNet	(TNBC) UNet	(MoNuSeg) UNet	(Kidney) Proposed	(TNBC) Proposed	(MoNuSeg) Proposed
BCE	0.8537/0.7489	0.7324/0.6559	0.7278/0.6029	0.9422/0.8994	0.8008/0.6698	0.7854/0.6675
Dice	0.8699/0.7631	0.7470/0.6690	0.7186/0.5991	0.9479/0.9044	0.8100/0.6828	0.7927/0.6607
WBCE	0.8408/0.7376	0.7360/0.6585	0.7168/0.5938	0.9460/0.9115	0.7930/0.6588	0.7919/0.6597
Focal	0.8622/0.7563	0.7228/0.6467	0.7399/0.6088	0.9346/0.8990	0.8175/0.6928	0.7940/0.6619
Tvesky	0.8793/0.7713	0.7573/0.6782	0.7423/0.6149	0.9286/0.9040	0.8346/0.7182	0.7810/0.6513
BCE+Dice	0.8479/0.7329	0.7250/0.6493	0.7321/0.6065	0.9472/0.9010	0.8087/0.6796	0.7950/0.6637
Proposed Loss	0.8878/0.7732	0.7653/0.6854	0.7460/0.6179	0.9579/0.9261	0.8517/0.7415	0.8065/0.6795

to the training and validation curve illustrated in Fig. 3.18, Fig. 3.19, and Fig. 3.20. The training and validation curves should be closer together for the model’s best representation, and they should be collapsed to one another for the best bias and variance. The model is robust during testing since both curves are closer to one another. The large gap between the training and validation curve indicates, the model is operating very well in training data, but

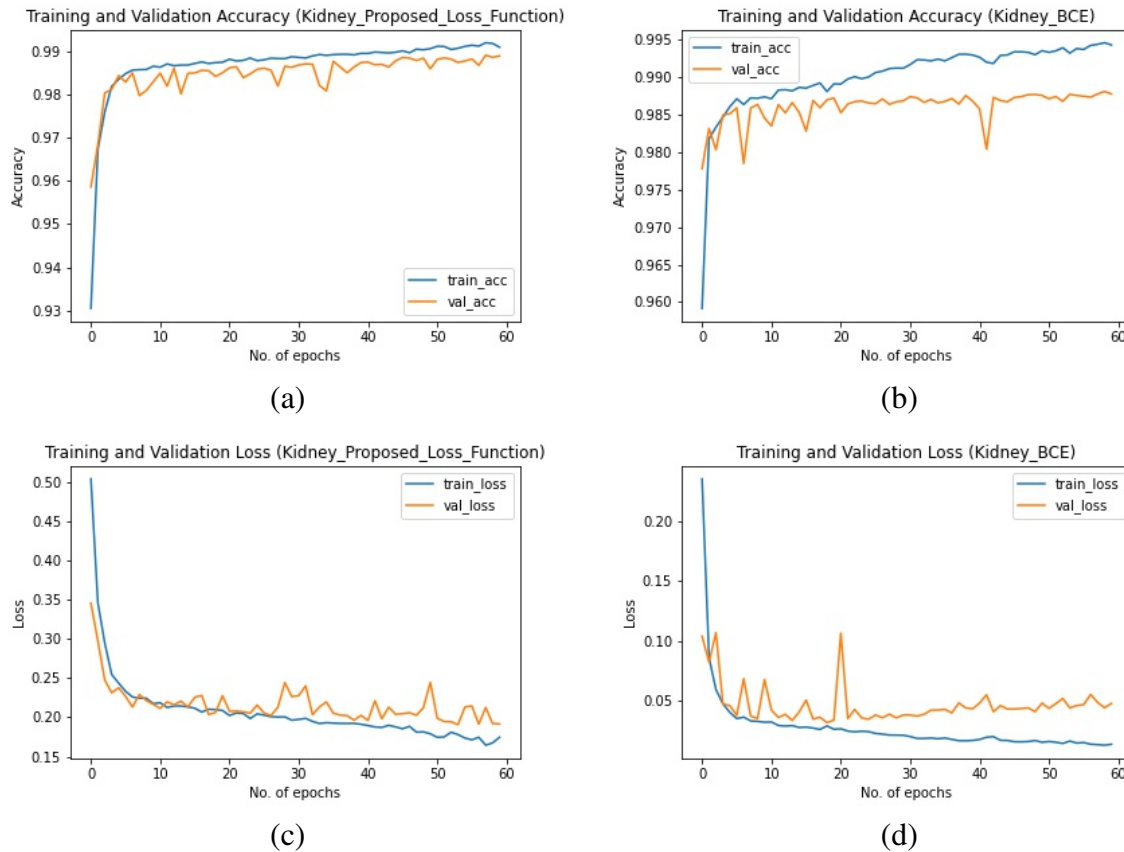
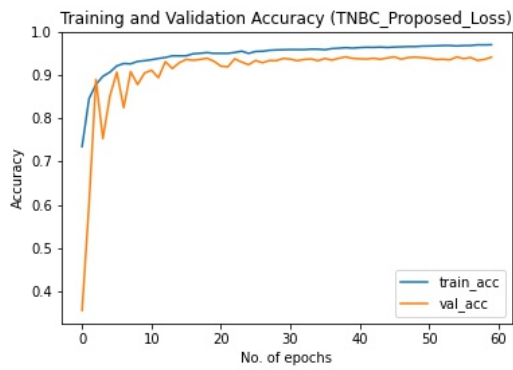


Figure 3.18: Accuracy and loss plot (Kidney Dataset), (a), (c) - Learning curve using proposed loss, (b), (d) - Learning curve using BCE loss

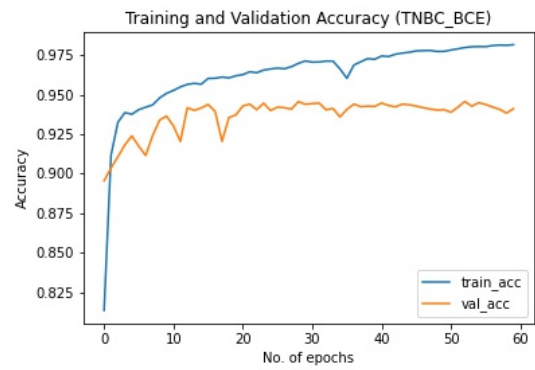
not as well as it does in validation data. The model has been generalized and expanded to operate with many forms of data using the proposed loss function. This ablation study helps in finding the best loss function for multi-organ histopathology datasets and it's imperative to include it in the proposed architecture.

3.10.2 Comparison with state-of-the-art segmentation methods

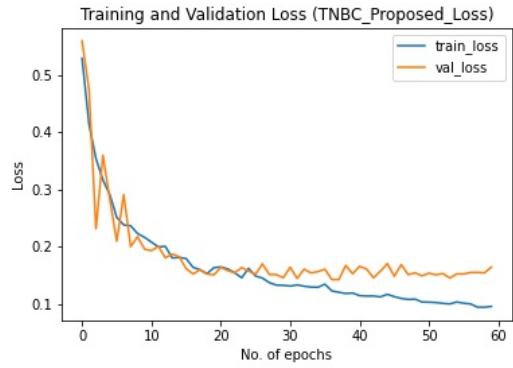
The proposed model has been evaluated against five different CNN models, which serve as a benchmark in the segmentation of biomedical images. We have evaluated the reference models and the proposed model based on the proposed loss function described in equation (3.14). Three histopathology datasets are used to evaluate the DSREDN model to five benchmark models in Table 3.14. Visual results shown in Fig.3.21 are the comparison of



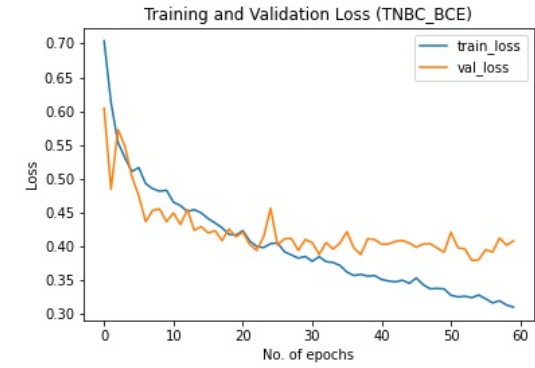
(a)



(b)

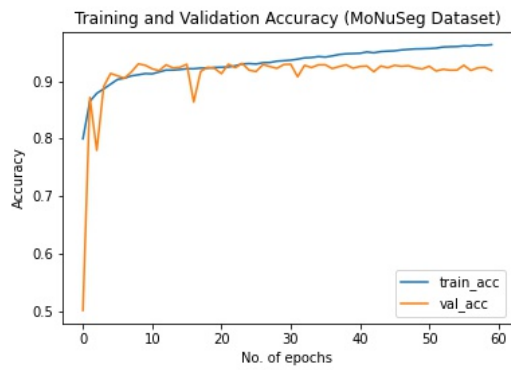


(c)

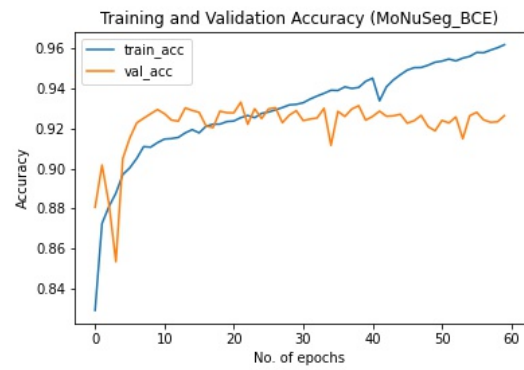


(d)

Figure 3.19: Accuracy and loss plot (TNBC Dataset), (a), (c) - Learning curve using proposed loss, (b), (d) - Learning curve using BCE loss



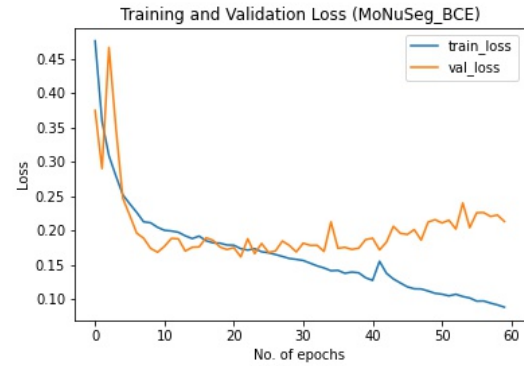
(a)



(b)



(c)



(d)

Figure 3.20: Accuracy and loss plot (MoNuSeg Dataset), (a), (c) - Learning curve using proposed loss, (b), (d) - Learning curve using BCE loss

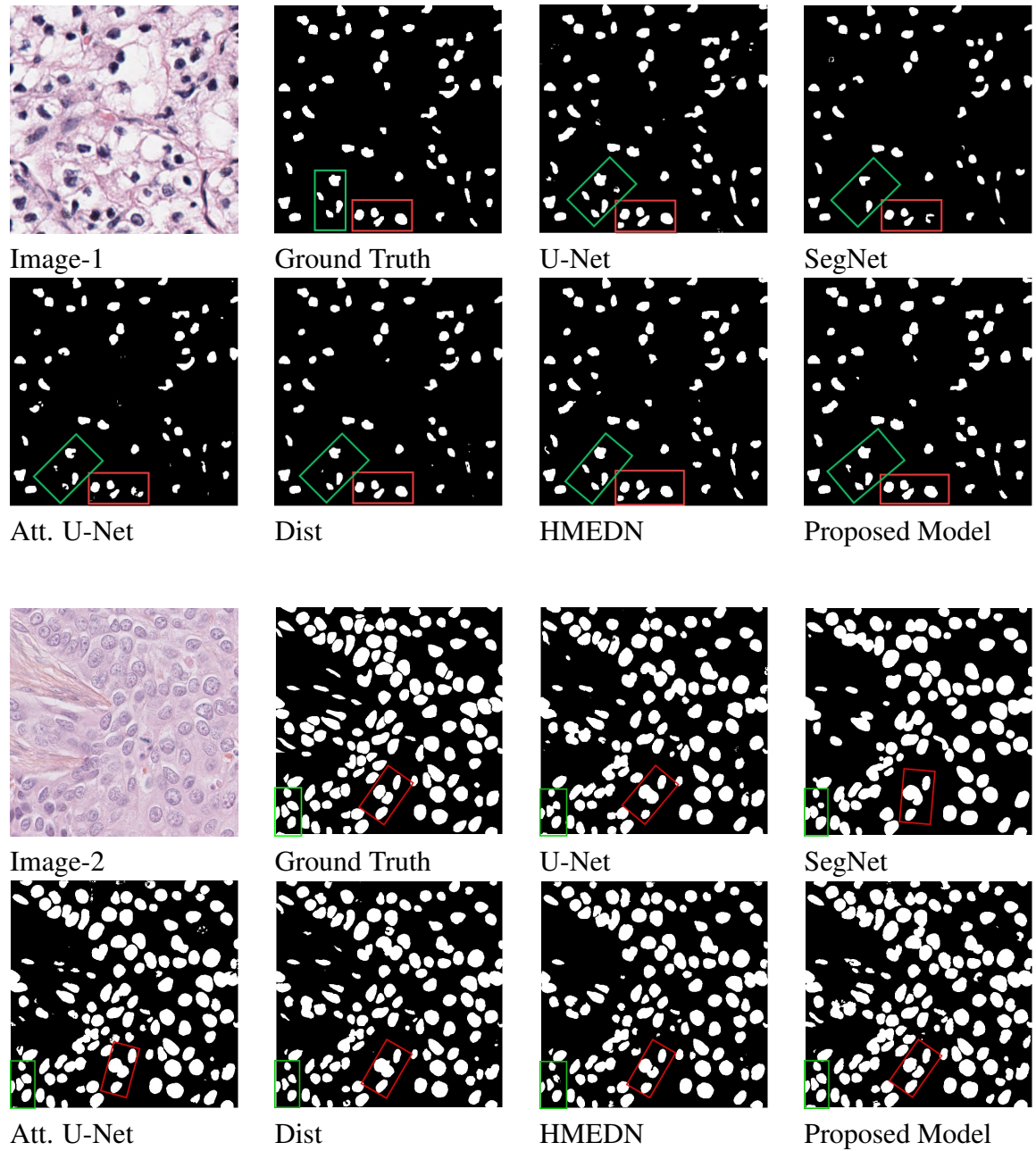


Figure 3.21: Comparison of predicted nuclear regions of different models

Table 3.14: Performance comparison of different models with three datasets

Model	F1 (Kidney Dataset)	AJI	F1 (TNBC Dataset)	AJI	F1 (MoNuSeg Dataset)	AJI
Unet (Ronneberger <i>et al.</i> , 2015)	0.8537	0.7489	0.7324	0.6559	0.7278	0.6029
SegNet (Badrinarayanan <i>et al.</i> , 2017)	0.8972	0.8304	0.7685	0.6434	0.7879	0.6514
Dist (Naylor <i>et al.</i> , 2019)	0.8992	0.8272	0.7516	0.6727	0.7795	0.6467
Att. UNet (Schlemper <i>et al.</i> , 2019)	0.9135	0.8590	0.7216	0.6194	0.7735	0.6315
HMEDN (Zhou <i>et al.</i> , 2020)	0.9262	0.8676	0.8124	0.7017	0.7865	0.6522
Proposed Model	0.9579	0.9261	0.8517	0.7415	0.8065	0.6795

predicted images of five state-of-the-art models with the proposed model of two images of the mentioned datasets. The red box and green box show how many nuclei are correctly identified, how many are partially identified, how many false detections in each model are present, and how many nuclei are not detected. The 23 convolution layers of the U-Net design consist of repeated applications of 3x3 unpadded convolution, 2x2 max-pooling for downsampling, 2x2 upsampling, and ReLU activation after each convolution layer. Sigmoid activation comes next, and then 1x1 convolution. In images 1 and 2, the U-Net architecture recognises 51 of the 57 nuclei and 45 of the 53 nuclei. This architecture does not clearly identify about 10 % to 15% nuclei. A few of the nuclei are grouped, while others are unpredicted. In images 1 and 2, there are two and six grouped nuclei are found. There are four more ducts and six new ducts, both of which are undesirable in images 1 and 2, respectively. The other four images almost exactly match the pattern that this design follows.

SegNet is an encoder-decoder architecture that is extremely effective in terms of memory and time for semantic segmentation of outdoor and interior scene. The decoder upsamples using the transmitted pool and its encoder’s lower resolution input to produce a sparse feature. Its encoder contains 13 convolution layers in encoder side, followed by a 13 decoder layer, batch normalised, 2x2 max-pooling, and ReLU activation. The feature is upsampled by SegNet’s that has modified decoder using the max-pooling indices before being convoluted with a trainable filter. The number of clearly recognised nuclei according to SegNet’s visual segmentation is (80-85 %), which is fewer than U-net. Compared to U-Net, overlapped nuclei are less common, and this model doesn’t pick up on any unwanted things. Extensions that are meaningful to the regular U-Net by adding an additional unit called the attention gate that is trained to suppress unimportant characteristics while highlighting the important ones by enhancing the decoder’s capabilities. This architecture uses attention coefficients to broaden the receptive field, which is essential for more accu-

rate semantic segmentation. Any other segmentation technique can readily incorporate their attention module. Although there are almost as many clearly identifiable nuclei as in the SegNet model, the number of partially detected nuclei is at its highest in this instance. This model fails to anticipate nearly 5% of the nuclei.

By developing a loss function based on intranuclear distance, a challenging boundary-related segmentation problem is addressed. They contrast their model, known as DIST, Mask R-CNN, U-Net, and U-Net+PP with three different datasets. All nuclei in this architecture have either partial or complete identification. 82-88% nuclei were clearly seen. Clustered nuclei segmentation is an issue that has been partially but not fully solved. An encoder-decoder architecture with meaningful connections was used to precisely detect the complicated boundaries for the segmentation of microscopic, MR, and CT images. Blurred images and vanishing borders are precisely detected using a dense connection path with dilated convolution blocks. The prediction accuracy in this architecture is marginally better than that of the Dist algorithm, however as prediction accuracy increased, HMEDN saw an increase in false detection situations. Two undesired items in image 2 and four more items in image 1 have been discovered. The best thing about the suggested model is that 90% of the nuclei are easily recognised, while the remaining nuclei are either grouped or just partially detected. In images 1 and 2, out of 57 and 53 nuclei, 55 and 48 nuclei were clearly visible. In the suggested model, fewer extra and unwanted ducts were found. The number of partially identified nuclei is extremely small. The morphology of the anticipated image is more similar to the actual image. The overlapping nuclei problem has been partially resolved in the suggested approach.

3.10.3 Computational complexity analysis

A comparison of the proposed DSREDN architecture's computational complexity to those of previous segmentation architectures is shown in Table 3.15. Performance is assessed using the F1 Score, AJI Score, and the total number of trainable parameters, which represent the complexity and length of the training process. The proposed model uses 4.3 million parameters which is the least among all reference models except HMEDN architecture. Dist architecture has approximately two times trainable parameter compared to proposed model and it has 7.7 million parameters. HMEDN has least number of FLOPs. Attention UNet has highest number of trainable parameters and FLOPs.

Table 3.15: Computational complexity comparison of different models

Model	Parameters(millions)	FLOPs(millions)
Unet (Ronneberger <i>et al.</i> , 2015)	31.3	62.7
SegNet (Badrinarayanan <i>et al.</i> , 2017)	18.8	47
Dist (Naylor <i>et al.</i> , 2019)	7.7	15.5
Att. UNet (Schlemper <i>et al.</i> , 2019)	31.9	63.7
HMEDN (Zhou <i>et al.</i> , 2020)	0.20	0.45
Proposed Model	4.3	8.17

3.11 Summary

Three new algorithms are proposed for segmentation of nuclei from histopathology images. A deep learning architecture named SCPP-Net was presented to segment the nuclear regions in histopathology images. The proposed SCPP-Net is composed of multiple stages of convolution layers, max-pooling layers and batch-normalization layers to learn new features. To separate nuclei from complex structured histopathology images with different histology and molecular characteristics was the first major concern. This issue was addressed by implementing an effective encoder–decoder architecture with the SCPP unit at the bottleneck layer. SCPP block extracted and learned deep features by varying four different dilation rates. The second major concern was to reduce the computational complexity and total trainable parameters, which was effectively handled by incorporating depth-wise separable convolution. Second proposed architecture called high resolution deep transferred ASPPU-Net that addressed automatic nuclei segmentation of histopathology images having a widely varied spectrum with a large number of artifacts. The implemented networks effectively leverage the strength of residual learning as well as encoder-decoder architecture by incorporating wide and deep network paths that strengthen the intermediate features. Promising results were obtained due to the effective use of the wide and deep network with ASPP at the bottleneck layer.

Finally, The proposed deep structured residual encoder-decoder network focuses on two key areas. First, it efficiently makes advantage of residual connections across the network and offers a wide and deep encoder-decoder path, which allows it to collect more localized information and pertinent context. In order to better train the suggested model and eliminate incorrect predictions, which are undesirable notably in healthcare applications, the disappearing border of discovered nuclei is addressed in step two. On three separate publicly accessible H&E stained histopathology datasets, the proposed architecture was tested. The

majority of popular performance measure metrics for comparing nuclei segmentation have been used, including F1-score, Aggregated Jaccard Index (AJI), the total number of parameters, and FLOPs. The suggested architecture outperformed five cutting-edge deep learning models on three separate histopathology datasets.

Chapter 4

Deep Learning Architecture for Classification of Histopathology Images

4.1 Introduction

In this Chapter, an efficient and robust end-to-end deep learning architecture called RenalNet for the classification of subtypes of RCC from kidney histopathology images is proposed. This chapter describes a new CNN module called multiple channel residual transformation (MCRT) to focus on the most relevant morphological features of RCC. Another novel module called group convolution deep localization (GCDL) that effectively integrates three different feature descriptors. These blocks contribute maximally to the proposed RenalNet model and complete the classification process. In addition, a new benchmark kidney histopathology dataset for the classification of subtypes of renal cell carcinoma is developed from the WSIs of TCGA. The organization of this Chapter is as follows: The methodology of the proposed RenalNet is described in detail with mathematical representation and block diagram in Section 4.2. Implementations and training are presented in Section 4.3 which describes the dataset used, training details, and quality metrics. The simulation results of the proposed RenalNet and other state-of-the-art classification techniques are described in Section 4.4.

4.2 Proposed Method

The proposed RenalNet model comprises of four important phases which are (1) Data preparation (2) Multiple channel residual transformation (3) Group convolution deep localization Network (4) Finally classification phase, where the model works on patches of whole slide images of TCGA dataset. Data preparation phase is carried out under the supervision of experienced pathologists and results of the proposed RenalNet are computed on the patches created from the WSI of TCGA. Multiple channel residual transformation enables the model to produce very fine spatial maps of learned features and focused on the main subject of input data. To capture and better extract the more representative set of features, a group convolution deep localization (GCDL) network has been employed. GCDL integrated with three different feature descriptors called strided convolution, attention mechanism, and aggregated depthwise separable convolution block. GCDL allows the network to generalize and adapt a variety of data.

Structure of proposed RenalNet model for classification of RCC subtypes and normal renal tumor from H&E stained histopathology data is shown in Fig 4.1. This study is different from others for several reasons: (1) Of course, the transfer learning approach transfers the powerful feature, the development of an end-to-end deep learning model that is intended to capture morphological and clinical features as well as overall tissue organization for the classification of RCC subtypes from kidney histopathological images have not been focused on. (2) The ultimate aim of using AI in a clinical workflow is speed and accuracy. The proposed method meet both requirements. (3) The introduced dataset is class balanced compared to any other RCC dataset. (4) Comparable experiments have been performed on multiple organ histopathology data to verify the validity of the proposed method. (5) The performance of networks leveraged by transfer learning of pre-trained ImageNet weights has been included.

4.2.1 Multiple channel residual transformation

Multiple channel residual transformation (MCRT) is composed of the convolutional unit, squeeze and excitation (SE) channel attention unit, and one identity connection. MCRT focus on two aspects: First, it achieves multiple equivalent receptive fields by utilizing different squeeze ratio, which effectively enriches the nuclear features at three different points within the network, and second, multiple parallel branches produce very fine spatial maps at very low computational complexity. The standard convolution operation in three

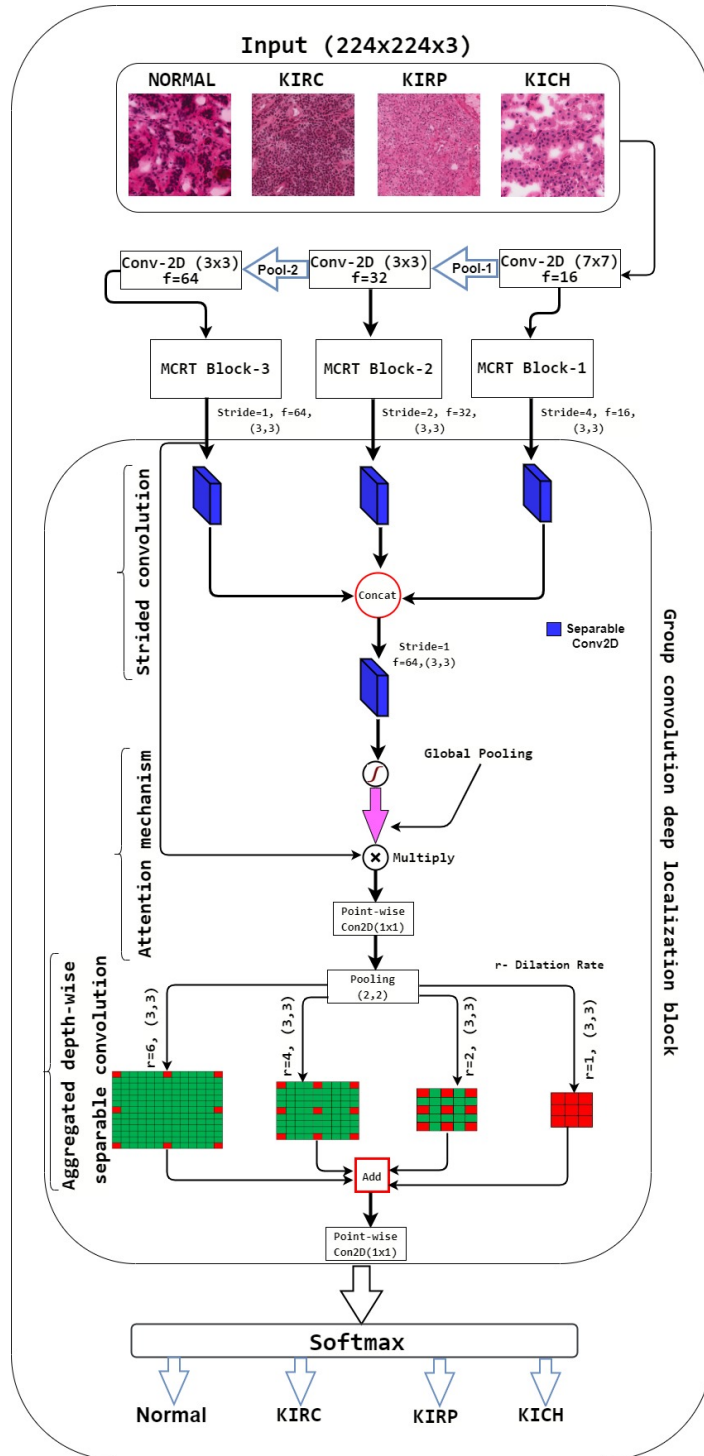


Figure 4.1: Structure of proposed RenalNet model for classification of RCC subtypes and normal renal tumor from H&E stained histopathology data

layers of MCRT operates on both spatial and channel-wise features within local receptive fields. SE channel attention unit (Hu *et al.*, 2020) in deep learning is a mechanism that

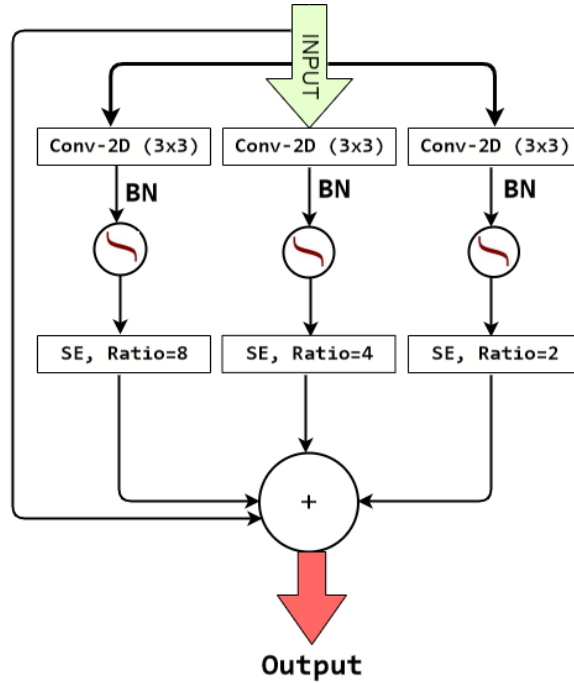


Figure 4.2: Multiple channel residual transformation network

focuses on the main subject of input data for some layer to create corresponding output by concentrating only on the relevant features. SE block re-calibrates the channel-wise feature by producing a channel descriptor across its spatial span. To generate channel-wise statistics, the first step of SE block is to squeeze the global spatial information into the channel parameters called channel descriptors. Here the input feature map is first converted into a vector using global average pooling operation. A dense layer with ReLU activation squeezes the layer to different ratio. Since channel attention mechanism concentrates on the most relevant channels (Woo *et al.*, 2018), (Hu *et al.*, 2020), Fu *et al.* (2020), using three different ratios, channel depth of f , $f/2$, $f/4$ has been achieved. By fusing the outputs of three SE blocks with different channel ratios, it is possible to further strengthen by capturing the most relevant content contributed by different channel depths. In this way, squeezed phase in MCRT utilizes varying squeeze ratios to strengthen the local semantic feature.

