# DEVELOPING LIGHTWEIGHT CNN MODELS FOR DETECTING LESIONS IN THE MEDICAL IMAGING

Thesis submitted in fulfillment of the requirements for the Degree of

of

## **DOCTOR OF PHILOSOPHY**

By

### SIVA SKANDHA SANAGALA

Roll Number.: E18SOE841



Department of Computer Science Engineering

#### BENNETT UNIVERSITY

(Established under UP Act No 24, 2016)

Plot Nos 8-11, Tech Zone II,

Greater Noida-201310, Uttar Pradesh, India.

Month: July, Year: 2022

## @ Copyright Bennett University, Greater Noida

Month: July, Year:2022

ALL RIGHT RESERVED



# TABLE OF CONTENTS

DE	CLAR	RATION	N BY THE SCHOLAR	IX
SU	PERV	ISORS	CERTIFICATE	X
PR	EFAC	E AND	ACKNOWLEDGMENT	XI
AB	STRA	CT		XIII
LIS	ST OF	ACRO	NYMS	XV
LIS	ST OF	SYMB	OLS	XIX
LIS	ST OF	FIGUE	RES	XXI
LIS	ST OF	TABLI	ES	XXVII
LIS	ST OF	ALGO	RITHMS	XXIX
1.	INTR	ODUC	TION	1
	1.1	STAT	ISTICS OF LESION	1
	1.2	NON-	INVASIVE TECHNIQUES IMAGING	2
		1.2.1	MAGNETIC RESONANCE IMAGING FOR CAROTID	2
			PLAQUE  COMPUTED TOMOGRAPHY IMAGING FOR CAROTID	2
		1.2.2	PLAQUE	2
		1.2.3	ULTRASOUND IMAGING FOR CAROTID PLAQUE	3
	1.3	STAT	ISTICAL METHODS FOR DETECTION OF LESION	4
	1.4	MACI	HINE LEARNING BASED LESION CLASSIFICATION	5
	1.5	DEEP	LEARNING BASED LESION CLASSIFICATION	7
		1.5.1	2D- CONVOLUTION NEURAL NETWORK	7
		1.5.2	PERFORMANCE EVALUATION METRICES	9
	1.6	TRAN	SFER LEARNING	11
		1.6.1	VGG-16 AND VGG-19	11
		1.6.2	INCEPTION V3	11
		1.6.3	RESNET	12
		1.6.4	DENSENET	13
		1.6.5	MOBILENET	13
		1.6.6	XCEPTIONNET	14
		1.6.7	ALEXNET	14

		1.6.8	SQUEEZENET	15
	1.7		RACTERIZATION OF THE LESION USING MACHINE NING	15
		1.7.1	CHARACTERIZATION OF THE MRI LESION USING MACHINE LEARNING	16
		1.7.2	CHARACTERIZATION OF THE CT LESION USING MACHINE LEARNING	16
		1.7.3	CHARACTERIZATION OF THE US LESION USING MACHINE LEARNING	17
	1.8	COM	PRESSED CNN ARCHITECTURES	18
	1.9	RESE	ARCH GAPS	19
	1.10	OBJE	CTIVES AND THESIS ORGANIZATION	20
2.			ATION AND CHARACTERIZATION OF STROKE USING RNING IN ULTRASOUND SCANS	21
	2.1	INTR	ODUCTION	21
	2.2	BACK	GROUND LITERATURE	23
	2.3	PATII	ENT DEMOGRAPHY AND DATASET	27
		2.3.1	MULTICENTRE PATIENT DEMOGRAPHICS	27
		2.3.2	ULTRASOUND IMAGE ACQUISITION FROM MULTICENTRE STUDY	27
		2.3.3	AUGMENTATION PROTOCOL	27
	2.4	PROP	POSED AI MODELS	28
		2.4.1	PROPOSED 3D-OPTIMIZED DEEP LEARNING MODEL	28
		2.4.2	TWO KINDS OF TRANSFER LEARNING ARCHITECTURES	29
		2.4.3	MACHINE LEARNING ARCHITECTURES	30
	2.5	RESU	LTS AND DISCUSSION	30
		2.5.1	3D-OPTIMIZATION	30
		2.5.2	BENCHMARKING OF THREE DL SYSTEMS AGAINST THE FOUR ML SYSTEMS AND TWO TRANSFER	
			LEARNING THE SYSTEM'S ABILITY TO GENERALIZE AT	31
		2.5.3	DCNN11A5*: MEMORIZATION VS. GENERALIZATION	33
		2.5.4	ATHEROMATIC <sup>TM</sup> PLAQUE SEPARATION INDEX USING DCNN11A5*	

	2.6	CHAR	RACTERIZATION OF THE STROKE LESIONS	34
	2.7	PERF	ORMANCE EVALUATION	36
		2.7.1	DIAGNOSTICS ODDS RATIO	36
		2.7.2	POWER ANALYSIS	36
		2.7.3	HARDWARE & TIMING ANALYSIS	38
		2.7.4	COHEN'S KAPPA ANALYSES	39
	2.8	DISCU	USSION	39
		2.8.1	CLINICAL IMPLICATIONS	39
		2.8.2	STRENGTH/WEAKNESS/EXTENSION	40
	2.9	CONC	CLUSION	40
3.	OPT	MIZEI	O MODEL FOR CLASSIFICATION AND	
			ERIZATION OF THE STROKE IN MULTICENTER	
			ENT	41
	3.1		ODUCTION	41
	3.2		GROUND LITERATURE	42
	3.3	DATA	SET	44
		3.3.1	ULTRASOUND IMAGE ACQUISITION AND PLAQUE DELINEATION	45
	3.4	PROP	OSED CNN& TL MODEL	46
		3.4.1	DL OPTIMIZATION	46
		3.4.2	TL OPTIMIZATION	46
		3.4.3	ML OPTIMIZATION	47
	3.5	RESU	LTS	48
		3.5.1	DL OPTIMIZATION	48
		3.5.2	TL OPTIMIZATION	49
		3.5.3	ML OPTIMIZATION	50
	3.6	CHAR	RACTERIZATION OF LESIONS USING STATISTICAL	
	3.0	AND A	AI METHODS	52
		3.6.1	MEAN FEATURE STRENGTH (MFS)	53
	3.7	PERF	ORMANCE EVALUATION METRICES	54
		3.7.1	DIAGNOSTICS ODDS RATIO (DOR)	54
		3.7.2	COHEN AND KAPPA ANALYSIS	54
		3.7.3	ATHEROMATIC PLAQUE SEPARATION INDEX	55

	3.8	DISC	USSION	55
		3.8.1	MIXED COHORT	55
		3.8.2	STRENGTH/ WEAKNESS/EXTENSIONS	56
	3.9	CONC	CLUSION	56
4.	OPT	MIZEI	MODEL FOR CLASSIFICATION AND	
	_		CRIZATION OF THE STROKE USING TRANSFER	
				59
	4.1		ODUCTION	59
	4.2		HODOLOGY FOR CLASSIFICATION OF STROKE USING ISFER LEARNING METHODS	61
		4.2.1	DEEP LEARNING ARCHITECTURE	61
		4.2.2	EXPERIMENTAL PROTOCOL	63
	4.3	RESU	LTS	63
		4.3.1	3D OPTIMIZATION OF TL ARCHITECTURES AND BENCHMARKING AGAINST CNN	63
		4.3.2	3D OPTIMIZATION OF MODIFIED UNET	65
		4.3.3	VISUALIZATION OF THE MODIFIED UNET	65
	4.4	PERF	ORMANCE EVALUATION	66
		4.4.1	POWER ANALYSIS	66
		4.4.2	RANKING OF AI MODELS	67
		4.4.3	BENCHMARKING THE PERFORMANCE WITH EXISTING	67
	4.5	SCIE	NTIFIC VALIDATION VS. CLINICAL VALIDATION	68
		4.5.1	SCIENTIFIC VALIDATION USING HEATMAPS	68
		4.5.2	CORRELATION ANALYSIS	69
	4.6	DISC	USSION	72
		4.6.1	ADVANTAGES OF TL MODELS	72
		4.6.2	GUI DESIGN	75
		4.6.3	STRENGTH/WEAKNESS/EXTENSIONS	76
	4.7	CONC	CLUSION	76
5.			ATION AND CHARACTERIZATION OF THE STROKE	
			BRID DEEP LEARNING MODELS	<b>78</b>
	5.1		ODUCTION	<b>78</b>
	5 2	DATA	SET	70

	5.2.1	ULTRASOUND IMAGE ACQUISITION FROM MULTICENTRE STUDY	86
	5.2.2	AUGMENTATION PROTOCOL	8
5.3	METI	HODOLOGY	8
	5.3.1	PART A: ONE HYBRID DEEP LEARNING ARCHITECTURE USING TWO SDL	8
	5.3.2	PART B: TEN HYBRID DEEP LEARNING ARCHITECTURES	8
	5.3.3	PART C: FIVE TRANSFER LEARNING MODELS - A CLASS OF DL.	8
	5.3.4	PART D: ONE SOLO DEEP LEARNING MODEL	8
	5.3.5	LOSS FUNCTIONS	8
5.4	EXPE	CRIMENTAL PROTOCOL	8
	5.4.1	ACCURACY BAR CHARTS FOR ALL EACH COHORT CORRESPONDING TO ALL HDLS	8
	5.4.2	MEAN FEATURE STRENGTH FOR EACH COHORT USING A CEL LOSS FUNCTION	8
	5.4.3	CROSS-VALIDATION PROTOCOL	8
	5.4.4	COMPUTATIONAL EFFICIENCY USING SUPERCOMPUTER FRAMEWORK	8
5.5	RESU	JLTS	8
	5.5.1	RESULTS OF HDL OPTIMIZATION ON THREE COHORTS	8
	5.5.2	RESULTS OF 17 AI MODELS USING THREE COHORTS FOR BEST LOSS FUNCTION AND BEST AUGMENTATION	8
5.6	PERF	ORMANCE EVALUATION	9
	5.6.1	ROC ANALYSIS	9
	5.6.2	POWER STUDY	9
	5.6.3	ATHEROMATIC PLAQUE SEPARATION INDEX	9
	5.6.4	STATISTICAL TEST	9
5.7	DISC	USSION	9
	5.7.1	A SHORT NOTE ON HARDWARE RESOURCES	9
	5.7.2	A SHORT NOTE ON HDL VS. DL AND HDL VS. TL	9
5.8	CONG	CLUSION	q

6.			IGHT CNN MODEL FOR IDENTIFICATION OF LESION SING CT SCANS	96
	6.1	INTR	ODUCTION	96
		6.1.1	MULTI-NODULE CLASSIFICATION	97
	6.2	BACE	KGROUND LITERATURE	98
	6.3	DATA	ASET	99
		6.3.1	PATIENT DEMOGRAPHICS	99
		6.3.2	AUGMENTATION	10
		6.3.3	DATASET FOR MULTINODULE DETECTION	10
	6.4		POSED MODEL FOR LUNG LESION CLASSIFICATION CHARACTERIZATION	10
		6.4.1	MODIFIED CNN	10
		6.4.2	TRANSFER LEARNING ARCHITECTURE	10
		6.4.3	MACHINE LEARNING ARCHITECTURE	10
		6.4.4	PROPOSED MODEL FOR MULTI NODULE DETECTION	10
	6.5	RESU	ULTS	10
		6.5.1	RESULTS FOR MULTI-NODULE LESION CLASSIFICATION	10
		6.5.2	CHARACTERIZATION OF THE LUNG LESION	11
		6.5.3	HEAT MAPS	11:
	6.6	PERF	FORMANCE EVALUATION	11
		6.6.1	POWER STUDY	11
		6.6.2	DIAGNOSTICS OF ODDS RATIO	11
		6.6.3	RANKING	11
		6.6.4	MANN-WHITNEY TEST AND PAIRED T-TEST	11
	<b>6.7</b>	DISC	USSION	12
		6.7.1	BENCHMARKING	12
		6.7.2	HIDDEN LAYER VISUALIZATION	12
		6.7.3	MEMORIZATION VS. GENERALIZATION	12
		6.7.4	EFFECT OF TRAINING ON DATASIZE	12
	6.8	CON	CLUSION	12
7.	LIG	HT CN	N MODEL FOR IDENTIFICATION OF LESION IN	
	WIL	SON D	ISEASE USING MRI SCANS	12
	7 1	INTD	ODUCTION	12

	7.2	BACK	GROUND LITERATURE	126
	7.3	DATA	SET AND PRE-PROCESSING	128
		7.3.1	DATASET	128
		7.3.2	DATA AUGMENTATION PRE-PROCESSING: SKULL AND BACKGROUND	128
		7.3.3	REMOVAL	128
	7.4	METH	HODOLOGY	129
		7.4.1	GLOBAL ARCHITECTURE: TRANSFER LEARNING	129
		7.4.2	ALEXNET	132
		7.4.3	RESNET50	133
		7.4.4	DENSENET161	134
		7.4.5	XCEPTIONNET	135
		7.4.6	INCEPTION-V3	135
		7.4.7	MODIFIED UNET	136
	7.5	RESU	LTS	136
		7.5.1	TL OPTIMIZATION	136
		7.5.2	CHARACTERIZATION	138
	7.6	DISCU	USSION	139
		7.6.1	BENCHMARKING	139
		7.6.2	A SHORT NOTE ON WD CHARACTERIZATION	140
		7.6.3	POWER ANALYSIS	141
	7.7	CONC	CLUSION	141
8.	COM	PRESS	ION OF THE OPTIMIZED CNN MODEL USING GENETIC	
	ALG	ORITH	M	143
	8.1	INTRO	ODUCTION	143
		8.1.1	INTRODUCTION TO GENETIC ALGORITHM	144
		8.1.2	INTRODUCTION TO PRE-TRAINED MODELS	145
	8.2	BACK	GROUND LITERATURE	148
	8.3	METE	HODOLOGY	150
		8.3.1	CHROMOSOME REPRESENTATION	151
		8.3.2	FITNESS FUNCTION	152
		8.3.3	SELECTION	154
		8.3.4	CROSSOVER	154
		8.3.5	MUTATION	155
	8.4	RESU	LTS	156

## TABLE OF CONTENTS

		8.4.1	DATASETS	156
		8.4.2	STORAGE SPACE COMPARISON	156
		8.4.3	INFERENCE TIME COMPARISON	160
		8.4.4	COMPRESSION OF PRE-TRAINED ARCHITECTURES TRAINED ON THE LIDC-IDRI DATASET	161
		8.4.5	COMPRESSION CNN TRAINED ON MEDICAL DATASET (LIDC-IDRI)	161
	8.5	DISC	USSION	165
		8.5.1	PRINCIPAL FINDING	165
		8.5.2	A SPECIAL NOTE ON THE COMPRESSION DL MODEL	166
	8.6	CON	CLUSION	166
9.	CON	CLUSI	ONS AND FUTURE WORK	169
		9.1	CONCLUSION	169
		9.2	FUTURE WORK	172
	BIBI	LOGR	APHY	173
	LIST	OF PI	IRLICATIONS	193

**DECLARATION BY THE SCHOLAR** 

I hereby declare that the work reported in the Ph.D. thesis entitled "**Developing**"

lightweight CNN models for detecting lesions in the medical Imaging"

submitted at Bennett University, Greater Noida, India, is an authentic record

of my work carried out under the supervision of Dr. Suneet Kr. Gupta and

Dr Vijaya Kumar Koppula. I have not submitted this work elsewhere for

any other degree or diploma. I am fully responsible for the contents of my Ph.D.

Thesis.

Signature of the Scholar

Name: Siva Skandha Sanagala

Department of Computer Science Engineering

Bennett University, Greater Noida, India.

Date: 08-07-2022

ix

SUPERVISORS' CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled "Developing

lightweight CNN models for detecting lesions in the medical Imaging",

submitted by Siva Skandha Sanagala at Bennett University, Greater Noida,

**India**, is a bonafide record of his original work carried out under our supervision. This

work has not been submitted elsewhere for any other degree or diploma.

Signature of Supervisor

Dr. Suneet Kr. Gupta

Bennett University, India.

Date: 08-07-2022

Signature of Co-Supervisor

Dr. Vijaya Kumar Koppula

CMRCET, Hyderabad

Date: 08-07-2022

X

#### PREFACE AND ACKNOWLEDGMENT

The objective of this thesis is to study the application of deep learning for the classification and characterization of a lesion in medical imaging. Firstly, it has been shown that convolution neural networks can classify and characterize the lesions and their components. Secondly, several attempts have been made to compress popular CNN architectures to develop low-order models for classification and characterization.

First and foremost, I would like to thank my divine Guru Raghavendra Swami for his grace and for providing me with the strength to embark on this difficult endeavor and the chance to register in a PhD programme at Bennett University, and for making me capable of completing its prerequisites.

I owe a debt of gratitude to my parents Rama Hanuman and Vijaya Lakshmi for their unwavering support and unconditional love. Without their affection, it would not have been possible to make it here. A very special thanks to my loving and caring wife, Sri Maha Lakshmi, for her love, affection, continuous support, and encouragement. I am grateful to my life's love, daughter Aadhya, for bringing me joy in life. I thank my Aunt Annapurna, Mother-in-law B Padma, Father-in-Law B Vijay Kumar, brothers-in-law B Manikanta, and B Chandrasekhar. I thank my grandparents, Late Sri Lakshmi Narasimha Sastry, and Kali Parameshwary, for their blessings. I also thank all my relatives for their moral support, best wishes, and love.

My heartfelt special thanks and gratitude to my supervisor, Dr. Suneet Kr. Gupta, and Dr. Vijaya Kumar Koppula. Dr. Suneet Kumar's vast scientific knowledge, as well as his remarkable personal traits and experience, have always helped me to carry out this research work with ease. Dr. Vijaya Kumar Koppula has piqued my interest in the subject of applying Deep Learning to categorize lesions, and he has generously shared his enthusiasm for enhancing the research in a better way. It was a huge honor for me to get an opportunity to collaborate and work with them. Both have taught me the fundamentals of carrying out research in a scientific and efficient way.

I express my sincere thanks to Prof. Deepak Garg, the Dean, School of Computer Science Engineering and Technology, Dean, International Affairs and Corporate Outreach, for the continual source of encouragement and support for me. I would also like to thank members

of my Research Advisory Committee, Dr. Simranjit Singh, and Dr. Souradyuti Ghosh for their valuable remarks and assistance at different phases of my work. I owe my thanks to senior Dr. Mohit Agarwal for introducing and briefing NVIDIA computing and enlightening me on a variety of deep learning models. I am grateful to Prof. Jasjit Suri and his team for providing me with their deep insights on health care and computer-aided systems for diagnosis.

I'd like to express my gratitude to every one of the department faculty and staff of Computer Science Engineering (CSE) at Bennett University (The Times of India Group) Greater Noida for their assistance and support. I am grateful to the Chancellor, Vice-Chancellor, Bennett University, the Registrar, Dean (Academic), Dean (School of Engineering & Applied Sciences), and Dean (R & C) for their moral support and for providing me with all the essential facilities during my candidacy.

I appreciate the continued assistance and encouragement from CMR College of Engineering & Technology secretary and correspondent Shri. Ch Gopal Reddy, Principal Major Dr. V A Narayana, and Head of the Department Dr. Vijaya Kumar Koppula in completing my PhD. All members of the CSE department's faculty and non-teaching personnel for their inspiration and support.

I am thankful to my friends G. Hemanth Kumar, K. Purna Chandu, L. Suresh Kumar, Nitin Lomte, and Ch. Raj Kishore for emotional support and balance in my life. Special thanks to Mohit Agarwal, my senior, and Archana Bathula, my junior, for their motivation and bringing in confidence in me. Finally, I thank all my teachers from my childhood to PhD submission. Without them, I would have been unable to experience the taste of knowledge.

I thank all the people in my life for their encouragement and inspiration in making my dreams a reality.

Siva Skandha Sanagala

Department of Computer Science Engineering
Bennett University, Greater Noida, India

Date:

#### **ABSTRACT**

Classification and characterization of the lesions in medical imaging are getting interested in the field of computer-aided diagnostics (CAD) systems for radiology. A num ber of machine learning (ML) algorithms are proposed for classification and characterization. All the existing models suffer from inter-observer variability. CNN models are based on Deep Neural Networks, which are distinct from typical Machine Learning methods such as k-NN, Decision-Trees, and so on. Moreover, the performance of Deep Neural Network-based approaches is better as compared to traditional Machine Learning approaches as these models extract the features from training data automatically. In the past, the researchers have proposed many CNN architectures such as VGG16, VGG19, Inception V3, MobileNet, ResNet50, etc. for the classification of 1000 class ImageNet datasets. These models can also be utilized for the classification of other datasets by transfer learning. While pre-trained CNN models work well, the high number of parameters required makes them computationally intensive.

In this thesis, we proposed several simplified and optimized CNN models for the classification and characterization of the lesions in the carotid artery, lung, and brain (Wilson disease). The datasets are taken from the medical practitioners at different geological locations and several publicly available datasets. We got the ground truth from the radiologist for real-time data. Then we augmented the data into folds. In these cohorts, we ran all of the recommended models, including CNN models, Transfer Learning (TL) models, and Hybrid CNN models. We achieved an accuracy of 95.66% in carotid using optimized CNN and 99.45% accuracy using hybrid CNN, 98.9% in lung, and 91.2% in Wilson disease. We validated our models and hypothesis on the characterization of the multicenter study. We compared the performance of the AI models on various performance metrices and then ranked the models based on the grading schema for the best classifier. We tested the performance of the models on various cross-validation protocols and on different hardware resources. We achieved a prediction time on the local computer of less than 2 sec for all the datasets.

In the next step, we characterized the lesions using a novel deep learning method called mean feature strength (MFS). It is validated using statistical approaches, histograms, fractal dimensions, and image processing techniques such as higher-order spectrum. We hypothesized that symptomatic lesions are dark, patchy, and contain high grayscale medium (GSM) compared to the asymptomatic lesion which appears bright and contains low GSM. Whereas in lung lesions the benign lesions appear dark and patchy, and malignant lesions are bright and smooth. All the hypothesis about the lesions

is validated at carotid plaques, lung nodules, and Wilson disease lesions. We evaluated the risk of the lesions using DL-based models and generated the heatmaps and validated them with correlation analysis with ground truth. We build a CAD system for detection of the risk of the lesions which gives the type of lesion and heatmaps of the lesions.

Finally, we compressed the proposed models using genetic algorithms and implemented them in edge devices. The proposed approach demonstrated the reduction in the storage space of AlexNet by 87.62%, 80.97%, and 86.20% corresponding to the data sets MNIST, CIFAR-10, and CIFAR-100, respectively. Further, for the same three datasets, VGG16, ResNet50, and SqueezeNet, the system average compression was 91.15%, 78.42%, and 38.40%, respectively. In addition to that, the inference time of the models using the proposed strategy was significantly improved with an average of the three datasets of ~35.61%, 9.23%, 73.76%, and 79.93% corresponding to AlexNet, SqueezeNet, ResNet50, and VGG16 models. Further, our method when applied to the proposed CNN using the LIDC-IDRI dataset showed a 90.3% reduction in the storage space and 37% inference time. All the experiments are validated with power analysis, diagnostics odds ratio, ROC Analysis, Mann- Whitney and paired t-test.

### LIST OF ACRONYMS

AC Attenuation Coefficient

AI Artificial Intelligence

ALZ Alzheimer

AP Average Pooling

APSI Atheromatic<sup>TM</sup> Plaque Separation Index

ASCVD Atherosclerotic Cardiovascular Disease

AUC Area Under the Curve

BN Batch Normalization

BRB Bottleneck Residual Blocks

BT Brain Tumour

CAD Computer Aided Design

cCNN Convolution Neural Network

CEL Cross-Entropy Loss

CL Convolution Layer

CNN Convolution Neural Network

CSV Comma-Separated Value

CT Computed Tomography

CTA Computed Tomography Angiography

CUS Carotid Ultrasound Scans

CV Cross-Validation

CVD Cardiovascular Disease

DCNN Deep Conventional Convolutional Neural Networks

DE Dual Energy

DL Deep Learning

DNA Deoxyribonucleic Acid

DOR Diagnostics Odds Ratio

DT Decision Tree

DT Decision Tree

DWI Diffusion-Weighted Imaging

DWT Discrete Wavelet Transform

DY Diagnostic Yield

EAI Enhanced Activity Index

EL Extreme Learning

FCN Fully Connected Network

FD Fractal Dimension

FFT Fast Fourier Transform

FM Feature Map

FOS First Order Statics

FPR False Positive Rate

GA Genetic Algorithm

GAP Global Average Pooling

GBDT Gradient Boosting Decision Tree

GGO Ground-Glass Opacification

GLCM Gray Level Co-Occurrence Matrix

GLDS Gray Level Different Statistics

GSM Grayscale Median

GUI Graphical User Interface

HDL Hybrid Deep Learning

HL Hinge Loss

HOGs Histograms Of Oriented Gradients

HOS Higher Order Spectrum

HU Hounsefield Unit

ICA Internal Carotid Artery

ILSVRC ImageNet Large-Scale Visual Recognition Challenge

IPH Intraplaque Hemorrhage

IST Instituto Superior Técnico

IV3 Inceptionv3

*k*-NN *K*-Nearest Neighbor

LBP Local Binary Pattern

LCC Lesion Characterization and Classification

LDA Linear Deterministic Analysis

LisD Lisbon Dataset

LonD London Dataset

LRNC Lipid-Rich/Necrotic Core

LTE Law's Texture Energy

MCI Mild Cognitive Impairment

mCNN Modified Convolution Neural Network

MFS Mean Feature Strength

MixD Mixed Dataset

MPP Mean Of the Positive Pixels

MRI Magnetic Resonance Imaging

MSEL Mean-Squared-Error Loss

MUV M-Mode Ultrasound Videos

NB Naïve Bayes

NORD National Organization for Rare Disease

PCC Pearson Correlation Coefficient

PEM Performance Evaluation Metrices

PGLVQ Parametrized Generalized Learning Vector Quantization

QDA Quadratic Deterministic Analysis

RC Residual Connection

RF Radio Frequency

RF Random Forest

RF Random Forest

RMM Rayleigh Mixture Model

ROC Receiver Operating Characteristics

ROI Region Of Interest

SACI Symptomatic And Asymptomatic Carotid Index

SDL Solo DL

SGD Stochastic Gradient Descend

SGLDM Spatial Gray Level Dependence Matrices

SOM Self-Organized Maps

SS Single Source

SVM Support Vector Machine

TIA Transient Ischemic Attack

TL Transfer Learning

TPR True Positive Rate

UPP Uniformity Of the Positive Pixels

### LIST OF ACRONYMS

US Ultrasound

USA United States of America

VGG Visual Geometry Group

WD Wilson's Disease

WHO World Health Organization

WMH White Matter Hyperintensity

### LIST OF SYMBOLS

D<sub>fd</sub> Fractal Dimension

N Number Of Miniature Pieces

S Scaling Factor

x<sub>i</sub> I<sup>th</sup> Sample of X Variable

y<sub>i</sub> I<sup>th</sup> Sample of Y Variable

 $\overline{x}$  Mean of X Variables

 $\bar{y}$  Mean of Y Variables

*K*<sub>a</sub> Cohen And Kappa

P<sub>0</sub> Relative Observed Agreement Among Raters

Pe Hypothetical Probability of Agreement

K Size of Image M x N

 $w_{out}$  Width of the Output

 $w_{in}$  Width Size of Input

*f* Filter Size

s Stride

L(a, y) Loss Function

y Expected Output

a Predicted Output

TP True Positive

TN True Negative

FP False Positive

FN False Negative

 $f_1$  F1 Score

TPR True Positive Rate

FPR False Positive Rate

z\* Standard Z- Table

*p* Data Proportion

MoE Margin of Error

MFS AF Mean Feature Strength Of AF

MFS<sub>AS</sub> Mean Feature Strength Of AS

 $\bar{\eta}(m)$  Mean Accuracy

η(m, k) Accuracy of M Model at K Combination

CEL Cross Entropy Loss

MSE Mean Squared Error

S<sub>a</sub>(x) Swish Activation Function

 $Z_{1-\beta}$  Type II Error

σ Standard Deviation

Δ Mean Difference

γ The Number of Hidden Units

δ Number of Hidden Units

Fc Full Corrected

w Weight

## LIST OF FIGURES

Figure Number	Caption	Page Number
1.1	Imagining Machines of (a) MRI, (b) CT, and (c) Carotid ultrasound scanning using a linear probe (MRI/CT imagescourtesy of Dr. Luca Saba, Italy, and US image, courtesy of AtheroPoint, CA, USA).	4
1.2	VGG16 and VGG19 architectures, CONV: Convolution Layer, FC: fully connected network.	11
1.3	InceptionV3 architecture.	12
1.4	ResNet Architecture.	13
1.5	DenseNet Architecture with three dense blocks and three transition blocks, followed by fully connected network. Post processing is the SoftMax.	13
1.6	MobileNet Architecture, BRB: bottleneck and residual blocks.	14
1.7	XceptionNet Architecture.	14
1.8	AlexNet architecture.	15
1.9	SqueezeNet Architecture.	16
1.10	Local assessment of plaque echo-morphology, in terms of (a-c) hypo echogenicity and (d) heterogeneity [72]	18
2.1	Sample (a) symptomatic (b) asymptomatic US cut sections of the carotid plaque.	28
2.2	DCNN11 architecture (CL: convolution layer, APL: average pooling layer).	29
2.3	Transfer learning using "ImageNet" for plaque classification.	29
2.4	The machine learning architecture adopted in this study.	30
2.5	Left: 3-D optimization showing the best DCNN and augmentation combination. This corresponds to 11 DCNN layers and a 5-fold augmentation and is denoted as DCNN11A5*. Right: Bar chart showing optimized location.	31
2.6	(a) Comparison of the accuracy rates of all AI methods at the optimization point, (b) ROC curves of all AI methods.	32
2.7	Memorization vs. generalization corresponding to DCNN11A5* using K10 protocol.	33
2.8	3-D plot of APSI vs. features vs. accuracy for four types of ML classifiers.	34
2.9	Mean feature strength of DCNN11A5*.	35
2.10	Histogram of symptomatic vs. asymptomatic plaques for the 5-fold augmentation data.	35

2.11	3-D plot of Bispectrum. (a) symptomatic plaque (b) asymptomatic plaque.	36	
3.1	Perspective of clusters on carotid plaque study.	44	
011	(a) Sample ultrasound scans of 4 symptomatic patients. Row 1,		
3.2	Row3 represents full scan Row 2, Row 4 represents corresponding delineated plaque of London and Lisbon dataset	45	
	(b) Sample ultrasound scans of 4 asymptomatic patients. Row 1, Row3 represents full scan Row 2, Row 4 represents corresponding delineated plaque of London and Lisbon dataset.		
3.3	The global architecture of the DCNN model.	47	
3.4	The global architecture of the TL model.	47	
3.5	Online and offline ML Architecture.	48	
3.6	3-D optimization curves of three centers.	48	
3.7	3-D optimization of VGG19.	49	
3.8	ML 3-D Optimization using (a) LBP Features, (b) HOG Features (c) Hu-Moments.	50	
3.9	ROC curves of three centers.	51	
3.10	(a) Grayscale intensity histograms of three datasets. (b) 3-D Bispectrum of corresponding datasets.	53	
3.11	MFS of three centers at each layer in optimized DCNN.	54	
	Online AI architecture of the Atheromatic <sup>TM</sup> 2.0TL study (TL:		
4.1	Transfer Learning, DL: Deep Learning, and Grad-CAM: Gradient-weighted Class Activation Mapping).	60	
	Global TL architecture using 10 different TL models ((i-ii) Visual		
	Geometric Group-16, 19 (VGG16, 19); (iii) Inception V3 (IV3);		
4.2	(iv-v) DenseNet121, 169; (vi) XceptionNet; (vii) ResNet50; (viii)	61	
	MobileNet; and (ix) AlexNet, and (x) SqueezeNet), te, stands for		
	testing and tr stands for training, FN: fine-tune networks.		
4.3	Modified Unet architecture.	62	
4.4	3D bar chart representation of the AI model accuracy vs. augmentation folds.	64	
4.5	(a) Visualization of the intermediate layers of Modified Unet on symptomatic class (b) asymptomatic class.	66	
4.6	ROC comparison of 12 AI models (10 TL and 2 DL).	68	
	Heat maps of the symptomatic plaque (left) and asymptomatic		
4.7	plaque (right).	70	
	(a), (b) and (c) are the symptomatic image heatmaps vs. original		
4.8	images, (d), (e), and (f) are asymptomatic image heatmaps vs.	<b>71</b>	
- • •	original images (Red color arrow represents the important regions).	7.1	
4.0	Correlation of AI (Modified Unet) and the three biomarkers -	71	
4.9	FD, GSM, and HOS.	<b>71</b>	

4.10	GUI screenshot of the Atheromatic <sup>™</sup> 2.0TL system.	<b>76</b>	
	Representation of 11 HDL models (i) fusion of Inception and		
5.1	ResNet (1 HDL model) and (ii) fusion of 10 ML models with	<b>79</b>	
	CNN, leading to 10 HDL models.		
5.2	Sample (a) symptomatic (b) asymptomatic US cut sections of the	81	
3.2	carotid plaque.	01	
5.3	HDL-IRN model using two SDL (Inception+ResNet).	81	
	Cascading of ML models with SDL for the HDL2 to HDL11		
5.4	model design architecture, a class of Atheromatic™ 2.0HDL. The	02	
5.4	feature extraction process is discussed below showing four CL	83	
	and four APL layers, leading to vector design in the end.		
5.5	Generalized transfer learning architecture.	83	
5.6	Solo deep learning architecture.	84	
	Optimization of 11 HDL models using three kinds of loss function		
5.7	on three kinds of cohorts (LonD, LisD, and MixD). (a) CE, (b)	00	
5.7	HL, (c) MSE. HDL7 is best for CEL and MSE, while HDL2 is	89	
	best for HL.		
	Comparison between 17 AI models over multicenter cohorts		
5.8	using the best loss function (CEL vs. HL vs. MSEL) and best	90	
	augmentation data for each of the HDL 17 categories.		
	Mean feature strength (taken over HDL2-HDL11) for 5		
5.9	augmentation states corresponding to (a) LonD, (b) LisD, and (c)	<b>92</b>	
	MixD cohorts (Blue: symptomatic, Orange: asymptomatic).		
5.10	ROC analysis for 17 AI model using the three cohorts: (a) LonD,	93	
3.10	(b) LisD, and (c) MixD.	93	
6.1	Sample of LIDC-IDRI dataset [34].	100	
6.2	Sample of Images of ELCAP Dataset (a) Single Nodule (b) Multi	101	
0.2	Nodule (c) Healthy.	101	
6.3	The global architecture of deep learning architecture.	102	
6.4	Global TL architecture.	103	
6.5	ML system architecture.	104	
6.6	Proposed CNN Architecture.	104	
	(a) Original image and Activations after (b) First convolution		
6.7	layer (c) Second convolution layer (d) Third convolution layer	105	
	and (e) fourth convolution layer.		
6.8	Training accuracy graph with epochs.	106	
6.9	Validation accuracy graph with epochs.	106	
6.10	3D surface plots of optimization of cCNN and mCNN. The blue	107	
0.10	color dot represents the optimization point.	10/	
6.11	(a) Comparison of proposed AI models at an optimization point	107	
0.11	(Aug5x). (b) ROC Curves of 22 AI Techniques with AUC.		

	Specificity for traditional machine learning algorithms i.e.,			
6.12	decision tree, logistic regression, k-NN, LDA, Naive Bayes,	107		
	Random Forest, and Proposed model.			
	Sensitivity for traditional machine learning algorithms i.e.,			
6.13	decision tree, logistic regression, k-NN, LDA, Naive Bayes,	108		
	Random Forest, and Proposed model.			
	ROC-AUC for healthy lung images with 3 features i.e., HSV, Hu-			
6.14	moments, Haralick features as input in machine learning	109		
	algorithms.			
	ROC-AUC for healthy lung images with 3 features i.e., HSV, Hu-			
6.15	moments, Haralick features as input in machine learning			
	algorithms.			
	Cohen Kappa for traditional machine learning algorithms i.e.,			
6.16	decision tree, logistic regression, k-NN, LDA, Naive Bayes,	110		
	Random Forest, and Proposed model.			
6.17	The architecture of VGG16.	110		
6.18	The architecture of Inception V3.	111		
6.19	Accuracy comparison among Inception V3, VGG16 and	111		
0.17	proposed CNN model.	111		
6.20	Inference time comparison among Inception V3, VGG16 and	112		
0.20	proposed CNN model.	112		
6.21	Space comparison among Inception V3, VGG16 and proposed	112		
0.21	CNN model.	112		
6.22	Histogram of (a) healthy and (b) malignant lung nodule.	113		
6.23	Fractal dimension of (a) healthy and (b) malignant.			
6.24	Higher-order spectra of (a) healthy and (b) malignant. The red	114		
0.27	arrow is pointing to the peak of the spectrum.	117		
6.25	Heat Maps of the malignant and healthy class generated from	114		
0.23	VGG-19 Grad-Cam Technique.	117		
6.26	DOR vs. Accuracy comparison of 22 AI techniques. The red color	116		
0.20	arrow represents the highest DOR (mCNN).	110		
6.27	Intermediate layer visualization top (mCNN) DL architectures.			
6.28	Memorization vs. generalization of mCNN11 at Augment Fold 5.			
6.29	The training time of mCNN11 at augmented fold 5 with varying	122		
0.27	CV protocols.	122		
7.1	Control (left) and WD (right). Both images show the skull with	129		
	removal of background.	130		
7.2	Samples of six patient segmented MRI scans of Control dataset.			
7.3	Samples of six patient segmented MRI scans of the diseased	131		
	dataset.			
7.4	Global Architecture of our study.	132		
7.5	AlexNet Architecture.	133		

7.6	Resnet50 architecture.				
7.7	DenseNet161 architecture.				
7.8	Proposed architecture of XceptionNet.				
7.9	Modified architecture of InceptionV3.				
7.10	Modified Unet architecture.				
7.11	Optimization of TL and Modified Unet with Augmentation.	137			
7.12	ROC analysis of six AI models.				
7.13	Mean Feature Strength of Modified Unet.				
7.14	(a) Mean Bispectrum analysis (b) Higher-order spectra of the Wilson disease.				
8.1	The flowchart of GA.	145			
8.2	AlexNet architecture [203].				
8.3	Architecture of VGG16 [204].	147			
8.4	Architecture of ResNet50.	147			
8.5	The architecture of SqueezeNet.	148			
8.6	The architecture of the fire module.	148			
8.7	Process for generating chromosome for a three-layer CNN				
8.8	Two randomly generated chromosomes (a) and (b) for the CNN architecture presented in Figure 8.7				
8.9	The fitness values of randomly generated chromosomes (green color dots indicate the accuracy of chromosome greater than 60%, blue color dots indicate chromosome accuracy greater than 40%, and less than 40% represented as red color).				
8.10	Crossover operation.				
8.11	Sample images of LIDC- IDRI datasets for four parents.	157			
8.12	Variations in accuracy (%) values while executing GA for 10 epochs over different datasets (i.e., CIFAR-100, CIFAR-10 and MNIST) for (a) VGG16 (b) ResNet50 (c) SqueezeNet (d) AlexNet.				
8.13	Variations in fitness values while executing GA for 10 epochs over different datasets (i.e., CIFAR-100, CIFAR-10 and MNIST) for (a) VGG16 (b) ResNet50 (c) SqueezeNet (d) AlexNet.				
8.14	Variations in fitness values vs. Accuracy while executing GA for 10 epochs over different datasets with weight(w) is equal to 0.99 (i.e., CIFAR-100, CIFAR-10 and MNIST) for (a) VGG16 (b) ResNet50 (c) SqueezeNet (d) AlexNet.				
8.15	(a) Input image, (b) activation before compression, and (c) activation after compression in the first set of the convolution layer in VGG16 on the MNIST dataset.				
8.16	Hidden units before and after compression in a customized CNN on the LIDCIDRI dataset.				