The second step of the process called excitation operation that uses a fully connected layer to capture channel-wise dependencies. MCRT extract features by fusing the information of four parallel paths including one skip connection as shown in Fig. 4.2. The

enhanced feature using multiple squeeze ratio followed by convolution, batch normalization, and ReLU in three parallel paths are added with input using identity connection. Let the transformation $T : X \mapsto Y$, Where $X \in \mathbb{R}^{H \times W \times C}$, $Y \in \mathbb{R}^{H \times W \times C}$ H, W, and C are the height, width, and depth of the input feature.

Let $X = [x^1, x^2, x^3, \dots, x^C]$ be the input of MCRT and $K^i = [k_1^i, k_2^i, k_3^i, \dots, k_C^i]$ be the filter kernel for first layer convolution operation. The output $Y^i = [y_1^i, y_2^i, y_3^i, \dots, y_C^i]$ is convolution of input with kernel $k_c^i = [(k_c^i)^1, (k_c^i)^2, (k_c^i)^3, \dots, (k_c^i)^C]$ shown in equation (4.1).

$$y_c^i = k_c^i \star X = \sum_{j=1}^C (k_c^i)^j \star x^j \quad (4.1)$$

Where $i = 1, 2, 3$ represents three parallel paths of MCRT. $(k_c^i)^j$ is a 2D spatial kernel and therefore represents a single channel of K_c^i .

To generate channel-wise statistics, the first step is to squeeze the global spatial information into the channel parameters called channel descriptors. The c^{th} element of S_{sq}^i , Where $S_{sq}^i = [s_1^i, s_2^i, s_3^i, \dots, s_C^i] \in \mathbb{R}^C$ is denoted as s_c^i is calculated by shrinking the transformed output Y through its spatial dimensions ($H \times W$) and expressed in equation (4.2).

$$s_c^i = \frac{1}{(H \times W)} \sum_{a=1}^H \sum_{b=1}^W y_c^i(a, b) \quad (4.2)$$

The second step of the process called excitation operation that uses a fully connected layer to capture channel-wise dependencies. The excitation operation expressed in equations (4.3) and (4.4), uses a simple gating mechanism with a sigmoid activation function where $W_1^i \in \mathbb{R}^{\left(\frac{C}{r} \times C\right)}$, $W_2^i \in \mathbb{R}^{(C \times \frac{C}{r})}$.

$$\widetilde{S}_{sq}^i = \sigma(W_2^i ReLU(W_1^i S_{sq}^i)) \quad (4.3)$$

$$\widetilde{y}_c^i = \widetilde{s}_{sq(c)}^i \times y_c \quad (4.4)$$

The Final output of MCRT module as

$$\widetilde{Y} = X + \widetilde{Y}^1 + \widetilde{Y}^2 + \widetilde{Y}^3 \quad (4.5)$$

The gradient equation for back propagation to update the weight is defined as

$$\frac{\partial \widetilde{Y}}{\partial X} = \frac{\partial (X + \widetilde{Y}^1 + \widetilde{Y}^2 + \widetilde{Y}^3)}{\partial X} = 1 + \frac{\partial \widetilde{Y}^1}{\partial X} + \frac{\partial \widetilde{Y}^2}{\partial X} + \frac{\partial \widetilde{Y}^3}{\partial X} \quad (4.6)$$

Where $\frac{\partial \widetilde{Y}^1}{\partial X}$, $\frac{\partial \widetilde{Y}^2}{\partial X}$, $\frac{\partial \widetilde{Y}^3}{\partial X}$ represents gradient of outputs from three parallel path with respect to the input.

4.2.2 Group convolution deep localization network

The extracted features from the MCRT block at three different stages within the network are fed to the group convolution deep localization (GCDL) block. GCDL block is composed of three highly efficient sub-sections called strided convolution block, attention mechanism, and aggregated depthwise separable convolution block. GCDL block is intended to improve the generalization power of the proposed CNN model and add more nonlinearity to the representation learned by the previous layer with limited computational complexity.

The term stride defines the step size of the kernel when sliding through the image. The purpose of using strided convolution is to scale each stage output feature into similar dimensions. There are other methods to do this but this is advantageous in the sense that strided convolution performs pooling as well as convolution operation. In a convolution operation, a kernel gets multiplied by the image pixel value with a given stride. Since strided convolution performs both pooling and convolution, pooling alters the dimension as per requirement whereas the convolution process has some learnable parameters, which enhance the feature learning capability. Let X_1 , X_2 , and X_3 are the input feature received from different stages of the network and passed it to three parallel strided convolution layer. The results of these parallel strided convolution are concatenated channel-wise and is expressed in equation (4.7).

$$X_{Concat} = \left(X_1 * K_1^{3 \times 3} \right)_{stride=4} \textcircled{\text{C}} \left(X_2 * K_2^{3 \times 3} \right)_{stride=2} \textcircled{\text{C}} \left(X_3 * K_3^{3 \times 3} \right)_{stride=1} \quad (4.7)$$

where $K_1^{3 \times 3}$, $K_2^{3 \times 3}$, and $K_3^{3 \times 3}$ are the kernels used in the strided convolution at three stages, *-convolution operation, and $\textcircled{\text{C}}$ -concatenation operation.

The resulting concatenated feature pass through another separable convolution operation with kernel $K_4^{3 \times 3}$ to combine the effect of three different stages. equation (4.8) represents

the output feature vector (X_{out}) of strided convolution block.

$$X_{out} = \left(X_{Concat} * K_4^{3 \times 3} \right)_{stride=1} \quad (4.8)$$

Further, X_{out} is passed through ReLU activation and after that global pooling is carried out to improve the awareness of global information. The obtained feature vector is denoted as X_{out}^{GP} and it is the overall effect of all previous operations.

The c^{th} element of (X_{out}^{GP}) , ($X_{out}^{GP} \in \mathbb{R}^C$), denoted as $(X_{out}^{GP})_c$ is calculated by shrinking the transformed output through its spatial dimensions ($H \times W$) and expressed in equation (4.9).

$$(X_{out}^{GP})_c = \frac{1}{(H \times W)} \sum_{a=1}^H \sum_{b=1}^W (X_{out})_c(a, b) \quad (4.9)$$

In attention mechanism, the output feature vector of MCRT block-3 is X_3 gets magnified by the factor X_{out}^{GP} . X_3 is obtained after three subsequent convolution operations of feature space 16, 32, 64 and enriched nuclear feature by the MCRT block-3. The data of X_3 contains the highest contextual information about the nuclear region of the whole of the network. X_{out}^{GP} has the overall effect of all previous operations. The magnified vector expressed in equation (4.10) is multiplication of X_{out}^{GP} and X_3 and it contributes in object localization. Attention also helps to extract contextual information of granular tumor region and improve the classification accuracy. The effect of attention mechanism is studied separately in Section 4.4.1.

$$X_{AM} = (X_{out}^{GP} \times X_3) \quad (4.10)$$

Depthwise separable convolution is very effective in parameter reduction of CNN model, which includes both depthwise convolution and pointwise convolution. In depthwise convolution, the approach is to split the input tensor and kernel into channels. Here, each of the filter channels is applied to only one input channel and the results of convolution of the corresponding input feature with the corresponding channel are stacked together. For the depthwise part, dilated convolution of four different dilation rates is used and results are aggregated. At the final stage, this helps us to effectively capture multi-level contextual features with enlarged fields-of-view. In pointwise convolution, a filter of the unit dimension of a certain depth is applied over the entire image. He *et al.* (2019) used two separate kernels complete the process of depthwise convolution and pointwise convolution. Visualization of detailed computational process in depthwise convolution is shown in Fig.4.3. The need of D filter of size $K \times K \times 1$ applied on input image of size $H \times W \times D$ yields

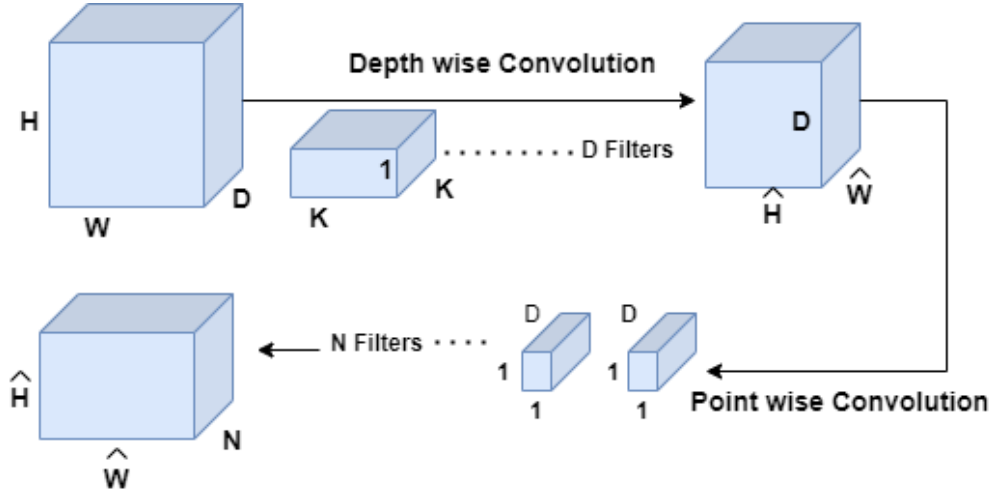


Figure 4.3: Visualization of depth-wise separable convolution

feature size of $\hat{H} \times \hat{W} \times D$. Convolution of N filters of $1 \times 1 \times D$ with $\hat{H} \times \hat{W} \times N$ results $\hat{H} \times \hat{W} \times N$. Computational complexity needed in depth-wise separable convolution is $(K \times K \times 1)(\hat{H} \times \hat{W} \times D) + (1 \times 1 \times D)(\hat{H} \times \hat{W} \times N)$, which is significantly less than standard convolution.

4.3 Implementation and Training

Section 4.3.1 describes the newly introduced dataset for the classification of kidney histopathology images. Two additional liver datasets to check the generalization of the proposed deep-learning model are also mentioned in Section 4.3.1. Training details, quality metrics, deep learning tools used in this case, and a list of important hyper-parameters are presented in Section 4.3.2.

4.3.1 Novel kidney histopathology dataset

Table 4.1: Datasets class distribution at 40X

New introduced Dataset	Total Cases		Extracted ROIs (1024x1024)		(Tabibu <i>et al.</i> , 2019)
	Training	Test	Training	Test	Total Extracted ROIs:Training+Test
Normal	164	29	942	158	-
KIRC	157	28	942	158	1027
KIRP	149	26	942	158	303
KICH	107	19	942	158	254
Total Cases=679		Total ROIs=4400		Total ROIs=1584 (Variable Size)	

The introduced kidney histopathology dataset contains normal renal tissue (Non-cancerous), Renal Clear Cell Carcinoma (KIRC), Renal Papillary Cell Carcinoma (KIRP), and Kidney Chromophobe (KICH) histopathology images of the Renal Cell Carcinoma (RCC). Digital hematoxylin and eosin-stained WSIs of 950 different cases from the cancer genome atlas data portal (<https://gdc.cancer.gov/>) have been acquired . Out of 950 cases we have considered 679 cases, where morphological and clinical features are highly relevant and expressive. The experimental results has been computed and analyzed on the patches created from the WSI of TCGA. The patches creation phase was carried out under the supervision of experienced pathologists who performs routine diagnoses. It has been focused and cropped only on those parts of WSI which have histologically defined clinical features. A patch was considered to be cancerous if the majority of the area had an abnormal region. Created TCGA kidney dataset has a total of 4400 patches belonging to four classes of kidney tissue extracted from 679 different cases. All extracted patches are of size (1024x1024) and each of four classes namely Normal, KIRC, KIRP, and KICH have an equal number of samples.

The processed data from WSI of TCGA used by (Tabibu *et al.*, 2019), (Zhu *et al.*, 2021) is not publicly accessible and there exist high class imbalance. It will not be a fair assessment of a deep learning network if, in multi-class classification, one class contains the majority of total data. However, the newly introduced dataset is class balanced and will be made accessible for the reproducibility of the results. Before any data augmentation, the dataset has been split to avoid overfitting and obtained 4400 patches were slidewise distributed into training (85%), and testing (15%). The validation set was created by randomly selecting patches from the training set. All the images are resized to 224x224 and normalized to zero mean and unit variance before passing them to deep learning architecture for training. Results are evaluated on 632 test images.

The testing phase was performed by the patches of different cases and it is from those WSI which was completely unseen for the model. Table 4.1 gives the complete overview of the data collection process, including the number of cases and number of images obtained per class. To verify the robustness of the deep learning model on novel kidney histopathology data, the performance of state-of-the-art models is evaluated on two additional liver histopathology datasets Aatresh *et al.* (2021a), obtained from Kasturba Medical College (KMC), Mangalore, MAHE, Manipal, Karnataka, India, and the cancer genome atlas data portal (<https://gdc.cancer.gov/>).

Table 4.2: Details of hyper-parameters

Optimizer	Adam with initial learning rate of 0.001
Batch size	4
Convolution filter	16, 32, 64
Dilation rates	1,2,4,6
Total epochs	65
Reduce learning rate	Factor=0.5, patience=5, min-delta=0.0001, Monitor-Validation acc.
Early stop	Patience=30, min-delta=0.0001, Monitor-Validation acc. restore-best-weights

4.3.2 Training details

Implementation and training of the proposed RenalNet architecture were conducted on Dell-G4G3NSM workstation with 8 GB NVIDIA QUADRO P4000 GPU. TensorFlow 2.4.1 and Python-3 environment have been used to speed up the entire process. The proposed model was trained on the training set for a given range of hyper parameters and evaluated the efficiency of the model on the test and validation datasets. After creating patches of size (1024x1024x3) from whole slide image, these patches were resized to (224x224x3) and given as input to the proposed RenalNet model.

Adam optimizer (Kingma and Ba, 2014), with a default learning rate of 0.001 has been used. Categorical-Crossentropy (Togacar *et al.*, 2020), (Aatresh *et al.*, 2021a) is used to optimize the loss and the softmax classifier (Jiang *et al.*, 2019) at the final layer completes the classification process. These are the important controlling hyper-parameters that have been tuned: (1) Reduce learning rate of factor = 0.5, Patience = 5, Min-delta = 0.0001, and monitored the validation accuracy. (2) Batch size = 4 (3) The size of the convolution filter = 16, 32, and 64. (4) Dilation rates = 1,2,4,6 (5) Total Epoch = 65. The details of hyper-parameters are summarized in the Table 4.2. Quality matrices used in this study are precision *Precision* (Tabibu *et al.*, 2019), (Togacar *et al.*, 2020), *recall* (Aatresh *et al.*, 2021a), *F1-score* (Zhu *et al.*, 2021), and *Accuracy* Jiang *et al.* (2019).

The process of optimizing loss function quantifies the error between the output of the network and the corresponding labels. Cross-entropy loss followed by sigmoid activation is called sigmoid Cross-Entropy loss which is used for two-class classification problems. Similarly, cross-entropy loss followed by softmax activation is called softmax cross-entropy that is useful for multiclass classification problems. Here, softmax cross-entropy loss function is used for the classification of RCC images into four categories namely Normal/Non-

cancerous, KIRP, KIRC, and KICH. To prove the worthiness and efficacy of the proposed RenalNet model, there were carefully trained and tested all models for three trials on each datasets. The final results based on an average of three trials have been calculated by initializing random weights in each trial. Class-wise and the overall performance metrics of the model were measured to analyze the performance of the network. Overall accuracy is not the average of class-wise accuracy, it is model accuracy based on total TP, TN, FP, and FN instances belonging to all classes.

4.4 Results and Discussions

In this work, the existing deep-learning models have been trained, and developed a fast and accurate RenalNet architecture, which has the potential to classify subtypes of RCC from kidney histopathology images. Further, the proposed RenalNet architecture reduces the need for extensive computations by incorporating highly optimized modules in the network. The analysis of other variations of the proposed network is presented through an ablation study in Section 4.4.1. The quality metrics and the prediction of the proposed RenalNet model are compared with other classification networks like DenseNet, ShuffleNet, MobileNetV3, ESPNetV2, BHCNet, ResNet (Deep tuned on Imagenet weights), BreastNet, and LiverNet in Section 4.4.2. The performance of networks leveraged by transfer learning of pre-trained weights from the ImageNet dataset is also included. 5-fold cross-validation to analyze how the model is working other than a particular set of data is presented in Section 4.4.3. Statistical and computational complexity analysis of the proposed RenalNet and other benchmark models are presented in Sections 4.4.4 and 4.4.5 respectively.

4.4.1 Ablation study

The proposed RenalNet is composed of two main modules. (I) Multiple channel residual transformation (MCRT). (II) Group convolution deep localization (GCDL) network. GCDL block contains three highly efficient sub-modules called strided convolution block, attention mechanism, and aggregated depthwise separable convolution block. To demonstrate the effectiveness of the proposed RenalNet, we have measured the strength of individual modules as well as some combination of these by detaching the main module from the proposed RenalNet. All these different versions of the network were trained on the same environment and performance was compared with the proposed RenalNet, where all these important

modules were jointly worked. The overall test accuracy, and other quality metrics of different versions are listed in Table 4.3. Computational complexity of different variations were expressed in terms of the total number of parameters and floating-point operations (FLOPs). Visual performance of each variations are analysed in Fig 4.4 by using activation maps which shows how accurately the model separates the malignant tissue regions from the background.

Comparison of activation maps of different variations: The classification performance depends on the probability score assigned to the relevant region. An activation map is a simple and straight-forward visualization technique to show the assigned probability to the relevant regions. This activation map is generated at the final stage of the feature extraction part of the model. CNN-1, where the complete GCDL block is detached from the proposed network, assigns a high probability score for most of the nuclear regions but also assigns a low probability value to background pixels in some cases. CNN-3 and CNN-4 avoid such conditions. CNN-3 measures the strength of GCDL with aggregated depth-wise separable convolution block alone while CNN-4 evaluates GCDL capability with attention mechanism part only. CNN-4 assigns a very high probability to granular tumor region compared to any other variations but images that have larger nuclear regions are not detected properly. This shortcoming with CNN-4 is somewhat solved with CNN-3. The proposed method utilizes the combined effect of CNN-1, CNN-3, CNN-4, such that it is helpful in avoiding false-positive cases. In CNN-2 high probability values are less in number but help to reduce false-positive cases to some extent. By using same channel ratio in three parallel SE blocks as in CNN-5, CNN-6, and CNN-7, as the ratio increases, the true positive cases increases, at the same time it assigns low probability values for background regions. The proposed model

Table 4.3: Overall quality comparison of eight variants of proposed RenalNet MCRT: Multiple channel residual transformation, GCDL: Group convolution deep localization, ADSC:Aggregated depthwise separable convolution, AM: Attention mechanism, SE with excitation ratio (R)= 2, 4, 8.

Model Variations	MRRT	GCDL	ADSC	AM	R=2	R=4	R=8	F1 Score	ACC	Parameters	FLOPs (G = 10 ⁹)
CNN-1	✓	×	×	×	✓	✓	✓	0.8924	0.8924	182,886	2.58G
CNN-2	×	✓	✓	✓	×	×	×	0.8801	0.8813	56,500	0.609G
CNN-3	✓	×	✓	×	✓	✓	✓	0.8969	0.8971	208,582	2.65G
CNN-4	✓	×	×	✓	✓	✓	✓	0.8957	0.8955	202,038	2.7G
CNN-5	✓	✓	✓	✓	✓	×	×	0.8994	0.9003	219,964	2.71G
CNN-6	✓	✓	✓	✓	×	✓	×	0.8981	0.8987	211,816	2.71G
CNN-7	✓	✓	✓	✓	×	×	✓	0.9093	0.9098	207,742	2.71G
Proposed	✓	✓	✓	✓	✓	✓	✓	0.9165	0.9167	213,174	2.71G

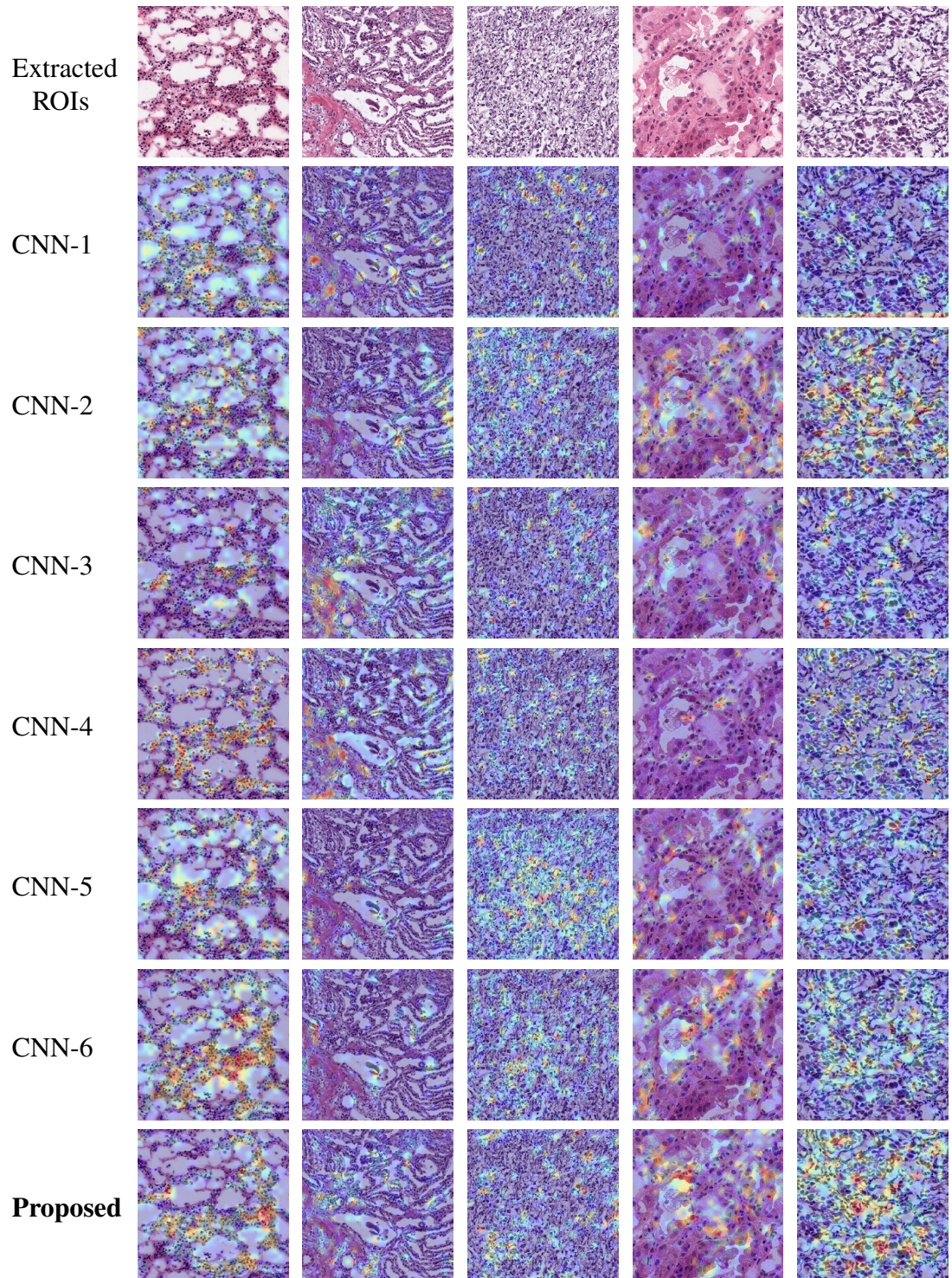


Figure 4.4: Visual performance comparison of different CNN variations using activation maps. (Red: Very high probability score regions, Orange: Medium probability score regions, Light blue: Low probability score regions.)

uses multiple channels ratio, assigns high probability values to nuclear regions, segregates tissue regions from the background properly and, enhances the level of predictions.

Without group convolution deep localization: This variant of the proposed model attains an accuracy of 89.24%, and an F1 score of 89.24%. GCDL block contains three efficient sub-sections called strided convolution block, attention mechanism, and aggregated depth-wise separable convolution block. The strength of each sub-module in GCDL is also measured. There is performance drop observed after detaching the GCDL block and its constituent parts from the proposed model. Very small variations in computational complexity were observed however its contribution to the proposed model cannot be ignored. The produced activation map of different variations indicated that the introduced GCDL block was useful in improving the generalization power of the proposed CNN model to learn and adapt other histopathology data.

Without multiple channel residual transformation: MCRT module composed of three parallel branches of the convolutional unit, squeeze and excitation (SE) channel attention unit, and one skip connection. The squeezed phase in three SE units utilizes different squeeze ratios to strengthen the local semantic feature. This block contributes maximally to the proposed RenalNet model and completes the classification process. Without MCRT results in a drop of 3.54 %, and 3.64% in accuracy, and F1-score respectively from the proposed model. Table 4.3 and Fig. 4.4 shows that the proposed RenalNet is the best performing combination among all modifications made in the base model. This ablation study helps us to choose the best performing combination among all variations. The two important modules togetherly making the proposed RenalNet very effective in feature extraction with minimum computational complexity.

4.4.2 Comparison with state-of-the-art classification methods

The results of the proposed model have been compared with the benchmark models called DenseNet (Huang *et al.*, 2017), ShuffleNet (Zhang *et al.*, 2017), MobileNetV3 (Howard *et al.*, 2019), ESPNetV2 (Mehta *et al.*, 2018), BHCNet (Jiang *et al.*, 2019), ResNet (Deep tuned on Imagenet weights) Tabibu *et al.* (2019), BreastNet (Togacar *et al.*, 2020), and LiverNet (Aatresh *et al.*, 2021a). All models were evaluated on three different histopathology datasets: (1) Introduced TCGA kidney dataset (2) KMC Liver dataset (3) TCGA Liver dataset. The proposed model substantially outperforms the other models in terms of *precision*, *recall*, *F1 score*, and *accuracy*. All benchmark models and the proposed model have been experimented for three trials on each histopathology dataset and evaluated by ran-

domly initializing the weights in each trial. The reported result in the table is the average of three independent trials of each class. The *F1* score, and *accuracy* of the proposed model are 0.9165, and 0.9167 respectively for the new TCGA kidney dataset. Overall *accuracy* for TCGA kidney dataset, the RenalNet model achieved the highest 17% margin on ESPNetv2 while the lowest 4.12 % margin on BHCNet. For KMC Liver dataset, *F1*, and *accuracy* of the proposed model are 0.9726, and 0.9724 respectively. The classwise *Precision*, *Recall*, *F1*, *Accuracy*, and overall performance of all benchmark models and proposed model are presented in Table 4.4, Table 4.5, and Table 4.6 for TCGA kidney dataset, KMC Liver dataset, and TCGA Liver dataset respectively.

Table 4.4: Performance metrics comparison of proposed RenalNet with other competitive model on introduced TCGA kidney dataset

Metrics	Class	DenseNet (2016)	ShuffleNet (2018)	MBNetV3 (2019)	ESPNetv2 (2019)	BHCNet (2019)	Pretrained ResNet(2019)	BreastNet (2020)	LiverNet (2021)	RenalNet (Proposed)
Precision	KICH	0.8038	0.8544	0.8055	0.7056	0.8818	0.8341	0.8331	0.8788	0.9327
	KIRC	0.8961	0.8771	0.9150	0.8620	0.9288	0.8637	0.9063	0.9072	0.9477
	KIRP	0.7223	0.8236	0.8402	0.7209	0.8197	0.7983	0.8400	0.8346	0.8651
	Normal	0.8361	0.8736	0.8245	0.7172	0.8765	0.8398	0.8497	0.8412	0.9244
	Overall	0.8145	0.8572	0.8463	0.7514	0.8767	0.8340	0.8573	0.8654	0.9175
Recall	KICH	0.8438	0.9029	0.9071	0.7890	0.8776	0.8438	0.8734	0.8966	0.9536
	KIRC	0.8882	0.9324	0.9008	0.8607	0.9388	0.9219	0.9472	0.9240	0.9494
	KIRP	0.7742	0.7911	0.7489	0.6371	0.8396	0.7447	0.8101	0.7721	0.8881
	Normal	0.7362	0.8016	0.8185	0.6983	0.8460	0.8291	0.7953	0.8607	0.8755
	Overall	0.8106	0.8571	0.8439	0.7463	0.8755	0.8349	0.8565	0.8633	0.9167
F1 Score	KICH	0.8215	0.8778	0.8525	0.7422	0.8791	0.8386	0.8519	0.8857	0.9426
	KIRC	0.8921	0.9039	0.9075	0.8608	0.9335	0.8918	0.9258	0.9144	0.9484
	KIRP	0.7470	0.8066	0.7904	0.6673	0.8290	0.7703	0.8245	0.7983	0.8761
	Normal	0.7818	0.8358	0.8209	0.7061	0.8604	0.8344	0.8203	0.8508	0.8992
	Overall	0.8106	0.8560	0.8428	0.7441	0.8755	0.8337	0.8556	0.8623	0.9165
Accuracy	KICH	0.9082	0.9372	0.9214	0.8634	0.9393	0.9187	0.9235	0.9425	0.9710
	KIRC	0.9462	0.9504	0.9541	0.9303	0.9668	0.9441	0.9620	0.9572	0.9741
	KIRP	0.8691	0.9051	0.9013	0.8433	0.9135	0.8892	0.9140	0.9024	0.9372
	Normal	0.8976	0.9214	0.9108	0.8554	0.9314	0.9177	0.9135	0.9245	0.9509
	Overall	0.8106	0.8571	0.8439	0.7463	0.8755	0.8349	0.8565	0.8633	0.9167

Comparison of validation accuracy of state-of-the-art models, validation loss of three trials, confusion matrix of best trial, and ROC-AUC of test data are shown in Fig.4.5, Fig.4.6, and Fig.4.7 for each dataset respectively. The learning curves indicate the proposed model is working very well for training data as well as validation data. The proposed model shows the best learning behavior and generalizations to work on different types of histopathology data. ROC-AUC plot indicated the measure of the ability of a classifier between different labels. The higher is ROC-AUC value better the performance of the model. In the ROC-AUC viewpoint, the proposed model is best for each dataset.

Performance of networks leveraged by transfer learning of pre-trained weights from

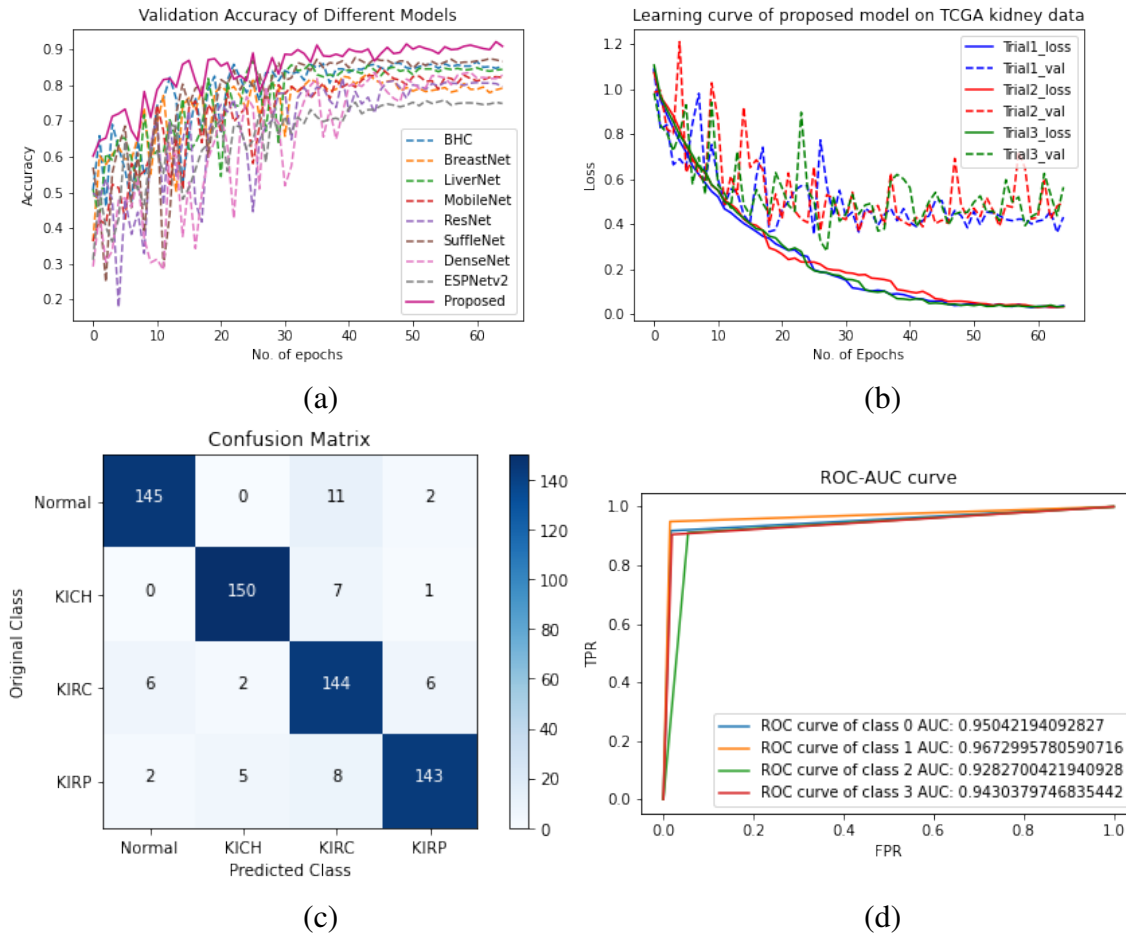


Figure 4.5: Learning curve of RenalNet on introduced TCGA kidney dataset (a) Comparison of validation accuracy of state-of-the-art models (b) Training and validation loss of three trials (c) Confusion matrix of best trial (d) ROC-AUC of test data for four class classification class-0 (KICH) class-1 (KIRC), class-2 (KIRP), class-3 (Normal)

the ImageNet dataset is shown in Table 4.7. Here all the layers are frozen except very few layers at the top are trained for 30 epochs. The quality metrics and the prediction of the proposed RenalNet model trained from scratch are compared with other classification networks like VGG-16, ResNet-50, ResNet-150, DenseNet, InceptionNet, InceptionResNetV2, MobileNetV2, and NasNet, which is using pre-trained weights from the ImageNet dataset. From the overall *accuracy* point of view for TCGA kidney dataset, the RenalNet model achieved the highest margin of 11.93% on NasNet while the lowest margin of 4.65% on DenseNet. Among all examinations, CNN models like BHCNet, LiverNet, BreastNet, and the proposed RenalNet model achieve acceptable results without transfer learning.

Table 4.5: Performance metrics comparison of proposed RenalNet architecture with other competitive model on KMC liver dataset

Metrics	Class	DenseNet (2016)	ShuffleNet (2018)	MBNetV3 (2019)	ESPNetv2 (2019)	BHC-Net (2019)	Pretrained ResNet(2019)	BreastNet (2020)	LiverNet (2021)	RenalNet (Proposed)
Precision	HCC-1	0.9953	1	0.8701	1	0.9950	0.9907	1	1	1
	HCC-2	0.9117	0.9051	0.9330	0.8392	0.9364	0.9120	0.8756	0.9350	0.9195
	HCC-3	0.8128	0.8377	0.7868	0.8170	0.8562	0.8245	0.8735	0.8834	0.9865
	HCC-4	1	1	0.9869	1	1	1	1	0.9697	1
	Overall	0.9299	0.9357	0.8942	0.9140	0.9469	0.9318	0.9373	0.9470	0.9765
Recall	HCC-1	1	1	1	0.9952	0.9619	1	1	1	1
	HCC-2	0.9167	0.9416	0.8083	0.8458	0.9083	0.85	0.9083	0.8875	1
	HCC-3	0.9041	0.9	0.8416	0.8458	0.9666	0.9125	0.8708	0.9208	0.9125
	HCC-4	0.8	0.8133	0.86	0.9333	0.8733	0.92	0.9333	0.9733	0.98
	Overall	0.9052	0.9137	0.8775	0.9050	0.9275	0.9206	0.9281	0.9454	0.9731
F1 Score	HCC-1	0.9976	1	0.9299	0.9976	0.9780	0.9953	1	1	1
	HCC-2	0.9133	0.9225	0.8594	0.8420	0.9191	0.8796	0.8909	0.9102	0.9580
	HCC-3	0.8552	0.8672	0.8116	0.8306	0.9065	0.8651	0.8704	0.9001	0.9480
	HCC-4	0.8888	0.8969	0.9160	0.9654	0.9299	0.9573	0.9647	0.9702	0.9898
	Overall	0.9137	0.9216	0.8792	0.9089	0.9334	0.9243	0.9315	0.9451	0.9740
Accuracy	HCC-1	0.9988	1	0.9619	0.9988	0.9892	0.9976	1	1	1
	HCC-2	0.95	0.9547	0.9261	0.9095	0.9547	0.9333	0.9369	0.95	0.975
	HCC-3	0.9130	0.9214	0.8892	0.9011	0.9428	0.9190	0.925	0.9416	0.9714
	HCC-4	0.9642	0.9666	0.9726	0.9880	0.9773	0.9857	0.9881	0.9892	0.9964
	Overall	0.9130	0.9214	0.875	0.8988	0.9321	0.9178	0.925	0.9404	0.9714

Table 4.6: Performance metrics comparison of proposed RenalNet architecture with other competitive model on TCGA liver dataset

Metrics	Class	DenseNet (2016)	ShuffleNet (2018)	MBNetV3 (2019)	ESPNetv2 (2019)	BHCNet (2019)	Pretrained ResNet(2019)	BreastNet (2020)	LiverNet (2021)	RenalNet (Proposed)
Precision	Normal	0.9871	0.8938	0.9783	0.9255	0.9901	0.9450	0.9743	0.9685	0.9615
	Primary	0.9484	0.8071	0.6805	0.8425	1	0.7577	0.9326	0.9115	0.9827
	Secondary	0.8668	0.6209	0.5925	0.6851	0.8464	0.6388	0.7979	0.8976	0.9714
	Overall	0.9341	0.7739	0.7504	0.8177	0.9455	0.7805	0.9016	0.9259	0.9719
Recall	Normal	1	0.9066	0.9	0.98	1	1	1	0.9866	1
	Primary	0.9055	0.5777	0.8166	0.7222	0.8833	0.7666	0.8444	0.9111	0.95
	Secondary	0.9142	0.8666	0.4571	0.7904	1	0.5714	0.8952	0.8666	0.9714
	Overall	0.9399	0.7837	0.7246	0.8308	0.9611	0.7793	0.9132	0.9214	0.9738
F1 Score	Normal	0.9934	0.8893	0.9375	0.9515	0.9950	0.9713	0.9869	0.9770	0.9804
	Primary	0.9260	0.6516	0.7424	0.7762	0.9377	0.7617	0.8862	0.9098	0.9661
	Secondary	0.8872	0.7210	0.5161	0.7332	0.9160	0.6031	0.8434	0.8793	0.9714
	Overall	0.9360	0.7540	0.7320	0.8203	0.9495	0.7787	0.9055	0.9221	0.9726
Accuracy	Normal	0.9954	0.9241	0.9586	0.9655	0.9965	0.9793	0.9908	0.9839	0.9862
	Primary	0.9402	0.7609	0.7655	0.8275	0.9517	0.8022	0.9103	0.9264	0.9724
	Secondary	0.9448	0.8367	0.7931	0.8620	0.9551	0.8183	0.9195	0.9425	0.9862
	Overall	0.9402	0.7609	0.7600	0.8275	0.9517	0.8000	0.9103	0.9264	0.9724

The proposed RenalNet, trained from scratch achieved the highest accuracy for renal cell carcinoma (RCC) detection and two additional histopathology datasets.

Table 4.7: Performance metrics comparison of proposed RenalNet trained from scratch with other models using pre-trained weights from the ImageNet weight on introduced kidney dataset

Metrics	Class	VGG-16	ResNet-50	ResNet-150	DenseNet	Inception	Inception-ResnetV2	MobileNetV2	NasNet	RenalNet (Trained from scratch)
Precision	KICH	0.9231	0.7964	0.7964	0.875	0.8562	0.8431	0.8259	0.8258	0.9327
	KIRC	0.9068	0.8742	0.8650	0.9295	0.8152	0.8820	0.9032	0.8312	0.9477
	KIRP	0.7391	0.7814	0.7615	0.8072	0.7549	0.7654	0.7559	0.7362	0.8651
	Normal	0.8471	0.8451	0.8543	0.8734	0.7953	0.8525	0.8506	0.7987	0.9244
	Overall	0.8540	0.8243	0.8193	0.8712	0.8054	0.8358	0.8339	0.7980	0.9175
Recall	KICH	0.7595	0.8417	0.8417	0.8417	0.8291	0.8164	0.8101	0.8101	0.9536
	KIRC	0.9240	0.8797	0.8924	0.9177	0.8101	0.8987	0.8861	0.8417	0.9494
	KIRP	0.8607	0.7468	0.7278	0.8481	0.7215	0.7848	0.8038	0.7595	0.8881
	Normal	0.8417	0.8291	0.8164	0.8734	0.8607	0.8417	0.8291	0.7785	0.8755
	Overall	0.8465	0.8243	0.8196	0.8702	0.8053	0.8354	0.8322	0.7975	0.9167
F1 Score	KICH	0.8333	0.8184	0.8185	0.8581	0.8424	0.8295	0.8179	0.8179	0.9426
	KIRC	0.9154	0.8770	0.8785	0.9235	0.8127	0.8903	0.8945	0.8365	0.9484
	KIRP	0.7953	0.7637	0.7443	0.8271	0.7378	0.7750	0.7791	0.7476	0.8761
	Normal	0.8444	0.8370	0.8349	0.8734	0.8267	0.8471	0.8397	0.7884	0.8992
	Overall	0.8471	0.8241	0.8191	0.8705	0.8049	0.8354	0.8328	0.7976	0.9165
Accuracy	KICH	0.9240	0.9066	0.9066	0.9303	0.9224	0.9161	0.9098	0.9098	0.9710
	KIRC	0.9573	0.9382	0.9383	0.9620	0.9066	0.9446	0.9477	0.9177	0.9741
	KIRP	0.8892	0.8844	0.875	0.9113	0.8718	0.8861	0.8860	0.8718	0.9372
	Normal	0.9225	0.9193	0.9193	0.9367	0.9098	0.9240	0.9208	0.8955	0.9509
	Overall	0.8465	0.8243	0.8196	0.8702	0.8053	0.8354	0.8322	0.7974	0.9167

Table 4.8: 5-Fold cross validation average quality metrics comparison of proposed RenalNet with other competitive models on introduced kidney dataset

Metrics	Class	DenseNet (2016)	ShuffleNet (2018)	MBNetV3 (2019)	ESPNetv2 (2019)	BHCNet (2019)	Pretrained ResNet(2019)	BreastNet (2020)	LiverNet (2021)	RenalNet (Proposed)
Precision	KICH	0.8292	0.8740	0.7845	0.6990	0.8820	0.8432	0.8174	0.8707	0.9116
	KIRC	0.8750	0.8602	0.8831	0.7903	0.8850	0.8793	0.8695	0.8983	0.9005
	KIRP	0.8080	0.8019	0.7873	0.7050	0.8223	0.7732	0.8041	0.8333	0.8748
	Normal	0.8460	0.8961	0.8162	0.7344	0.8839	0.8496	0.7968	0.8911	0.9229
	Overall	0.8395	0.8581	0.8178	0.7322	0.8683	0.8363	0.8219	0.8733	0.9024
Recall	KICH	0.8572	0.8845	0.8736	0.81	0.8818	0.8490	0.8545	0.8909	0.9063
	KIRC	0.8836	0.8963	0.8181	0.8054	0.9045	0.8572	0.8490	0.9054	0.9373
	KIRP	0.7736	0.8045	0.7636	0.6481	0.81	0.7490	0.7563	0.8045	0.8491
	Normal	0.8263	0.8372	0.7972	0.6527	0.8718	0.8773	0.79	0.8818	0.9118
	Overall	0.8352	0.8557	0.8132	0.7291	0.8670	0.8332	0.8125	0.8706	0.9011
F1 Score	KICH	0.8396	0.8777	0.8258	0.7500	0.8814	0.8438	0.8310	0.8785	0.9077
	KIRC	0.8777	0.8766	0.8466	0.7971	0.8929	0.8656	0.8523	0.8989	0.9172
	KIRP	0.7894	0.8025	0.7715	0.6720	0.8154	0.7605	0.7758	0.8179	0.8614
	Normal	0.8329	0.8649	0.8059	0.6889	0.8777	0.8619	0.7911	0.8856	0.9169
	Overall	0.8349	0.8554	0.8124	0.7270	0.8668	0.8329	0.8125	0.8702	0.9008
Accuracy	KICH	0.9182	0.9384	0.9070	0.8647	0.9404	0.9216	0.9120	0.9388	0.9541
	KIRC	0.9382	0.9366	0.9259	0.8977	0.9457	0.9336	0.9263	0.9486	0.9572
	KIRP	0.8961	0.9016	0.8888	0.8418	0.9086	0.8818	0.8907	0.9104	0.9320
	Normal	0.9179	0.9347	0.9045	0.8538	0.9393	0.9293	0.8959	0.9434	0.9588
	Overall	0.8352	0.8557	0.8132	0.7291	0.8670	0.8332	0.8125	0.8706	0.9011

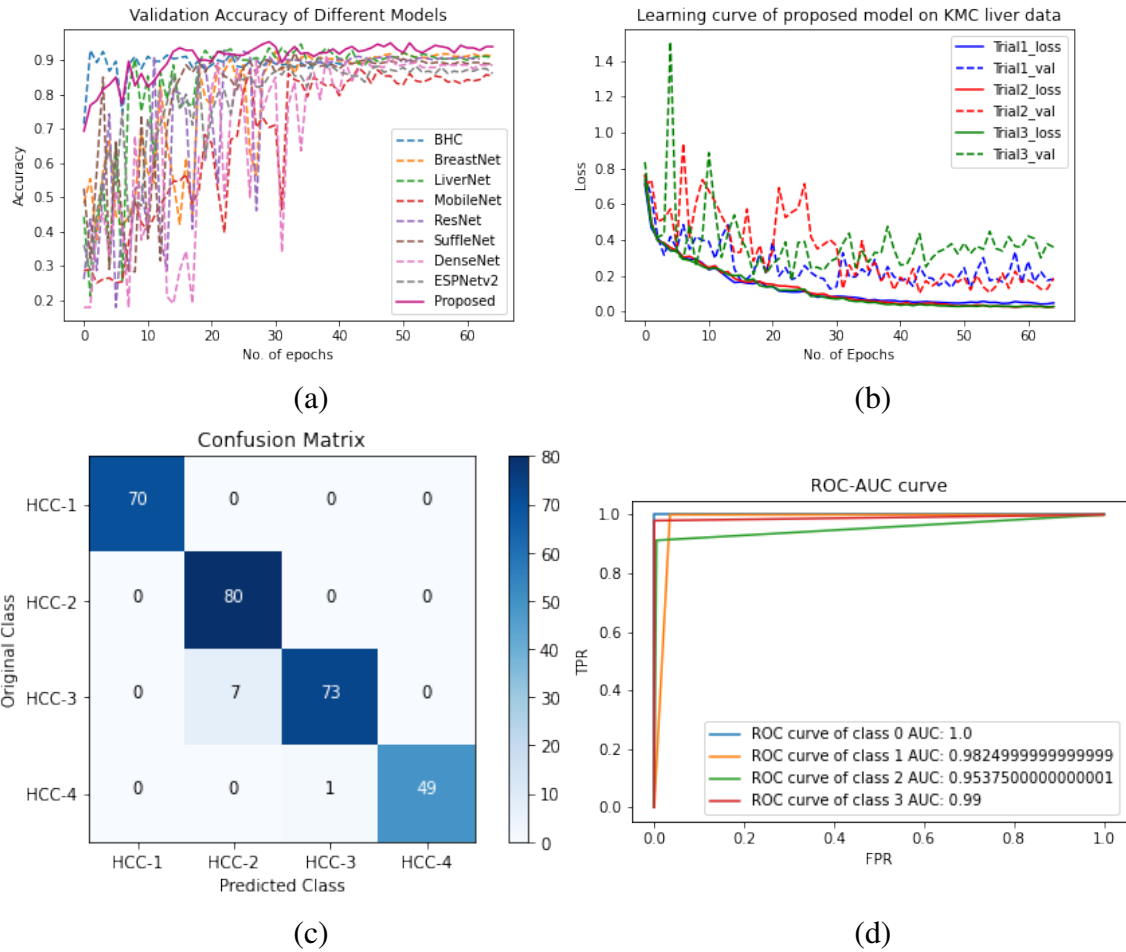


Figure 4.6: Learning curve of RenalNet on KMC liver dataset (a) Comparison of validation accuracy of state-of-the-art models (b) Training and validation loss of three trials (c) Confusion matrix of best trial (d) ROC-AUC of test data for four class classification class-0 (HCC-1) class-1 (HCC-2), class-2 (HCC-3), class-3 (HCC-4)

4.4.3 Cross-validation

In the proposed dataset, the training set and test set are slide-wise separated such that the test set was completely unseen for the model. However it is possible that some cases in the test set have same content as in training set. In this way, the model might not produce the true result. To avoid such issues model was tested using 5-fold cross-validation. This allows us to analyse how the model is working apart from a particular set of test data. To do so, 5 folds of 880 images of each class were created to carry out 5-fold cross-validation. For first-run fold-2 to fold-5 is used to train the model while fold-1 is to test the model. In the

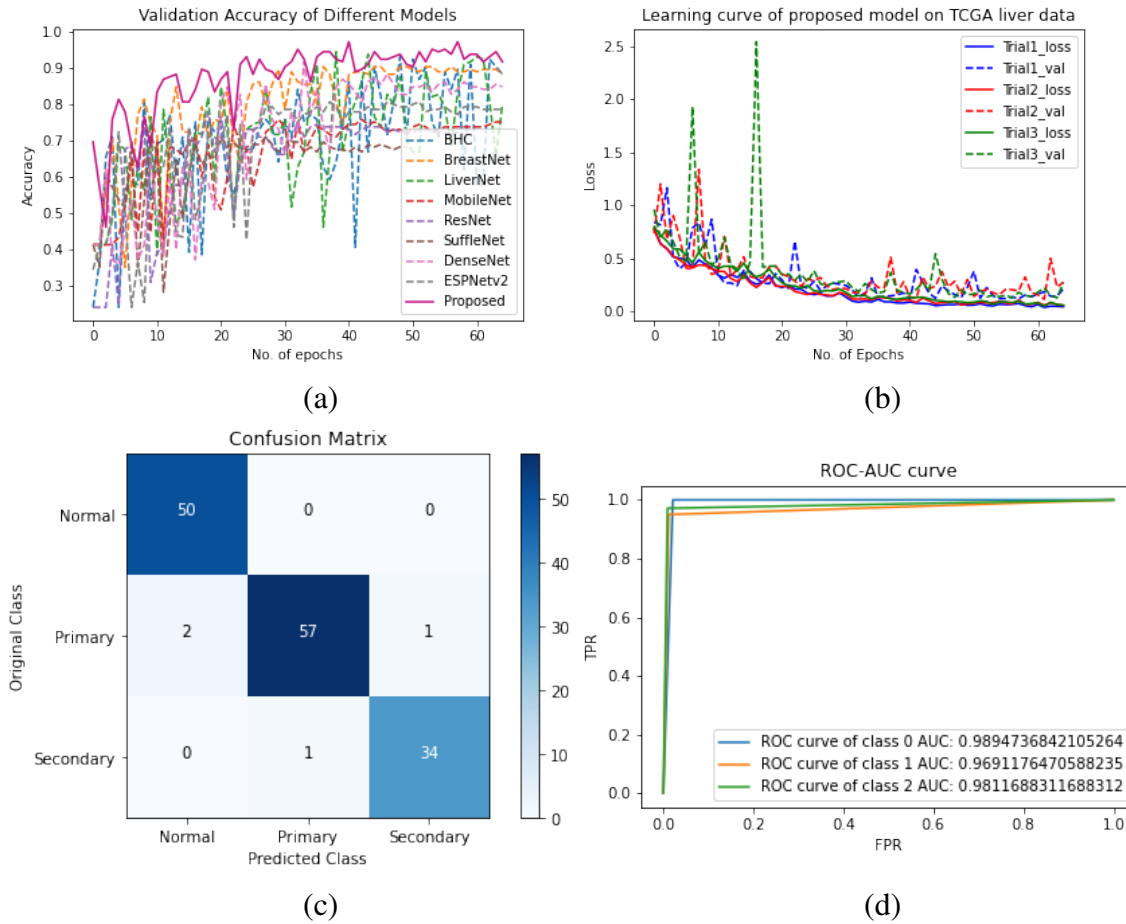


Figure 4.7: Learning curve of RenalNet on TCGA liver dataset (a) Comparison of validation accuracy of state-of-the-art models (b) Training and validation loss of three trials (c) Confusion matrix of best trial (d) ROC-AUC of test data for three class classification class-0 (Normal) class-1 (Primary), class-2 (Secondary)

second run fold-1, fold-3 to fold-5 is used to train the model while fold-2 is to test the model. This process has been repeated till the last fold where fold-5 is for testing and the remaining fold is for training the model. The average results of all five runs for the proposed model and all benchmark models are included in Table 4.8. The proposed model attains almost similar result given in Table 4.4. BreastNet (Togacar *et al.*, 2020), MobileNetV3 (Howard *et al.*, 2019) and ESPNetV2 (Mehta *et al.*, 2018) show considerable variations in the result of Table 4.4 and Table 4.8. The Fig.4.8 shows the learning curves of the proposed model for the 5-fold cross-validation. Validation accuracy of all 5-folds of the proposed model and reference models, and validation accuracy of five different runs of the proposed model

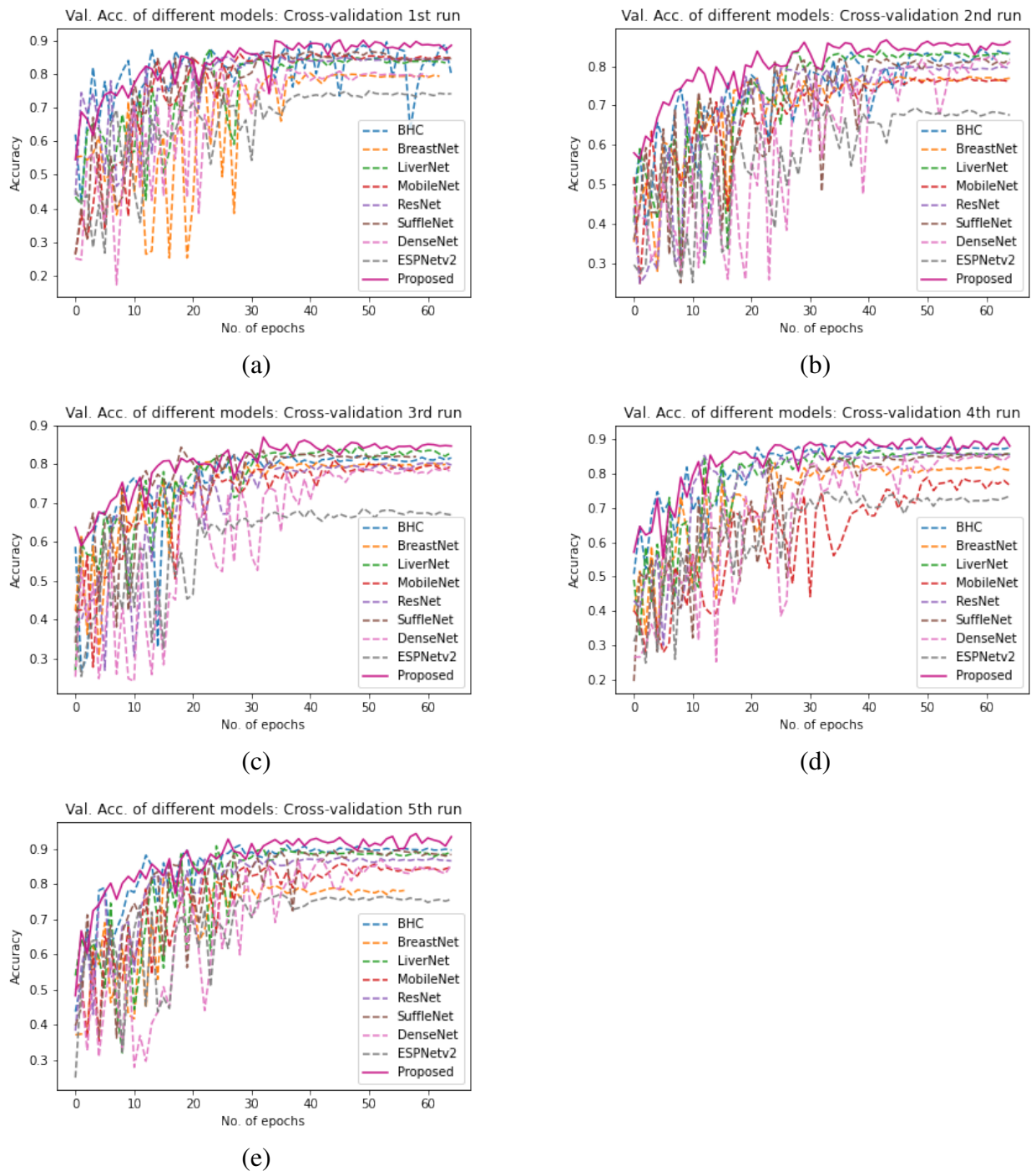


Figure 4.8: 5-Fold cross-validation learning curve of RenalNet on newly introduced TCGA kidney dataset. (a) to (e) Comparison of validation accuracy of state-of-the-art models of five different fold

shows almost same learning behavior, this suggests that the proposed model is generalized independent of the dataset.

4.4.4 Statistical analysis

Fig.4.9 is the box plot of *Accuracy* of all comparison models, and the proposed model for each dataset. It is a very good measure to check the dispersion of data between minimum to maximum via first quartile, median, and second quartile. Each of the boxes in the box plot contains a class-wise score and overall score of a model. Performance variability of the class-wise and overall score depends on the statistical distribution of values obtained by all classification models for three datasets. The above box plot indicates that the proposed RenalNet outperforms the other models in all three histopathology datasets.