## LIST OF FIGURES

0 17	Hidden units before and after in (a) ResNet50, (b) VGG16, (c)	165
8.17	SqueezeNet and (d) AlexNet on the LIDC-IDRI dataset.	105

## LIST OF TABLES

Table Number	Caption					
2.1	Similarities and differences between imaging modalities using AI.					
2.2	Six types of DCNN models consisting of different combinations of convolution, average pooling, and dense layers.					
2.3	DCNN Layers vs. Augmentation Folds vs. Accuracy (%).	31				
2.4	Benchmarking Table of seven artificial intelligence methods for plaque classification.					
2.5	APSI vs. features vs. accuracy for four types of ML classifiers.	34				
2.6	DORs for all the AI models using different features.	37				
2.7	Comparison between LC and SC specifications.	38				
2.8	Timing analysis and time reductions for the SC against the LC.	39				
2.9	Cohen's kanna analysis for AI methods (increasing order of					
3.1	Centers vs. Layers vs. Augmentation vs. Accuracy.	<b>49</b>				
3.2	Centers vs. TL vs. Augmentation vs. Accuracy.					
3.3	Machine learning techniques Feature vs. Model vs. Augmentation vs. Accuracy.					
3.4	Histogram Lobe Area.					
3.5	DOR and Ka Analysis at optimization point (Increasing order at individual centers).					
3.6	APSI of AI methods used in this study.					
4.1	Modified Unet architecture parameters.					
4.2	Accuracies of 10 TL and 2 DL models for 6 augmentations.					
4.3	TL systems vs. DL Systems.	65				
4.4	Ranking of the AI models.	67				
4.5	Correlation analysis.					
4.6	Euclidean distance between biomarker pairs.					
4.7	Benchmarking Table.					
4.8	Comparison of TL models.					
4.9	Similarities and differences between the TL models.	73 74				
5.1	Accuracies of three cohorts under unbalanced vs. balanced vs. optimized conditions for HDL7 (SDL+DT) using CEL and K10 protocol.					
5.2	Accuracy comparison between 11 HDLs and 6 Non-HDLs (1 DL + 5 TL).					
5.3	Atheromatic Plaque Separation Index for 12 HDL models.					
6.1	Details of the augmentation.					
6.2	Hyper-parameter tuning.					
6.3	Accuracy comparison of AI models.					

6.4	Performance Comparison with Other Machine Learning.			
6.5	DOR vs. Accuracy (Descending order of DOR).			
6.6	Attributes and range values.			
6.7	Gradient scheme for ranking the AI models.			
6.8	Mann-Whitney test between AI models.			
6.9	Mann-Whitney test between ML features.			
6.10	Paired sample t-test between AI models.			
6.11	Benchmarking Table with existing work.			
7.1	3D-Optimization of TL architectures and Modified U-Net.			
7.2	Benchmarking Table with existing systems.			
8.1	Parameters used for Genetic Algorithm.	156		
8.2	Comparison of statistics with different models in terms of size on MNIST dataset.			
8.3	Comparison of statistics with different models in terms of size on the CIFAR-10 dataset.			
8.4	Comparison of statistics with different models in terms of size on the CIFAR-100 dataset.			
8.5	Performance comparison in terms of inference time after compression on MNIST dataset.			
8.6	Performance comparison in terms of inference time after compression on the CIFAR-10 dataset.	163		
8.7	Performance comparison in terms of inference time after compression on the CIFAR-100 dataset.	163		
8.8	Comparison of static with different models and datasets for comparison using 1-point crossover and recombination crossover.	163		
8.9	Set of hyper parameters.	164		
8.10	Comparison of 9-layer customized CNN model before and after compression.	164		
8.11	Inference time (in sec) of prediction on different OS.	166		

## LIST OF ALGORITHMS

1:	Generation of the Chromosome	142
----	------------------------------	-----

### **CHAPTER 1**

#### INTRODUCTION

Lesions may occur at any part of the body [1] and it is developed due to abnormal mass or swelling in the tissue. There are two types of lesions: benign lesions which are non-cancerous and malignant lesions which are cancerous. Malignant lesions refer to the cells which grow out of control and invade other cells [1]. Infections, genetic inheritance, environmental factors, and poor lifestyles strengthen malignant lesions [2] and these factors also damage Deoxyribonucleic acid (DNA) and lead to cancer. However, all these DNA damages will be identified by the cells automatically and repaired. If the cells are damaged, they cannot repair themselves and it is so-called "cell of death." Cancer occurs when these damaged cells spread abnormally. In this study, we considered the lesions formed at the carotid, lung, and Wilson disease. Wilson disease is a rare disease effected by hereditary. Due to the recent outbreak of SARS-Cov2 called as Covid-19 causing the highest death rates. This research work contains the study of these diseases.

#### 1.1 Statistics of Lesion

Among all cardiovascular disease (CVD), lung cancer causes higher mortality rates. CVD's leading cause is atherosclerosis which occurs due to lesions in the major arteries such as carotid [3]. There are several risk factors for CVD and atherosclerosis, including hypertension, hyperlipidemia, diabetes, obesity, smoking, and a sedentary lifestyle [4]. The process of atherosclerosis is characterized by excessive plaque deposits in the walls of the arteries [5]. In the case of an unstable plaque, the associated fibrous cap breaks, causing a form of thrombosis [6]. The thrombosis leads to embolism which in turn causes a blockage in blood flow resulting in a stroke or myocardial infarction [7].

According to the World Health Organization (WHO), cancer is the second biggest cause of mortality in the world, accounting for one out of every five fatalities [8, 9]. Lung cancer is the most frequent kind of cancer in both men and women. [10]. The early

identification and diagnosis of lung cancer decreases the death rate. [11].

Wilson's disease (WD) is one of the rare and it appears one in 30,000 to 40,000 peoples worldwide as per the report published by National Organization for Rare Disease (NORD). Moreover, the main cause for WD is deposition of excessive copper in the lungs and brains [12]. As per the study published in 2005, there are 9000 people affected by WD in the United States (US) [13]. There are many methods such as blood tests, and histology to diagnose the WD but these approaches are very slow and unreliable [14, 15]. We can also diagnose the WD disease by finding the ATP7B genes in the genetic analysis [16]. Moreover, most of the methods used for WD identification were non-imaging and time-consuming processes [17].

### 1.2 Non-Invasive Techniques- Imaging

#### 1.2.1 Magnetic Resonance Imaging for Carotid Plaque

The Magnetic Resonance Imaging (MRI) contains four components: i) a magnetic coil, ii) a gradient coil, iii) a radio frequency (RF) coil, and iv) a shim coil. A set of magnet coils made of superconducting metal-alloy produce powerful magnetic fields when a current passes through it. It also generates negligible heat, which is subsequently cooled by cryogenic helium. There are three concentric gradient coils located at the primary magnet coil. These gradient coils change the magnetic field's direction along the x, y, and z-axes. RF coils are mounted concentric to the main magnetic field and serve as antennae to send RF energy to the tissue and receive the induced RF signals generated by the tissue. The MRI requires the aid of a homogeneous magnetic field to localize the region of interest. Shim coils were used to homogeneously adjust the magnetic field by controlling it via the computer. The human body contains 80% water. Thus, when a patient lies in the bore of the MRI machine, the hydrogen atoms in the tissue align themselves with the magnetic field of the MRI machine. The resultant MR signals are localized and construct an image using gradient field magnets. Inhomogeneity correction in MRI for carotids can be adapted [18]. Figure 1.1 (a) shows the MRI machine representation.

### 1.2.2 Computed Tomography Imaging for Carotid Plaque

Computed Tomography (CT) imaging gives the region of interest (ROI) of tissue without the need to superimpose on the adjacent structure. It uses an X-ray radiation attenuation coefficient (AC) when imaging organs. AC was calculated as the density of the X-ray beam attenuated by the tissue. The CT machine mainly consists of an X-ray tube with slip-ring

#### 1.2 Non-Invasive Techniques- Imaging

technology and a detector. The detector continuously rotates around the patient during the diagnosis. This technology is called Helical CT, which has replaced the conventional CT (step and shoot) acquisition technique. The images acquired via CT are three-dimensional and allow multiplanar reconstruction. In a CT machine, the "pitch" parameter indicates the quality of the image acquired. If the pitch is less than one, then the acquired image is considered high-quality and oversampled but increases the patient's radiation exposure. Typically, the pitch value should be between 1 and 1.5 [19].

The thickness of the image slice depends on the X-ray collimator. The longer the collimator is exposed, the thicker the slice. However, the amount of radiation exposure increases and fails to plan reconstruction. Multiple detector array CT scanners overcome this problem. These scanners contain multiple detector arrays, and the collimator spacing is quite wide. Thus, it improves the imaging quality and slice thickness. Dual-source CT scanners have recently been developed, containing dual X-ray sources and detectors; these detectors are perpendicular to each other and rotate around the patient body. This setup cuts the acquisition time in half while still producing high-quality images.

Recent research used two single-source (SS) and two dual-energy (DE) CT scanners to assess the picture quality and radiation dosage supplied to patients during computed tomography angiography (CTA) of the supra-aortic arteries. Figure 1.1 (b) illustrates a typical CT machine. Several types of energy levels are sometimes considered for the optimization of carotid plaque characterization [20].

Several applications for carotid plaque characterization and microbleeds were shown in CT. Semi-automated wall measurements were computed in CT. Baradaran *et al.* [21] recently showed the calcium volume measurement using 74 CT scans. Saba *et al.* [22] showing using a color scale rather than a traditional grayscale increase the readers' diagnostic confidence, accuracy, and inter-observer agreement in the diagnosis of internal carotid artery dissection on non-contrast CT.

#### 1.2.3 Ultrasound Imaging for Carotid Plaque

Ultrasound (US) machines have a duplex scanner and a transducer with a linear broadband width of 4–7 MHz (multifrequency) and a resolution of 20 pixels/mm. Ultrasounds operate on the principle of the piezoelectric effect. The machine contains a transducer, a CPU, and a monitor. The transducer contains piezoelectric crystals, which act as the senders and receivers of ultrasound signals. The piezoelectric crystals generate sound waves when a current flows through them, and they generate a voltage when they receive a vibration

(ultrasonic wave) [23]. When the ultrasound waves fall on the ROI tissue, the tissue reflects the ultrasonic waves (echo), and the transducer catches the reflected signals. It then constructs and displays an image in the monitor based on these signals.

Some prerequisites are necessary for successful image normalization. In this investigation, the following conditions were met: (1) Dynamic range was used. (2) The average of the frames was calculated. (3) Because the ultrasonic beam was not attenuated when it traveled through the blood, the time gain compensation curve that was sloping through the tissues was positioned vertically across the vessel's lumen. This was done to match the brightness of the adventitia on the anterior and posterior walls. (4) The gain has been tweaked. (5) A linear transfer curve was used for post-processing. (6) The ultrasonic beam was kept at a 90-degree angle with the artery wall. (7) The picture's minimum depth was employed to guarantee that the plaque took up a significant portion of the image. (8) The probe was adjusted such that the adventitia close to the plaque was evident as a normalizing hyperechoic band. The ultrasound machine at Atheropoint<sup>TM</sup> is shown in Figure 1.1 (c).



**Figure 1.1:** Imagining Machines of (a) MRI, (b) CT, and (c) Carotid ultrasound scanning using a linear probe (MRI/CT images-courtesy of Dr. Luca Saba, Italy, and US image, courtesy of AtheroPoint, CA, USA).

#### 1.3 Statistical Methods for Detection of Lesion

Several authors have used manual and statistical methods for lesion characterization and classification (LCC) [24]. Some of the statistical methods for the detection of lesions are discussed below:

a) Grayscale Median: grayscale median (GSM) is a popular biomarker for the characterization of the lesion. GSM is a measure of echogenicity of the lesion. Its value is determined by taking the grayscale values in the segmented lesion. In ultrasound

#### 1.4 Machine learning-based lesion classification

imaging, GSM would be used for risk assessment and quantifying echo density. Medical practitioners use photo editing software like Adobe photoshop for thresholding the segmented lesion. They will calculate GSM from histograms, this GSM value differentiates the lesion with echo lucent and echogenic. Lesions with less GSM value will cause stroke or malignant.

b) Fractal Dimension: Fractal dimension (FD) is another popular biomarker for the characterization of the lesion's surface behavior. FD will be measured using the following equation 1.1. Here "N" is called the number of miniature pieces of the given figure, "s" is the scaling factor, and "Dfd" is the fractal dimension. The larger the value of D<sub>fd</sub> greater the amount of patchiness or roughness of the surface existed in the lesion.

$$D_{fd} = \frac{\log N}{\log S} \tag{1.1}$$

c) Pearson Correlation coefficient: Pearson correlation coefficient (PCC) gives the statistical relationship between two continuous variables. In medical diagnostics, the person correlation coefficient is used to measure the correlation between the interesting features. PCC gives the magnitude of correlation and direction of the relationship; it is calculated using equation 1.2. Here "r" is the correlation coefficient, "xi" is the ith sample of the x variable, "yi" is the ith sample of the y variable,  $\bar{x}$ , and  $\bar{y}$  are the mean of x and y variables. In medical imaging, PCC is used for correlation between automatic measurement and expert measurements.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$
(1.2)

d) Cohen and Kappa statistics: Cohen's kappa analysis measures inter-rater agreement for categorical items in cases involving more than one rater kappa values should be interpreted as follows: 0 indicates no agreement, 0.01–0.20 suggests modest agreement, 0.21–0.40 indicates reasonable agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates strong agreement, and 0.81 indicates perfect agreement, according to Cohen [25]. The equation for Cohen's kappa is given in equation 1.3. Here "k" is related to the kappa constant, "P<sub>0</sub>" gives the relative observed agreement among raters, and "P<sub>e</sub>" gives the hypothetical probability of agreement.

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{1.3}$$

# 1.4 Machine learning-based lesion classification

Most of the medical imaging classification is based on machine learning techniques.

Among all support vector machines (SVM), k-nearest neighbor (k-NN) was used by most of the researchers. Some of the methods in ML used by researchers are explained below:

#### a) Support Vector Machine:

The fundamental goal of a support vector machine (SVM) is to create a hyperplane that divides a collection of training data into positive and negative samples. Acharya *et al.* [26] proposed a classification technique based on 346 US scans in which texture, Local Binary Pattern (LBP), and Law's Texture Energy (LTE) features were extracted using SVM. The authors achieved an accuracy of 83%, a sensitivity of 84.4%, and a specificity of 79.7%. The same authors in [27] classified CT artery images from 20 patients as symptomatic and asymptomatic by extracting LBP features and wavelet features with SVM. Han *et al.*[28] extracted 2D and 3D Haralick texture features from the LIDC-IDRI dataset and fed these features in an SVM for classifying the CT lung nodule. Moreover, the authors have achieved the area-under-the-curve 0.9441on 3D Haralick and 0.9372 on 2D Haralick texture features. Most of the research work in medical imaging uses SVM classifiers.

#### b) K-Nearest Neighbor:

Another popular classifier in medical image processing is the k-nearest neighbor (*k-NN*). Its classification does not depend on the training data points. It is a lazy classifier that uses data distribution. Here the class of the data point is predicted based on the weighted distance from its k-nearest neighbor.

In the carotid US, Christodoulou *et al.* [29] extracted gray level different statistics (GLDS), spatial gray level dependence matrices (SGLDM), and first-order statics (FOS) texture features from the 330 scans and fed to *k-NN* achieved the highest accuracy of 66.3%. In lung CT, Lennartz *et al.* [30] extracted Entropy, kurtosis, mean of positive pixels (MPP), skewness, uniformity, and uniformity of positive pixels (UPP) were extracted and input to k-NN, which obtained an accuracy of 84 percent. In an MRI scan, the brain. In MRI brain, alkubeyyer *et al.* [31] extracted LBP, gray level co-occurrence matrix (GLCM), and discrete wavelet transform (DWT) and fed it to *k-NN* achieving an accuracy of 87%.

#### c) Naïve Bayes:

The Naïve Bayes (NB) classifier is a probabilistic classifier that employs the Bayes theorem. It calculates the likelihood of a class based on a collection of characteristics. The probability is computed using equation 1.4. Zhang *et al.* [32] compared the classification accuracies of manually segmented plaques with plaques segmented by SNAP. When the manually segmented plaque was employed, the accuracy of NB, SVM, RF, gradient

#### 1.5 Deep learning-based lesion classification

boosting decision tree (GBDT), and ANN was more than 88 percent. However, when SNAP was used, the same classifiers achieved higher than 78% accuracy (p<0.0001).

$$P(y|x_1,x_2,x_3,x_4,...,x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)*P(y)}{P(x_1)P(x_2)P(x_3)....P(x_n)}$$
(1.4)

#### d) Other ML methods:

In addition to the above classifiers, there are some classifiers such as self-organized maps (SOM) [29, 33], and logistic regression (LR) [34] used by researchers that use statistical techniques. Some classifiers use ensemble techniques such as Adaboost [35, 36], Quadratic deterministic analysis (QDA) [37], linear deterministic analysis (LDA) [37], decision tree (DT) [38], and random forest (RF) [39, 40] are used for classification of the lesions at lung, carotid, coronary, and brain. Among all these classifiers ensemble classifiers increase in popularity due to the achievement of higher accuracies and efficient performances.

# 1.5 Deep learning-based lesion classification

Deep learning (DL) provides a flexible and efficient model for the detection of the lesion in medical imaging. Many researchers worked on the classification of the lesion at the thyroid [41], skin [42], liver [43], prostate [44], and ovaries [45].

#### 1.5.1 2D Convolution Neural Network

Convolution neural network (CNN) [46] plays a vital role in image processing with AI. Some of the researchers are implementing 2D -CNN in the classification of medical imaging. The major drawbacks of the ML were overcome by the introduction of CNN. It uses automatic feature extraction. There are three types of CNN networks based on the movement of filters. They are 1-dimension CNN (1D-CNN), 2-dimension CNN (2D-CNN), and 3-dimension CNN (3D-CNN). In1D-CNN the filter will move only in one direction (either up or down), in 2D-CNN filter moves in two directions (both up or down, left or right), whereas in 3D-CNN filter moves in three directions i.e., left or right, front or back, top or bottom. Among all these CNN models 2D-CNN is efficient and requires fewer hardware requirements than 3D-CNN. Each layer of CNN contains a certain number of filters, it convolves on the input image and gives the "feature map (FM)". Initially, all these filters are assigned with random weights, in the beginning of the learning, these filters start to learn the secondary features like edges, corners, shape, color, etc. After that, they start learning complex patterns in the input image. The convolution process is using equation 1.5. Here "I" represents the 2D input and "K" represents the filter matrix with size "m x n".

$$(1*K)[i,j] = \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} I(i-p,j-q)K[p,q]$$
(1.5)

All of the weights in each layer were randomly initialized at the start of the training. The model then adjusts the weights using the back-propagation technique, considering the validation and training errors. These weights are modified in such a manner that the intended result may be predicted. In general, pooling layers come after each convolution or batch of convolution layers.