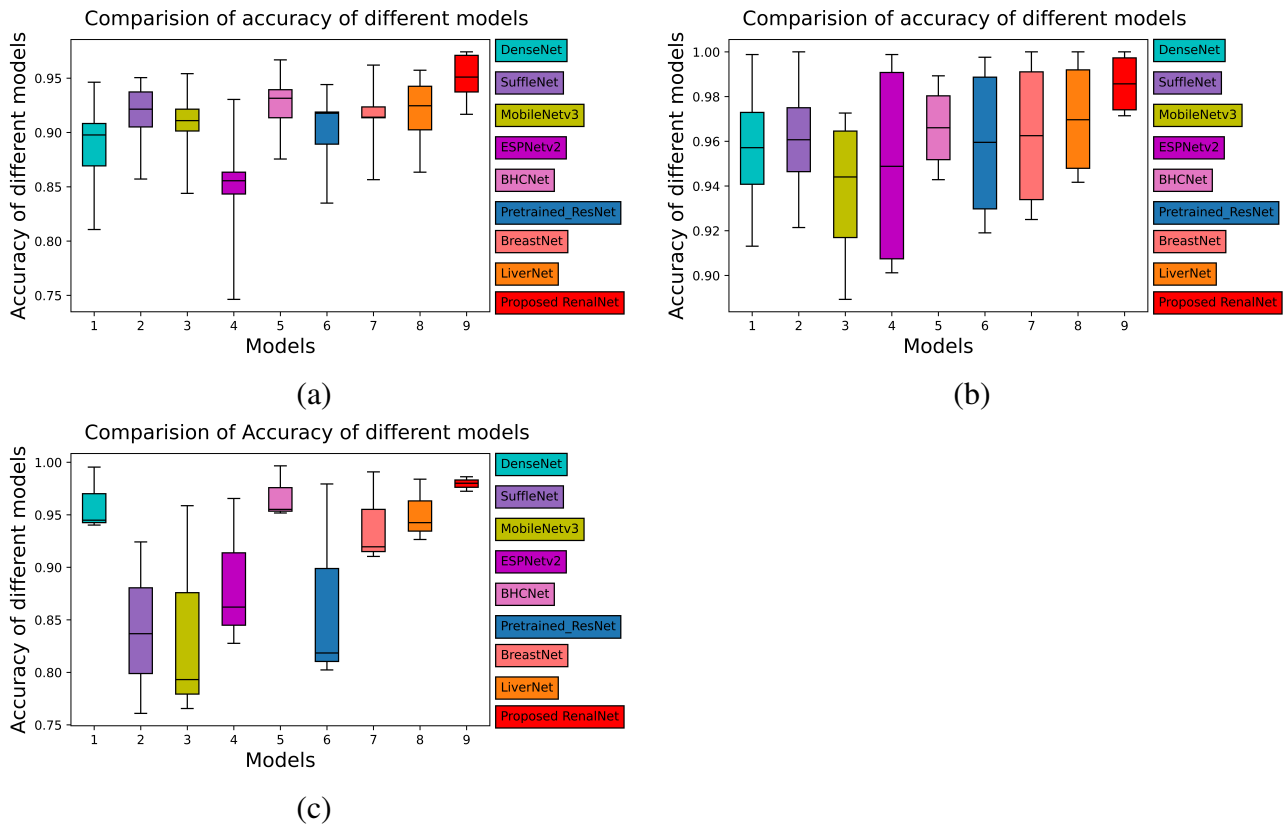


Figure 4.9: Box Plot: (a), (b), (c) Accuracy of different models for new TCGA kidney dataset, KMC Liver dataset, TCGA Liver dataset.

4.4.5 Computational complexity analysis

From the computational complexity viewpoint, the proposed RenalNet model showed effectiveness in comparison with other competitive models. Computational complexity is expressed in terms of the total trainable parameters and floating-point operations (FLOPs) shown in Table 4.9. The proposed model utilizes the least number of trainable parameters with respect to any other comparable models. The MobileNetV3 (Howard *et al.*, 2019), and ESPNetV2 (Mehta *et al.*, 2018) utilized least number of FLOPs which is 0.564 billion and 0.527 billion, respectively, while proposing model utilizes 2.71 billion FLOPs.

Training and prediction time is also a good measure of the complexity of a model. The proposed model and other benchmark models have been trained for 65 epochs in the same training and test environment. These values are given in Table 4.9. The time required for training and prediction of the proposed model is 1.227 hrs and 12.73 sec respectively, which is minimum compared to any other model.

Table 4.9: Computational complexity comparison of architectures

Model	Parameters (M = 10^6)	FLOPs (G = 10^9)	Training Time (Hours)	Prediction Time in Second (Test Set)
DenseNet(2016)	7.0416 M	5.70 G	2.365	27.50
ShuffleNet(2018)	23.8230 M	6.69 G	2.058	15.77
MobileNetV3(2019)	2.2762 M	0.564 G	1.679	22.83
ESPNetV2 (2019)	3.9134 M	0.527 G	1.986	21.71
BHCNet(2019)	0.3020 M	4.27 G	1.986	20.95
Resnet(2019) [Imagenet weights]	23.5960 M	7.75 G	1.462	16.02
BreastNet(2020)	0.6061 M	5.68 G	1.535	20.31
LiverNet(2021)	0.5740 M	3.72 G	2.112	26.66
Proposed RenalNet	0.2131 M	2.71 G	1.227	12.73

4.5 Summary

This chapter describes a fast and accurate deep learning (DL) based automated system for the detection of renal cell carcinoma (RCC) from kidney histopathology images. The proposed RenalNet model experimented for four class (Normal, KIRP, KIRC, and KICH) classification of patches of WSIs of TCGA kidney dataset. Extraction of a class-specific representative set of features was possible due to the utilization of optimized transformation modules at three different resolutions within the network. The results of the proposed model have been compared with the most recent DL models trained from scratch as well as networks leveraged by transfer learning of pre-trained weights. The experimental result

indicated that proposed model gave better results compared to the most recent existing DL models in terms of quality metrics, computational complexity, and training and prediction time.

Chapter 5

Deep Learning Architectures for Grading of Histopathology Images

5.1 Introduction

This chapter describes two newly proposed deep learning framework which is robust and computationally efficient. The proposed architectures are fully automated and designed for Renal Cell Carcinoma (RCC) classification and grade prediction from kidney histopathology images. A new benchmark RCC dataset of H&E stained kidney histopathology images is created for grading of renal tumors, obtained from the Department of Pathology, KMC Mangalore, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India. The organization of this chapter is as follows: The methodology of the two proposed models is described in detail with mathematical representation and block diagram in Sections 5.2 and 5.5. Implementations and training are presented in Sections 5.3 and 5.6 which describes the dataset used, training details, and quality metrics. The simulation results of proposed architectures and other state-of-the-art deep learning techniques are described in Sections 5.4 and 5.7.

5.2 Proposed Architecture-1

The proposed Renal Cell Carcinoma Grading Network (RCCGNet) comprises of three foremost stages that are (1) Data preparation (2) Shared channel residual (SCR) block (3) Finally, grading phase where the network differentiates the five different grades of renal tu-

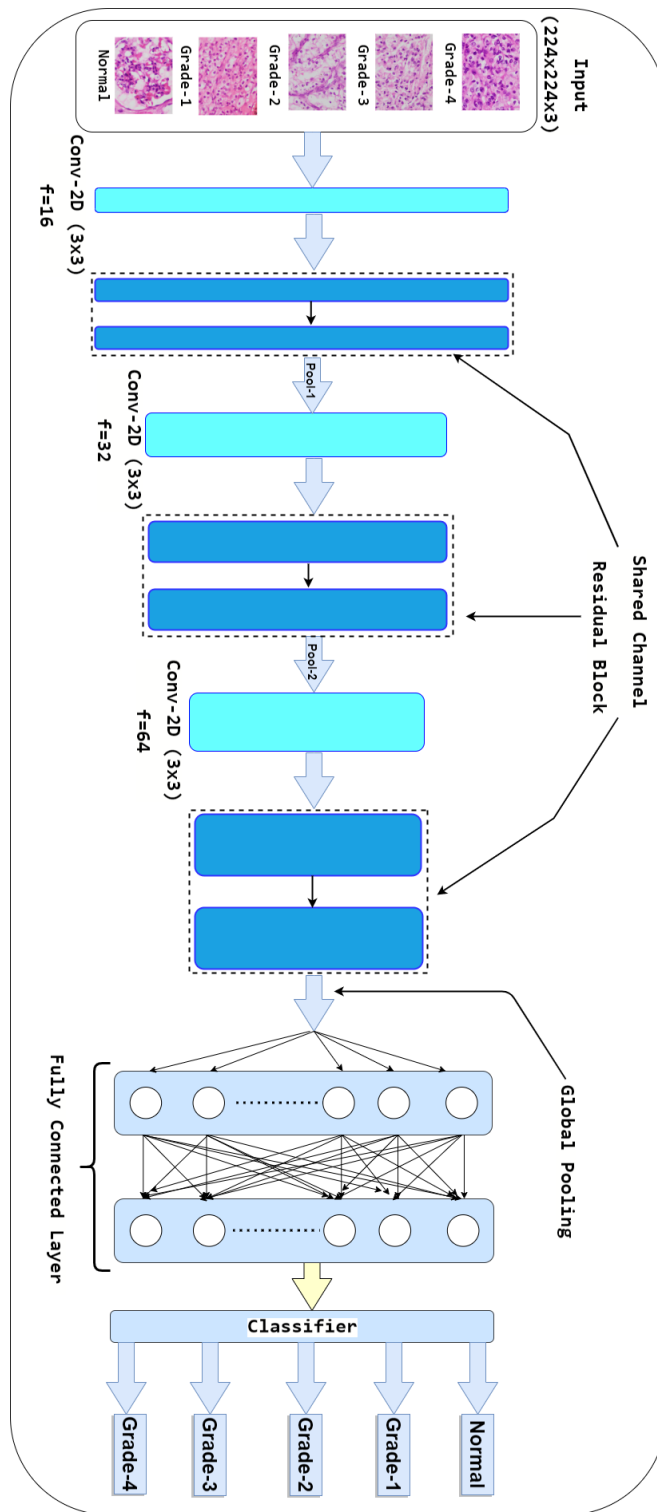


Figure 5.1: Structure of the proposed RCCGNet for grading of renal cell carcinoma from kidney histopathology images

mors by using a softmax non-linear activation function. The proposed model is designed by integrating SCR block, Conv2D, and fully connected layers to extract the features from the kidney histopathology images. The structure of the proposed RCCGNet for grading of RCC from kidney histopathology images has been shown in Fig.5.1.

5.2.1 Shared channel residual block

This section describes the designed architecture of the proposed shared channel residual (SCR) block. The SCR block employed in the proposed RCCGNet contains convolutional layers, a new and efficient method of sharing the information between two different layers, a simple gating mechanism to focus on the most relevant morphological features of RCC, and a residual connection. The first stage of the SCR block accepts an input, which undergoes a convolution operation. After that, both the input and the feature map obtained after the convolution operation are divided into two groups channel-wise. In the next step of operation, group-1 of the lower layer information is concatenated with group-2 of the higher layer feature map. Similarly, group-1 of higher layer information is also concatenated with group-2 of lower layer data. In this way, the SCR block with two concatenated layers forms two parallel paths and gives the advantage of skip connection with the enhanced features by sharing the channel between two layers. The proposed SCR network is different from previous work (Zhang *et al.*, 2017), (Ma *et al.*, 2018), (Yang, 2021), (Yang *et al.*, 2022), (Kang *et al.*, 2017), (Kumawat *et al.*, 2022), which shuffles the channel within a layer, whereas the SCR block shares the information between two different layers.

The resultant feature of the SCR block is the combination of lower-order features and higher-order features. Further, the SCR block allows to learn feature maps associated with different versions of the input with two parallel paths and one skip connection. Following are the advantages of the proposed SCR block over the other related existing methods. (1) The proposed method is scalable since the feature map can be split into any number of groups and also it can be shared with any higher-layer feature map. The architecture, in this study, is designed especially for grading of RCC histopathology application. (2) SCR block can be incorporated into any other deep learning architecture to improve accuracy. (3) Information sharing between different layers, a part of the previous layer data is used as a skip connection which strengthens the local semantic features at different stages of the network. (4) There is no additional computational complexity involved in sharing the information between two different layers. In Fig.5.2, the input feature map X is transformed to Y by convolution operation, expressed in equation (5.1), Where $F_{lr(3 \times 3)}$ is a convolutional

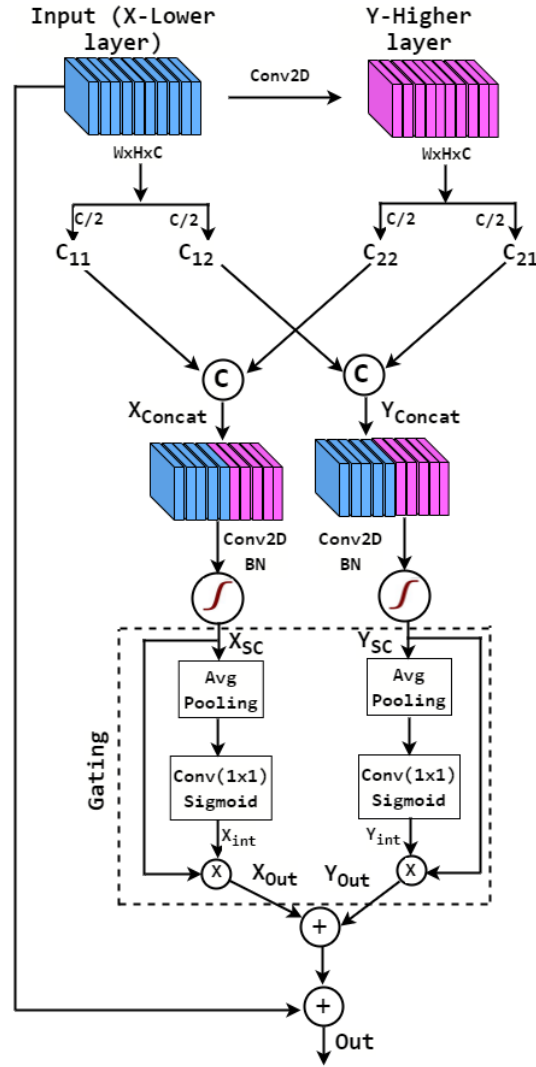


Figure 5.2: Shared channel residual block

operator of kernel size (3×3) .

$$X \in \mathbb{R}^{H \times W \times C} \xrightarrow{F_{tr(3 \times 3)}} Y \in \mathbb{R}^{H \times W \times C} \quad (5.1)$$

The kernel $K = [k_1, k_2, k_3, \dots, k_C]$ is an optimizable feature extractor, applied at each channel and each image position. k_c refers to the C^{th} kernel of K . Here $C = 16, 32$, and 64 channels are used at three different stages of the network. Y is the result of convolution

operation (*) of input feature map X with kernel K represented in equation (5.2).

$$Y = [y_1, y_2, \dots, y_C] = X * K \quad (5.2)$$

equation (5.3) represents convolution of a single feature map in X .

$$y_c = k_c * X = \sum_{s=1}^C k_c^s * x^s \quad (5.3)$$

Where $k^c = [k_c^1, k_c^2, \dots, k_c^C]$, $X = [x^1, x^2, x^3, \dots, x^C]$, $y_c \in \mathbb{R}^{H \times W}$, and k_c^s is a single channel of k_c .

Now X and Y are divided into $C_{11}, C_{12}, C_{21}, C_{22}$ channel-wise, where $C_{11}, C_{12}, C_{21}, C_{22} \in \mathbb{R}^{H \times W \times C/2}$ are shown below in equations (5.4) to (5.7).

$$C_{11} = [x^1, x^2, x^3, \dots, x^{\frac{C}{2}}] \quad (5.4)$$

$$C_{12} = [x^{\frac{C}{2}+1}, x^{\frac{C}{2}+2}, x^{\frac{C}{2}+3}, \dots, x^C] \quad (5.5)$$

$$C_{21} = [y^1, y^2, y^3, \dots, y^{\frac{C}{2}}] \quad (5.6)$$

$$C_{22} = [y^{\frac{C}{2}+1}, y^{\frac{C}{2}+2}, y^{\frac{C}{2}+3}, \dots, y^C] \quad (5.7)$$

The lower order features and deeper layer features are shared and concatenated at three different stages of the network and represented using the below-mentioned equations (5.8), and (5.9), Where \odot represents concatenation operation.

$$X_{Concat} = C_{11} \odot C_{22} = [x^1, x^2, \dots, x^{\frac{C}{2}}, y^{\frac{C}{2}+1}, \dots, y^C] \quad (5.8)$$

$$Y_{Concat} = C_{11} \odot C_{22} = [y^1, y^2, \dots, y^{\frac{C}{2}}, x^{\frac{C}{2}+1}, \dots, x^C] \quad (5.9)$$

The concatenated features are fed to a sequence of operations called Conv2D, Batch Normalization (BN), and Rectified Linear Unit (ReLU) and presented in equations (5.10), (5.11).

$$X_{SC} = ReLU [F_{tr(3 \times 3)} (X_{Concat})] \quad (5.10)$$

$$Y_{SC} = ReLU [F_{tr(3 \times 3)} (Y_{Concat})] \quad (5.11)$$

A kind of gating mechanism is also deployed effectively and uniquely in the SCR block to focus on the most relevant morphological features of RCC and enable contextual information for the network while producing predictions. X_{SC} is given as input to average pooling where the kernel used is $(H \times W)$, so that it is possible to extract the global information of each channel. The resulting feature is convoluted point-wise followed by sigmoid activation. The sigmoid activation assigns scores to each channel according to the relevance of global content associated with the channels. This way, it is giving attention to highly variable nuclear features present in different grades of kidney histopathology images. The extracted global feature calculated from equations (5.12) and (5.13) is multiplied by input gating signal X_{SC} and Y_{SC} respectively, and represented in equations (5.14) and (5.15).

$$X_{int} = \sigma [f_{tr(1 \times 1)} [Avg. Pool_{H \times W} \{X_{SC}\}]] \in \mathbb{R}^{1 \times 1 \times C} \quad (5.12)$$

$$Y_{int} = \sigma [f_{tr(1 \times 1)} [Avg. Pool_{H \times W} \{Y_{SC}\}]] \in \mathbb{R}^{1 \times 1 \times C} \quad (5.13)$$

$$X_{out} = X_{int} \times X_{SC} \quad (5.14)$$

$$Y_{out} = Y_{int} \times Y_{SC} \quad (5.15)$$

The final output of the SCR block expressed in equation (5.16) is obtained by adding original input X to the output generated by the gating mechanism.

$$Out = X_{out} + Y_{out} + Input(X) \quad (5.16)$$

5.3 Implementation and Training

Section 5.3.1 describes the newly introduced dataset of five categories namely Normal/Non-cancerous (Grade-0), Grade-1, Grade-2, Grade-3, and Grade-4. To show the proposed model is generalized and independent of the dataset, it has experimented with one additional well-established data called BreakHis dataset for eight class-classification and it is mentioned in Section 5.3.1. Training details, a list of important hyper-parameters, and quality metrics used for grading purpose is described in Section 5.3.2.

5.3.1 Novel RCC grading dataset

The introduced KMC kidney histopathology dataset includes non-cancerous (Grade-0) and cancerous (Grade-1 to Grade-4) images of the Renal Clear Cell Carcinoma. These images were collected from October 2020 to December 2022 as a part of a clinical study at Department of Pathology, Kasturba Medical College (KMC) Mangalore, Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India. This project was approved by the institutional ethics committee Kasturba Medical College (KMC), Mangalore, protocol no-IEC KMC MLR 02/2022/57. The conducted research reported in this article involving human participants was in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Regarding data, informed consent from all the patients was obtained for conducting the experiments, and the personal details were protected.

The samples were collected by surgical (open) biopsy (SOB) of kidney tissue, stained with hematoxylin and eosin (H&E). These histopathology slides are classified into Normal or Non-cancerous (Grade-0), Grade-1, Grade-2, Grade-3, Grade-4 and labeled through a careful clinical study in the department of pathology in KMC. The procedure used by the pathologists here is the most commonly used paraffin procedure. The complete preparation procedure includes steps such as fixation, dehydration, clearing, infiltration, embedding, and trimming (Mescher, 2013). After the preparation process, the pathologist visually identified the tumoral areas in each slide under a microscope. An olympus BX-50 system microscope with a relay lens and a magnification of 3.3X coupled to a olympus digital color camera DP-22 is used to obtain digitized images from the kidney tissue slides. This camera uses a 1/1.8" colour CCD (Charge-Coupled Device) with pixel size $3.69\mu\text{m}\times 3.69\mu\text{m}$ and a total pixel number of $1920\times 1440\times 3$.

This dataset has been annotated by a group of pathologists at KMC, Manipal. Pre-setting was kept automatic and except for the selection of area, everything was the same for all images. Non-overlapping square patches were extracted and resized to a size of 224×224 and normalized to zero mean and unit variance before passing them through deep learning architecture for training and testing. The entire dataset is splitted into training and test sets. Approximately 80% of the total images were considered for training, while the rest 20% for the test. The validation set was created by cropping non-overlapping regions of training patches. The data augmentation techniques, horizontal and vertical flipping were used on training set to further increase the diversity. After random cropping and data augmentation, 3442 patches belonging to five classes were obtained as training set. Out of these, 693

patches are of Normal/Non-cancerous grade, 708 patches belong to grade-1 cancer, 648 patches belong to grade-2 cancer, 735 patches belong to grade-3 and, 648 patches belong to grade-4 kidney cancer. Data augmentation on the test set has not been applied and model evaluation performed on an original test set. The details of the grade distribution of the proposed data are presented in Table 5.1.

Table 5.1: Introduced dataset grades distribution

Type	Training Patches	Test Patches
Grade0	162	41
Grade1	108	27
Grade2	99	25
Grade3	96	24
Grade4	112	28
Total	577	145

BreaKHis dataset: To verify the validity of the proposed RCCGNet on the proposed novel dataset, we also analyzed the performance of the RCCGNet on the BreaKHis dataset. BreaKHis dataset (Togacar *et al.*, 2020) is a well-established publicly available breast cancer histopathology dataset used in various state-of-the-art deep learning models. BreaKHis is also an imbalanced dataset. To overcome the data imbalance problem, data augmentation techniques of horizontal flip, vertical flip, and rotation were performed on training and test sets. For the BreaKHis dataset, the details of class distribution of original training and test patches and after data augmentation are shown in Table 5.2.

5.3.2 Training details

The proposed RCCGNet and all comparison models were implemented on a Dell-G4G3NSM workstation with 8 GB NVIDIA QUADRO P4000 GPU and 64 GB RAM. All of the training processes performed on Python-3, TensorFlow 2.4.0, Keras an open source platform. Categorical-Crossentropy (Togacar *et al.*, 2020), (Aatresh *et al.*, 2021a) is used to optimize the loss and the softmax classifier (Jiang *et al.*, 2019) at the final layer completes the classification process. Experimentation of the proposed RCCGNet consists of three stages: training, validation, and testing. For a given range of hyperparameters, the network was trained and measured the performance based on the validation and test sets. The proposed RCCGNet model is more efficient with the Adam optimizer (Kingma and Ba, 2014), with

Table 5.2: BreaKHis dataset at 40X magnification

Class	Original Training Patches	Original Test Patches	Augmented Training Images	Augmented Test Images
Adenosis	100	16	298	34
Fibroadenoma	219	35	657	90
Phyllodes Tumor	105	19	287	39
Tubular Adenoma	128	23	380	60
Ductal Carcinoma	752	112	2256	267
Lobular Carcinoma	138	18	414	45
Mucinous Carcinoma	174	31	522	69
Papillary Carcinoma	126	19	378	42
Total	1742	273	5192	646

Table 5.3: Details of hyper parameters

Optimizer	Adam with initial learning rate of 0.001
Batch size	4
Convolution filter	16, 32, 64
Total epochs	65
Reduce learning rate	Factor=0.5, patience=5, min-delta=0.0001, Monitor-Validation acc.
Early stop	Patience=30, min-delta=0.0001, Monitor-Validation acc., restore-best-weights

the default learning rate of 0.001. Three important parameters namely reduced learning rate, model checkpoint and early stopping were used for the effective training. Reduced learning rate helps to schedule the learning rate with the progress of training. A measure can be chosen so that we were able to monitor validation accuracy, validation loss, etc. For this purpose, we have monitored validation accuracy, and if it is not improving in 5 successive epochs with a minimum change of 0.0001, this callback reduces the current learning rate. The new learning rate will be half of the previous learning rate and this process continues till the last epoch. Another callback is model checkpoint, which helps in saving the model weight, which is the best fit for the data. Validation accuracy is continuously monitored and the model checkpoint saves the best weight on the provided file path. The early stopping method helped to decide the total number of epochs that we have trained the model for given data. Less training probably leads to underfitting, while excessive training may cause

overfitting. After 65 epochs there is no improvement in performance has been observed for all reference models. In this way, all reference models and proposed model were trained for 65 epochs. Another hyper parameter is batch size, larger batch size allows to parallelize computations to a greater degree but lead to poor generalization. The proposed model giving satisfactory results for batch size 4. The important controlling hyperparameter that has been tuned is shown in Table 5.3. Quality matrices used in this study are *Precision* (Tabibu *et al.*, 2019), (Togacar *et al.*, 2020), *recall* (Aatresh *et al.*, 2021a), *F1-score* (Zhu *et al.*, 2021), and *Accuracy* Jiang *et al.* (2019).

5.4 Results and Discussions

The performance of the proposed RCCGNet is compared with the models that uses transfer learning approach as well as deep learning model trained from scratch. This study includes examination of most recent models through two different organ histopathology datasets called KMC kidney dataset and BreakHis dataset. The analysis of other variations of the proposed network is presented through an ablation study in Section 5.4.1. The quality metrics and the prediction of the proposed RCCGNet model are compared with other classification networks like ResNet50, InceptionResNetV2, NASNet, Shufflenet, BHCNet, BreastNet, LiverNet, and Vision Transformer in Section 5.4.2. The performance of networks leveraged by transfer learning of pre-trained weights from the ImageNet dataset is also included. Three-fold cross-validation, statistical analysis, and computational complexity analysis are also presented in the below Sections.

5.4.1 Ablation study

The core idea of RCCGNet lies in the proposed shared channel residual (SCR) block. The SCR block shares the channel with the next higher-order layer. The proposed model contains two paths of shared information with a simple gating mechanism, residual connections, convolutional, and fully connected layers. To measure the effectiveness of the main components of RCCGNet, these components are detached from the proposed model and made different variations of proposed model. These variations were trained in the same environment. Visual performance of each variation using intermediate features and activation map was compared with the proposed RCCGNet where all the important modules jointly worked. The overall test accuracy, and other quality metrics of different versions

on the KMC kidney dataset and BreakHis dataset are listed in Table 5.4 and Table 5.5 respectively. Comparison of computational complexity of different variations was expressed in terms of the total number of parameters and floating-point operations (FLOPs). Comparison of intermediate features with and without channel sharing between layers for KMC kidney dataset is shown in Fig. 5.3 . Visual comparison of different CNN variations using activation map for KMC kidney dataset is shown in Fig. 5.4.

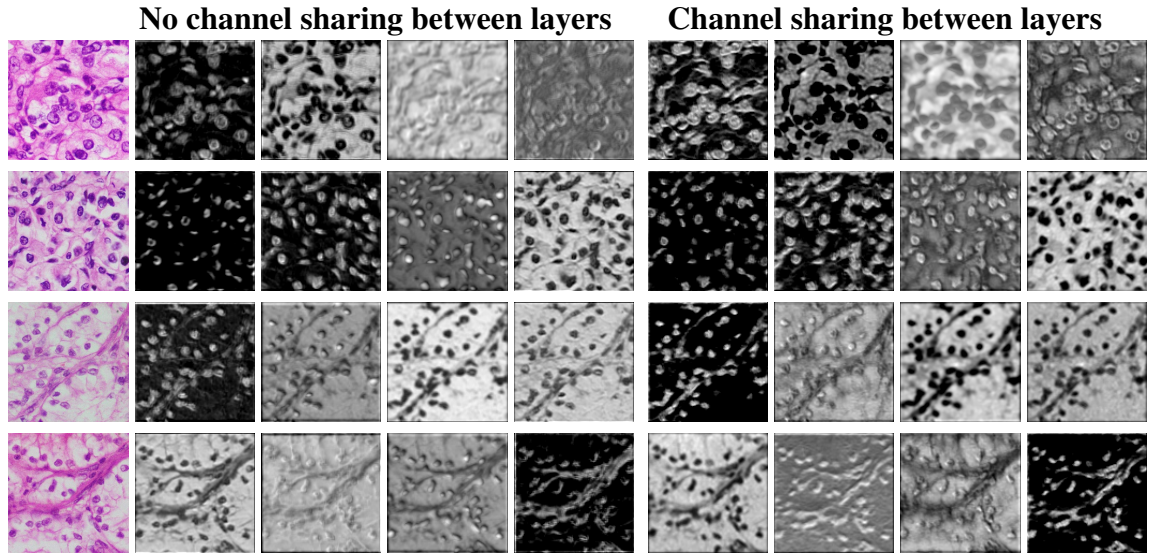


Figure 5.3: Comparison of intermediate features with and without channel sharing between layers

Table 5.4: Overall quality comparison of three variants of proposed RCCGNet on new KMC Kidney dataset

Model Variations	Channel sharing	Gating mechanism	Proposed SCR block	Conv2D, FC layers	Precision	Recall	F1 Score	ACC	Parameters	FLOPs (G = 10 ⁹)
Variation-1	×	✓	✓	✓	0.9047	0.9009	0.9014	0.9032	365,125	4.48G
Variation-2	✓	×	✓	✓	0.8886	0.8844	0.8843	0.8864	343,173	4.47G
Variation-3	×	×	×	✓	0.8962	0.8838	0.8840	0.8873	147,397	1.68G
RCCGNet	✓	✓	✓	✓	0.9333	0.9216	0.9229	0.9249	365,125	4.48G

SCR block without channel sharing: To verify the validity of the channel sharing method in the proposed SCR block, the extracted intermediate features of the proposed model, where the information is shared with the higher-order layer and a version of the proposed model which does not share the information between layers were analyzed and compared.