#### a) Pooling Layer:

The pooling layer is used for dimensionality reduction, there are three types of pooling techniques available, they are 1) max-pooling, 2) min-pooling and 3) average-pooling. In max-pooling they will consider the maximum values in FM, in min-pooling, it will consider the minimum values in FM, whereas in average pooling the average of the values will be considered. The output FM from the pooling layer was calculated using the following equation 1.6. Here  $w_{out}$  is the width of the output,  $w_{in}$  is the width of the input, "f" is the filter size, and "s" is the stride.

$$\mathbf{w}_{\text{out}} = \frac{\mathbf{w}_{\text{in}} \cdot \mathbf{f}}{\mathbf{s}} + 1 \tag{1.6}$$

#### b) Forward propagation in CNN:

In CNN, the input picture is scaled to a fixed size before being convolved with a filter. The filter will be used to extract the key features from the training dataset in this case. In forwarding propagation, the filters will learn the patterns in the input image.

#### c) Backpropagation in CNN:

The loss in backpropagation is determined using equation 1.7, which is shown below. The weights and bais will be adjusted following the gradient loss. Every epoch will go through the whole process, including forwarding and backward propagation. This process will be repeated until the loss reaches a global minimum.

$$L(a,y) = -(y\log(a) + (1-y)\log(1-a))$$
(1.7)

Here L(a, y) is the loss function, "y" is the expected output and "a" is the predicted output for each test case.

#### d) Tuning hyperparameters:

There are a number of hyper parameters that must be tuned in a CNN model. The number of nodes in each layer, the number of hidden layers, the activation function, the dropout

#### 1.5 Deep learning-based lesion classification

rate, the learning rate, the momentum, the batch size, and the number of epochs are just a few examples. Changes in these parameters will affect the performance of the model. So hyperparameter tuning is a crucial role in developing CNN models. Generally, most researchers use grid techniques for hyperparameter tuning.

#### 1.5.2 Performance Evaluation metrics

Performance evaluation metrics (PEM) are evaluated for every CNN model for measuring the performance of the model and comparing it with the existing techniques or with machine learning techniques. The following are the various evaluation metrics used for medical image processing and build for computer-aided diagnostics systems (CAD).

#### a) Accuracy:

Accuracy is the main metric for evaluating the performance of the model, higher accuracy models are higher efficient models. The accuracy of any model is calculated using the following equation 1.8.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
 (1.8)

#### b) F1-Score:

It is also a famous performance metrics for measuring the performance of the model. It is calculated by using the harmonic mean of precision and recall. It is calculated using the following equation 1.9, 1.10, and 1.11.

$$f_1 = \frac{2 \cdot \text{precision*recall}}{\text{precision + recall}}$$
 (1.9)

Where precision and recall are calculated as follows

$$precision = \frac{TP}{TP + FP}$$
 (1.10)

$$recall = \frac{TP}{TP + FN} \tag{1.11}$$

Here "TP" means true positive, "FP" means false positive, and "FN" means false negative.

# c) The area under the curve (AUC) – Receiver Operating Characteristics (ROC):

One of the most significant measures in medical imaging is the AUC-ROC, which indicates how well a model distinguishes between positive and negative classifications. True positive rate (TPR) (y-axis) and false-positive rate (FPR) (x-axis) are shown, with TPR and FPR

computed using equations 1.12 and 1.13. A better model would have a good area under the curve, which would be close to 1. The model's AUC score of 0.5 suggests that it is unsuitable for classification.

$$TPR = \frac{TP}{TP + FN} \tag{1.12}$$

$$FPR = \frac{FP}{FP + FN} \tag{1.13}$$

# d) Diagnostics odds ratio:

Diagnostics odds ratio (DOR) is another major performance metrics in medicine. It is used to distinguish the subjects with target disorders from subjects without the disorder. DOR is calculated using equation 1.14. It takes the value between 0 to infinity. Highest DOR (not infinity) is associated with better test performance. If the model DOR is less than 1 implies that it is going in the wrong direction and predicting oppositely. Sensitivity and specificity are calculated using equation 1.15 & 1.16.

$$DOR = \frac{\text{sensitivity*specificity}}{(1-\text{sensitivity})*(1-\text{specificity})}$$
(1.14)

sensitivity= 
$$\frac{TP}{TP+FN}$$
 (1.15)

specificity= 
$$\frac{TN}{TN+FP}$$
 (1.16)

#### e) Cohen-kappa analysis:

Cohen-kappa analysis measures inter-rater agreement for categorical items in cases involving more than one rate. Cohen suggested that kappa values should be interpreted as follows:  $\leq 0$  indicates no agreement, 0.01-0.20 indicates slight agreement, 0.21-0.40 indicates fair agreement, 0.41-0.60 indicates moderate agreement, 0.61-0.80 indicates substantial agreement, and  $\geq 0.81$  is considered perfect agreement [25]. The equation for Cohen's kappa is given in supplementary material. Cohen-kappa (k) is calculated using the following equation 1.17.

$$k = \frac{(p_0 - p_e)}{1 - p_e} \tag{1.17}$$

Here  $p_0$  is the average of the agreement,  $p_e$  is the sum of the correct agreement with the incorrect agreement.

#### f) Power analysis:

DL models require sufficient data samples for better prediction. Power analysis gives the number of samples required for measuring the performance metrics. It can be calculated

using statistical tools or using the conventional formula shown in equation 1.18.

$$n = \left[ \left( z^{*2} \right) x \left( \frac{(\hat{p})^*(1-\hat{p})}{\text{MoE}^2} \right) \right]$$
 (1.18)

Here n is the number of samples required,  $z^*$  denotes the standard z- table,  $\hat{p}$  represents data proportion, and MoE standards for margin of error.

# 1.6 Transfer learning

The issue with DL models is overcome with the transfer learning models (TL). TL models use pre-trained models trained on the natural images, these weights will be transferred to the model and retrained with the target labels. Due to this technique, the hardware issues with DL will be overcome with the TL model. Many of the researchers in lesion classification use various pre-trained models which are discussed below.

#### 1.6.1 VGG-16 and VGG-19

Visual Geometry Group (VGG-16) is a popular pre-trained model developed by Simonyan *et al.* [47] to increase the neural networks' depth by adding a number of 3x3 convolution filters. The purpose of VGGx is to design a very deep CNN for complex pattern understanding in the input features, typically adapted for object recognition in medical imaging and computer vision. The architecture of the VGG-16 and 19 is as shown in Figure 1.2, where the input block accepts the image of size 224x224. VGG-19 is three layers more than VGG-16 (not shown in the figure). Few applications of VGG 16 and 19 can be seen for the classification of Wilson [48], and Covid-19 pneumonia [49] disease.

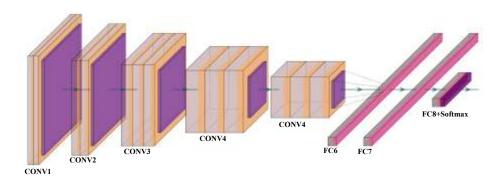


Figure 1.2: VGG16 and VGG19 architectures, CONV: Convolution Layer, FC: fully connected network.

# 1.6.2 InceptionV3

InceptionV3 (IV3), is version 3 of the Inception stage and was first developed by Szegedy *et al.* [50]. This model was developed to overcome the computational cost and low parameter count. This model can handle big data. Thus, this model has overall high

efficiency. Inception V3 achieves accuracy greater than 78.1% when using the ImageNet dataset. The architecture model contains several blocks. The blocks contain convolution and max-pooling layers. In the architecture given in Figure 1.3, DL1 to DL6 represents the depth-wise convolution, C1 represents the initial convolution block, T1 to T3 represents the transition layer, and D1 to D4 represents the batch normalization blocks. In inception V3 architecture each block in the top row represents the repeated process of row 2 and row 3. In row 2 each block represents the repeated process of row 3. Each convolution layer is fused with a 1x1 convolution filter with stride 1 and padding 0. First, it increases the feature map (FM) size, then a 3x3 convolution layer with stride 1 and padding 1 is added. It reduces the FM depth; the resultant FM and the initial FM is fused to give each block in row 2.

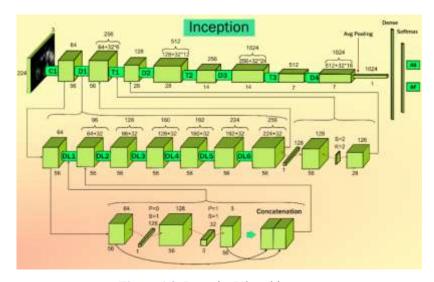


Figure 1.3: InceptionV3 architecture.

#### **1.6.3** ResNet

He *et al.* [51] from the Microsoft research proposed ResNet architecture for solving the vanishing gradient problem. It contains residual blocks. Residual blocks contain skip connections. These skip connections skip some layers from training and connect directly to the output. The advantage of these connections is to skip the layers, so that model will learn complex patterns. Unlike other TL models, this model is trained on the CIFAR-10 data set. Figure 1.4 represents the ResNet architecture. In the architecture, two 3x3 convolution layers are paired together. The output of these pairs and their input is fused and fed to the next pair. Here, the number of filters is in increasing order from 64 to 512. At the end of the last 3x3 convolution layer with 512 filters and an added flatten layer for vectorization of the 2D features, the output is predicted using the SoftMax activation function.

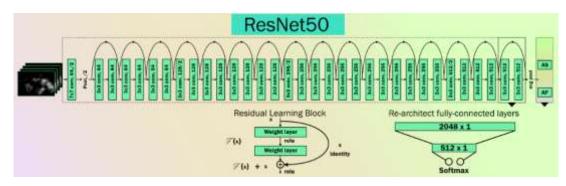
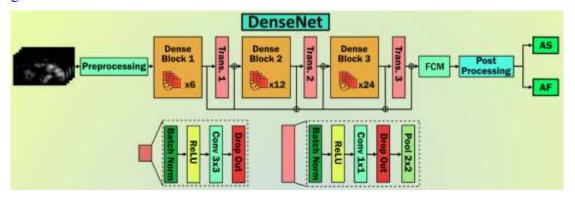


Figure 1.4: ResNet Architecture.

#### 1.6.4 DenseNet

Huang *et al.* [52] proposed DenseNet architecture for solving vanishing gradient problems in deep neural nets. In this model, dense blocks were introduced, it contains a pool of convolution layers with 3x3 filters to 1x1 filters followed by batch normalization, every layer uses the "ReLu" activation function. Each of these dense blocks was concatenated with previous block output and input using transition blocks. Each transition block contains a convolution and pooling layer with 2x2 to 1x1 filters with dropout layers. This concertation of blocks preserves the feature propagation nature. In addition, the author proposed architectures (DenseNet-121, 169, 201, and 264) to increase the dense block. Figure 1.5 shows the DenseNet architecture.



**Figure 1.5:** DenseNet Architecture with three dense blocks and three transition blocks, followed by a fully connected network. Post-processing is the SoftMax.

#### 1.6.5 MobileNet

Howard *et al.* [53] from Google had developed MobileNet architecture. The main inspiration for MobileNet comes from the IV3 network. It aims to solve resource constraint problems such as working on edge devices like NVIDIA Jetson (<a href="www.nvidia.com">www.nvidia.com</a>) or Rasberry Pi (from Rasberry Pi Foundation, UK). This architecture is a small, low latency, and low power model. This was the first computer vision model developed for TensorFlow for mobile devices. It contains 28 layers and uses the TFlite (database) library. Figure 1.6

presents the architecture of MobileNet architecture. This model contains bottleneck residual blocks (BRB), also referred to as inverted residual blocks used for reducing the number of training parameters in the model.

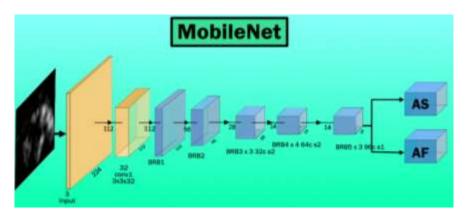


Figure 1.6: MobileNet Architecture, BRB: bottleneck and residual blocks.

#### 1.6.6 XceptionNet

Chollet *et al.* [54] from Google proposed modifying IV3 by replacing inception modules with modified depth-wise separable convolution layers. This architecture contains 36 layers. In comparison with IV3, XceptionNet is lightweight and contains the same number of parameters as in IV3. This architecture outperforms the InceptinV3 with Top-1 accuracy of 0.790, and Top-5 accuracy of 0.945. Figure 1.7 will represent the architecture of XceptionNet.

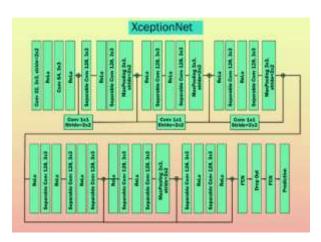


Figure 1.7: XceptionNet Architecture.

#### 1.6.7 AlexNet

Alex Krizhevsky *et al.* [55] proposed AlexNet in 2012 for solving the complicated ImageNet challenges. It is the first CNN architecture built for solving complex computer vision problems. This architecture achieves a top-5 error rate of 15.3%. This architecture

#### 1.7 Characterization of the lesion using machine learning

shifts the paradigm of AI entirely. It has five convolution layers, followed by max-pooling with two fully linked networks, and accepts 256x256 picture input. The SoftMax layer is the last output layer. The sample architecture is shown in Figure 1.8.

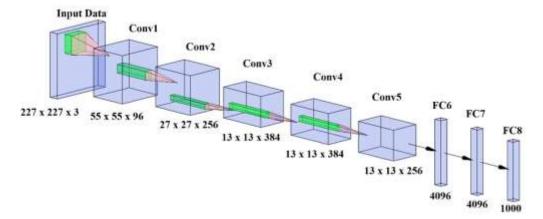


Figure 1.8: AlexNet architecture.

#### 1.6.8 SqueezeNet

Landola *et al.* [56] proposed a relatively 50x times smaller model than the AlexNet architecture. Nevertheless, the authors achieved 82.5% in top-5 accuracy on ImageNet. This model contains a novel "Fire Module". It contains a 1x1 filtered squeeze convolution layer fed to "Expand Module," which contains a mix of 1x1 to 3x3 filters for convolution. The squeeze layer (Fire Module) helps to reduce the number of input channels to 3x3. The architecture of the SqueezeNet and Fire Module is shown in Figure 1.9. In this study, we transferred trained weights to SqueezeNet initial layers and fed our cohort at the end layer.

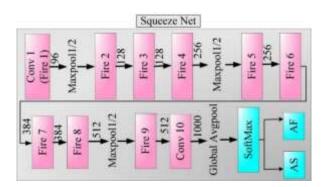


Figure 1.9: SqueezeNet Architecture.

# 1.7 Characterization of the lesion using machine learning

Many researchers attempted to characterize the lesion using machine learning in different modalities such as MRI, CT, and the US. Most of their work is based on handmade features and only on the carotid. They quantify the components using manual methods and applied

ML techniques for validating. Some of the researcher's views on characterization using different modalities are explained below.

#### 1.7.1 Characterization of the MRI lesion using machine learning

Winn et al. [57] characterized 20 carotid fibrous caps in T2-weighted MRI using four observers (radiologists). The ROI was generated by comparing T2-weighted MRI against histological images taken from carotid endarterectomy. The AUC came out to be **0.75.** The authors further characterized fibrous cap or rupture, having detection and characterization evaluated as 90% and 98%, respectively. Murata et al. [58] designed a motion-sensitized driven equilibrium prepared rapid gradient echo (3-D MERGE) acquisition protocol to characterize carotid plaque using black blood MRI. The researchers looked at 194 carotid arteries from 97 people (70 men and 27 women, mean age of 60 years). The researchers discovered 136 plaques, with 68 (50%) being inside 2-D MRI coverage, 46 (33.8%) somewhat outside, and 22 (16.2%) completely outside.

Zhang *et al.* [59] the usefulness of employing simultaneous non-contrast angiography and intraplaque hemorrhage (SNAP) to identify the lipid-rich/necrotic core (LRNC) from 1436 images produced by 3T MRI and 3-D SNAP sequence was evaluated. The authors compared the classification accuracies of manually segmented plaques with plaques segmented by SNAP. The accuracies of NB, SVM, RF, gradient boosting decision tree (GBDT), and ANN were higher than 88% when the manually segmented plaque was used. However, when SNAP was used, the same classifiers achieved higher than 78% accuracy (*p*<0.0001).

#### 1.7.2 Characterization of the CT lesion using machine learning

Zhu *et al.* [60] investigated CT Angiogram (CTA) for quantitative features to predict 10 years' worth of atherosclerotic cardiovascular disease (ASCVD) risk using 117 CTA scans. The authors built two semi-automated linear regression models with continuous and dichotomous features. The models yielded ASCVD risk scores of  $18.87\% \pm 13.26\%$  and  $18.39\% \pm 11.6\%$  (p<0.0001), and the mean biases between observed ASCVD and predicted ASCVDs were  $-1.954\% \pm 10.88\%$  and  $-1.466\% \pm 12.04\%$ , respectively. The most accurate prediction for ischemic stroke depends on the optimal spatial coverage of the carotid plaque. Arora *et al.* [61] studied the optimal coverage area required to distinguish between stroke and non-stroke patients. The authors found that 20 mm coverage on each side of the carotid bifurcation offers optimal results. Further, the authors used a CT-automated classification algorithm, multivariate, and univariate analysis on different

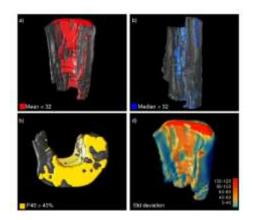
#### 1.7 Characterization of the lesion using machine learning

components of plaque obtained from 136 patients. Acharya *et al.* [27] classified CT carotid artery images from 20 patients as symptomatic and asymptomatic by extracting LBP features and wavelet features with SVM. Their classifications were 88% accurate. Further, they characterized the plaque using the Atheromatic index (AtheroPoint, CA, USA).

# 1.7.3 Characterization of the US lesion using machine learning

Several ML methods have been tried on various organs such as thyroid gland [62], liver [63], ovarian [64], carotid plaque characterization [65], coronary [66], stroke risk [67]. Seabra *et al.* [68], mainly to overcome the weakness due to subjectivity with data acquisition and operator-dependent selection. In their characterization model, the authors applied a labeling procedure using graph cuts, which allow them to identify, locate, and quantify vulnerable plaque. The authors tested their method on five patients who had undergone 2-D ultrasound scans and validated their synthetic data method. The same authors in [69] proposed a Bayesian technique for estimating the volume inside plaque by removing speckle noise in the ROI. The authors compared the proposed method with a gold standard, and the quantified volumes P40 (40% hypoechogenic voxels) were 56.76 mm<sup>3</sup> and 50.62 mm<sup>3</sup>, respectively. Afonso *et al.* [70] developed a computer-aided diagnosis tool for measuring plaque rupture risk using an activity index and an enhanced activity index (EAI). The authors were able to characterize the plaque echogenicity using the EAI.