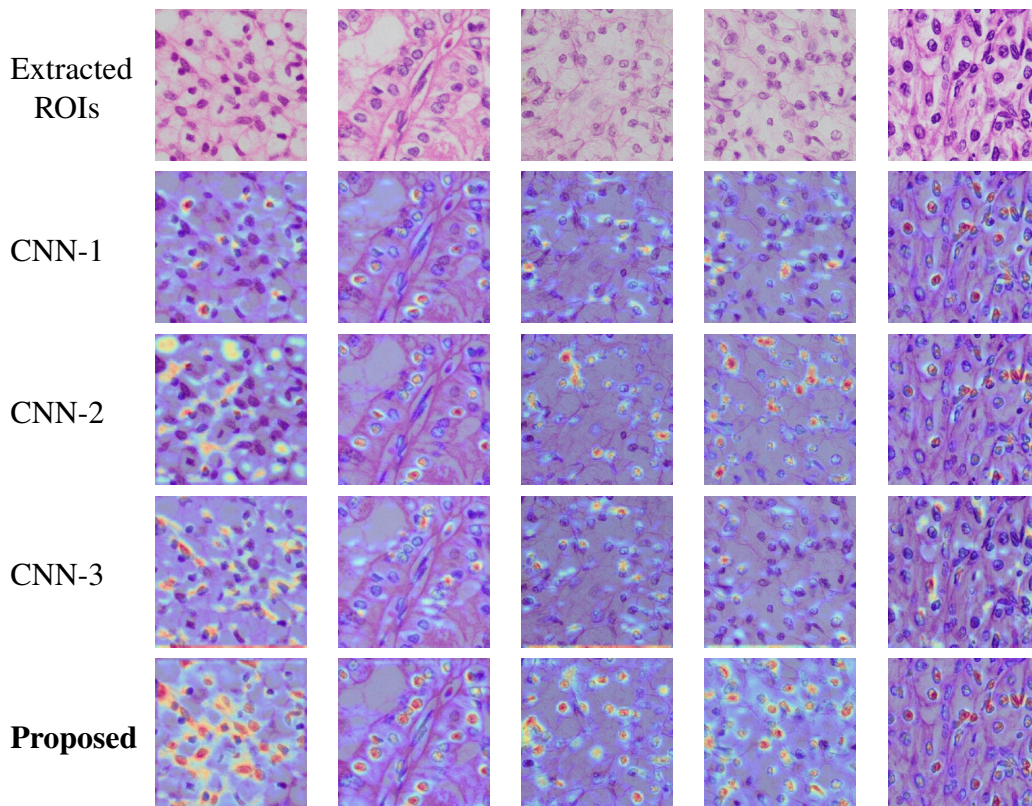


Figure 5.4: Visual comparison of different CNN variations using activation maps. (Red: Very high probability score regions, Orange: Medium probability score regions, Light blue: Low probability score regions.)

Table 5.5: Overall quality comparison of three variants of proposed RCCGNet on BreakHis dataset

Model Variations	Channel sharing	Gating mechanism	Proposed SCR block	Conv2D, FC layers	Precision	Recall	F1 Score	ACC	Parameters	FLOPs (G = 10 ⁹)
Variation-1	×	✓	✓	✓	0.8799	0.8588	0.8649	0.8839	365,125	4.48G
Variation-2	✓	×	✓	✓	0.8919	0.8578	0.8725	0.8792	343,173	4.47G
Variation-3	×	×	×	✓	0.7849	0.7562	0.7644	0.7987	147,397	1.68G
RCCGNet	✓	✓	✓	✓	0.9062	0.8842	0.8890	0.9009	365,125	4.48G

The second stage of both CNN versions produce 32 feature maps, out of which 4 feature maps of both cases are analyzed for four different images of different grades. The effect of channel sharing between layers can be observed in Fig. 5.3. Proposed SCR produces sharp and clear nuclear boundaries, which helps the deep learning model to predict the associated grades. The visuals of intermediate features also indicated that the proposed

SCR block better segregates the object and background region. Without channel sharing, 2.3% drop in test accuracy were observed. Further, a drop in the value of F1 score, recall, and precision was also observed, which are 2.3%, 2.24%, and 3.06% respectively. From these observations, it is clear that in the proposed SCR block, inter-channel information sharing boosts the grading scores and effective utilization of this module in the proposed work was found through this ablation study.

Effect of gating mechanism: Attention block helps the model to separate the most relevant nuclear region from background regions. The effect of removing the attention block can be seen in Fig. 5.4 of RCCGNet variation CNN-2. It also improves the network performance as the proposed RCCGNet produced higher values for both the datasets shown in Table 5.4 and Table 5.5. Thus, the contribution of the attention block can not be ignored and it's imperative to include it in the proposed new block.

Comparison of activation maps of different variations: The proposed model is composed of several blocks. This part demonstrates the extracted features of the aforementioned blocks using activation maps produced with the help of gradient-weighted class activation mapping (Grad-CAM). This activation map is generated at the final stage of the feature extraction part of the model. The classification model in deep learning assigns a probability score to the final stage of the classifier based on extracted critical features. An activation map is a simple and straight-forward visualization technique to show the assigned probability to the relevant regions. How well each variation of the proposed model segregate the tissue regions from background pixels can be visualized by using activation maps shown in Fig. 5.4. Activation maps produced in CNN-1, where the model is not sharing inter-channel information, high probability regions are less, and some critical nuclear regions are missed to assign any probability. In CNN-2, where the gating mechanism is absent, it assigns some probability scores to irrelevant regions for some grades, which is not desirable. False-positive cases are maximum in CNN-2 compared to any other variations. Total number of detected nuclei are more in CNN-2 compared to CNN-1. In CNN-3, the complete SCR block is absent. The effect of removing the proposed SCR block can be observed in CNN-3 variation. Only few tissue regions get identified in the CNN-3 variation and the total detected nuclei are less compared to any other variations. Many relevant nuclear regions are assigned by lesser probability scores which results the number of partially detected nuclei being more in CNN-3. Whereas the proposed RCCGNet, where all the important components jointly worked, segregates tissue regions from background pixels better than compared to all other versions. The number of not detected and partially detected nuclei are

very less in number in the proposed RCCGNet. In the RCCGNet, most of the tissue regions are clearly identified with high probability scores for all grades of images.

Table 5.4, Fig. 5.3, and Fig. 5.4 shows that the proposed RCCGNet is the best performing combination among all modifications made in the base model. This ablation study helps us to choose the best performing combination among all variations. The three important components togetherly making the proposed RCCGNet very effective in feature extraction with minimum computational complexity.

Table 5.6: Performance metrics comparison of proposed model with other competitive models on KMC kidney dataset

Metrics	Grade	Transfer learning approach			End-to-end trained deep learning networks					
		ResNet50 (2016)	IncResV2 (2016)	NASNet (2018)	ShuffleNet (2018)	BHCNet (2019)	BreastNet (2020)	LiverNet (2021)	ViT (2021)	RCCGNet (Proposed)
Precision	0	0.8478	0.8260	0.9523	1	0.9743	0.9210	0.9523	0.9743	0.9756
	1	0.8076	0.8095	0.8571	0.9523	0.8181	0.9230	0.92	0.8260	0.8846
	2	0.5937	0.6	0.7666	0.6451	0.75	0.6666	0.6896	0.8421	0.9047
	3	0.7272	0.75	0.4827	0.6	0.75	0.8333	0.7142	0.5666	0.7241
	4	0.7037	0.6071	0.9259	0.96	0.96	1	1	0.8709	1
	Overall	0.7360	0.7185	0.7969	0.8315	0.8508	0.8688	0.8552	0.8160	0.8978
Recall	0	0.9512	0.9268	0.9756	0.9756	0.9268	0.8536	0.9756	0.9268	0.9756
	1	0.7777	0.6296	0.4444	0.7407	1	0.8888	0.8518	0.7037	0.8518
	2	0.76	0.84	0.92	0.8	0.84	0.96	0.8	0.64	0.76
	3	0.3809	0.4285	0.6666	0.7142	0.5714	0.7142	0.7142	0.8095	1
	4	0.6785	0.6071	0.8928	0.8571	0.8928	0.8571	0.8928	0.9642	0.8928
	Overall	0.7097	0.6864	0.7799	0.8175	0.8462	0.8547	0.8469	0.8088	0.89606
F1 Score	0	0.8965	0.8735	0.9638	0.9876	0.9500	0.8860	0.9638	0.9500	0.9756
	1	0.7924	0.7083	0.5853	0.8333	0.9	0.9056	0.8846	0.76	0.8679
	2	0.6666	0.7	0.8363	0.7142	0.7924	0.7868	0.7407	0.7272	0.8260
	3	0.5	0.5454	0.56	0.6521	0.6486	0.7692	0.7142	0.6666	0.84
	4	0.690	0.6071	0.9090	0.9056	0.9259	0.9230	0.9433	0.9152	0.9433
	Overall	0.7093	0.6869	0.7709	0.8186	0.8434	0.8541	0.8493	0.8038	0.8906
Accuracy	0	0.9366	0.9225	0.9788	0.9929	0.9718	0.9366	0.9788	0.9718	0.9859
	1	0.9225	0.9014	0.8802	0.9436	0.9577	0.9647	0.9577	0.9154	0.9507
	2	0.8661	0.8732	0.9366	0.8873	0.9225	0.9084	0.9014	0.9154	0.9436
	3	0.8873	0.8943	0.8450	0.8873	0.9084	0.9366	0.9154	0.8802	0.9436
	4	0.8802	0.8450	0.9647	0.9647	0.9718	0.9718	0.9788	0.9648	0.9788
	Overall	0.7464	0.7183	0.8028	0.8380	0.8661	0.8591	0.8662	0.8239	0.9014

5.4.2 Comparison with state-of-the-art models

The performance of the proposed RCCGNet is compared with the models that uses transfer learning approach as well as deep learning model trained from scratch. Transfer learning methods applies the pre-trained weights of ImageNet dataset. This study includes examination of most recent models through two different histopathology datasets called KMC kidney dataset and BreakHis dataset. ResNet50 network (2016) (He, 2016); Inception-ResNetV2 (2016) (Szegedy *et al.*, 2016); and NASNet (2018) (Zoph *et al.*, 2017); utilizes

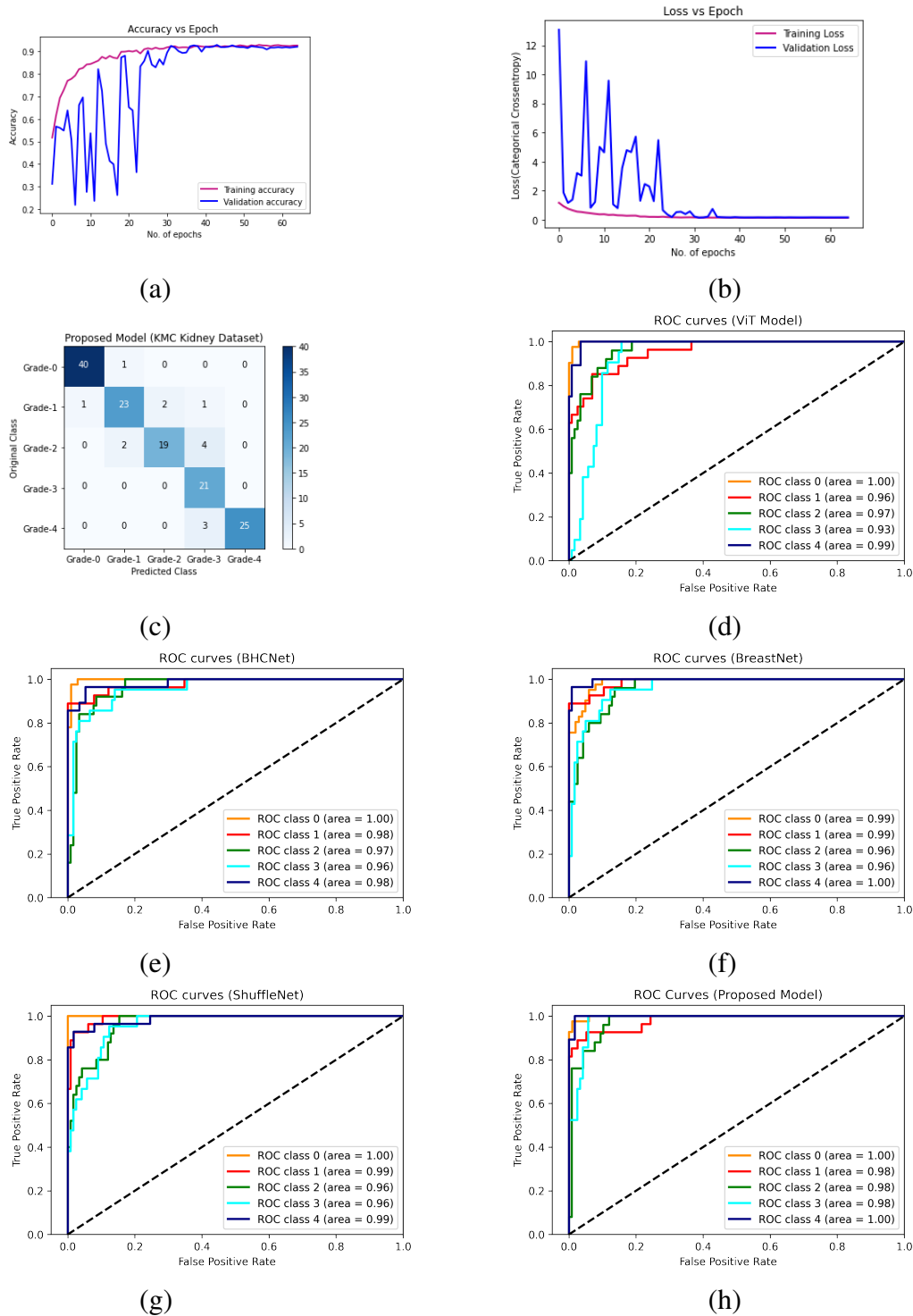


Figure 5.5: Learning curve of RCCGNet on introduced KMC kidney dataset (a) Training and validation accuracy of the proposed model (b) Training and validation loss of proposed model (c) Confusion matrix of proposed model (d)-(h) Receiver operating characteristic (ROC) curve of five best-performing state-of-the-art models using one versus rest approach.

transfer learning approach, where all the layers are frozen except the few top layers. A global average pooling layer, a dense layer, and a softmax layer were added in place of the final stage of the above models and used to fine-tune the model. The comparison table also includes models, Shufflenet (2018) (Zhang *et al.*, 2017), BHCNet (2019) (Jiang *et al.*, 2019), BreastNet (2020) (Togacar *et al.*, 2020), and LiverNet (2021) (Aatresh *et al.*, 2021a), which are end-to-end trained deep learning networks. The comparison table contains a Vision Transformer (2021) (Dosovitskiy and Beyer *et al.*, 2021), where an image is interpreted as a sequence of patches and processed by a standard transformer encoder. Models trained from scratch does not have any pre-trained weights and all weights and biases were randomly initialized. Experimental results indicate that transferring ImageNet weights to classify the RCC images is not much beneficial. The data preparation phase, morphology, and histopathologic structure of RCC images are completely different from the ImageNet dataset. It would probably be more advantageous if it is possible to transfer the features of histopathology data. Embedding efficient and lightweight CNN module as in (Jiang *et al.*, 2019), (Togacar *et al.*, 2020), (Aatresh *et al.*, 2021a) gained satisfactory improvement over transfer learning techniques. Table-5.6 shows performance of different models for grading kidney tissues from KMC kidney histopathology dataset.

Table 5.7: Overall quality comparison of the proposed model with other competitive models (BreakHis dataset-Eight Class)

Metrics	Transfer learning approach			End-to-end trained deep learning networks					RCCGNet (Proposed)
	ResNet50 (2016)	IncResV2 (2016)	NASNet (2018)	ShuffleNet (2018)	BHCNet (2019)	BreastNet (2020)	LiverNet (2021)	ViT (2021)	
Precision	0.7309	0.7678	0.7147	0.7897	0.8671	0.7868	0.8361	0.7415	0.9062
Recall	0.6726	0.7642	0.7362	0.7507	0.8446	0.7313	0.7979	0.7083	0.8842
F1 Score	0.6954	0.7549	0.7026	0.7579	0.8504	0.7492	0.8156	0.7170	0.8890
Accuracy	0.7384	0.7879	0.7755	0.8018	0.8823	0.7988	0.8529	0.7956	0.9009

To avoid any biased advantage we used a diverse set of quality metrics which is *precision*, *recall*, *F1 score*, and *accuracy* which includes class-wise and the overall score of the model. Overall accuracy is not the average of class-wise accuracy, it is model accuracy based on total TP, TN, FP, and FN instances belonging to all classes. These values are calculated using scikit-learn library in Python and it is possible that the overall metric is lower than the lowest metric of an individual class. For KMC kidney dataset the proposed RCCGNet achieve 90.14% classification *accuracy* which is the best among any comparison model. *Precision*, *recall*, and *F1-score* values of the proposed RCCGNet are 89.78%, 89.60%, and 89.06% respectively. Resnet-50 and InceptionResNetV2 utilizes

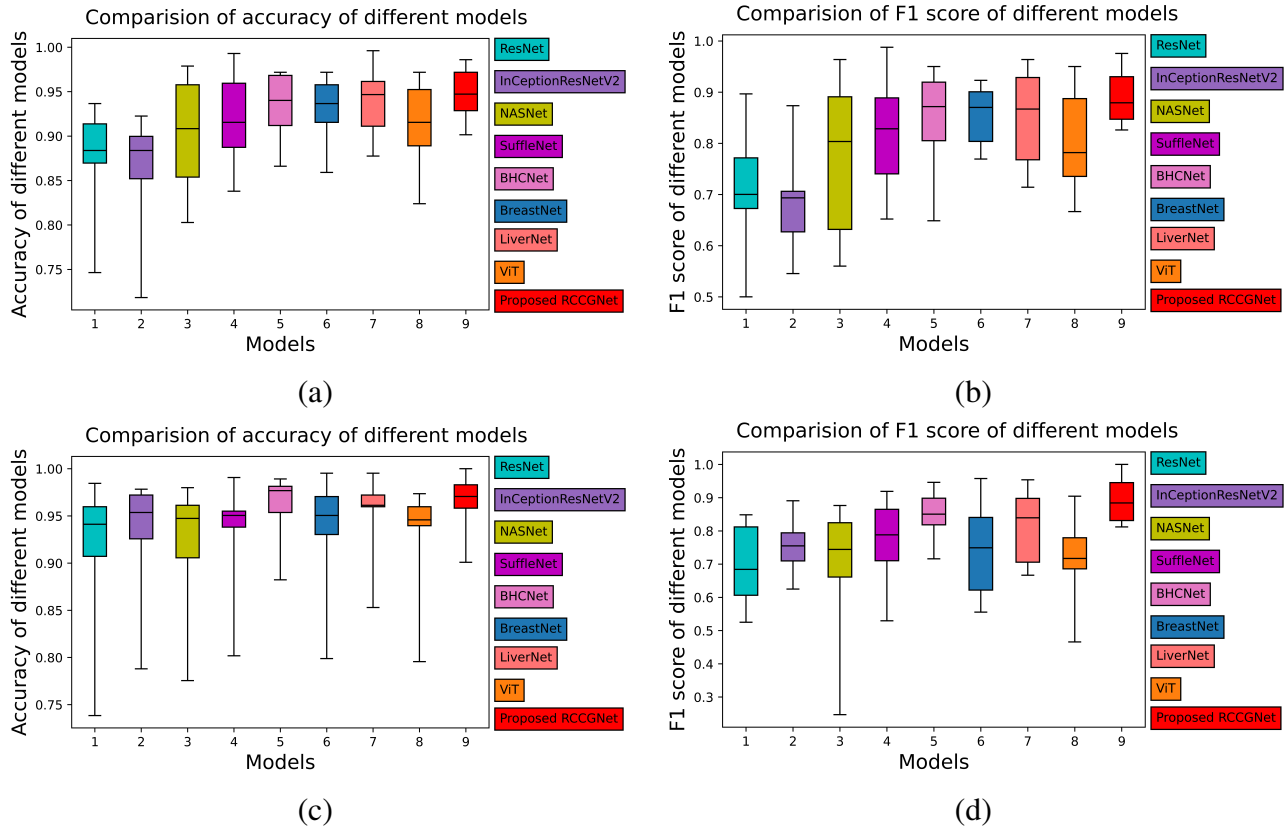


Figure 5.6: Box Plot: (a), (b) Accuracy, and F1 Score of different models for new KMC kidney dataset. (c), and (d) Accuracy and F1 Score of different models for BreakHis dataset.

transfer learning technique show poor performance among all comparison models. Fig.5.5 shows the learning curves of the proposed RCCGNet. In Fig.5.5 (a) and (b), training curves and validation curves are very close to each other. This indicates training data best represents the validation data. The model which has training and validation curves closer to each other is robust during testing. The performance of a classifier can also be described by the confusion matrix itself. Fig.5.5 (c) is the classification report of the proposed model expressed in terms of the confusion matrix. Out of 142 instances in the test set proposed model predicts 128 cases correctly which is the maximum compared to any reference models. In Grade-0 and Grade-3 case there is minimum misclassified samples in the proposed model.

Receiver operating characteristic (ROC) curve and area under the curve (AUC) indicate how much a classification model is capable to distinguish different grades. It is possible that some model which is not the best-performing model can perform better for a particular

class. The average AUC of the proposed model is best compared to all reference models. The ROC-AUC curve of five best-performing state-of-the-art models using the one versus rest approach is shown in Fig.5.5 (d) to (h). To verify the validity of RCCGNet on the proposed kidney KMC dataset, experiments were conducted on a well-established BreKHis histopathology dataset having eight classes, including all the breast cancer subcategories. Table-5.7 shows the performance of different models for the classification of breast tissues from BreKHis histopathology dataset. For the BreKHis dataset, the proposed RCCGNet achieve 90.09% classification *accuracy*. *Precision*, *recall*, and *F1-score* values of the proposed RCCGNet are 90.62%, 88.42%, and 88.90% respectively. On BreKHis dataset also the proposed RCCGNet is the best-performing model, this suggests that the proposed model is generalized independent of the dataset.

5.4.3 Cross-validation

To avoid any possible overfitting issue the proposed model and all reference models are tested using 3-fold cross-validation. Three folds of each class image were made to accomplish this. Models are trained for the first run using fold 2 and fold 3 and tested using fold 1. Models are trained in the second run using folds 1 and fold 3 and tested in fold 2. Similarly, for third run the models are trained using fold 1 and fold 2 and tested using fold 3. Table 5.8 lists the average outcomes of all three runs for the proposed model and the benchmark models. The suggested model produces nearly identical results to those shown in Table 5.6. ResNet (He, 2016) and ShuffleNet (Zhang *et al.*, 2017) show considerable variations in the result of Table 5.6 and Table 5.8.

5.4.4 Statistical analysis

Depicting data using box plot is a standard method where it distribute the whole data points between five regions. It is a very good measure to check the dispersion of data between minimum to maximum via first quartile, median, and second quartile. *Accuracy*, and *F1 score* values of different models on both datasets is presented using box plot in Fig.5.6. Each of the boxes in the box plot contains a grade-wise score and overall score of a model. Performance variability of the grade-wise and overall score depends on the statistical distribution of values obtained by all classification models for both datasets. Box plot shows that two metrics, *accuracy*, and *F1 score* of RCCGNet, the median value, first quartile value, and third quartile value are higher compared to the reference models.

Table 5.8: 3-Fold cross validation average quality metrics comparison of proposed RCCGNet with other competitive models (KMC kidney dataset)

Metrics	Grade	Transfer learning approach			End-to-end trained deep learning networks					
		ResNet50 (2016)	IncResV2 (2016)	NASNet (2018)	ShuffleNet (2018)	BHCNet (2019)	BreastNet (2020)	LiverNet (2021)	ViT (2021)	RCCGNet (Proposed)
Precision	0	0.7917	0.8754	0.8891	0.8360	0.8963	0.9522	0.9345	0.8974	0.9390
	1	0.6759	0.7307	0.6851	0.6919	0.8017	0.8241	0.8511	0.7606	0.9320
	2	0.6795	0.6635	0.7829	0.7921	0.8840	0.7792	0.8305	0.8185	0.9361
	3	0.6755	0.6215	0.7057	0.6499	0.8414	0.7588	0.8033	0.6980	0.8933
	4	0.7369	0.7955	0.8904	0.7668	0.9390	0.8167	0.8765	0.8233	0.9475
	Overall	0.7119	0.7373	0.7906	0.7473	0.8725	0.8262	0.8592	0.7996	0.9296
Recall	0	0.8475	0.8124	0.8621	0.8816	0.9803	0.8913	0.9018	0.9014	0.9607
	1	0.6814	0.7555	0.7925	0.6962	0.8592	0.7925	0.8148	0.7703	0.8888
	2	0.6920	0.7265	0.6968	0.7333	0.7884	0.8031	0.9115	0.7904	0.9357
	3	0.4583	0.575	0.6833	0.55	0.8	0.7333	0.7916	0.7083	0.9416
	4	0.7930	0.7064	0.8711	0.8432	0.8644	0.9007	0.8646	0.8146	0.8931
	Overall	0.6944	0.7152	0.7812	0.7408	0.8585	0.8242	0.8569	0.7970	0.9240
F1 Score	0	0.8127	0.8414	0.8746	0.8569	0.9365	0.9203	0.9152	0.8993	0.9490
	1	0.6710	0.7355	0.7253	0.6935	0.8286	0.8053	0.8301	0.7645	0.9035
	2	0.6792	0.6537	0.7306	0.7612	0.8283	0.7858	0.8668	0.8037	0.9327
	3	0.5249	0.5710	0.6881	0.5952	0.8193	0.7430	0.7881	0.7016	0.9148
	4	0.7631	0.7428	0.8805	0.8002	0.9001	0.8491	0.8702	0.8171	0.9190
	Overall	0.6902	0.7089	0.7798	0.7415	0.8625	0.8207	0.8541	0.7972	0.9238
Accuracy	0	0.8877	0.9153	0.9308	0.9168	0.9626	0.9569	0.9544	0.9432	0.9709
	1	0.8754	0.8989	0.8824	0.8850	0.9335	0.9292	0.9376	0.9113	0.9653
	2	0.8892	0.8639	0.9184	0.9209	0.9445	0.9263	0.9513	0.9334	0.9763
	3	0.8725	0.8696	0.8988	0.8753	0.9418	0.9156	0.9294	0.9002	0.9709
	4	0.9044	0.9057	0.9542	0.9197	0.9626	0.9390	0.9502	0.9293	0.9695
	Overall	0.7146	0.7268	0.7924	0.7589	0.8726	0.8336	0.8615	0.8088	0.9265

Table 5.9: Computational complexity comparison of architectures

Model	Parameters (M = 10 ⁶)	FLOPs (G = 10 ⁹)
ResNet-50(2015)	23.60 M	7.75 G
InceptionResNetV2(2016)	54.34 M	13 G
NASNet (2017)	4.275 M	1.15 G
ShuffleNet (2018)	23.83 M	6.69 G
BHCNet(2019)	0.3034 M	4.46 G
BreastNet(2020)	0.6064 M	5.68 G
LiverNet(2021)	0.5740 M	3.72 G
ViT(2021)	45.26 M	33.03 G
Proposed RCCGNet	0.3651 M	4.48 G

5.4.5 Computational complexity analysis

The computational complexity of all the classification models is expressed in terms of total trainable parameters and floating-point operations (FLOPs). These values for the proposed RCCGNet and all reference models are shown in Table 5.9. The proposed RCCGNet uses 0.3651 million parameters which is the least among all reference models except BHCNet. BHCNet has a marginal difference in trainable parameters compared to the proposed RCCGNet and it uses 0.3034 million parameters. NASNet uses the least number of FLOPs.

InceptionResNetv2 uses the highest number of trainable parameters and FLOPs.

5.5 Proposed Architecture-2

Convolutional neural networks (CNNs) have gained a lot of popularity in recent years in all computer vision tasks including histopathology image analysis. A general CNN includes block with convolution layers followed by batch normalization and a non-linear activation. With f as input, h as kernel and A referring to activation function, convolution block can be explained efficiently by equations (5.17), (5.18).

$$g(x, y) = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} f[k, l]h[x - k, y - l] \quad (5.17)$$

$$Conv_features = A(Batch_Norm(g(x, y))) \quad (5.18)$$

Multiple of these blocks cascaded to each other helps in obtaining unique and distinguishing features in the feature map. A max-pooling layer follows a set of convolution blocks in the network. This layer concentrates on important features besides reducing the size of the feature map leading to faster convergence. The proposed architecture uses a SE-ResNet model by (Hu *et al.*, 2020) containing 14 layers as the base model. This architecture uses squeeze and excitation channel attention block cascaded with the convolution layers. Histopathology image-based cancer grading requires the algorithm to detect nucleus of various shapes and sizes. Therefore using a convolution block with a fixed kernel size (as used in SE-ResNet) won't be effective. To overcome this issue, instead of a simple ResNet block, the Inception ResNet module (Szegedy *et al.*, 2016) has been used. Inception ResNet block extract features using different kernel sizes, enabling the network to detect nucleus for all the grades. An additional inception ResNet block follows the SE-Inception ResNet module. This block enhances the features obtained from the channel attention unit. To determine the final probability for the classes a hyper-column network at the end of the model has been used. This block concatenates features from different parts of the network focusing on low-level features necessary for classification. The complete overview of the architecture is shown in the Fig. 5.7.

In this work, the suggested architecture is used for grading (Normal, Grade-1, Grade-2, Grade-3, and Grade-4) and classification (Normal, KIRC, KIRP, KICH) in kidney cancer tissue. These five different associated grades of KIRC and four subtypes of RCC are fed

to the deep learning network shown in the Fig.5.7. The input is first passed through a single convolution neural network with a kernel size of 7x7 to learn the low-level features. The obtained feature map iteratively passes through the SE-Inception-ResNet block and the Inception-ResNet block. After every three iterations, the feature map passes through a max-pooling layer with a kernel size of 3x3 and strides of 2. The kernel size chosen helps in maintaining correlation with the neighboring pixel during the pooling operation. The output of these pooling operations is passed through the hyper-column to obtain the final feature map. The final feature map obtained from the hyper-column passes through a global average pooling layer to get a vector of size 112. This feature vector passes through a dense layer of size 128, activated by the Relu activation function. To avoid over-fitting the model, a dropout layer with a dropping rate of 0.3 has been used. The final probabilities using a dense layer have been determined with the same size as the number of classes, activated by the softmax activation function. Given $x_i \in X$ with X being the final vector from dense layer and N being the total number of classes, equation (5.19) demonstrates the softmax activation function.

$$Softmax(x_i) = \frac{e^{-x_i}}{\sum_{j=0}^N e^{-x_j}} \quad (5.19)$$

The class with the maximum probability after softmax activation is the predicted class. Based on the input data provided, the proposed network produces five different associated grades of KIRC (Normal, Grade-1, Grade-2, Grade-3, and Grade-4) for the KMC kidney dataset, and for the TCGA dataset, it produces four subtypes (Normal, KIRC, KIRP, KICH) of RCC.

5.5.1 Inception-ResNet block

The inception block used in the proposed model is similar to InceptionNet (Szegedy *et al.*, 2016). Fig.5.8 provides a graphical overview of this block. This block has four sub-blocks operating in parallel to each other. The first sub-block involves a convolution using a kernel of size 1x1, enabling the network to learn channel-wise features from the input. The second sub-block contains two convolution operations cascaded to each other, with kernel sizes of 1x1 and 3x3 in that order. The third sub-block is similar to the second operation, with convolution using a kernel size of 5x5 instead of 3x3. These kernels of sizes 3 and 5 enable the network to learn features at different scales from the same input. Therefore the nucleus of different shapes and sizes for different cancer grades are appropriately detected. The

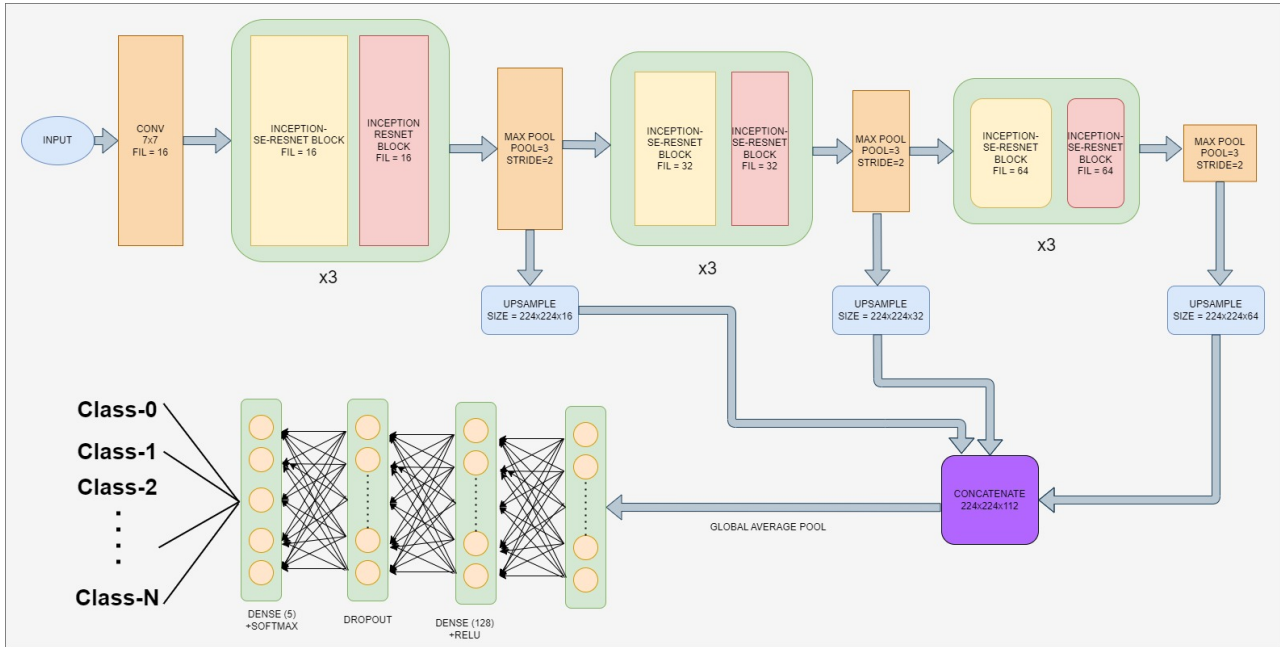


Figure 5.7: Complete overview of the proposed architecture

final sub-block performs a max-pooling operation followed by convolution with kernel size 1×1 . This sub-block provides attention to the sharpest features present in the input feature map. A batch normalization layer and a Relu activation succeed each convolution layer in the block. The convolution operations use one-fourth of the number of filters present in the input feature map. The concatenation of the outputs from all four sub-blocks creates the feature map of the inception network. To efficiently learn the identity function, the input to this feature map creating an Inception-ResNet block has been added. Using a fourth of the filters during the convolution drastically reduces the number of parameters and flops involved in this operation, despite increasing the number of layers. Hence it concluded that the Inception-ResNet block makes the model a lot more efficient than the simple ResNet block.

5.5.2 Squeeze and excitation block

In the SE-Inception-ResNet block, the feature map obtained from the cascaded Inception unit passes through a squeeze and excitation (SE) channel attention unit (Hu *et al.*, 2020). The input feature map is first converted into a vector using global average pooling operation. A dense layer with Relu activation squeezes the layer to one-eighth of its size. Another dense layer excites these squeezed features to the pooling vector size. This feature vector passes

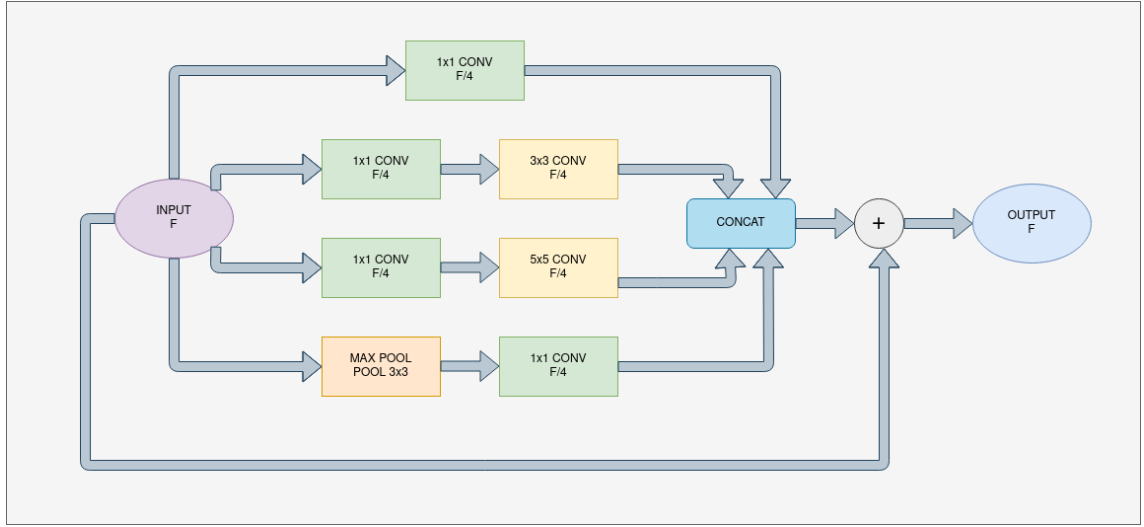


Figure 5.8: Inception-ResNet block

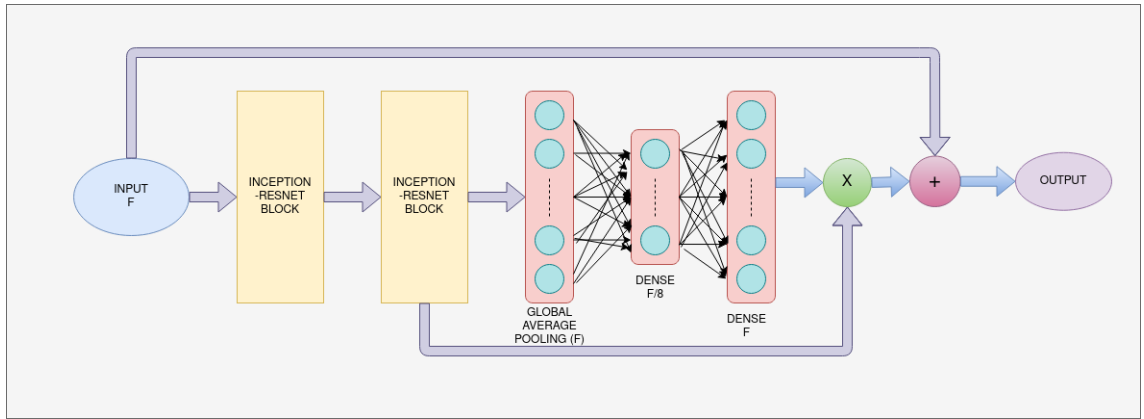


Figure 5.9: Inception-SE-ResNet block

through a sigmoid activation function to provide the final output of the SE block. Given $F \in R^{X \times Y \times Z}$ be the feature map from the Inception unit, $G \in R^{Z \times 1}$ be the global average pooling features, and $S \in R^{Z/8 \times 1}$ and $E \in R^{Z \times 1}$ be the squeeze and excitation features respectively. The equations (5.20), (5.21), and (5.22) explain the operation.

$$G(z) = \sum_{x \in X} \sum_{y \in Y} F(x, y, z) \quad (5.20)$$

$$S = Relu(W_s G + B_s) \quad (5.21)$$

$$E = Sigmoid(W_e S + B_e) \quad (5.22)$$

Where $W_s \in R^{Z/8 \times Z}$ and $B_s \in R^{Z/8 \times 1}$ are the weight and bias of the squeeze layer and $W_e \in R^{Z \times Z/8}$ and $B_e \in R^{Z \times 1}$ are the weight and bias of the excitation layer. The squeeze and excitation distinguish the filters in the feature map containing necessary features with the filters which add noise to the model. The sigmoid layer assigns a high value (around 1) to crucial filters and low values (around 0) to noisy filters. This channel attention is multiplied to the input feature map, suppressing the noisy filters and carrying forward only the necessary features. Fig. 5.9 provides a complete overview of the used Inception-SE-ResNet block.

5.5.3 Hyper-column

The final feature map for the model using the hypercolumn technique (Hariharan *et al.*, 2014) has been obtained. Feature maps from three different parts of the network have been taken, as shown in Fig. 5.7. The first map has 16 filters, the second has 32 filters and, the third has 64 filters. These feature maps are first Up-Sampled to the same height and width using transpose convolution (Dumoulin and Visin, 2016). The concatenation of the up-sampled feature maps along the depth axis provides the final feature map with 112 filters.

5.6 Implementation and Training

Experimentation of the proposed RoCNN consists of three stages: training, validation, and testing. For a given range of hyperparameters, the network was trained and measured the performance based on the validation and test sets. Section 5.6.1 describes the dataset used for the classification and grading of kidney histopathology images. In the Sections 5.6.2, and 5.7 discussed in detail procedure followed in conducting this experiment.

5.6.1 Dataset

The proposed model is evaluated on two different histopathology datasets. Dataset-1 called the KMC Kidney cancer dataset contains five different grades of renal clear cell carcinoma (KIRC) namely Normal, Grade-1, Grade-2, Grade-3, and Grade-4. Dataset-2 called the TCGA dataset contains four sub-types of Renal Cell Carcinoma (RCC) Called normal renal tissue (Non-cancerous), Renal Clear Cell Carcinoma (KIRC), Renal Papillary Cell Carcinoma (KIRP), and Kidney Chromophobe (KICH) histopathology images of the RCC. The total number of training and test patches of each grade have been presented in Table 5.1.

About 80% of the total images were used for training, and the remaining 20% were used for testing. To fine-tune and develop the model, the validation set was created by cropping non-overlapping regions of training patches. Since the dataset size is small, to increase the diversity further the training set goes through a series of data augmentation like horizontal flipping and vertical flipping. After data augmentation, 3432 training patches belonging to 5 classes for training has been obtained. There is no data augmentation performed on the test set and the test set belongs to different patient IDs such that the training set is completely independent of the test set.

We computed and analyzed our experimental results on another dataset acquired from The Cancer Genome Atlas (TCGA) data portal (<https://gdc.cancer.gov/>). The patches creation phase was carried out under the supervision of experienced pathologists and we cropped only those parts of WSI which have histologically defined clinical features. Created TCGA kidney dataset has a total of 4400 ROIs belonging to four classes of kidney tissue extracted from 679 different histopathological slides. All extracted ROIs are of size (1024x1024) and each of the four classes namely Normal, KIRC, KIRP, and KICH have an equal number of samples. We split the dataset before any data augmentation to avoid over-fitting issues and obtained 4400 patches that were slidewise distributed into training (85%), and testing (15%). The validation set was created by cropping non-overlapping regions of training patches. Results are evaluated on 632 test images. The classwise distribution of the TCGA dataset is presented in Table 5.10.

Table 5.10: TCGA dataset class distribution

Grade	Extracted ROIs for Training	Extracted ROIs for Test
Normal	942	158
KIRC	942	158
KIRP	942	158
KICH	942	158
Total	3768	632

5.6.2 Training details

The proposed model and all reference models run on the training and validation set, and evaluated the model’s effectiveness using the test set for a specified range of hyper parameters using a Dell-G4G3NSM workstation with 8 GB NVIDIA QUADRO P4000 GPU and 64 GB RAM. TensorFlow 2.4.0 and a Python-3 environment have been used. The proposed

RoCNN model is more efficient with the Adam optimizer (Kingma and Ba, 2014), with the default learning rate of 0.001. During the training process, the proposed model and all the comparison models fit the prepared training data. The validation data monitors the training process continuously and it is helpful to develop and fine-tune the models. All the models run for 75 iterations. The basis for choosing this number of iterations is that all the models saturate by 75 epochs on both datasets. The important controlling hyper parameter that has been tuned is shown in Table 5.11. Quality matrices used in this study are *Precision* (Tabibu *et al.*, 2019), (Togacar *et al.*, 2020), *recall* (Aatresh *et al.*, 2021a), *F1-score* (Zhu *et al.*, 2021), *Cohen’s Kappa coefficient* Maier-Hein and Reinke *et al.* (2022), and *accuracy* Jiang *et al.* (2019).

Table 5.11: Details of hyper parameters

Batch size	4
Optimizer	Adam ($\eta = 0.001$)
Dropout	0.3
Regularizers	L2 ($\lambda = 0.003$)
Convolution filter	16, 32, 64
Epochs	75
Reduce learning rate	Factor=0.5, patience=5, min-delta=0.0001

5.7 Results and Discussions

We trained the proposed model as well as all the comparison models on the training dataset. These trained models made their predictions on the test data, and we calculated five quality metrics to compare the performance of the models. These metrics include Precision, Recall, F1 score, Cohen’s Kappa coefficient, and Accuracy. For proper comparison, besides calculating the overall score for all these metrics, we also calculated the values for individual classes. The overall score is based on the total number of True Positive, True Negative, False Positive, and False Negative instances belonging to all classes. The overall score is not the average class-wise or grade-wise score. It is possible that the average score may be lower than the lowest score for a given class or grade. We experimented with two well-known histopathological datasets. The analysis of other variations of the proposed network is presented through an ablation study in Section 5.7.1. The proposed model has been compared with the models namely, MobileNetV2 (Sandler *et al.*, 2019), BreastNet (Togacar *et al.*, 2020), LiverNet (Aatresh *et al.*, 2021a), EfficientNetV2 (Tan and Le, 2021), Vision

Transformer (Dosovitskiy and Beyer et al., 2021), NFNet (Brock *et al.*, 2021), and Hameed et al.(Hameed *et al.*, 2022) on both multi-grade and multi-class datasets in Section 5.7.2. Computational complexity analysis of the proposed RoCNN and other benchmark models is presented in Section 5.7.4.

5.7.1 Ablation study

To prove the efficacy of the proposed model, an ablation study on the different variations of the proposed model has also been conducted. BHCResNet (Jiang *et al.*, 2019) proved to be the best-performing model because it performs better on the datasets used in this study and other multi-class datasets. Therefore we chose that to be our base model and made changes to improve the performance further. Table 5.12 gives an overview of all the modifications made in the base model and the results we got in each stage. To maintain uniformity, we followed the same methods for calculating results as used for calculating the results of other models. For each CNN, we took three trials and calculated the results on the test data. Table 5.12 reports the average result of three trials. CNN 1 in Table 5.12 describes the base model (BHCResNet (Jiang *et al.*, 2019)). This model attains an accuracy of 87.32% and an F1 score of 86.75% with around 300K parameters. In CNN2 (Table 5.12) we replaced the expensive cascaded convolution blocks used in SE-ResNet block with cascaded Inception-ResNet block, using reduced number of filters, creating a SE-Inception-ResNet block (Fig. 5.9). Inception-ResNet blocks are useful in extracting features at different layers, essential for detecting nuclei with different shapes and sizes. CNN2 (74.5K parameters) used less than one-fourth of the parameters used in CNN1, and the Accuracy and F1 score improved by 2% compared to CNN1. To extract the complete performance of the Inception block we placed an extra Inception-ResNet block after the SE-Inception block Fig.5.7, creating CNN3 (Table 5.12). CNN3 attained significant improvement of greater than 1% in both accuracy and F1 score compared to CNN2. To avoid losing information in the middle layers of the model, we added a hyper-column layer in CNN3, creating the final model. The proposed model got an accuracy of 91.54%, F1 score of 90.59%, and 89.30% kappa coefficient with just 183K parameters. This is an improvement of around 1.5% over CNN3 and around 4% improvement over the base model while requiring a lesser number of parameters. Table 5.12 demonstrates that of all the base model modifications, the proposed model is the most efficient. This ablation study helps us to choose the best-performing combination among all variations. The two important modules together make the proposed model very effective in feature extraction with minimum computational complexity.

Table 5.12: Summary of the ablation study

Metrics	Grade	CNN-1 (BHCResNet)	CNN-2 (Inception SE)	CNN-3 (CNN-2+InceptionRes)	Proposed RoCNN (CNN-3+HyperColumn)
Precision	G0	0.9722	1	1	0.9756
	G1	1	0.9231	0.9230	0.9259
	G2	0.75	0.8571	0.9047	0.9090
	G3	0.8095	0.6667	0.7143	0.7692
	G4	0.8181	1	0.9259	0.9615
	Overall	0.8699	0.8894	0.8936	0.9082
Recall	G0	0.8573	0.9756	0.9756	0.9756
	G1	0.8889	0.8889	0.8888	0.9259
	G2	0.84	0.72	0.76	0.8
	G3	0.8095	0.9524	0.9523	0.9523
	G4	0.9643	0.8928	0.8928	0.8928
	Overall	0.8712	0.8859	0.8939	0.9093
F1 Score	G0	0.9091	0.9876	0.9876	0.9756
	G1	0.9412	0.9056	0.9056	0.9259
	G2	0.7924	0.7826	0.8261	0.8510
	G3	0.8095	0.7843	0.8163	0.8510
	G4	0.8852	0.9434	0.9091	0.9259
	Overall	0.8675	0.8807	0.8889	0.9059
Kappa	G0	0.8754	0.9827	0.9827	0.9657
	G1	0.9283	0.8840	0.8840	0.9085
	G2	0.7450	0.7409	0.7928	0.8216
	G3	0.7765	0.7388	0.7790	0.8219
	G4	0.8541	0.9304	0.8873	0.9085
	Overall	0.84	0.8666	0.8754	0.8930
Accuracy	G0	0.9507	0.9929	0.9929	0.9859
	G1	0.9788	0.9648	0.9648	0.9718
	G2	0.9225	0.9296	0.9437	0.9507
	G3	0.9436	0.9225	0.9366	0.9507
	G4	0.9507	0.9789	0.9648	0.9718
	Overall	0.8732	0.8944	0.9014	0.9154
Parameters	-	303,423	74,523	115,999	183,303

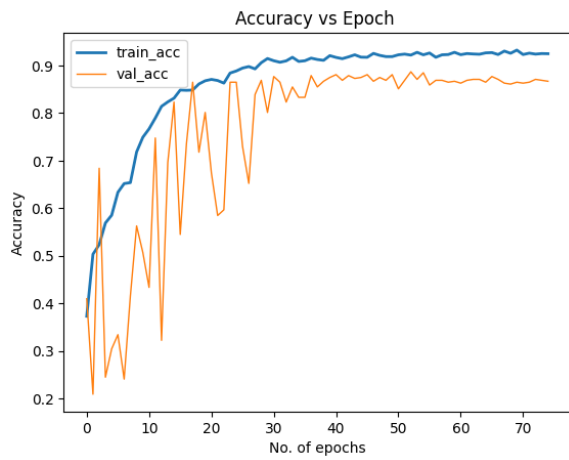
5.7.2 Comparison with state-of-the-art models

We evaluated the proposed model and seven comparison models namely, MobileNetV2, BreastNet, LiverNet, EfficientNetV2, Vision Transformer, NFNet, and Hameed et al. on both multi-grade and multi-class datasets. We present the compiled results in Table 5.13 and Table 5.14. From the Table, our proposed model shows a significant improvement compared to any reference models and achieves 90.59% F1-score, 89.30% kappa coefficient, and 91.54% accuracy on the KMC kidney dataset. On the TCGA dataset also the proposed model achieves 90.14% F1-score, 86.92% kappa coefficient, and 90.19% accuracy, which is the highest among all comparison models. Fig. 5.10 shows the training plots for grade prediction of the Kidney KMC dataset. The leftmost of Fig. 5.10(a) is the training accuracy vs. the number of epochs and validation accuracy vs. epochs. The loss (categorical cross-entropy) vs. the number of epochs plot in Fig. 5.10(b) shows the convergence of loss for the model. We observe that the training and validation curve in Fig. 5.10(a) and Fig. 5.10(b)

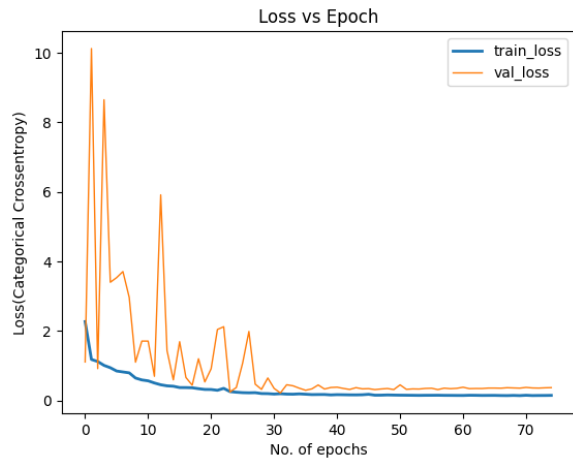
are closer to each other which shows the non-overfit property of the model. Fig. 5.10(c) is a confusion matrix that shows how many instances are correctly predicted over all test cases. We also present the receiver operating characteristic (ROC) curve and area under the curve (AUC) of all reference models and the suggested model in Fig. 5.10(d-h). Fig. 5.11 shows the training plots for the multi-class classification on the TCGA dataset. The learning curves indicate the proposed model is working very well for training data as well as validation data. The proposed model shows the best learning behavior and generalizations to work on different types of histopathology data. ROC-AUC plot indicated the measure of the ability of a classifier between different labels. The higher is ROC-AUC value better the performance of the model. From the ROC-AUC viewpoint, our model is best for each dataset.

Table 5.13: Comparative study of detection of grades using KMC kidney dataset

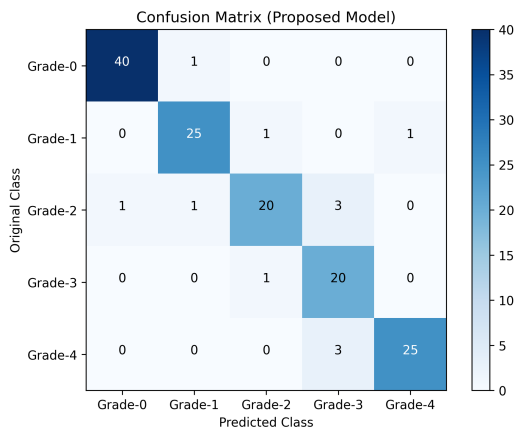
Metrics	Grade	MobileNetV2 (2019)	BreastNet (2020)	LiverNet (2021)	EfficientNetV2 (2021)	ViT (2021)	NFNet (2021)	Hameed et al. (2022)	RoCNN (Proposed)
Precision	G0	0.8904	1	0.9523	0.9762	0.9473	1	1	0.9756
	G1	0.7894	0.9259	0.92	0.9047	0.88	0.8260	0.9259	0.9259
	G2	0.8717	0.7303	0.6896	0.7407	0.8125	0.6333	0.8	0.9090
	G3	0.6734	0.6667	0.7172	0.6363	0.5862	0.6190	0.6818	0.7692
	G4	0.8974	0.9615	1	0.8667	0.7941	0.9259	0.6944	0.9615
	Overall		0.8245	0.8570	0.8552	0.8249	0.8040	0.8008	0.8204
Recall	G0	0.9701	0.9512	0.9756	1	0.8780	1	0.7805	0.9756
	G1	0.6666	0.9259	0.8518	0.7037	0.8148	0.7037	0.9259	0.9259
	G2	0.85	0.76	0.8	0.8	0.52	0.76	0.8	0.8
	G3	0.825	0.7619	0.7143	0.6667	0.8095	0.6190	0.7142	0.9523
	G4	0.7608	0.8928	0.8928	0.9286	0.9643	0.8928	0.8928	0.8928
	Overall		0.8145	0.8584	0.8469	0.8198	0.7973	0.7951	0.8227
F1-Score	G0	0.9285	0.975	0.9638	0.9879	0.9114	1	0.8767	0.9756
	G1	0.7229	0.9259	0.8846	0.7916	0.8461	0.76	0.9259	0.9259
	G2	0.8607	0.7450	0.7407	0.7692	0.6341	0.6909	0.8	0.8510
	G3	0.7415	0.7111	0.7142	0.6511	0.68	0.6190	0.6976	0.8510
	G4	0.8235	0.9259	0.9434	0.8965	0.8709	0.9090	0.7812	0.9259
	Overall		0.8154	0.8566	0.8493	0.8193	0.7885	0.7958	0.8163
Kappa	G0	0.8988	0.9652	0.9489	0.9830	0.8773	1	0.8349	0.9657
	G1	0.6648	0.9085	0.8588	0.7501	0.8117	0.7091	0.9085	0.9085
	G2	0.8330	0.6893	0.6802	0.7176	0.5758	0.6174	0.7572	0.8216
	G3	0.6828	0.6570	0.6647	0.5889	0.6137	0.5529	0.6437	0.8219
	G4	0.7854	0.9085	0.9304	0.8700	0.8353	0.8872	0.7188	0.9085
	Overall		0.7818	0.8398	0.8303	0.8035	0.7599	0.7769	0.7785
Accuracy	G0	0.9579	0.9859	0.9788	0.9929	0.9507	1	0.9366	0.9859
	G1	0.9033	0.9718	0.9577	0.9296	0.9437	0.9154	0.9718	0.9718
	G2	0.9538	0.9084	0.9014	0.9154	0.8944	0.8802	0.9295	0.9507
	G3	0.9033	0.9084	0.9155	0.8944	0.8873	0.8873	0.9084	0.9507
	G4	0.9369	0.9718	0.9788	0.9577	0.9437	0.9647	0.9014	0.9718
	Overall		0.8277	0.8732	0.8661	0.8450	0.8098	0.8239	0.8239