Christodoulou *et al.* [29] Using a self-organizing map (SOM) and k-nearest neighbor, they categorized 330 carotid ultrasound images (CUS) as symptomatic and asymptomatic, extracting textural characteristics, spatial grey level dependence matrix (SGLDM), and grey level differential statistics (GLDS). The researchers were able to reach a diagnostic yield of 70% using first-order statistics (DY). In [71], the author extracted ten different Law texture energy features and fractal dimension features from CUS. The author then fed these features into SOM and *k*-NN for classification, achieving 73.1% and 68.8% accuracy rates, respectively. The echo morphology of the lesion is shown in Figure 1.10.



**Figure 1.10:** Local assessment of plaque echo-morphology, in terms of (a-c) hypoechogenicity and (d) heterogeneity [72]

# 1.8 Compressed CNN architecture

The availability of vast computing power has piqued interest in the field of deep learning. Researchers employed a large number of convolution layers to boost classification accuracy to the maximum. When these pre-trained models are placed on edge devices with limited memory and processing capacity, however, this becomes an issue.

Gong et al. [73] found that dense layers contribute 90% of storage in CNNs, and offer Vector quantization approaches to reduce the number of nodes in the dense layer. The authors demonstrate experimentally that by sacrificing 1% accuracy, the model may be compressed by 16-24%. Cheng et al. [74] discussed the use of pruning to compress the Convolution Neural Network. During the pruning phase, unnecessary weights are removed, resulting in a large reduction in the size of the model. The best compression on AlexNet was 4.94, which was calculated by dividing the original model parameters by the compressed model parameters. Thinet is a compressed model that is created by removing any filters whose sum of weights is less than a certain threshold [75]. ResNet50 parameters are reduced by 66 percent, but accuracy is reduced by 4.46 percent using this filter pruning method. Model compression is accomplished by Anwar et al. [76] by utilizing pruning to accelerate the model. Evolutionary particle filtering and activation sum voting are two strategies outlined by the authors for selecting the best candidates for pruning. Using the CIFAR-10 dataset as an example, they proved that the network may be pruned by 70% with less than 1% accuracy loss using this method. Han et al. [77] take a similar method, reducing the weights of convolution and Dense layers that are smaller than a certain threshold. Li et al. [78] proposed a compression process that involves pruning convolution layer filters that have minimal effects on the model's accuracy. The authors ranked the

#### 1.9 Research Gaps

filters based on the L1-norm and removed low-ranking filters from each layer. Liu *et al.* [79] proposed a network slimming technique that prunes unimportant channels during training using sparsity-based regularisation by calculating the channel sparsity and its scaling factor is nearly equal to zero.

Choudhary *et al.* [80] addressed different model compression and acceleration strategies that had previously been described in the literature, as well as doing their experiment on AlexNet utilizing the MNIST and CIFAR-10 datasets. From the structural, algorithmic, and implementation viewpoints, Zhang *et al.* [[81] investigated numerous strategies for CNN compression and acceleration. Layer decomposition and network pruning may be used to decrease duplication in CNN at the structural level, according to the authors. Furthermore, the fast Fourier transform (FFT) may be employed for model acceleration at the algorithm level.

# 1.9 Research Gaps

- Research showed that deep learning models such as VGG16 [82], AlexNet [83], and Inception V3 [84] models perform the classification of the lesion in medical imaging. All these models contain a large number of layers and are faced with the highest false-positive rate. Depending on the imaging modalities the deep learning application is varying. So, there is a need to develop optimized models that work on all the imaging modalities with lightweight and faster prediction times.
- Most of the researchers applied deep learning techniques for the segmentation and classification of the lesion [85, 86]. The most important use of medical imaging is the characterization of the lesion [87]. This characterization of the lesion was done by the medical practitioners manually by calculating GSM, morphological textures, and calcium percentage. Some of the researchers applied machine learning techniques for the detection of GSM levels [88]. However, these ML techniques suffer from handcrafted images so there is a need for automatic feature extraction for unbiased characterization.
- The availability of the cohorts of different imaging techniques is limited and contains a smaller number of images. So, the augmentation of the images using geometrical operations such as rotation, flip, and tilt. There are no such attempts to take care of the effect of augmentation on the performance of the model. Along with the classification, we are presenting the risk of the disease.

# 1.10 Objectives and thesis organization

Our objective in this thesis is three-fold. We have first shown how effectively the light-weighted and optimized CNN models for classification of the lesion in stroke, lung, and Wilson disease with improved accuracy and highest diagnostic odds ratio.

Secondly, characterizing the lesion using a novel deep learning technique called mean feature strength and validating with the existing statistical and image processing techniques. Finally, creating a computer-aided design (CAD) system using multi-center trained deep learning models, hybrid CNN models, and pre-trained CNN models for disease characterization and generation of the heatmaps.

The thesis is organized in the following order: chapter 2 describes the classification and characterization of stroke using carotid lesions, chapter 3 tells about the characterization of the carotid lesion in the multicenter paradigm, and chapter 4 brings the transfer learning in the characterization of the carotid lesion, chapter 5 gives the classification and characterization of the Wilson disease lesion using MRI scans, chapter 6 explains about the classification and characterization of the lung lesions using CT scans for covid and nodules. Finally, chapter 7 gives the conclusion and future work of the thesis.

# **CHAPTER 2**

# CLASSIFICATION AND CHARACTERIZATION OF STROKE USING OPTIMIZED DEEP LEARNING TECHNIQUES

#### 2.1 Introduction

Stroke is the third leading cause of mortality in the United States of America (USA) [89]. According to World Health Organization (WHO) statistics, cardiovascular disease (CVD) leads to 17.9 million deaths each year [90]. Atherosclerosis is the fundamental cause of CVD by forming complex plaques associated with the arterial walls [23].

Strokes may occur when atherosclerotic plaques in the internal carotid artery (ICA) break and embolize the brain. Only a small percentage of plaques are unstable and rupture, resulting in a 1-2 percent yearly stroke incidence in asymptomatic individuals with >80% stenosis. Thus, operating on all patients with >80% stenosis will result in many unnecessary operations. Additionally, the procedure is linked to a 3% preoperative stroke rate. Some plaques are unstable due to a large lipid core, a thin fibrous cap, and low collagen content (vulnerable). Therefore, they are more likely to rupture, producing symptoms (symptomatic or hyperechoic or unstable plaque) compared with the more stable ones, which have a small lipid core, a thick fibrous cap, and a large amount of collagen that tend not to produce symptoms (asymptomatic or hypoechoic or stable plaque) [91]. Therefore, it is important to characterize the plaque early, especially when it is turning symptomatic or likely to be unstable leading to rupture with subsequent stroke.

Several modalities exist for imaging carotid plaque, such as magnetic resonance imaging (MRI) [92], computed tomography (CT) [93], and ultrasound (US) [94]. Ultrasound offers essential advantages because it is non-invasive, radiation-free, and portable [95]. In addition, features like compound and harmonic imaging are now available on standard equipment at a resolution of 0.2 mm [95]. However, visual classification of

plaques as stable or unstable using ultrasound images is challenging due to the plaque tissues' inter-variability [96].

All methods describe a trial-and-error approach toward feature extraction and are manually adapted for classification. Thus, they are ad-hoc and highly variable [97]. Thus, there is a clear need to design and develop automated feature extraction approaches to characterize tissues such as atherosclerotic plaque into symptomatic and asymptomatic types.

Deep learning (DL) technology has dominated all walks of life, particularly in radiological imaging [98, 99]. This technology provides an alternative to the ML strategies, especially: (i) the ability to generate a down-sampled representation of the original pattern automatically (so-called feature maps), and (ii) dynamically adjust the variations in the grayscale contrast via the neural network layer of the DL architecture [100]. Lekadir *et al.* [101] developed a CNN model for the classification of the plaque components by extracting 90,000 patches from the 50 *in-vivo* ultrasound images and achieved a 0.90 correlation coefficient. The purpose of this study is to develop and design an automated carotid plaque characterization and classification system into binary classes, namely symptomatic and asymptomatic types via the deep learning (DL) framework implemented on a supercomputer.

We hypothesize that symptomatic plaque has tissue characteristics such as (a) hypoechoic regions, having a low grayscale median (GSM) as a result of a large lipid core, low calcium or intraplaque hemorrhage (IPH), and (b) more chaotic (heterogeneous) representation in the ultrasound scans [102, 103] because of the frequent presence of neovascularization alternating with areas of collagen or lipid. This is contrary to asymptomatic plaques, which are often hyperechoic with higher GSM, because they have higher and diffuse collagen content, often with calcification and a small lipid core. We designed a novel carotid plaque tissue characterization and classification system using DL components of artificial intelligence (AI) based on such a hypothesis. The design overcomes the ML weaknesses such as (i) manual feature extraction and (ii) classification. The classification system's accuracy is determined using a K10 (90% training and 10% testing) cross-validation protocol. The characterization of the plaque is accomplished by (a) computing the mean feature strength (MFS) at different DL layers [104] and (b) fractal dimension exhibiting quantification of the randomness in these plaque images [105]. Subsequently, we benchmark the DL system against the previously developed ML system

#### 2.2 Background Literature

(on the same cohort), and finally, the system speed is optimized using the supercomputer framework.

# 2.2 Background Literature

The existing work on carotid plaque characterization using ultrasound with AI techniques is primarily focused on the machine learning paradigm. A handful of the studies are focused on using DL. Our study is the first of its kind that uses the TL paradigm embedded with heatmaps for PTC. The section briefly presents the works on PTC. Detailed tabulation is described in the discussion section.

Seabra *et al.* [68] used graph cut techniques for the characterization of 3D ultrasound. It allows for the detection and quantification of the vulnerable plaque. The same set of authors in [69] estimated volume inside the ROI plaque using the Bayesian technique. They compared the proposed method with a gold standard and achieved better results with GSM<32. In [35], they used ML-based techniques to classify 146 US scans using "the leave one out" cross-validation technique in the Adaboost classifier framework that yielded an accuracy and sensitivity of 99.2% and 100%, respectively. In [36], they characterized the plaque components such as lipids, fibrotic, and calcified using the Rayleigh mixture model (RMM).

Afonso *et al.* [70] proposed a CAD tool (AtheroRisk<sup>TM</sup>, AtheroPoint, Roseville, CA, USA) to characterize plaque echogenicity using activity index and enhanced activity index (EAI). The authors achieved an AUC of 64.96, 73.29, and 90.57% for the degree of stenosis, activity index, and enhanced activity index. This AtheroRisk<sup>TM</sup> CAD system was able to measure the plaque rupture risk. Christodoulou *et al.* [29] classified carotid US using a self-organizing map (SOM) and statistical k-nearest neighbor classifiers by feeding texture features extracted using first-order statistics. The authors had the best accuracy of 70%. In [71], the same authors extracted Law's texture energy features and fractal dimension features and fed them to SOM and k-NN classifiers. They achieved an accuracy of 73.1% and 68.8%, respectively. Loizou *et al.* identified and segmented the carotid plaque in M-mode ultrasound videos (MUV) using a snake algorithm [106]. In [107], the authors studied the variations of the texture features such as spatial gray level dependence matrices (SGLD), and gray level difference statistic (GLDS) in the MUV framework and classified them using an SVM classifier. Doonan *et al.* [108] studied the relationship between textural and echo density features of carotid plaque by applying the PCA-based feature selection

technique. The authors showed a moderate coefficient of correlation (r) between these two features having the range from 0.211 to 0.641.

AtheroPoint has designed Atheromatic <sup>TM</sup> 1.0<sub>ML</sub>, where Acharya *et al.* [109] were able to classify 346 Carotid US into symptomatic *vs.* asymptomatic using texture-based features, local binary patterns (LBP), and law's texture features. They fed these features to the SVM classifier and achieved an accuracy of 83%. In [110], the same authors extracted texture features and fed them to Atheromatic 1.0<sub>ML</sub> and achieved an accuracy of 83.7%. In [111], they extracted texture features from 99 patients' carotid US scans, trained them using an SVM classifier with RBF kernel, and used cross-validation K10 protocol, yielding an accuracy of 91.7%. Further, they created an index for separating symptomatic *vs.* asymptomatic carotid plaque called the symptomatic and asymptomatic carotid index (SACI). Along the lines of ML designs, Molinari *et al.* [37] using bidimensional empirical mode decomposition and entropy characteristics, they suggested a data mining framework for the categorization of symptomatic and asymptomatic plaque, with an accuracy of 91.43 percent.

Demirer *et al.* [112] developed a GUI-based toolbox using a deep neural network (DNN) for 2-D/3-D segmentation and classification, as well as for integration with third-party scripting languages. The researchers used this tool to annotate 1843 arteries and 294 coronary CTA atherosclerotic plaques in 23 days. This graphical user interface (GUI) used NoSQL as its database, while TensoFlows was used for DNN. In other work, Lee *et al.* [113] developed a DL model and used it to classify 6,556-lumen images of OCT scans using CNN. The authors achieved sensitivity and specificity rates of 84.8% and 91.4% (fibrolipidic) and 97.8% and 95.7% (fibrocalcific). Table 1 represents the similarities and differences in imaging modalities.

Table 2.1: Similarities and differences between imaging modalities using AI.

	(a) Difference between MR, CT, and US for TCCA using AI							
	MRI	CT	US 2-D/3-D					
Ref	[39, 58, 59]	[27, 38, 40]	[26, 29, 34, 35, 37, 67, 70, 71, 109-111,					
			114-118]					
	Segmentation: Bayes		Segmentation: Manual, Simple Linear					
ML	Clustering, Structural	Segmentation:	Iterative clustering, Real Adaboost,					
IVIL	Support Vector	Automatic	Bayesian,					
	Machines, Manual,		Bayesian,					

Segmentation, 3-D MARGE		Extracted Features: Multiresolution Features, Bi-dimensional empirical mode
MARGE		
Extracted Features: Surface disruption features, SIFT, Morphological, Intensity Features of inversion recovery (IR), Intensity features of reference acquisition (REF), TOF-MRA and Black Blood MRI (BBMRI), histological	Extracted Features: Texture features and relative position of pixels, Local Binary Patterns (LBP) and wavelet transform, Texture features	decomposition, and entropy features, texture features, automatic, Rayleigh Mixture Model, Histogram, Texture, Morphological, Monogenic, Wavelet energies, co-occurrence matrix, 1st order statistics, Multilevel binary morphological, second-order statistics spatial gray level dependence matrices, ACRS Clinical, discrete wavelet and higher-order spectra, 2-D DWT, Degree of stenosis, DWT with Averaging, Envelop Radio-Frequency, statistical features, fractal dimension, laws texture energy, Fourier power spectrum, Spatial Based Plaque Feature, 3-D Plaque Feature Extraction, Neighborhood Gray Tone Difference Matrix, Quadratic Programming Feature Selection, Minimal Redundancy Maximal Relevance, Mutual Information Quotient, Spectral Conditional Mutual Information, Cramer's V test, neighborhood gray-tone difference matrix, 3-D Fractal Dimension
Classification: RF, t- test and logistic regression, SSVM, NB, SVM, RF, GBDT, ANN, KNN, SVM, DT,	Classification: RF, SVM RBF, DT, NB, LR, NN	Classification: SVM and Probabilistic Neural Network, SVM RBF, SVM Polynomial, linear, LibSVM, DT, AtheroRisk, Adaboost, self-organizing map, KNN, ANN, 3-D Blanket, SVM with 3-DUS
Performance Metrics:	Performance	Performance Metrics:
AUC Ranges: 0.95;	Metrics:	AUC Ranges: 0.649, 0.732, 0.905; ACC
ACC Ranges (%) :87,	ACC Ranges (%):	Ranges (%): 85, 91.43, 82.4, 83.5, 73.7,
88, 76, 87.5, 90	83.1, 88, 69	77.18, 76, 91.7, 83.7, 90.66, 83.7, 99.2,

	Misclassification Rate		73.1, 73.72, 68.8, 69.3, 81.82, 80.38, 81			
	(%): 9.6					
	[119], [120],	[121], [122],	, [123], [124]			
	Segmentation: U-NET,					
	DeepMAD					
	morphological active	Segmentation:	Segmentation: U-Net, Dilatad U-Net, 3-			
	contours with the	LVO	D U-Net			
	iterative framework, 3-					
	D Level set, Automatic					
	Extracted Features:	Extracted				
DL	Morphological features	Features:	Extracted Features: Automatic			
	Worphological leatures	Automatic				
	Classification:	Classification:	Classification: Optimized CNN,			
	DeepMAD	DeepSymNet,	Dynamic CNN,			
	Весрічні	Faster R-CNN	Byllamic Civiv,			
		Performance				
	Performance Metrics:	Metrics:	Performance Metrics:			
	ACC Ranges (%): 99.1,	AUC Ranges: 0.88,	ACC Ranges (%):95.66, Dice			
	92.6, 89.16	ACC Ranges (%):	Coefficient Ranges: 96.6, 84			
		83				

#### (b) The similarity between MR, CT, and US for TCCA using AI

- All the segmentation and classification were attempted on 2-D slices without considering the 3-D spatial information.
- All the modalities have implemented segmentation of the wall as their first step.
- The centreline algorithm was adapted to extract the orthogonal slices to the blood flow for all three modalities.
- All the modalities have attempted tissue characterization.
- ML has been attempted on all three modalities for TCCA.
- SVM, RF, and DT are the common ML classifiers adapted by three modalities.
- DL has been attempted on all three modalities for TCCA.
- U-Net is the most common architecture used in the DL framework for all three modalities.
- CNN is the most popular architecture tried for all modalities.
- TL is the least adopted among all the architectures.
- Accuracy and AUC are the common performance metrics for all the three modalities

# 2.3 Patient Demography and Dataset

The main components of this section primarily consist of (i) patient demographics and exclusion criteria, (ii) ultrasound data acquisition, and (iii) plaque delineation and data augmentation.

#### 2.3.1 Multicentre Patient Demographics

Two sets of demographics were collected from two geographical areas of the world. The first was collected from St. Mary's Hospital, Imperial College, London, UK, symbolized as LonD. The patients with contralateral cerebral hemispheric/retinal symptoms, as well as those who had a transient ischemic attack (TIA), were included only if they had been asymptomatic for more than six months. Note that this data was taken from our previous study (Nicolaides *et al.* [125]), which consisted of 196 *symptomatic* and 150 *asymptomatic* cases. The study had a mean age of 69.9±7.8 years of which 61% were male who had ICA stenosis of 50% to 99%. Note that these patients did not have any previous symptoms or neurological abnormalities. The second data was collected from Instituto cardiovascular de Lisboa, Lisbon, Portugal, symbolized as LisD. This was approved by the ethical committee and the data taken from our previous study (Sanches *et al.* [126]). It consists of 50 *symptomatic* and 110 *asymptomatic* scans with a mean age of patients 67.5±0.77 years.

#### 2.3.2 Ultrasound Image Acquisition from Multicentre Study

For LonD, all the scans were acquired using an ATL HDI-3000 duplex scanner, Advanced Technology Laboratories, Seattle, WA, the USA with a linear broadband width 4-7 MHz transducer. The mean resolution was 20 pixels per millimeter. Region-of-Interest (ROI) of the scans was delineated and normalized using a medical practitioner with a mouse and "Plaque Texture Analysis Software". For LisD, a Philips HDI 5000 machine with a 5-12 MHz broadband linear array transducer was used to obtain these scans. The mean resolution was 25 pixels per millimeter. ROI was delineated and normalized using MATLAB by the medical practitioner. We combined LonD and LisD to form a new cohort called mixed cohort labeled MixD. It contained 246 (196+50) symptomatic, 260 (150+110) asymptomatic scans. Figure 2.1 shows the representative images of these cohorts.

#### 2.3.3 Augmentation Protocol

Both cohorts are moderate in size. Thus, we have used augmentation techniques for increasing the cohorts into 2x, 3x, 4x, 5x, and 6x folds as used in our previous work [48, 127-129]. Initially, we took the unbalanced cohort and then balanced the classes labeled as

balanced fold (1x) and then augmented up to six folds (6x). We used "Augmenter" for these purposes consisting of geometrical operations such as random flipping and rotation of 180 degrees.

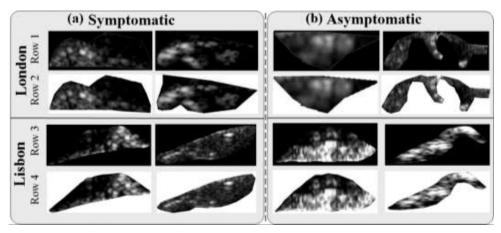


Figure 2.1: Sample (a) symptomatic (b) asymptomatic the US cut sections of the carotid plaque.

# 2.4 Proposed AI Models

In this section, we discussed the novel technique of deep learning optimization using varied augmentation folds. This section contains proposed models and 3D optimization, characterization, result, and discussion.

#### 2.4.1 The proposed 3D-optimized deep learning model

The deep convolutional neural comprises four convolution layers as shown in Figure 2.2, which are followed by an average pooling layer; thus, there are nine layers in total. Then, a layer is flattened to convert the 2-D feature map into a 1-D feature map. This layer is followed by two thick layers of 128 nodes each. The "SoftMax" layer, which has two nodes, one representing the symptomatic class and the other representing the asymptomatic class, is the final output. We chose the "ReLu" activation function for all the n-1 layers because it converges the solution faster than the "sigmoid" and "tanh" activations functions. Equation 2.1 is the categorical cross-function used for the experimentation of all models.

Loss = 
$$-[(y_i \times \log a_i) + (1 - y_i) \times \log(1 - a_i)]$$
 (2.1)

where  $y_i$  is the class label for input and  $a_i$  is the predicted probability of the class being  $y_i$ .

The DCNN model's performance depends on how many layers and hyperparameters are tuned. Therefore, we considered several configurations of DCNN that consisted of different combinations of convolution, average pooling, and dense layers. Doing this required 3-D optimization between the accuracy, DCNN layers, and folds of the

augmentation. Table 2.2 shows the six DCNN models generated through this process.

**Table 2.2:** Six types of DCNN models consisting of different combinations of convolution, average pooling, and dense layers.

	Column1	Column2	Column3	Column4	
R#	DCNN Type	Convolution 2D Layers	Average Pooling Layers	Dense Layers	
R1	DCNN5	1	1	3	
R2	DCNN7	2	2	3	
R3	DCNN9	3	3	3	
R4	DCNN11	4	4	3	
R5	DCNN13	5	5	3	
R6	DCNN15	6	6	3	

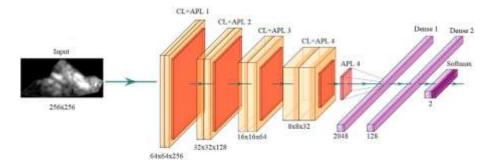


Figure 2.2: DCNN11 architecture (CL: convolution layer, APL: average pooling layer).

#### 2.4.2 Two kinds of Transfer Learning Architectures

Transfer learning (TL) is a kind of deep learning classification that requires less computation power and execution time than other techniques. According to previous work, CNN involving TL has played a vital role in classification applications. When TL is employed, the training weights and parameters of a model are inherited from the pre-trained network. In the present study, we used "Inception V3" and "VGG16" architectures, which have already been trained using "ImageNet". Figure 2.3 illustrates the TL paradigm adopted for plaque classification.

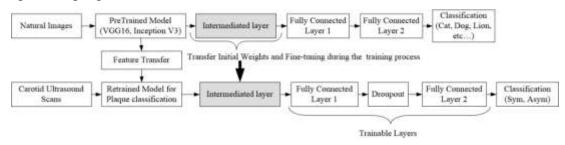


Figure 2.3: Transfer learning using "ImageNet" for plaque classification.

# 2.4.3 Machine Learning Architectures

As part of the current AI study, we designed four kinds of ML systems to benchmark the proposed system about the DCNN system. The four ML systems chosen were Naïve Bayes (NB), *k*-Nearest Neighbor (*k*-NN), Support Vector Machine (SVM), and Decision Tree (DT). The main features extracted were Haralick, Hu-moments, histogram-based, local binary patterns (LBPs), and histograms of oriented gradients (HOGs). The online ML architecture shows the same trends that were adapted in our previous ML systems (Figure 2.4).

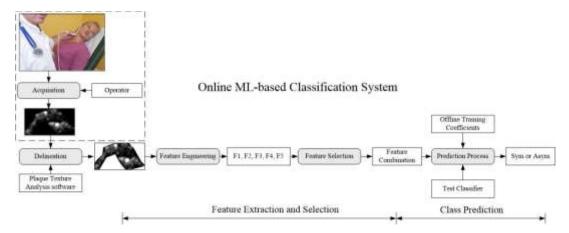


Figure 2.4: The machine learning architecture adopted in this study.

#### 2.5 Results and Discussion

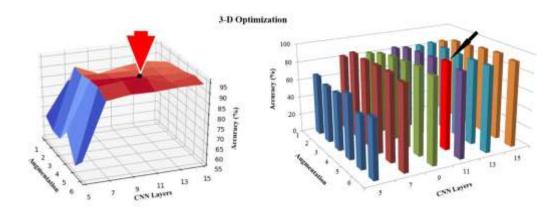
#### 2.5.1 3D-Optimization

The goal of this experiment was to find the best DCNN model under augmentation conditions. We developed 36 types of DCNN combination models under different augmentations (six types of CNN multiplied by six types of augmentation yielded 36 DCNN models). Each combination used the K10 cross-validation (CV) protocol (90% training and 10% testing and, thus, a cyclic rotation of 10 combinations). Each CV protocol conducted up to 1000 epochs, which were computed empirically after ensuring the flatness of the loss and validation functions in the DCNN classifier. Thus, 360,000 epochs (6×6×10×1000) were run during the 3-D optimization process. We used a supercomputer to complete such an extensive process. Table 2.3 shows the 3-D optimization results with varying numbers of CNN layers (columns) under different augmentation conditions (rows). The value corresponding to each model and augmentation condition is the accuracy (given as percentages). The optimization plot is shown in Figure 2.5 (left), while the bar charts representing the accuracies are shown in Figure 2.5 (right). In Figure 2.5 (left), the

optimization point is indicated by the black dot on the 3-D surface plot. As more DCNN layers and augmentation folds are added, the accuracy of CADx increases until it begins to fall gradually when 5-fold augmentation is exceeded. The optimization point (i.e., when there are 11 CNN layers and 5-fold augmentation) is called DCNN11A5\*.

R#	Column1	Column2	Column3	Column4	Column5	Column6	Column7
	Model	Balanced	Augm 2x	Augm 3x	Augm 4x	Augm 5x	Augm 6x
R1	DCNN5	66.49	62.67	62.74	70.328	57.43	63.66
R2	DCNN7	83.21	93.43	92.83	93.99	94.24	91.33
R3	DCNN9	82.74	88.61	92.22	93.41	95.41	92.91
R4	DCNN11	84.24	90.6	92.12	92.99	95.66*	92.66
R5	DCNN13	82.9	91.99	92.49	92.18	92.27	93.27
R6	DCNN15	84.24	90.24	91.29	93.91	95.4	93.16

Table 2.3: DCNN Layers vs. Augmentation Folds vs. Accuracy (%).



**Figure 2.5:** Left: 3-D optimization showing the best DCNN and augmentation combination. This corresponds to 11 DCNN layers and a 5-fold augmentation and is denoted as DCNN11A5\*. Right: Bar chart showing optimized location.

# 2.5.2 Benchmarking of three DL systems against the four ML systems and two Transfer learning

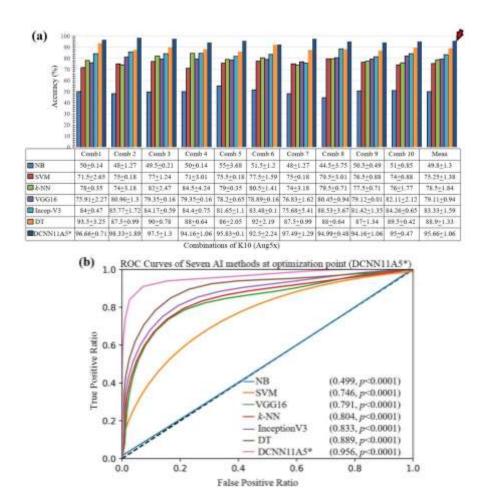
Benchmarking was conducted for DCNN11A5\* against the transfer learning base Inception V3, VGG16, and four ML systems (NB, *k*-NN, SVM, and DT). This was accomplished by using the K10 protocol, as discussed in the methodology section. The benchmarking results are shown in Table 2.4.

 Table 2.4: Benchmarking of seven artificial intelligence methods for plaque classification.

Row#	Column1 Column2		Column3	Column4	Column5				
KOW#	Model	Accuracy (%)	AUC	<i>p</i> -value	Confidence Interval				
	3 Types of Deep Learning Systems								
R1	VGG16	79.11+1.84	0.791	< 0.0001	77.7969 to 80.4371				

R2	InceptionV3	83.33+3.35	0.833	< 0.0001	80.9375 to 85.7345				
R3	DCNN11*	95.66±1.55%	0.956	< 0.0001	94.3502 to 96.9738				
	4 Types of Machine Learning Systems								
R4	NB	49.5+1.76	0.499	< 0.0001	48.0000 to 51.2625				
R5	SVM	74.65+2.44	0.746	< 0.0001	73.3781 to 77.1219				
R6	k-NN	80.4+2.19	0.804	< 0.0001	76.0977 to 80.9023				
R7	DT	88.90+1.03	0.889	< 0.0001	87.2156 to 90.5844				

Figure 2.6 (a) displays the complete summary of the mean accuracies and the corresponding standard deviations for the seven AI techniques (three DCNN methods and four ML methods). Figure 2.6 (b) illustrates the ROC curves for all seven types of AI systems. The receiver operating characteristics curve represents the relationship between sensitivity (true positive rate) and false-positive rate. The area under the curve (AUC) validates our hypothesis.



**Figure 2.6:** (a) Comparison of the accuracy rates of all AI methods at the optimization point, (b) ROC curves of all AI methods.

# 2.5.3 The system's ability to generalize at DCNN11A5\*: Memorization vs. Generalization

The CADx system's ability to generalize is an essential aspect of DCNN. This generalization point was computed by varying the percentage of cohort training data for the DCNN11A5\* system (11 DCNN layers with a 5-fold augmentation). K10 cross-validation was used for this purpose. Figure 2.7 shows the memorization vs. generalization curve (the small bar at each point represents the standard deviation). The accuracy increases as the percentage of the training data increase up to 50%. As the training data exceeds 50% (the so-called point of inflection), the accuracy dips and slightly fluctuates thereafter. Our results are very similar to previous generalization protocols for carotid plaque data sets.

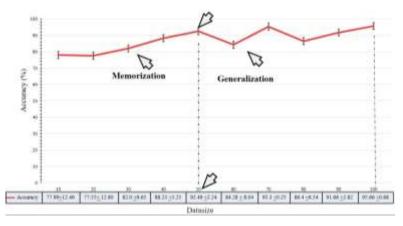


Figure 2.7: Memorization vs. generalization corresponding to DCNN11A5\* using K10 protocol.

# 2.5.4 Atheromatic<sup>™</sup> Plaque Separation Index Using DCNN11A5\*

Atheromatic<sup>™</sup> plaque separation index (APSI) evaluates a classifier's ability to determine how well symptomatic plaque has been separated from asymptomatic plaque. It is computed based on the MFS values derived from the DL and ML systems. Equation. 2.2 provides the formula used to compute APSI. Transfer learning is not considered in the separation index because we cannot deduce the MFS of the classes. Figure 2.8 shows the APSI values of four ML classifiers (NB, SVM, *k*-NN, DT); the corresponding values are represented in Table 2.5.

$$APSI_{K10}^{AImodel}(\%) = \left(\frac{|MFS_{AF} - MFS_{AS}|}{MFS_{AF}}\right) * 100$$
 (2.2)

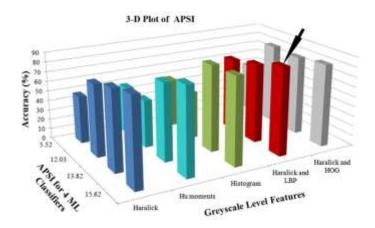


Figure 2.8: 3-D plot of APSI vs. features vs. accuracy for four types of ML classifiers.

Table 2.5: APSI vs. features vs. accuracy for four types of ML classifiers.

Classifier	Attribute	Hu Moments	Haralick & HOG	Haralick	Histogram	Haralick & LBP
	Sym	10.25 <u>+</u> 3.25	210.73 <u>+</u> 0.83	5247 <u>+</u> 11.57	0.27 <u>+</u> 0.09	2200 <u>+</u> 3.18
Naïve	Asym	10.85 <u>+</u> 1.25	185.36 <u>+</u> 0.59	4521.41 <u>+</u> 77.3	0.32 <u>+</u> 0.1	1850.75 <u>+</u> 21.2
Bayes	APSI	5.52	12.03	13.82	15.62	15.9
	Accuracy (%)	49.2 <u>+</u> 1.41	49.9 <u>+</u> 1.58	48.2 <u>+</u> 1.70	49.2 <u>+</u> 2.8	49.5 <u>+</u> 1.76
	Sym	10.25 <u>+</u> 3.25	210.73 <u>+</u> 0.83	5247 <u>+</u> 11.57	0.27 <u>+</u> 0.09	2200 <u>+</u> 31.8
SVM	Asym	10.85 <u>+</u> 1.25	185.36 <u>+</u> 0.59	4521.41 <u>+</u> 77.3	0.32 <u>+</u> 0.1	1850.75 <u>+</u> 21.2
SVIVI	APSI	5.52	12.03	13.82	15.62	15.9
	Accuracy (%)	48.2 <u>+</u> 0.91	82.20 <u>+</u> 1.65	73.1 <u>+</u> 1.75	48.05 <u>+</u> 1.02	74.65 <u>+</u> 2.44
	Sym	10.25 <u>+</u> 3.25	210.73 <u>+</u> 0.83	5247 <u>+</u> 11.57	0.27 <u>+</u> 0.09	2200 <u>+</u> 31.8
k-NN	Asym	10.85 <u>+</u> 1.25	185.36 <u>+</u> 0.59	4521.41 <u>+</u> 77.3	0.32 <u>+</u> 0.1	1850.75 <u>+</u> 21.2
W-1414	APSI	5.52	12.03	13.82	15.62	15.9
	Accuracy (%)	78.45 <u>+</u> 1.60	80.4 <u>+</u> 1.76	81.2 <u>+</u> 1.2	87.55 <u>+</u> 1.64	80.4 <u>+</u> 2.19
	Sym	10.25 <u>+</u> 3.25	210.73 <u>+</u> 0.83	5247 <u>+</u> 11.57	0.27 <u>+</u> 0.09	2200 <u>+</u> 31.8
DT	Asym	10.85 <u>+</u> 1.25	185.36 <u>+</u> 0.59	4521.41 <u>+</u> 77.3	0.32 <u>+</u> 0.1	1850.75 <u>+</u> 21.2
DI	APSI	5.52	12.03	13.82	15.62	15.9
	Accuracy (%)	88.25 <u>+</u> 0.7	83.65 <u>+</u> 1.47	87.70 <u>+</u> 0.8	88.30 <u>+</u> 1.47	88.90 <u>+</u> 1.03

# 2.6 Characterization of The Stroke Lesions

We used plaque characterization to validate our hypothesis that symptomatic plaque is hypoechoic (dark) and patchy (irregular) and contains many side lobes. We did this by using the MFS computed by the optimized DCNN11A5\* system. Figure 2.9 shows that the MFS of symptomatic plaque is higher than the MFS of asymptomatic plaque.

# a) Hypothesis - 1: MFS using DCNN11A5\*

MFS is the mean of the activation across all images within each class of a cohort. We computed the MFS of all four convolutions and average pooling layers. The output of these layers is a 2-D feature map, which was subsequently converted into a 1-D feature map. Furthermore, we computed the 1-D feature map for dense layers. The mean of the features

in the symptomatic class was higher than the MFS of the asymptomatic class, thus validating our hypothesis.

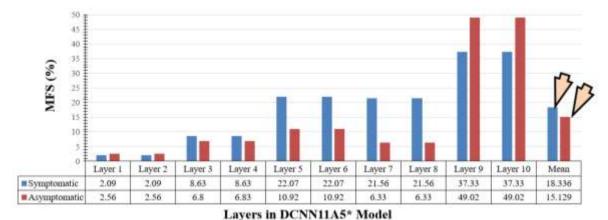


Figure 2.9: Mean feature strength of DCNN11A5\*.

We computed the intensity distribution of the 5-fold augmentation data to validate the MFS. Figure 2.10 shows the intensity distribution within the cohort. Here, A1 and A2 are the lobes of the symptomatic plaque, while B1 and B2 are the lobes of the asymptomatic plaque. The area of A1 is 10.28% greater than the area of B1. Meanwhile, the area of A2 is 15.89% greater than the area of B2. This confirms that MFS (symptomatic) > MFS (asymptomatic).

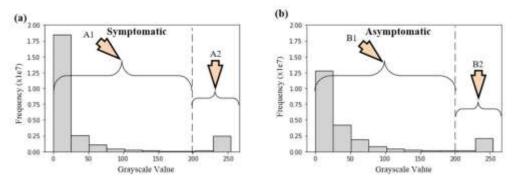


Figure 2.10: Histogram of symptomatic vs. asymptomatic plaques for the 5-fold augmentation data.

#### b) Hypothesis Validation 2: Bispectrum strength computation

Bispectrum (**B**) falls in the category of higher-order spectra and is calculated by varying the Radon transform of images from 0 to 180 degrees at 15-degree intervals. The Radon transform was applied to all images from the cohort and further calculations of the mean B values of the two classes. It was observed that the B value for the symptomatic class was **5.4%** higher than that of the asymptomatic class (*p*-value<0.0001).

Figure 2.11 illustrates the 3-D Bispectrum strength of symptomatic and asymptomatic plaques and shows that B(symptomatic) > B(asymptomatic). This finding

validates our hypothesis that symptomatic plaque is more hypoechoic and patchier than asymptomatic plaque.

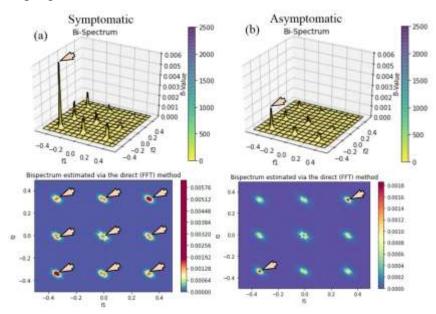


Figure 2.11: 3-D plot of Bispectrum. (a) symptomatic plaque (b) asymptomatic plaque.