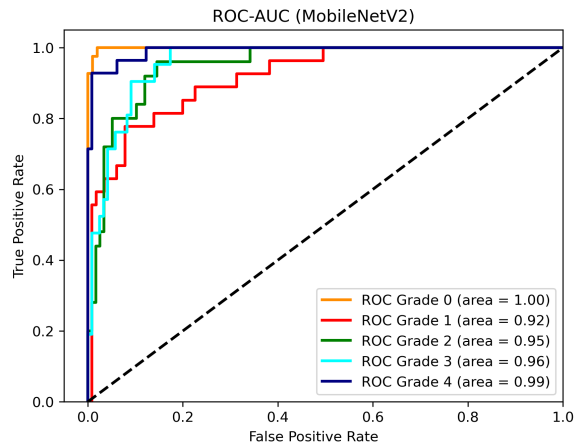
(a)



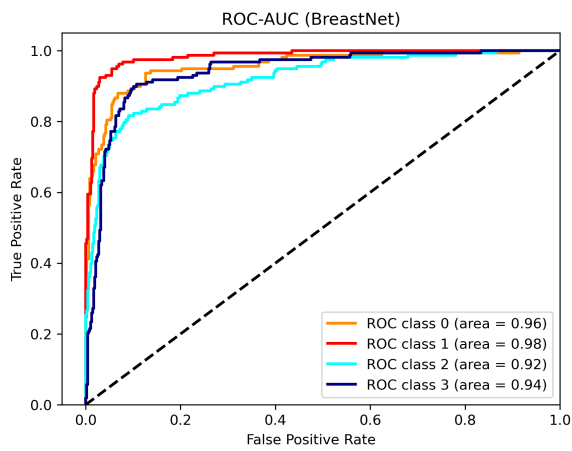
(b)



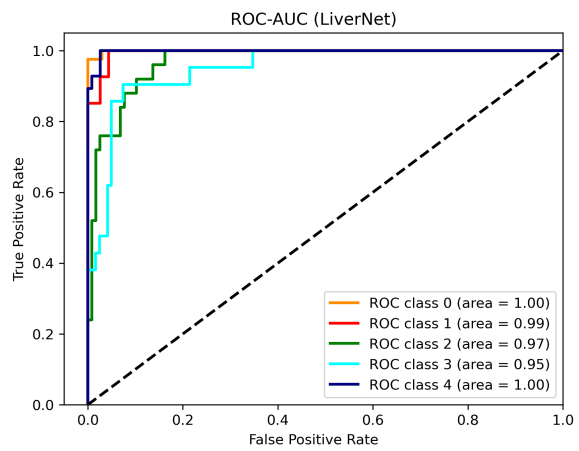
(c)



(d)



(e)



(f)

Figure 5.10: Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on KMC kidney dataset (*cont.*)

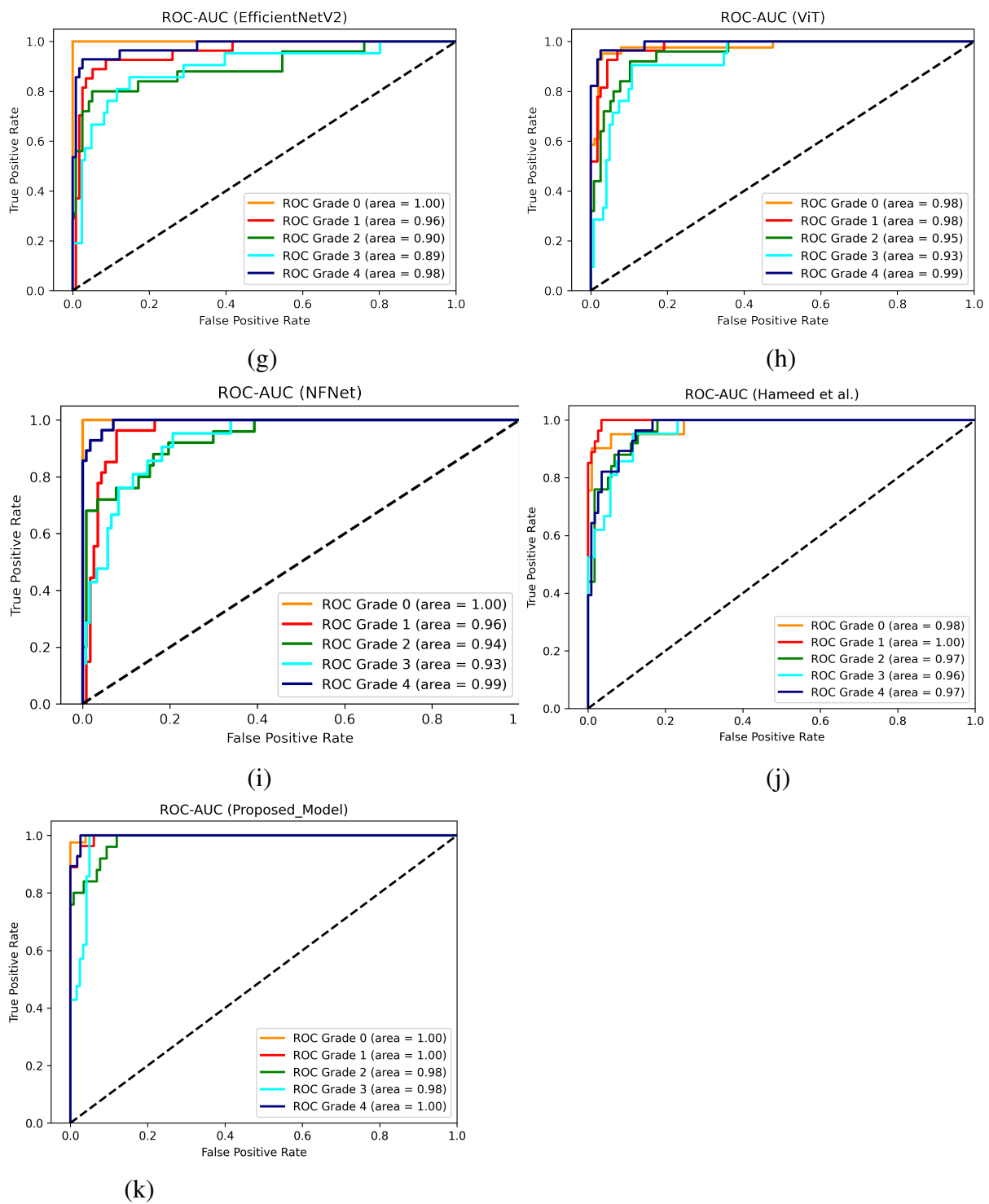
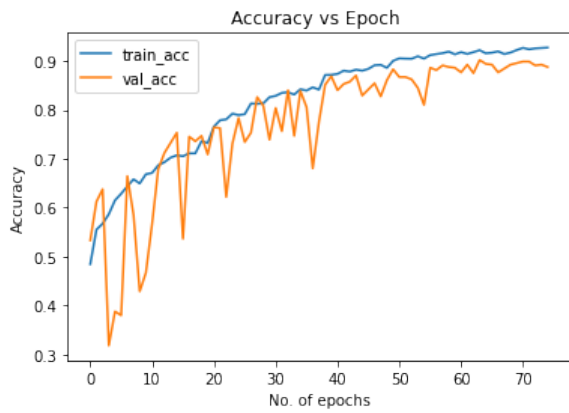
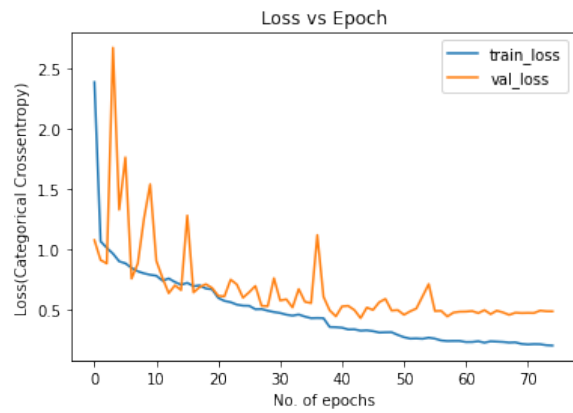


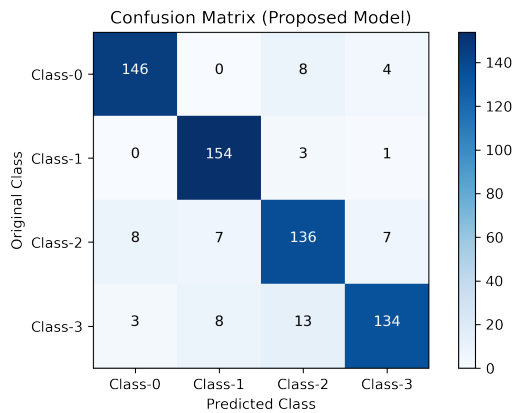
Figure 5.10: Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on KMC kidney dataset



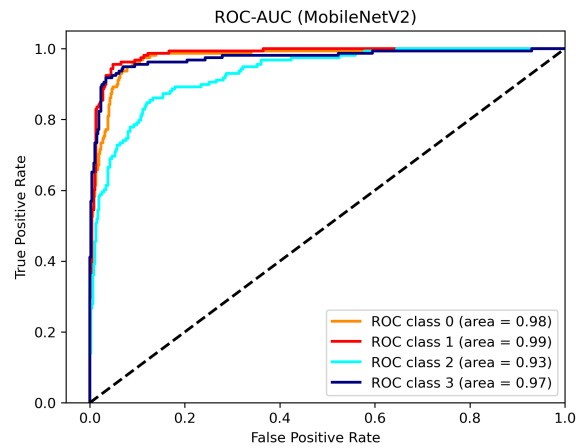
(a)



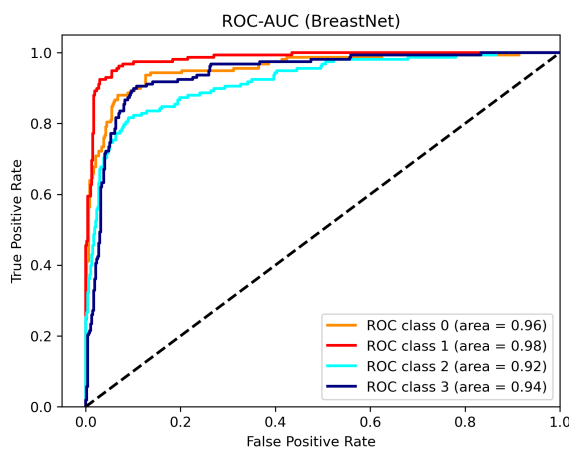
(b)



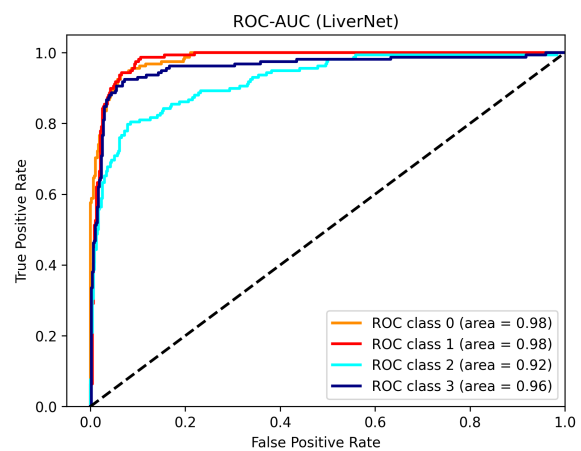
(c)



(d)

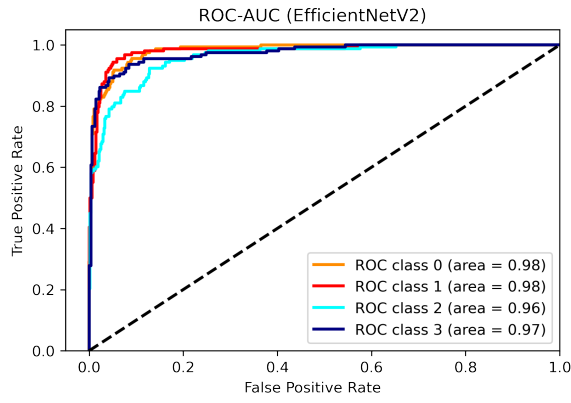


(e)

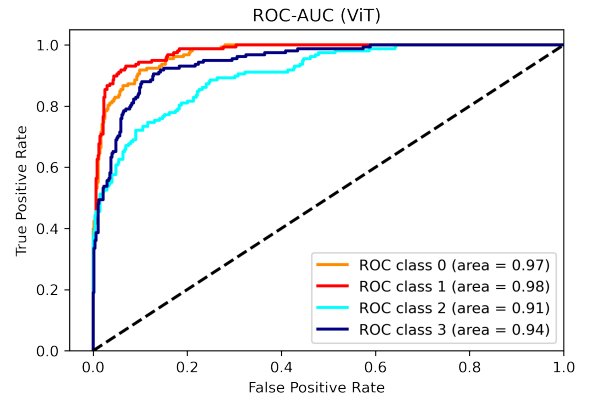


(f)

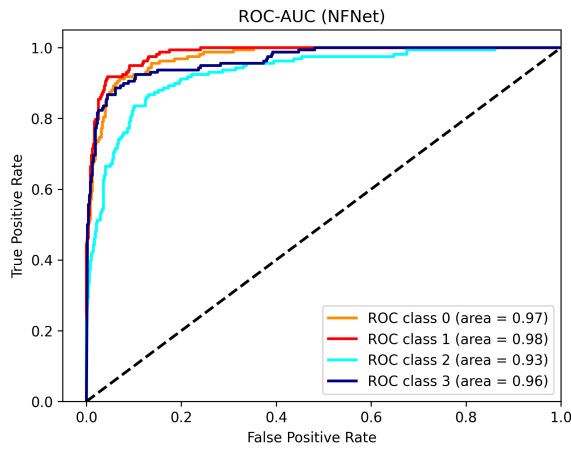
Figure 5.11: Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on TCGA dataset (*cont.*)



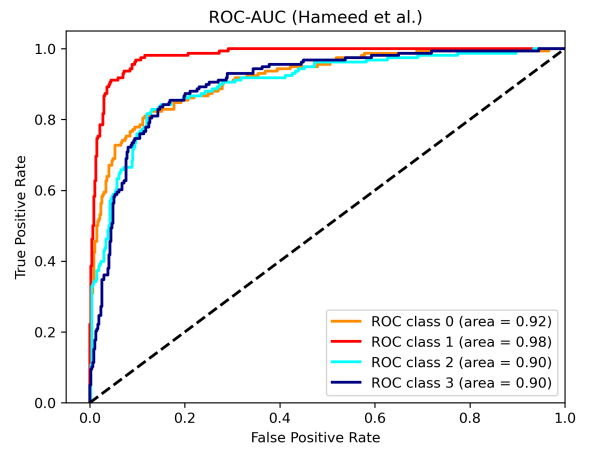
(g)



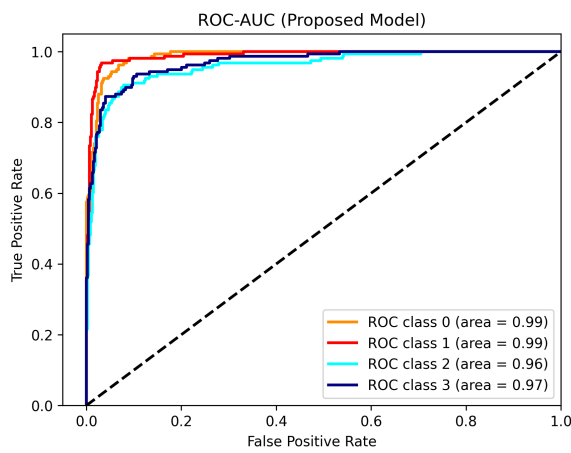
(h)



(i)



(j)



(k)

Figure 5.11: Learning curves, confusion matrix, and receiver operating characteristic (ROC) curves on TCGA dataset

Table 5.14: Comparative study of prediction of classes on TCGA dataset

Metrics	Class	MobileNetV2 (2019)	BreastNet (2020)	LiverNet (2021)	EfficientNetV2 (2021)	ViT (2021)	NFNet (2021)	Hameed et al. (2022)	RoCNN (Proposed)
Precision	C0	0.8222	0.8106	0.8381	0.8562	0.8620	0.875	0.7151	0.9299
	C1	0.9012	0.9230	0.8703	0.8698	0.7946	0.8579	0.8860	0.9112
	C2	0.8421	0.8455	0.8142	0.8620	0.7622	0.7712	0.7430	0.85
	C3	0.9044	0.7777	0.8662	0.9006	0.7798	0.8670	0.7088	0.9178
	Overall	0.8675	0.8393	0.8472	0.8722	0.7996	0.8428	0.7633	0.9022
Recall	C0	0.9367	0.8670	0.9177	0.9050	0.7911	0.8417	0.7784	0.9240
	C1	0.9240	0.9113	0.8924	0.9303	0.9303	0.9177	0.8860	0.9747
	C2	0.7088	0.7278	0.7215	0.7911	0.6898	0.7468	0.6772	0.8607
	C3	0.8987	0.8417	0.8607	0.8607	0.7848	0.8670	0.7088	0.8481
	Overall	0.8671	0.8370	0.8481	0.8718	0.7990	0.8433	0.7626	0.9019
F1-Score	C0	0.8757	0.8379	0.8761	0.88	0.8250	0.8580	0.7454	0.9270
	C1	0.9125	0.9172	0.8812	0.8990	0.8571	0.8868	0.8860	0.9419
	C2	0.7697	0.7823	0.7651	0.8250	0.7242	0.7588	0.7086	0.8553
	C3	0.9015	0.8058	0.8635	0.8802	0.7823	0.8670	0.7088	0.8816
	Overall	0.8649	0.8364	0.8465	0.8711	0.7972	0.8427	0.7622	0.9014
Kappa	C0	0.8306	0.7814	0.8323	0.8385	0.7700	0.8119	0.6557	0.9027
	C1	0.8828	0.8898	0.8410	0.8639	0.8043	0.8474	0.8481	0.9216
	C2	0.7015	0.7168	0.6929	0.7700	0.6383	0.6801	0.6173	0.8067
	C3	0.8689	0.7412	0.8181	0.8415	0.7094	0.8227	0.6118	0.8441
	Overall	0.8228	0.7827	0.7974	0.8291	0.7321	0.7911	0.6835	0.8692
Accuracy	C0	0.9335	0.9161	0.9351	0.9382	0.9161	0.9303	0.8670	0.9636
	C1	0.9557	0.9588	0.9398	0.9477	0.9224	0.9414	0.9430	0.9699
	C2	0.8939	0.8987	0.8892	0.9161	0.8686	0.8813	0.8607	0.9272
	C3	0.9509	0.9003	0.9319	0.9414	0.8908	0.9335	0.8544	0.9430
	Overall	0.8671	0.8370	0.8481	0.8718	0.7990	0.8433	0.7626	0.9019

Table 5.15: Computational complexity comparison of the suggested architecture against various comparison architecture

Models	Parameters (M = 10 ⁶)	Flops (B = 10 ⁹)
MobileNetV2 (2021)	2.264 M	0.613 B
BreastNet (2020)	0.6061 M	5.68 B
LiverNet (2021)	0.5740 M	3.72 B
EfficientNetV2 (2021)	20.33 M	5.75 B
Vision Transformer (2021)	86 M	16.85 B
NFNet (2020)	71.5 M	12 B
Hameed et al. (2022)	2.24 M	5.61 B
Proposed RoCNN	0.183 M	2.64 B

5.7.3 Cross-validation

In the used datasets, the training set and test set are slide-wise separated such that the test set was completely unseen for the model. However, it is possible that some cases in the test set have the same content as in the training set. In this way, the model might not produce the true result. To avoid such issues model was tested using cross-validation. This allows us to analyze how the model is working apart from a particular set of test data. Three folds of 722 images of each grade were created to carry out 3-fold cross-validation for the KMC kidney dataset. For the first-run fold-2 and fold-3 are used to train the model while fold-1 is to test

Table 5.16: Average quality comparison of three-fold KMC kidney dataset

Metrics	Grade	MobileNetV2 (2019)	BreastNet (2020)	LiverNet (2021)	EfficientNetV2 (2021)	ViT (2021)	NFNet (2021)	Hameed et al. (2022)	RoCNN (Proposed)
Precision	G0	0.8039	0.9117	0.9130	0.9420	0.8493	0.8714	0.9333	0.8648
	G1	0.9375	0.8809	0.9473	0.7818	0.72	0.6785	0.8139	0.9148
	G2	0.7143	0.8444	0.8541	0.8529	0.8919	0.9117	0.7174	0.9743
	G3	0.7368	0.75	0.8157	0.7407	0.7692	1	0.75	0.8139
	G4	0.9286	0.8723	0.8571	0.7719	0.8372	0.7627	0.7924	0.9743
	Overall		0.8242	0.8519	0.8775	0.8179	0.8135	0.8449	0.8014
Recall	G0	1	0.9117	0.9264	0.9559	0.9117	0.9104	0.8235	0.9411
	G1	0.5555	0.8222	0.8	0.9555	0.8	0.8444	0.7777	0.9555
	G2	0.8	0.9047	0.9762	0.6904	0.7857	0.775	0.7857	0.9047
	G3	0.6666	0.75	0.775	0.5	0.75	0.475	0.75	0.875
	G4	0.9285	0.8723	0.8936	0.9361	0.7659	0.9782	0.8936	0.8085
	Overall		0.7901	0.8522	0.8742	0.8076	0.8026	0.7966	0.8061
F1-Score	G0	0.8913	0.9117	0.9197	0.9489	0.8794	0.8905	0.8750	0.9014
	G1	0.6977	0.8505	0.8674	0.86	0.7579	0.7524	0.7954	0.9348
	G2	0.7547	0.8735	0.9111	0.7631	0.8354	0.8378	0.7500	0.9382
	G3	0.7	0.75	0.7948	0.5970	0.7595	0.6440	0.7500	0.8434
	G4	0.9285	0.8723	0.875	0.8461	0.8	0.8571	0.8400	0.8837
	Overall		0.7944	0.8516	0.8736	0.8030	0.8064	0.7964	0.8021
Kappa	G0	0.8401	0.8772	0.8880	0.9287	0.8299	0.8463	0.8303	0.8605
	G1	0.6478	0.8178	0.8402	0.8240	0.6989	0.6868	0.7500	0.9195
	G2	0.6986	0.8458	0.8909	0.7196	0.8035	0.8082	0.6945	0.9259
	G3	0.6509	0.7005	0.7554	0.5350	0.7126	0.6008	0.7005	0.8110
	G4	0.9110	0.8415	0.8441	0.8045	0.7544	0.8175	0.7985	0.8589
	Overall		0.7657	0.8225	0.8484	0.7853	0.7643	0.7652	0.7607
Accuracy	G0	0.9295	0.9504	0.9545	0.9710	0.9297	0.9369	0.9338	0.9421
	G1	0.9084	0.9463	0.9545	0.9421	0.9049	0.8949	0.9256	0.9752
	G2	0.9084	0.9545	0.9669	0.9256	0.9462	0.9495	0.9090	0.9793
	G3	0.9155	0.9173	0.9339	0.8884	0.9215	0.9117	0.9173	0.9463
	G4	0.9718	0.9504	0.9504	0.9338	0.9256	0.9369	0.9338	0.9587
	Overall		0.8169	0.8595	0.8801	0.8305	0.8140	0.8151	0.8099

the model. In the second run fold-1, and fold-3 are used to train the model while fold-2 is to test the model. We repeat the process till the last fold where fold-3 is for testing and the remaining fold is for training the model. Similarly, five folds of 4400 images of each class were created to carry out 5-fold cross-validation for the TCGA dataset. The average results of all runs for the proposed model and all benchmark models are included in Table 5.16 and Table 5.17 for the KMC kidney dataset and TCGA datasets respectively. BreastNet Togacar *et al.* (2020), and LiverNet Aatresh *et al.* (2021a) show considerable variations in the result of (Table 5.13, Table 5.16), and (Table 5.14, Table 5.17). The proposed model attains almost similar results given in Table 5.13 and Table 5.14, this suggests that the proposed model is generalized independent of the dataset.

Table 5.17: Average quality comparison of five-fold TCGA dataset

Metrics	Class	MobileNetV2 (2019)	BreastNet (2020)	LiverNet (2021)	EfficientNetV2 (2021)	ViT (2021)	NFNet (2021)	Hameed et al. (2022)	RoCNN (Proposed)
Precision	C0	0.8883	0.9161	0.9255	0.8571	0.7555	0.7865	0.8033	0.8421
	C1	0.7827	0.8877	0.8235	0.7732	0.7901	0.9103	0.8093	0.9505
	C2	0.7846	0.7289	0.7051	0.7905	0.7590	0.8578	0.7609	0.9245
	C3	0.9359	0.7909	0.9014	0.9014	0.8358	0.9241	0.7556	0.8630
	Overall	0.8479	0.8310	0.8389	0.8306	0.7851	0.8697	0.7823	0.8950
Recall	C0	0.8681	0.6954	0.7909	0.8454	0.9272	0.9045	0.8727	0.9454
	C1	0.95	0.7909	0.8909	0.9454	0.8727	0.8772	0.7909	0.8727
	C2	0.6954	0.9045	0.8045	0.6863	0.5727	0.7954	0.7090	0.8909
	C3	0.8636	0.8772	0.8318	0.8318	0.7636	0.8863	0.7590	0.8591
	Overall	0.8443	0.8170	0.8295	0.8272	0.7841	0.8659	0.7829	0.8920
F1	C0	0.8781	0.7907	0.8529	0.8512	0.8326	0.8414	0.8366	0.8908
	C1	0.8583	0.8365	0.8559	0.8507	0.8293	0.8935	0.7999	0.9100
	C2	0.7373	0.8073	0.7516	0.7347	0.6528	0.8254	0.7341	0.9074
	C3	0.8983	0.8319	0.8652	0.8652	0.7980	0.9048	0.7573	0.8610
	Overall	0.8430	0.8166	0.8314	0.8255	0.7782	0.8663	0.7820	0.8923
Kappa	C0	0.8381	0.7331	0.8089	0.8021	0.7690	0.7835	0.7790	0.8515
	C1	0.8048	0.7861	0.8053	0.7940	0.7686	0.8589	0.7343	0.8816
	C2	0.6567	0.7335	0.6613	0.6545	0.5577	0.7702	0.6496	0.8773
	C3	0.8662	0.7719	0.8227	0.8227	0.7347	0.8740	0.6762	0.8148
	Overall	0.7924	0.7560	0.7727	0.7697	0.7121	0.8212	0.7106	0.8561
Accuracy	C0	0.9397	0.9097	0.9318	0.9261	0.9068	0.9147	0.9147	0.9420
	C1	0.9216	0.9227	0.925	0.9170	0.9102	0.9477	0.9011	0.9568
	C2	0.8761	0.8920	0.8670	0.8761	0.8477	0.9159	0.8715	0.9545
	C3	0.9511	0.9113	0.9352	0.9352	0.9034	0.9534	0.8784	0.9307
	Overall	0.8443	0.8170	0.8295	0.8272	0.7841	0.8659	0.7829	0.8920

5.7.4 Computational complexity

Computational complexity is expressed in terms of the total trainable parameters and floating-point operations (FLOPs) shown in Table 5.15. The number of parameters employed in the proposed architecture is 0.183 Million, which is the minimum among all state-of-the-art models. In terms of flops as well, our model provides an efficient solution. The proposed model utilizes the least number of FLOPs except (Sandler *et al.*, 2019). The efficient Inception-ResNet module employed in the proposed model is the main reason for the improved performance with a significantly lower number of parameters. Inception-ResNet module extracts necessary histopathology-specific features using only a fraction of the parameters used in the convolution block.

5.8 Summary

In this chapter, two novel CNN architectures for grading and classification tasks have been proposed. This study is the first to propose an end-to-end automatic grading of kidney

cancer from kidney histopathology images that were not yet focused on. In addition, a novel kidney cancer dataset has been introduced, which is validated by skilled medical experts. To verify the validity, all comparable models including the proposed model experimented with additional well-established datasets called BreakHis dataset for eight class-classification and TCGA dataset for four class-classification. Comparable experiments include the most recent deep learning models trained from scratch as well as transfer learning techniques using pre-trained weights of the ImageNet. Performance is evaluated based on the metrics Precision, Recall, F1-score, Accuracy, and computational complexities based on the total number of parameters and floating point operations (FLOPs). In the ablation study, different versions of the proposed model were trained by removing a significant component each time, and found that each component removed, had a deteriorating effect on quality metrics. The introduced methodology of histopathology analysis and the results presented in this study would encourage medical and machine learning researchers and professionals and in turn will benefit the patients for early diagnosis and precise decisions on the course of their treatment.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This chapter summarizes the experimental results obtained using various proposed algorithms and concludes by discussing the future scope of the research work. The various methods for nuclei segmentation, classification, and grading of histopathological images were reviewed. Further, methods based on deep CNN's were effectively exploited and proposed new architectures for nuclei segmentation, classification of subtypes of Renal cell carcinoma (RCC), and grading of RCC images from histopathological data. These architectures help in automating the diagnosis process by avoiding manual intervention in all stages. As a part of this study, two novel benchmark datasets have been introduced. A dataset for the classification of subtypes of RCC and another dataset for the grading of RCC with five different grades from kidney histopathology images has been introduced.

Deep learning models for the nuclei segmentation of histopathology images address two major problems. To separate nuclei from complex structured histopathology images with different histology and molecular characteristics was the first major concern. This issue was addressed by implementing an effective encoder–decoder architecture with the separable convolution pyramid pooling (SCPP) unit at the bottleneck layer. SCPP block extracted and learned deep features by varying four different dilation rates. The second major concern was to reduce the computational complexity and total trainable parameters, which was effectively handled by incorporating depth-wise separable convolution.

Another CNN model called high resolution deep transferred ASPPU-Net was proposed that addressed automatic nuclei segmentation of histopathology images having a widely varied spectrum with a large number of artifacts. The implemented networks effectively

leverage the strength of residual learning as well as encoder-decoder architecture by incorporating wide and deep network paths that strengthen the intermediate features. Promising results were obtained due to the effective use of the wide and deep network with ASPP at the bottleneck layer.

Deep structured residual encoder-decoder network with a novel loss function introduces a strong encoder–decoder with multiple paths and improved discrimination capability and is able to retrieve relevant and compact textural information. The proposed loss function better represents the validation data and using the new loss function the proposed encoder-decoder model is much generalized to work on different types of histopathology data. The tabular and visual segmentation results of the proposed framework were better in comparison to other benchmark deep learning models.

This work designed a fast and accurate deep learning (DL) based automated system for the detection of renal cell carcinoma (RCC) from kidney histopathology images. The proposed RenalNet model experimented for four class (Normal, KIRP, KIRC, and KICH) classification of patches of WSIs of TCGA kidney dataset. Extraction of a class-specific representative set of features was possible due to the utilization of optimized transformation modules at three different resolutions within the network. The results of the proposed model have been compared with the existing DL models trained from scratch as well as networks leveraged by transfer learning of pre-trained weights. The experimental result indicated that the proposed model gave better results compared to the most recent existing DL models in terms of quality metrics, computational complexity, training time, and prediction time.

A fully automated deep learning framework called a Renal Cell Carcinoma Grading Network (RCCGNet) for the detection of malignancy levels of renal cell carcinoma (RCC) in kidney histopathology images were proposed. This study is the first to propose an end-to-end automatic grading of kidney cancer from kidney histopathology images that were not yet focused on. Extraction of a class-specific representative set of features was possible due to the effective utilization of inter-channel information exchange at three different resolutions within the network. Residual connection, information sharing between different layers, and gating mechanism in shared channel residual network were attributed to the distinguishing performance. Experimental results show that the proposed RCCGNet has the potential to grade five different grades associated with kidney cancer histopathology images with better accuracy. The proposed approach is generalized and effectively works on multiple organ histopathology datasets. It reduces the need for extensive computational complexity.

And finally, a computationally efficient deep-learning architecture based on Convolutional Neural Networks (CNNs) to automate the grading and classification task for kidney cancer tissue was proposed. The proposed Robust CNN (RoCNN) architecture is capable of learning features at varying convolutional filter sizes because of the inception modules employed in it. Squeeze and Extract (SE) blocks are used to remove unnecessary contributions from noisy channels and improve model accuracy. To demonstrate that the proposed model is generalized and independent of the dataset, it has experimented on two well-known datasets, the KMC kidney dataset of five different grades and the TCGA dataset of four classes. Compared to the best-performing state-of-the-art model the accuracy of RoCNN shows significant improvement on both datasets.

6.2 Future Work

Even though the suggested approach achieved good results, adding a high-performance feature extraction module may enhance the feature space even more. The suggested approach can also be expanded to operate with more image modalities. In the future, we will focus on the considerable extension of the kidney dataset and other pathological datasets. Our future work will also focus on the hardware implementation of the RenalNet model for RCC detection from kidney histopathology images.

Bibliography

- Aatresh, A. A., K. Alabhya, S. Lal, J. Kini, and P. U. P. Saxena** (2021a). Livernet: efficient and robust deep learning model for automatic diagnosis of sub-types of liver hepatocellular carcinoma cancer from h&e stained liver histopathology images. *Int J Comput Assist Radiol Surg.*, **16**(9), 1549–1563.
- Aatresh, A. A., R. P. Yatgiri, A. K. Chanchal, A. Kumar, A. Ravi, D. Das, R. BS, S. Lal, and J. Kini** (2021b). Efficient deep learning architecture with dimension-wise pyramid pooling for nuclei segmentation of histopathology images. *Computerized Medical Imaging and Graphics*, **93**(101975).
- Albayrak, A. and G. Bilgin** (2019). Automatic cell segmentation in histopathological images via two-staged superpixel-based algorithms. *Medical and biological engineering and computing*, **57**(3), 653–665.
- Alom, M. Z., M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari** (2018). Improved inception-residual convolutional neural network for object recognition. *Neural Comput & Applic*, **32**, 279–293.
- Araujo, T., G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polonia, and A. Campilho** (2017). Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE*, **12**(6), 1–14.
- Badrinarayanan, V., A. Kendall, and R. Cipolla** (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(12), 2481–2495.
- Brock, A., S. De, S. L. Smith, and K. Simonyan** (2021). High-performance large-scale image recognition without normalization. <https://arxiv.org/abs/2102.06171>.

- Chanchal, A. K., A. Kumar, S. Lal, and J. Kini** (2021a). Efficient and robust deep learning architecture for segmentation of kidney and breast histopathology images. *Computers & Electrical Engineering*, **92**(107177), 1–14.
- Chanchal, A. K., S. Lal, and J. Kini** (2021b). High-resolution deep transferred asppu-net for nuclei segmentation of histopathology images. *International Journal of Computer Assisted Radiology and Surgery*, **16**, pages2159–2175.
- Chen, H., X. Qi, L. Yu, and P.-A. Heng**, Dcan: Deep contour-aware networks for accurate gland segmentation. In *Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2016. URL <https://arxiv.org/abs/1604.02677>.
- Chen, L.-C., G. Papandreou, F. Schroff, and H. Adam** (2017). Rethinking atrous convolution for semantic image segmentation. *Computer Vision and Pattern Recognition-arXiv*, **3**, 1–14.
- Chen, L.-C., Y. Yang, J. Wang, W. Xu, and A. L. Yuille**, Attention to scale: Scale-aware semantic image segmentation. In *Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2015. URL <https://arxiv.org/abs/1511.03339>.
- Chen, M., B. Zhang, W. Topatana, J. Cao, H. Zhu, S. Juengpanich, Q. Mao, H. Yu, and X. Cai** (2020). Classification and mutation prediction based on histopathology h&e images in liver cancer using deep learning. *npj Precision Oncology*, **4**(14), 1–7.
- Ciecholewski, M.** (2016). An edge-based active contour model using an inflation/deflation force with a damping coefficient. *Expert Systems with Applications*, **44**, 22–36.
- Coudray, N., P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyo, A. L. Moreira, N. Razavian, and A. Tsirigos** (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine*, **24**(10), 1559–1567.
- Cousty, J., G. Bertrand, L. Najman, and M. Couprie** (2010). Watershed cuts: Thinnings, shortest path forests, and topological watersheds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**(5), 925–939.

- Cruz-Roa, A., H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. C. Shih, J. Tomaszewski, F. A. Gonzalez, and A. Madabhushi** (2018). Accurate and reproducible invasive breast cancer detection in whole slide images: A deep learning approach for quantifying tumor extent. *Scientific Reports*, **7**(46450), 1–14.
- Delahunt, B., J. N. Eble, and L. Egevad** (2019). Grading of renal cell carcinoma. *Histopathology, Wiley Online Library*, **74**, 4–17.
- Dhanachandra, N. and Y. J. Chanu** (2020). An image segmentation approach based on fuzzy c-means and dynamic particle swarm optimization algorithm. *Multimedia Tools and Applications*, **79**, 18839–18858.
- Ding, H., Z. Pan, Q. Cen, Y. Li, and S. Chen** (2020). Multi-scale fully convolutional network for gland segmentation using three class classification. *Neurocomputing*, **380**, 150–161.
- Dosovitskiy, A. and L. Beyer et al.**, An image is worth 16x16 words: Transformers for image recognition at scale. *In Computer Vision and Pattern Recognition*. arXiv, 2021. URL <https://arxiv.org/abs/2010.11929>.
- Du, Z., W. Chen, Q. Xia, O. Shi, and Q. Chen** (2020). Trends and projections of kidney cancer incidence at the global and national levels, 1990-2030: a bayesian age-period-cohort modeling study. *Biomarker research*, **8**(16), 1–10.
- Dumoulin, V. and F. Visin**, A guide to convolution arithmetic for deep learning. *In Machine Learning (stat.ML), Machine Learning (cs.LG), Neural and Evolutionary Computing (cs.NE), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2016. URL <https://arxiv.org/abs/1603.07285>.
- Elazab, N., H. Soliman, S. El-Sappagh, S. M. R. Islam, and M. Elmogy** (2020). Objective diagnosis for histopathological images based on machine learning techniques: Classical approaches and new trends. *Mathematics*, **8**(11), 1–26.
- Elston, C. W. and I. O. Ellis** (1991). Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology*, **19**, 403–410.
- Fu, J., J. Xu, K. Tasaka, and Z. Chen**, Residual squeeze-and-excitation network for fast image deraining. *In Image and Video Processing (eess.IV), Computer Vision and Pattern*

(*cs.CV*), *Machine Learning (cs.LG)*, *FOS: Electrical engineering*. arXiv, 2020. URL <https://arxiv.org/abs/2006.00757>.

Gamarra, M., E. Zurek, H. J. Escalante, L. Hurtado, and H. San-Juan-Vergara (2019). Split and merge watershed: A two-step method for cell segmentation in fluorescence microscopy images. *Biomedical Signal Processing and Control*, **53**(101575), 1–12.

Gonzalez, R. C. and R. E. Woods (2008). Digital image processing. *Prentice Hall, New York, USA 3rd Edition*.

Graham, S., H. Chen, J. Gamper, Q. Dou, P.-A. Heng, D. Snead, Y. W. Tsang, and N. Rajpoot (2018). Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical Image Analysis*, **52**, 199–211.

Graham, S., Q. Dang Vu, S. E. Ahmed Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot (2019a). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, **58**, 1–15.

Graham, S., D. Epstein, and N. Rajpoot (2019b). Rota-net: Rotation equivariant network for simultaneous gland and lumen segmentation in colon histology images. *European Congress on Digital Pathology, Springer Cham*, 109–116.

Graham, S. and N. Rajpoot (2018). Sams-net: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 590–594.

Hameed, Z., B. Garcia-Zapirain, J. J. Aguirre, and M. A. Isaza-Ruget (2022). Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network. *Scientific Reports*, **12**(15600).

Hanif, M. and B. Muhammad Shehzad (2020). Competitive residual neural network for image classification. *ICT Express*, **6**(1), 28–37.

Hariharan, B., P. Arbeláez, R. Girshick, and J. Malik, Hypercolumns for object segmentation and fine-grained localization. In *Computer Vision and Pattern Recognition (cs.CV)*, *FOS: Computer and information sciences*, *FOS: Computer and information sciences*. arXiv, 2014. URL <https://arxiv.org/abs/1411.5752>.

- Hashemi, s. s. m., d. seyed raein, s. p. salehi, s. k. erdogmus, prabhu, ali, and g. warfield** (2019). Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection. *IEEE Access*, **7**, 1721–1735.
- He, K., X. Zhang, S. Ren, and J. Sun** (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**(9), 1904–1916.
- He, S. R.-J. S., Xiangyu Zhang** (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, 770–778.
- He, Y., J. Qian, and J. Wang** (2019). Depth-wise decomposition for accelerating separable convolutions in efficient convolutional neural networks. *Computer Vision and Pattern Recognition-arXiv*, **1**, 1–10.
- Hirra, I., M. Ahmad, A. Hussain, M. U. Ashraf, I. A. Saeed, S. F. Qadri, and A. M. Alghamdi** (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*, **9**, 24273–24287.
- Hong, S. K., C. W. Jeong, J. H. Park, H. S. Kim, C. Kwak, G. Choe, H. H. Kim, and S. E. Lee** (2011). Application of simplified fuhrman grading system in clear-cell renal cell carcinoma. *BJU International, National Library of Medicine*, **107**(3), 409–415.
- Howard, A., M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam**, Searching for mobilenetv3. In *Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2019. URL <https://arxiv.org/abs/1905.02244>.
- Hu, J., L. Shen, S. Albanie, G. Sun, and E. Wu** (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**(8), 2011–2023.
- Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger**, Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

- Ioffe, S.** and **C. Szegedy**, Batch normalization: Accelerating deep network training by reducing internal covariate shift. *In Machine Learning (cs.LG), FOS: Computer and information sciences, FOS: Computer and information sciences.* arXiv, 2015. URL <https://doi.org/10.48550/arxiv.1502.03167>.
- Irshad, H., L. M. Kouhsari, G. Waltz, O. Bucur, J. A. Nowak, F. Dong, N. W. Knoblauch,** and **A. H. Beck** (2015). Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. *Pac Symp Biocomput*, 294–305.
- Jadon, S.**, A survey of loss functions for semantic segmentation. *In 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB).* IEEE, 2020. URL <https://doi.org/10.1109%2Fcibcb48159.2020.9277638>.
- Jiang, Y., L. Chen, H. Zhang,** and **X. Xiao** (2019). Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module. *PLOS ONE*, **14**(3), 1–21.
- Kandel, I.** and **M. Castelli** (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express*, **6**(4), 312–315.
- Kang, G., X. Dong, L. Zheng,** and **Y. Yang**, Patchshuffle regularization. *In Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences.* arXiv, 2017. URL <https://arxiv.org/abs/1707.07103>.
- Karimi, S. E.** and **S. Davood** (2020). Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, **39**(2), 499–513.
- Khoshdeli, M., A. Borowsky,** and **B. Parvin** (2018). Deep learning models differentiate tumor grades from h&e stained histology sections. *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 620–623.
- Kingma, D. P.** and **J. Ba**, Adam: A method for stochastic optimization. *In Machine Learning (cs.LG), FOS: Computer and information sciences, FOS: Computer and information sciences.* arXiv, 2014. URL <https://arxiv.org/abs/1412.6980>.

- Kumar, N., V. Ruchika, S. Sharma, S. Bhargava, V. Abhishek, and A. Sethi** (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, **36**(7), 1550–1560.
- Kumawat, S., G. Kanojia, and S. Raman** (2022). Shuffleblock: Shuffle to regularize deep convolutional neural networks. *2022 National Conference on Communications*, 36–41.
- Lal, S., D. Das, K. Alabhya, A. Kanfode, A. Kumar, and J. Kini** (2021). Nucleisegnet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. *Computers in Biology and Medicine*, **128**(104075), 1–8.
- Lal, S., M. Russel, A. Desouza, A. Maneesh, G. Kanfode, K. Kumar, A. K. Perayil, J. Alabhya, and K. Chanchal** (2020). A robust method for nuclei segmentation of h&e stained histopathology images. *7th IEEE International Conference on Signal Processing and Integrated Networks (SPIN2020)*, 453–458.
- Li, C., Y. Tan, W. Chen, X. Luo, Y. Gao, X. Jia, and Z. Wang** (2020). Attention unet++: A nested attention-aware u-net for liver ct image segmentation. *IEEE International Conference on Image Processing (ICIP)*, 345–349.
- Iizuka, O., F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki** (2020). Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Scientific Reports*, **10**(1504), 1–11.
- Ma, N., X. Zhang, H.-T. Zheng, and J. Sun**, Shufflenet v2: Practical guidelines for efficient cnn architecture design. *In Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2018. URL <https://arxiv.org/abs/1807.11164>.
- Maier-Hein, L. and A. Reinke et al.** (2022). Metrics reloaded: Pitfalls and recommendations for image analysis validation. <https://arxiv.org/abs/2206.01653>.
- Malekijoo, M. J. and F. Amirhossein** (2019). Convolution-deconvolution architecture with the pyramid pooling module for semantic segmentation. *Multimedia Tools and Applications*, **78**, 32379–32392.
- Mehta, S., H. Hajishirzi, and M. Rastegari**, Dicenet: Dimension-wise convolutions for efficient networks. *In Computer Vision and Pattern Recognition (cs.CV), Machine Learn-*

ing (cs.LG), *Image and Video Processing*. arXiv, 2019. URL <https://arxiv.org/abs/1906.03516>.

Mehta, S., M. Rastegari, L. Shapiro, and H. Hajishirzi, Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network. *In Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2018. URL <https://arxiv.org/abs/1811.11431>.

Mescher, A. L. (2013). Junqueira’s basic histology: text and atlas. *McGraw-Hill Medical 13th ed. New York*, **12**.

Milletari, F., N. Navab, and S.-A. Ahmadi (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. *Fourth International Conference on 3D Vision (3DV), Stanford*, 565–571.

Naylor, P., M. Lae, F. Reyal, and T. Walter (2019). Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, **38**(2), 448–459.

Nogues, I., L. Lu, X. Wang, H. Roth, G. Bertasius, N. Lay, J. Shi, Y. Tsehay, and R. M. Summers (2016). Automatic lymph node cluster segmentation using holistically-nested neural networks and structured optimization in ct images. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, **9901**, Lecture Notes in Computer Science.

Pan, X., L. Li, H. Yang, Z. Liu, J. Yang, L. Zhao, and Y. Fan (2017). Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. *Neurocomputing*, **229**, 88–99.

Ricketts, C. J., A. A. D. Cubas, H. Fan, and C. C. Smith (2018). The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Reports*, **23**, 313–326.

Ronneberger, O., P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015.

- Salehi, S. S. M., D. Erdogmus, and A. Gholipour** (2017). Tversky loss function for image segmentation using 3d fully convolutional deep networks. *International Workshop on Machine Learning in Medical Imaging*, (10541), 379–387.
- Samaratunga, H., T. Gianduzzo, and B. Delahunt** (2014). The isup system of staging, grading and classification of renal cell neoplasia. *Journal of Kidney Cancer and VHL*, **1**(3), 26–39.
- Sandler, M., A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen** (2019). Mobilenetv2: Inverted residuals and linear bottlenecks. <https://arxiv.org/abs/1801.04381>.
- Santurkar, S., D. Tsipras, A. Ilyas, and A. Madry**, How does batch normalization help optimization? In *Machine Learning (stat.ML), Machine Learning (cs.LG), Neural and Evolutionary Computing (cs.NE), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2018. URL <https://arxiv.org/abs/1805.11604>.
- Schlemper, J., O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert** (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, **53**, 197–207.
- Seth, N., S. Akbar, S. Nofech-Mozes, S. Salama, and A. L. Martel** (2019). Automated segmentation of dcis in whole slide images. *Digital pathology ECDP 2019, Springer Cham*, **11435**.
- Shahidi, F., S. M. Daud, H. Abas, N. A. Ahmad, and N. Maarop** (2020). Breast cancer classification using deep learning approaches and histopathology image: A comparison study. *IEEE Access*, **8**, 187531–187552.
- Simonyan, K. and A. Zisserman** (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Song, T.-H., V. Sanchez, H. EIDaly, and N. M. Rajpoot** (2017). Dual-channel active contour model for megakaryocytic cell segmentation in bone marrow trephine histology images. *IEEE Transactions on Biomedical Engineering*, **64**(12), 2913–2923.
- Stutz, D., A. Hermans, and B. Leibe** (2017). Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding*, **166**, 1–27.

- Sudre, C. H., W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso** (2017). Generalized dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep learning in medical image analysis and multimodal learning for clinical decision support*, 240–248.
- Sugino, T., T. Kawase, S. Onogi, T. Kin, N. Saito, and Y. Nakajima** (2021). Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks. *Healthcare, MDPI*, **9**(8), 938.
- Sun, C., A. Xu, D. Liu, Z. Xiong, F. Zhao, and W. Ding** (2020). Deep learning-based classification of liver cancer histopathology images using only global labels. *IEEE Journal of Biomedical and Health Informatics*, **24**(6), 1643–1651.
- Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray** (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *in CA Cancer J Clin*, 1–41.
- Szegedy, C., S. Ioffe, V. Vanhoucke, and A. Alemi**, Inception-v4, inception-resnet and the impact of residual connections on learning. *In Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2016. URL <https://arxiv.org/abs/1602.07261>.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich**, Going deeper with convolutions. *In Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2014. URL <https://arxiv.org/abs/1409.4842>.
- Tabibu, S., P. K. Vinod, and C. V. Jawahar** (2019). Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning. *Scientific Reports*, **9**(10509), 1–9.
- Tan, M. and Q. V. Le** (2021). Efficientnetv2: Smaller models and faster training. <https://arxiv.org/abs/2104.00298>.
- Tian, K., C. A Rubadue, D. I. Lin, M. Veta, M. E. Pyle, H. Irshad, and Y. J. Heng** (2019). Automated clear cell renal carcinoma grade classification with prognostic significance. *PLoS ONE*, **14**(3), 1–16.

- Togacar, M., K. B. ozkurt, B. Ergen, and Z. Comert** (2020). Breastnet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. *Physica A: Statistical Mechanics and its Applications*, **545**(123592), 17.
- Veit, A., M. Wilber, and S. Belongie**, Residual networks behave like ensembles of relatively shallow networks. *In Computer Vision and Pattern Recognition (cs.CV), Artificial Intelligence (cs.AI), Machine Learning (cs.LG)*. arXiv, 2016. URL <https://arxiv.org/abs/1605.06431>.
- Wan, T., L. Zhao, H. Feng, D. Li, C. Tong, and Z. Qin** (2020). Robust nuclei segmentation in histopathology using asppu-net and boundary refinement. *Neurocomputing*, **408**, 144–156.
- Wang, F., M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang**, Residual attention network for image classification. *In Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2017. URL <https://arxiv.org/abs/1704.06904>.
- Washington, M. K., J. Berlin, P. Branton, L. J. Burgart, D. K. Carter, F. P. L, K. Halling, W. Frankel, J. Jessup, S. Kakar, B. Minsky, R. Nakhleh, and C. C. Compton** (2009). Protocol for the examination of specimens from patients with primary carcinoma of the colon and rectum. *Archives of pathology & laboratory medicine*, **133**, 1539–1551.
- Wei, B., Z. Han, X. He, and Y. Yin** (2017). Deep learning model based breast cancer histopathological image classification. *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 348–353.
- Wei, J. W., A. A. Suriawinata, and L. J. Vaickus** (2020). Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides. *JAMA Netw Open*, **3**(4), 1–11.
- Win, K. Y., S. Choomchuay, and K. Hamamoto** (2017). K mean clustering based automated segmentation of overlapping cell nuclei in pleural effusion cytology images. *International Conference on Advanced Technologies for Communications (ATC)*, 265–269.
- Woo, S., J. Park, J.-Y. Lee, and I. S. Kweon**, Cbam: Convolutional block attention module. *In Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and*

information sciences, FOS: Computer and information sciences. arXiv, 2018. URL <https://arxiv.org/abs/1807.06521>.

Xie, J., R. Liu, J. Luttrell, and C. Zhang (2019). Deep learning based analysis of histopathological images of breast cancer. *Frontiers in Genetics*, **10**(80), 1–19.

Xie, S., R. Girshick, P. Dollár, Z. Tu, and K. He, Aggregated residual transformations for deep neural networks. In *Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2016. URL <https://arxiv.org/abs/1611.05431>.

Xing, F., Y. Xie, and L. Yang (2016). An automatic learning-based framework for robust nucleus segmentation. *IEEE Transactions on Medical Imaging*, **35**(2), 550–566.

Yang, K., S. Chang, Z. Tian, C. Gao, Y. Du, X. Zhang, K. Liu, J. Meng, and L. Xue (2022). Automatic polyp detection and segmentation using shuffle efficient channel attention network. *Alexandria Engineering Journal*, **61**(1), 917–926.

Yang, Q.-L. Z. Y.-B., Sa-net: Shuffle attention for deep convolutional neural networks. In *Computer Vision and Pattern Recognition (cs.CV), Artificial Intelligence (cs.AI), FOS: Computer and information sciences*. arXiv, 2021. URL <https://arxiv.org/abs/2102.00240>.

Zhang, X. and T. Chen (2020). Attention u-net for interpretable classification on chest x-ray image. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 901–908.

Zhang, X., X. Zhou, M. Lin, and J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2017. URL <https://arxiv.org/abs/1707.01083>.

Zhang, Z., H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, Et-net: A generic edge-attention guidance network for medical image segmentation. In *Computer Vision and Pattern Recognition (cs.CV), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2019. URL <https://arxiv.org/abs/1907.10936>.

- Zhao, H., Y. Zhang, S. Liu, J. Shi, C. Change Loy, D. Lin, and J. Jia** (2018). Psanet: Point-wise spatial attention network for scene parsing. *European Conference on Computer Vision*, 267–283.
- Zhou, S., D. Nie, E. Adeli, Y. Gao, L. Wang, J. Yin, and D. Shen** (2018). Fine-grained segmentation using hierarchical dilated neural networks. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, **11073**, 488–496.
- Zhou, S., D. Nie, E. Adeli, J. Yin, J. Lian, and D. Shen** (2020). High-resolution encoder–decoder networks for low-contrast medical image segmentation. *IEEE Transactions on Image Processing*, **29**, 461–475.
- Zhou, Y., S. Graham, N. A. Koohbanani, M. Shaban, P.-A. Heng, and N. Rajpoot**, Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In *Image and Video Processing (eess.IV), Computer Vision and Pattern Recognition (cs.CV)*. arXiv, 2019. URL <https://arxiv.org/abs/1909.01068>.
- Zhu, M., B. Ren, R. Richards, M. Suriawinata, N. Tomita, and S. Hassanpour** (2021). Development and evaluation of a deep neural network for histologic classification of renal cell carcinoma on biopsy and surgical resection slides. *Scientific Reports*, **11**(7080), 1–9.
- Zoph, B., V. Vasudevan, J. Shlens, and Q. V. Le**, Learning transferable architectures for scalable image recognition. In *Computer Vision and Pattern Recognition (cs.CV), Machine Learning (cs.LG), Machine Learning (stat.ML), FOS: Computer and information sciences, FOS: Computer and information sciences*. arXiv, 2017. URL <https://arxiv.org/abs/1707.07012>.

Publications based on the Thesis

1. **Amit Kumar Chanchal**, Aman Kumar, Shyam Lal, and Jyoti Kini “Efficient and robust deep learning architecture for segmentation of kidney and breast histopathology images”, *Computers and Electrical Engineering*, 92 (2021): 107177, *Elsevier Publisher*, [\[Published\]](#), *Indexed by SCI, Thomson ISI, Scopus (Elsevier), 2021, Impact Factor: 3.818*
2. **Amit Kumar Chanchal**, Shyam Lal, and Jyoti Kini “High Resolution Deep Transferred ASPPU-Net for Nuclei Segmentation of Histopathology Images”, *International Journal of Computer Assisted Radiology and Surgery*, *Springer Publisher*, [\[Published\]](#), *Indexed by SCI, Thomson ISI, Scopus (Elsevier), 2021, Impact Factor: 2.924*
3. **Amit Kumar Chanchal**, Shyam Lal, and Jyoti Kini “Deep Structured Residual Encoder-Decoder Network with a Novel Loss Function for Nuclei Segmentation of Kidney and Breast Histopathology Images”, *Multimedia Tools and Applications*, *Springer Publisher*, [\[Published\]](#), *Indexed by SCI, Thomson ISI, Scopus (Elsevier), 2022, Impact Factor: 2.757*
4. **Amit Kumar Chanchal**, Shyam Lal, Ranjeet Kumar, Jin Tae Kwak, and Jyoti Kini “A novel dataset and efficient deep learning framework for automated grading of renal cell carcinoma from kidney histopathology images”, *Scientific Reports*, *Nature Publisher*, [\[Published\]](#), *Indexed by SCI, Web of Science, Scopus, Impact Factor: 4.996*
5. **Amit Kumar Chanchal**, Shyam Lal, and Jyoti Kini “Development of Robust CNN Architecture for Grading and Classification of Renal Cell Carcinoma Histology Images”, *IEEE Transactions on Artificial Intelligence*, *IEEE Publisher*, [\[Revision Submitted\]](#), *Indexed by SCI, Thomson ISI, Scopus.*
6. **Amit Kumar Chanchal**, Shyam Lal, and Jyoti Kini “A Novel Dataset and Efficient Deep Learning Framework for Classification of Subtypes of Renal Cell Carcinoma from Kidney Histopathological Images”, *IEEE Transactions on Emerging Topics on Computational Intelligence*, *IEEE Publisher*, [\[Under Revision\]](#), *Indexed by SCI, Thomson ISI, Scopus, Impact Factor: 4.851*

Book Chapter

1. **Amit Kumar Chanchal**, Aman Kumar, Kumar Alabhya, Shyam Lal, and Jyoti Kini, “Efficient Deep Learning Framework with Group Convolution for Segmentation of Histopathology Image.” In Intelligent Computing and Communication Systems, Algorithm for intelligent systems, Springer Singapore 2021, pp. 383-392, ISSN-2524-7565, https://doi.org/10.1007/978-981-16-1295-4_39 [*Published*]

Curriculum Vitae

Name : AMIT KUMAR CHANCHAL

Address : Nai Pokhar Road,
Near Madhya Vidyalaya Rajgir,
Rajgir,
Bihar - 803116, India.
Ph: +919557760612.

Email : amit.chanchal01@gmail.com

Educational Qualification : **M.Tech** in Electronics and Communication Engineering,
National Institute of Technical Teachers Training and Research,
(NITTTR), Chandigarh, Under Ministry of Education, Govt. of India
B.Tech in Electronics & Communication Engineering,
West Bengal University of Technology (WBUT),
Kolkata, India

Teaching Experience : **8 years**
Assistant Professor (14 July 2014 To 15 June 2018)
Electronics & Communication Department,
Moradabad Institute of Technology, Moradabad, UP, India.
Associate Professor (02 December 2022 To Till Date)
Computer Science and Engineering (AI/ML),
KIT's College of Engineering (An Autonomous Institute)
Kolhapur, Maharashtra , India.