# 2.7 Performance Evaluation

# 2.7.1 Diagnostics Odds Ratio

DOR was employed to distinguish between people with and without a specific disease (equation 2.3 and 2.4), taking a value between 0 and infinity. Higher DORs are associated with better test performance. A DOR of less than 1 implies that it is in the wrong direction and predicts the opposite outcome. Table 2.6 shows the increasing order of DOR for all classifiers using different kinds of features. Using the notation of sensitivity ( $S_e$ ) and specificity ( $S_p$ ), DOR is mathematically expressed as:

DOR= 
$$\frac{(S_e * S_p)}{(1-S_e)*(1-S_p)}$$
(2.3)

In this equation,  $S_e$  and  $S_p$  are mathematically defined as

Accuracy = 
$$\frac{TP+TN}{TP+TN+FP+FN}$$
;  $S_e = \frac{TP}{TP+FN}$ ;  $S_p = \frac{TN}{TN+FP}$   $TPR = \frac{TP}{TP+FN}$ ;  $FPR = \frac{FP}{FP+TN}$  (2.4)

where TP, FP, FN, and TN denote true positive, false positive, false negative, and true negative values, respectively.

#### 2.7.2 Power Analysis

We carried out a power analysis using MEDCALC 17.0 to test the validity of the selected

augmented data. The conventional formula is shown in equation 2.5

$$n = \left[ \left( z^* \right)^2 \times \left( \frac{\hat{p}(1-\hat{p})}{MoE^2} \right) \right] \tag{2.5}$$

Here, n is the number of samples required, z\* denotes the standard z-table,  $\hat{p}$  represents data proportion, and MoE stands for margin of error. We considered a 95% confidence interval with a 5% margin of error and 0.5 data proportion. Our dataset consisted of 2000 samples, which is **80.8%** higher than the required 384 samples.

Table 2.6: DORs for all the AI models using different features.

S N	Type	Model	TP	TN	FP	FN	Se	Sp	Pr	Re	DOR
1	ML	Naive Base with Haralick and HOG	10	99.7	0	100.3	0	1	0	0	0
2	ML	SVM with Histogram	10	99	1	100	0	0.99	0	0	0
3	ML	SVM with Hu moments	10	97	0	103	0	0.99	0	0	0.1
4	ML	NaiveBase with Haralick	106	0	94	0	0.99	0	0.53	1	0.1
5	ML	SVM with Haralick	43	94	0	63	0.41	0.2	1	0.41	0.17
6	ML	NaiveBase with Haralick and LBP	89.7	9.9	90.4	10	1	0	0.5	1	0.5
7	DL	CNN5	45.1	23.8	37.3	13.8	0.78	0.39	0.48	0.78	2.2
8	ML	NaiveBase with Hu moments	1	97	0	102	0.01	1	1	0.01	9.07
9	ML	KNN with Hu moments	82	74	23	21	0.8	0.76	0.78	0.8	12.49
10	ML	NaiveBase with Histogram	22	98	2	78	0.22	0.98	0.92	0.22	13.82
11	ML	KNN with Haralick and LBP	83.5	73.5	26.8	16.2	0.84	0.73	0.76	0.84	14.03
12	ML	KNN with Haralick and HOG	85.6	75.8	23.9	14.7	0.85	0.76	0.78	0.85	18.46
13	ML	SVM with Haralick and HOG	83.3	81.1	18.6	17	0.83	0.81	0.82	0.83	21.63
14	ML	KNN with Haralick	89	77	17	17	0.84	0.82	0.84	0.84	23.58
15	TL	VGG16	55	52	17	10	0.76	0.84	0.85	0.76	23.71
16	ML	DT with Haralick and HOG	85.3	82	16.4	16	0.84	0.83	0.84	0.84	26
17	ML	DT with Haralick and LBP	85.4 6	82.8 7	15.33	16.3	0.84	0.84	0.86	0.84	27.57
18	TL	Inception V3	85	82	17	16	0.84	0.83	0.83	0.84	37.31

19	ML	DT with Hu moments	90	85	12	13	0.87	0.88	0.88	0.87	48.56
20	ML	SVM with Haralick and LBP	50.2	100.	0	49.5	0.5	0.99	1	0.5	99.8
21	ML	KNN with Histogram	96	85	15	4	0.96	0.85	0.87	0.96	136
22	ML	DT with Haralick	99	87	7	7	0.93	0.93	0.93	0.93	171.7
23	DL	CNN13	29.7	30	2.22	2	0.93	0.93	0.93	0.93	176.5
24	ML	DT with Histogram	96	90	10	4	0.96	0.9	0.91	0.96	216
25	ML	CNN7	57.2	55.9	4.2	2.7	0.95	0.93	0.93	0.95	277.9
26	ML	CNN9	58.8	56	2.4	2.8	0.95	0.96	0.96	0.95	473.0
27	ML	CNN15	58	56.5	3.7	1.5	0.97	0.94	0.94	0.97	525.9
28	ML	CNN11	58	56.5	3.7	1.5	0.97	0.94	0.94	0.97	535.9

# 2.7.3 Hardware & Timing Analysis

Our experiment required extra computational power due to the nature of DCNN optimization, as it consisted of six DCNNs, six augmentation data sets, and 10 combinations using the K10 protocol. Therefore, we adopted a supercomputer framework. The specification comparison between the supercomputer and local computer is shown in Table 2.7.

Table 2.7: Comparison between LC and SC specifications.

Attribute	LC	SC
Processor Type	Intel Core i7-9750H CPU	Intel XenonES-2698 v4
Processor Speed	2.6 GHz	2.2 GHz
Architecture	1 Core	Dual 20-Core
RAM	8 GB	128 GB
Graphics Card	Not Included	Nvidia DGX128 GB
TFlops	Quad Core	8 x Tesla V 100 GPU with 960 TFlops
Cuda cores	1536	400,600
Tensor Cores	0	5,120 cores
Clock Speed	12 GB/sec	900 GB/sec
Visual	₹msi	

# SC: Supercomputer, LC: Local Computer

We compared our supercomputer (using NVIDIA DGX V100) timings against the local computer (having no NVIDIA card) timings. The reduced time values are provided in Table 2.8. Using the supercomputer, it took 5.30 hours to complete one combination of K10 with

1000 an epoch. Using a local computer, the same task took 72.22 hours or ~3 days.

Table 2.8: Timing analysis and time reductions for the SC against the LC.

Comparison between LC <sup>#</sup> vs. SC <sup>+</sup> (1-epoch in a sec)						
Aug. Fold	LC#	SC <sup>+</sup>	Reductio n (folds)			
2x Augm	150	19	7.89			
3x Augm	165	18	9.16			
4x Augm	210	19	11.05			
5x Augm	230	18	12.77			
6x Augm	260	20	13.00			

Augm - Augmentation.

#### 2.7.4 Cohen's Kappa analyses (Ka)

In circumstances when more than one ratter is involved, Cohen's kappa analysis is used to determine inter-rater agreement for categorical items. According to Cohen, kappa values should be interpreted in the following way: 0 means no agreement, 0.01-0.20 means a little agreement, 0.21-0.40 means reasonable agreement, 0.41-0.60 means moderate agreement, 0.61-0.80 means strong agreement, and 0.81 means full agreement. The equation for Cohen's kappa is given in supplementary material. Table 2.9 shows Cohen's kappa statistics. DCNN11A5\* presents the highest  $K_a$  value.

Table 2.9: Cohen's kappa analysis for AI methods (increasing order of Ka values).

AI Methods	NB	DCNN5	SVM	DT	k-NN	VGG16
Kappa (Ka)	0.05	0.16	0.50	0.54	0.57	0.59
AI Methods	IV3	DCNN13	DCNN7	DCNN15	DCNN9	DCNN11
Kappa (Ka)	0.61	0.87	0.88	0.90	0.91	0.93

#### 2.8 Discussion

# 2.8.1 Clinical Implications

One of the main clinical implications of this study is to offer a second opinion to vascular surgeons whose main objective is to decide if they should undergo stenting or endarterectomy [130] based on the plaque tissue characteristics. Since the grayscale contrast of the carotid scans are difficult to characterize using naked eyes when it comes to unstable vs. stable plaque, therefore, our Atheromatic 2.0 would be the ideal paradigm for making this decision. This tool can further benefit interventional cardiologists who are also keen to know the effect of carotid disease, which is the surrogate marker for coronary artery

disease [126]. Finally, since our online system is a few seconds long, this can be adapted in clinical settings which can be added to AtheroEdge<sup>TM</sup> 2.0 (AtheroPoint<sup>TM</sup>, Roseville, CA, USA) systems.

#### 2.8.2 Strengths/Weakness/Extensions

We optimized the conventional neural network to classify symptomatic and asymptomatic plaque using DCNN11A5\*. The system produced encouraging tissue characterization results by using MFS and bispectrum designs.

A noteworthy limitation of our study is that the original cohort is moderate in size, whereas deep learning systems ideally use large data samples. Moreover, the availability of the supercomputer (using NVIDIA DGX v-100) was limited during our study; thus, we implemented all the runs using six of the eight GPUs available at our facility. Transfer learning was implemented with reasonable accuracy to reduce the amount of training time needed. More sophisticated deep learning systems can be developed while keeping big data frameworks [131] in mind. Further, multi-center studies need to be conducted to examine the resultant effect on current classification and characterization paradigms.

#### 2.9 Conclusion

The present work is the first study to use AI strategies to classify and characterize plaque tissues by implementing DL using a supercomputer. The deep learning system's accuracy and AUC were 95.66% and 0.956 (p-value<0.0001), respectively. These results represent improvements of 7.01% over previous ML methods and 12.05% over other previous systems.

Further, the CADx system was benchmarked against four kinds of ML frameworks. We confirmed our hypothesis that symptomatic plaque is heterogeneous, dark (echolucent), and patchy by computing two novel strategies (i.e., by computing MFS using deep learning and bispectrum based on higher-order spectra). Our system's performance was evaluated using several methods (DOR, power analysis, supercomputer timing analysis, and Cohen's kappa analysis); it was also evaluated based on its ability to generalize the DCNN11A5\* system. All AI-based systems were validated using the most widely accepted facial biometric and animal datasets.

# **CHAPTER 3**

# OPTIMIZED MODEL FOR CLASSIFICATION AND CHARACTERIZATION OF THE STROKE IN MULTICENTER ENVIRONMENT

#### 3.1 Introduction

The variations in the plaque imaging led us to study the multicenter data analysis. Because we can identify how the pixel distribution varies in the scans at varied geological locations. It will further give plaque echogenicity changes in detection and characterization. Several studies on multicenter give a correlation of diseases with other symptoms which helps to detect the target precisely [132]. Among all the computer vision techniques for detecting plaque, Artificial Intelligence (AI) models provide flexible and reliable solutions to radiologists. Machine learning (ML) is a branch of AI that gives convenient solutions for identifying plaque. ML techniques need a handcrafted feature for predicting the risk. In ML with varied features, we will get the varied performance of the model. So, if we choose a multicenter study then ML models will be inconsistent. Deep learning (DL) provides a solution for that problem with automatic feature extraction by a deep convolutional neural network (DCNN) [133]. Although it provides automatic feature extraction for the multicenter study, we require an optimized point for the uniform study of all the centers. DCNN provides all the flexible environments for finding the optimization point. The only drawback of the DCNN is it requires higher computational power, which can be solved using transfer learning (TL). So in our study, we used these three techniques for comparison and contrast.

we previously worked on ML-based plaque detection using texture features [115], morphology-based features [118], and HOS & DWT features [115] with feature changes the performance of the AI model also changes which lead to different hypothesis on the

detection of the plaque. All these methods are unreliable and adhoc. In our current work, we are considering automatic feature extraction using DL techniques.

We are hypothesizing that symptomatic plaques are hypoechoic, and dark in ultrasound, which contains low collagen and calcium and contains high-level features. Whereas asymptomatic plaques are echogenic contain higher calcium and appear brighter in ultrasound. So we need to study multicenter ultrasound scans for variations of the features and echogenicity at an optimized point. Our study will give the variations/instability of symptomatic plaques at the different centers. This study also gives the variations of the intensity distributions at different centers. This internship led to an investigation into the high-level feature distributions. We conducted DL, TL, and ML optimization experiments for studying the centers from a uniform viewpoint. We classified the ultrasound plaques of the multicenter with three AI methods. This study is the first of its kind. DL 3-D optimization of the multicenter using a supercomputer paradigm.

# 3.2 Background Literature

Most of the research on carotid plaque characterization and detection is done by five groups. Namely Atheropoint Team, Cyprus Team, IST Team, and China Team. Figure 3.1 will give the perspective of the clusters. These four groups researched carotid plaque characterization and classification but all the teams' perspectives of the problem solving, and approaches are different. China group mostly works in MRI of carotid plaque [57, 134]. San Francisco group works on CT scans [135], and the IST team deals with the problem of atherosclerosis by calculating the risk index [68], whereas the. Their approach is based on echogenicity. The Cyprus team approach was using a statistical and clinical trial based. Atheropoint Team group approach was after the US scan by applying AI methods for early detection of the disease.

China Team works on the classification and characterization of carotid plaque using MRI imaging. In the beginning, they characterize the clinical study. Winn *et al.* characterize the MRI atherosclerotic fibrous caps using blinded clinical observations [57]. Ingersleben *et al.* characterized MRI carotid plaque at the bifurcation spot with the help of experienced radiologists [134]. Chu *et al.* characterize eight plaque types using cohen and kappa analysis [136]. Hatsukami *et. al.* classified vulnerable and nonvulnerable plaques using clinical oncology methods [24]. Zhao *et al.* did a multicenter study on Chinese patients

#### 3.2 Background Literature

from different locations using MRI with statistical measures such as a t-Test on high-risk plaques [137].

Classification and characterization of carotid plaques using CT scans were done by San Francisco Team. Wintermark *et. al.* developed a semi-automated CAD on CT scans using kappa analysis [138]. He also did a cross-sectional study to classify the acute carotid and nonacute carotid stroke patients using multivariate analysis [135]. Li *et al.* were able to characterize the relation between CT angiography with atherosclerotic cardiovascular disease score (ASCVD) using American heart association guidelines [139]. Zhu *et al.* built a semi-automated CAD for the characterization of carotid plaque using ASCVD with a multivariable linear regression model [60].

The above two teams worked on different image modalities of carotid plaque imaging. Here we will discuss three more teams who work on the same imaging technique i.e Ultrasound (US). Although all these teams worked on the same, their approach is different. First, we will discuss the Instituto Superior Técnico (IST) team. This team mainly focuses on finding CVD using a risk index [35, 70]. Seabra *et al.* use a 3D ultrasound technique for measuring the degree of stenosis and plaque volume using the surface rendering technique [68] and plaque inside the 3D volume [69]. He also classified carotid plaque using a machine learning technique Adaboost with texture features [35]. He also characterizes the homogeneous (asymptomatic) plaques with rayleigh distribution [36]. Afonso *et al.* proposed a CAD tool for measuring the plaque rupture risk using activity index and enhanced activity index (EAI) [70].

Cyprus team is also one of the major teams that worked on the classification and characterization of carotid plaque using statistical and clinical trials [106, 114]. Christodoulou *et al.* classified carotid ultrasound scans (CUS) using statistical neural self-organizing map (SOM) and k- Nearest neighbor (*k*-NN) by extracting texture features [29], and ten different Law texture energy features and fractal dimension features [71]. Kyriacou *et al.* developed a CAD system for predicting the period of stroke by using binary logistic regression and a support vector machine [34]. Nicolaides *et al* characterize the plaque using juxtraluminal components [140], asymptomatic carotid with Risk analysis [125], and plaque echogenicity using 3D ultrasound [141]. Atheromatic Team most of the work is on artificial intelligence application on classification and characterization of carotid plaques,

Acharya *et al.* classified 346 patients with texture-based characteristics using Local Binary Patterns (LBP) and Law's Texture Energy (LTE) atherosclerotic threat [109]. The same authors built the atheromatic 1.0 for plaque classification using SVM with discrete wavelet transform characteristics [110] and texture features [111]. Molinari *et al.* present a data mining approach that uses bidimensional empirical mode decomposition and entropy characteristics to classify symptomatic and asymptomatic plaque [37].

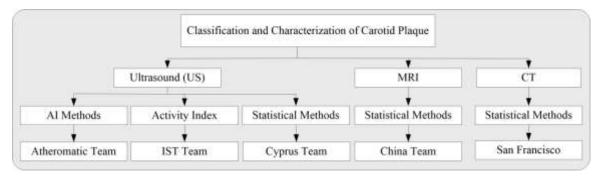


Figure 3.1: Perspective of clusters on carotid plaque study.

#### 3.3 Dataset

In our study we used two geological location datasets of the same disease, the patient selection process is discussed below, for both the dataset we took ethical committee approval from the corresponding institutions and patients' consent for this study. London dataset (LonD) extracted by Nicolaides *et al.* [142] contains 150 Symptomatic and 196 Asymptomatic patients. In this dataset the mean age of patients was 69.9±7.8 and 39% female, we included 50% to 99% of ICA stenosis patients who do not have any previous CORI symptoms and neurological abnormalities. In addition to the above, we included patients with contralateral cerebral hemispheric/retinal symptoms and also TIA patients if they are asymptomatic for more than six months. Along with St. Marry's Hospital, Imperial College, London, UK ethical committee approval we took consent from every patient for this study. In Figure 3.2 (a) Row, 1 and Row 2 represents symptomatic plaque whereas asymptomatic plaque is shown in Figure 3.2 (b) Row 1 and Row2.

Lisbon dataset (LisD) extracted by Sanches *et al.* [126] consists of 50 symptomatic and 110 asymptomatic scans, the mean age of patients was  $67.5 \pm 0.77$ , ethical committee approval taken from Instituto cardiovascular de liboa, Lisbon, Portugal. We included patients with focal transitory and neurological symptoms and the patients don't have any

symptoms about ipsilateral carotid territory in the past 6 months. A mixed dataset (MixD) is the balanced combination of LonD and LisD. Which contains 260 scans in each class.

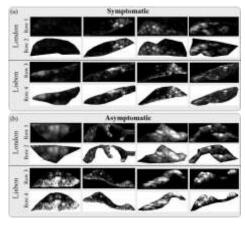
# 3.3.1 Ultrasound Image Acquisition and Plaque Delineation

Ultrasound scans of LD were taken from the Irving Laboratory for Cardiovascular Investigation and research at St. Mary's Hospital using standard image acquisition procedure with ATL HDI-3000 duplex scanner, advanced technology laboratories, Seattle, WA, USA. This machine was operated in a linear broadband width 4-7 MHz transducer. These images are having a resolution of 20 pixels/mm. Sample Images of full scans of LiD are shown in Figure 3.2 (a) Row 1 and Row 3.

Using "Plaque Texture Analysis software" developed by LifeQ Medical LTD was used for plaque delineation and normalization. ROI is selected using linear scaling with adventitia and blood as 2 reference points. The ROI of the plaque was delineated by medical practitioners using the mouse. Sample delineated plaques are shown in Figure 3.2 (a) Row 2 and Row 4

Ultrasound LiD is a color-flow duplex scan of both carotids. These scans are obtained using a Philips HDI5000 machine with a 5-12 MHz broadband linear array transducer. The resolution of the scans was 25 pixels/mm. Sample Images of full scans of LiD are shown in Figure 3.2 (b) Row 1 and Row 3.

In the plaque delenation, all these scans are normalized and linearly scaled. Here the intensities for adventitia and blood are in the range of 190-195 and 0-5. The ROIs are delineated by the medical practitioner using MATLAB by drawing around the plaque structure. These delineated plaques are resampled and smoothed using spline interpolation. Sample delineated plaques are shown in Figure 3.2 (b) Row 2 and Row 4.



**Figure 3.2:** (a) Sample ultrasound scans of 4 symptomatic patients. Row 1, Row3 represents full scan Row 2, Row 4 represents corresponding delineated plaque of London and Lisbon dataset (b) Sample ultrasound

scans of 4 asymptomatic patients. Row 1, Row3 represent full scan Row 2, and row 4 represents corresponding delineated plaque of London and Lisbon dataset.

# 3.4 Proposed CNN & TL Model

#### 3.4.1 DL Optimization

The global architecture of the DL frameworks is shown in Figure 3.3, we considered seven deep conventional convolutional neural networks (DCNN) for optimization over augmentation. All these seven models varied by a number of layers 5, 7, 9, 11, 13, 15, and 19. In all these models the number of combinations of convolution (CL) and average pooling (AP) varies but the number of dense layers and dropout layers are constant (2+1 Layers). We applied all these models to all the datasets augmentation folds using K10 cross-validation (90% training and 10% testing) to obtain an optimized point. So, the total number of experimentations is 1260 (3 datasets \* 6 Augmentations \*7 models \* 10 combinations).

In the DL architecture, we have x combinations of CL and AP layers, followed by a flattened layer that converts 2D Featuremaps to 1D Featuremaps. Then we have fully connected networks (so-called Dense Layer) with dropout. The output layer is the "softmax" layer. All the layers use the "ReLu" activation function. We adopted an "adam" optimizer for the learning protocol. We updated the weights using a categorical crossentropy loss function. Equation 3.1 gives a loss function where  $_i$  is the class label for input and  $a_i$  is the predicted probability of class being  $y_i$ .

Loss = 
$$-[(y_i \times \log a_i) + (1 - y_i) \times \log(1 - a_i)]$$
 (3.1)

#### 3.4.2 TL optimization

In addition to the DL models, we considered transfer learning models (TL) to reduce the hardware computation overheads in DL models. We considered Visual Geometric Group-19 (VGG19). Because this model is deep (contains 19 layers). The global architecture of VGG-19 is shown in Figure 3.4. We trained and tested VGG19 using the K10 protocol on all the augmentation folders to find the optimization point of VGG19 over augmentation. We implemented the VGG19 predefined model under the TL paradigm. i.e VGG19 is already turned and trained for classification Imagenet (natural images) those weights are transferred to the pre-trained model. In VGG19 under TL all the layers are

frozen with Imagenet weights except two dense layers. We fed our target labels to this VGG-19 for optimization.

#### 3.4.3 Machine Learning Framework for benchmarking

We benchmarked our DL methods with 6 ML methods such as logistic regression, linear deterministic analysis, support vector machines, decision tree, naïve bays, and random forest by extracting the Haralick, Hu-Moment, Histogram, LBP, and Histogram of Gradient, grayscale, mean pixel value, edge features from the carotid ultrasound scans, then fed to ML classifiers using K10 cross-validation protocol for predicting the class. Figure 3.5 will give the architecture of the ML paradigm

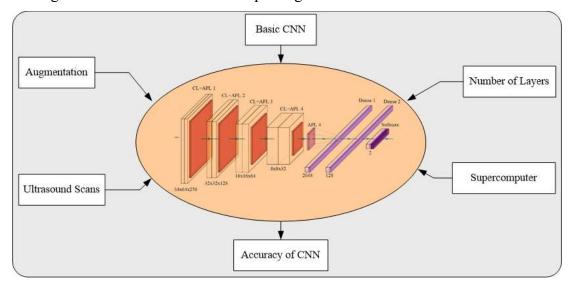


Figure 3.3: The global architecture of the DCNN model.

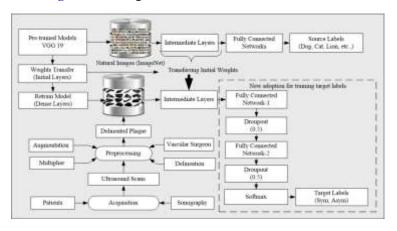


Figure 3.4: The global architecture of the TL model.

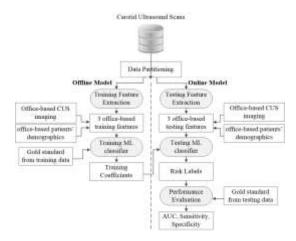


Figure 3.5: Online and offline ML Architecture.

# 3.5 Results

# 3.5.1 DL optimization

We optimized the DL with varied augmentations with varied layers on all three datasets (London, Lisbon, mixed). Figure 3.6 will represent the 3-D surface plots of three datasets. Table 3.1 represents the corresponding values.

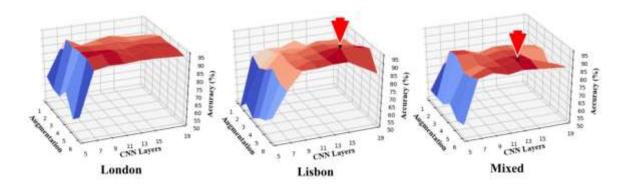


Figure 3.6: 3-D optimization curves of three centers.

Table 3.1: Centers vs. Layers vs. Augmentation vs. Accuracy.

	C1	C2	C3	C4	C5	C6	C7	C8	С9
R#	Augm x	Center	DCNN5	DCNN7	DCNN9	DCNN11	DCNN13	DCNN15	DCNN19
R1	Auroma	LonD	66.49	83.21	82.74	84.24	82.9	84.24	85.00
R2	Augm 1x	LisD	46.7	80.00	82.7	82.3	85.3	79.60	82.70
R3	1 X	MixD	56	72.1	77.5	74	73.6	72	73.5
R4	Auroma	LonD	62.67	93.43	88.61	90.6	91.99	90.24	89.16
R5	Augm 2x	LisD	54	75.9	84.6	85.6	84.6	87.8	86.3
R6	2.X	MixD	56	85.1	84.3	81.1	83.7	83.5	80.9
R7	A	LonD	62.74	92.83	92.22	92.12	92.49	91.29	93.33
R8	Augm 3x	LisD	49	79.7	89.2	87.2	89.8	90.3	87.1
R9	3X	MixD	56	86.6	86.3	83.8	85.7	84.6	82.9
R10	Augm	LonD	70.328	93.99	93.41	92.99	92.18	93.91	92.16

R11	4x	LisD	54.3	84.7	87.7	88.6	90.7	91.9	90.6
R12		MixD	61.6	80.2	83.2	85.1	86.2	85.6	76.2
R13	Auam	LonD	57.43	94.24	95.41	95.66*	92.27	95.4	92.83
R14	Augm 5x	LisD	58.2	80.5	88.3	91.6	91.2	96.9*	93.7
R15	JX	MixD	55.3	82.1	86.5	85.8	88.1*	87.5	81.2
R16	Auam	LonD	63.66	91.33	92.91	92.66	93.27	93.16	92.16
R17	Augm 6x	LisD	64.8	80.4	91.3	91.4	92.3	92.3	92.4
R18	UX	MixD	57.4	85.2	87.4	85	86.5	85.6	83.9

We choose the K10 cross-validation protocol for 3-D optimizing the deep convolutional neural network (DCNN). We took a number of epochs per combination is 1000 by experimental observation. So to complete 3-D optimization, we ran 12, 60, 000 runs (3x7x6x10x1000). Hence, we require massive computational power, so we used supercomputers. After optimization LonD optimized at (DCNN11, Augm5) with 95.66% accuracy, LisD at (DCNN15, Augm5) with 96.90% accuracy whereas MixD at (DCNN13, Augm5) with 88.10% accuracy. Here we observed that all the centers are optimized at augmentation fold 5.

# 3.5.2 TL optimization

We have used 12, 60, 000 massive runs for DL optimization. So it requires massive computations, to avoid this overhead we opted TL paradigm using the Visual Geometrical Group -19 (VGG19) model. For TL optimization also choose the K10 CV protocol with 1000 epochs. 3-D optimization of the TL is shown in Figure 3.7, corresponding values are shown in Table 3.2. LonD & MixD was optimized at augmentation fold 6x with 91.56% and 95.73 % accuracy, and LisD was optimized at augmentation fold 5x with 96.38% accuracy.

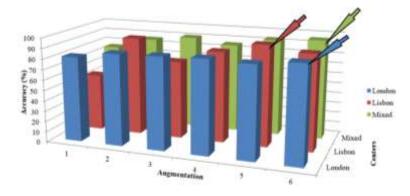


Figure 3.7: 3-D optimization of VGG19.

Table 3.2: Centers vs. TL vs. Augmentation vs. Accuracy.

Center	Augm 1x	Augm 2x	Augm 3x	Augm 4x	Augm 5x	Augm 6x

London	81.50	87.33	88.07	89.08	87.5	91.56*
Lisbon	54.66	94.45	74.66	87.16	96.38*	91.24
Mixed	75.69	86.14	90.62	85.78	93.334	95.73*

# 3.5.3 ML Optimization

We compared our DL model performance with well-established ML techniques. We extracted Local Binary patterns (LBP), Histogram of Oriented Gradients (HOG), and Humoments from the carotid ultrasound scans (CUS) then we fed these features to four machine learning classifiers such as *k*-nearest neighbor (*k*-NN), Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF) with K10 CV protocol. Figure 3.8 will show the 3-D bar representation of ML optimization against augmentations and features. The corresponding values are represented in Table 3.3.

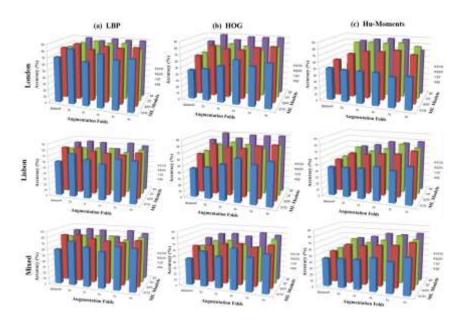


Figure 3.8: ML 3-D Optimization using (a) LBP Features, (b) HOG Features (c) Hu-Moments.

We computed the area under the curve (AUC) using the receiver operating characteristic curve (ROC). We observed that the order of AI methods varies on centers, for LonD the order is DL> TL>ML, LisD DL>TL>ML whereas in mixed TL>DL>ML. Figure 3.9 will represent the ROC curves of three dataset

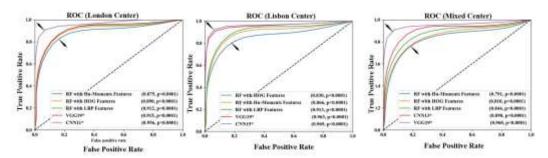


Figure 3.9: ROC curves of three centers.

Table 3.3: Machine learning techniques Feature vs. Model vs. Augmentation vs. Accuracy.

ML Tech	Center	Footumos	Accuracy (%)						
WIL Tech	Center	Features	Balanced	2x	3x	4x	5x	6x	
		LBP	70.00	83.75	66.66	80.00	73.00	76.66	
	London	HOG	45.00	48.75	55.83	66.87	59.50	66.25	
		HU- Moments	50.00	48.75	49.16	50.00	46.00	48.75	
		LBP	58.12	65.00	66.66	55.00	63.33	67.77	
SVM	Lisbon	HOG	40.00	46.66	45.55	60.00	58.00	55.00	
		HU- Moments	33.33	48.33	42.22	47.50	48.66	49.44	
		LBP	59.48	74.47	66.41	61.06	70.88	70.18	
	Mixed	HOG	44.80	58.95	52.62	67.68	55.58	62.61	
		HU- Moments	44.99	45.95	47.27	52.12	48.46	56.91	
		LBP	80.00	87.50	80.00	86.25	85.00	87.50	
	London	HOG	62.50	80.00	76.66	76.25	81.50	84.58	
		HU- Moments	57.50	68.75	74.16	75.00	79.00	74.16	
		LBP	78.11	81.66	81.11	77.50	82.66	82.77	
k-NN	Lisbon	HOG	66.66	78.33	65.55	73.33	73.33	69.44	
		HU- Moments	66.66	65.00	77.77	70.83	71.33	72.77	
		LBP	79.60	81.01	75.89	73.47	72.36	77.43	
	Mixed	HOG	61.14	63.93	68.00	70.05	65.55	70.84	
		HU- Moments	50.95	63.26	63.20	65.76	67.25	73.61	
		LBP	77.50	85.00	89.16	86.87	88.12	80.41	
	London	HOG	62.50	75.00	80.00	77.50	73.00	79.58	
		HU- Moments	32.50	82.50	84.16	83.75	85.50	81.66	
		LBP	79.48	90.00	84.44	81.66	88.33	86.11	
DT	Lisbon	HOG	66.66	65.00	75.55	70.83	62.00	71.11	
		HU- Moments	86.66	68.33	80.00	75.00	74.66	76.11	
		LBP	73.26	75.45	76.38	77.39	72.02	77.97	
	Mixed	HOG	56.34	66.24	68.89	64.34	61.08	75.40	
		HU- Moments	50.57	67.88	63.17	68.82	77.63	73.46	
RF	London	LBP	77.50	82.50	85.83	91.25*	89.12	89.58	
KI	London	HOG	77.50	82.50	82.50	88.75	89.00	90.00*	

	HU-	40.00	81.25	85.00	87.50*	87.50	71.25
	Moments						
	LBP	88.46	88.66	88.88	80.83	88.00	90.55*
Lisbon	HOG	66.66	78.33	71.11	77.50	80.66	82.22*
	HU-	86.66*	75.00	74.44	78.33	78.66	76.66
	Moments						
	LBP	80.57	84.21	82.76	72.58	84.46*	80.67
Mixed	HOG	66.15	75.12	76.92	73.51	80.94*	78.12
	HU-	49.60	65.18	69.03	75.71	75.07	78.87*
	Moments						

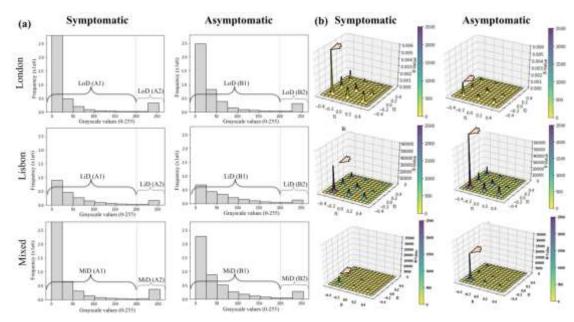
# 3.6 Characterization of lesions using statistical and AI methods

We hypothesized that the plaque's higher features depend on the distribution of the pixel intensity, with the help of a histogram we plotted the greyscale pixel intensity distributions. Figure x will represent the three datasets (centers) histogram. We computed the lobe area of each dataset and compared it. In London and Mixed datasets, A1 > A2 and B1 > B2. But in Lisbon dataset A1 < A2 and B1 > B2. The detailed lobe area is shown in Table 3.4.

	Lobe	Bin range	Lobe Area (pixel²)
atic	LonD (A1)		4371134.0
Symptomatic	LisD (A1)	0 to 200	1989160.0
	MixD (A1)		5238100.0
Syr	LonD (A2)		0355340.0
	LisD (A2)	201 to 255	2003020.0
	MixD (A2)		0406783.0
	LonD (B1)		4051470.0
ıtic	LisD (B1)	0 to 200	0189192.0
ma	MixD (B1)		4348156.0
ıpto	LonD (B2)		0345663.0
Asymptomatic	LisD (B2)	201 to 255	0161094.0
	MixD (B2)		0327508.0

Table 3.4: Histogram Lobe Area.

We further validated our hypothesis using a higher-order spectrum (BSpectrum) by varying radon angel from 0 to 180 degrees. In the Bspectrum also we got Bspectrum (Symptomatic) > BSpectrum (Asymptomatic) in LonD and MixD. Whereas in LisD Bspectrum (Symptomatic) < BSpectrum (Asymptomatic). This proves that grayscale pixel intensity distribution in LisD varies with other centers. Figure 3.10. Represents the 3-D Bspectrum representation



**Figure 3.10:** (a) Grayscale intensity histograms of three datasets. (b) 3-D Bispectrum of corresponding datasets.

# 3.6.1 Mean Feature Strength (MFS)

MFS is used for validating our hypothesis that symptomatic plaque is hypoechoic (darker) and has higher features. We computed MFS at the optimization point of all the centers. Figure 3.11 will give the MFS of three centers. We observed that all three centers show the same characteristic MFS (Sym) > MFS (Asym).

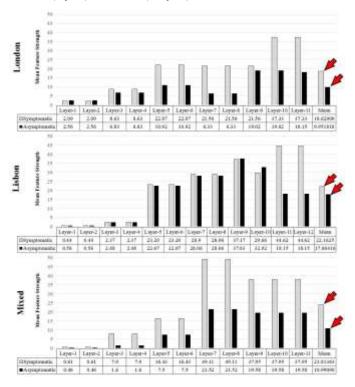


Figure 3.11: MFS of three centers at each layer in optimized DCNN.

#### 3.7 Performance Evaluation Metrics

# 3.7.1 Diagnostics Odds Ratio (DOR)

DOR is an effective diagnostic test for medical imaging. Which is a ratio of target disorder to subjects without it. It takes any value between 0 to infinity. Higher DOR means higher performance. We can calculate DOR from sensitivity (S<sub>e</sub>) and specificity (S<sub>p</sub>) shown in equation 3.2. Table 3.5 gives the DOR of AI methods used in this study in increasing order.

DOR= 
$$\frac{(S_e * S_p)}{(1-S_e)*(1-S_p)}$$
 (3.2)

# 3.7.2 Cohen and Kappa Analysis (Ka)

It measures the inter-rater agreement between categorical items.  $K_a$  is a strong measure to understand the agreement between them. If the  $K_a$  value is higher than 0.60 then there exists enough agreement between the two variables. Equation 3.3 gives the Cohen and kappa analysis. Table 3.5 gives the  $K_a$  analysis of our study

$$K_{a} = \frac{P_{ro} - P_{e}}{1 - P_{e}} \tag{3.3}$$

Here  $P_{ro}$  is the relative observed agreement among raters, and  $P_e$  is the hypothetical probability of chance agreement. All these performance evaluations are made at the optimization point of individual centers.

Table 3.5: DOR and Ka Analysis at optimization point (Increasing order at individual centers).

Center	AI Method	DOR	Ka
	CNN5	2.20	0.158
	RF With HOG	48.56	0.621
	CNN19	72.06	0.66
	TL- VGG19	98.56	0.731
	RF With Hu-moments	129.45	0.698
London	CNN13	176.51	0.873
	CNN15	182.36	0.9
	RF with LBP	216.00	0.767
	CNN7	277.91	0.883
	CNN9	473.05	0.901
	CNN11	535.98	0.931
	RF With HOG	2.08	0.639
	CNN5	26.59	0.669
	CNN7	42.85	0.733
Lisbon	RF with LBP	47.25	0.752
Lisbon	RF With Hu-moments	66.93	0.705
	CNN11	91.09	0.8
	CNN9	102.50	0.746
	CNN19	105.00	0.752

	CNN13	120.00	0.832
	TL- VGG19	158.40	0.879
	CNN15	211.00	0.765
	CNN5	0	0
	CNN7	24.21	0.605
	CNN19	25.91	0.619
	CNN11	30.10	0.649
	RF With HOG	38.92	0.669
Mixed	CNN15	48.70	0.732
	RF With Hu-moments	55.85	0.625
	CNN9	69.30	0.735
	RF with LBP	96.53	0.736
	TL- VGG19	143.00	0.849
	CNN13	214.15	0.767

# 3.7.3 Atheromatic Plaque Separation Index (APSI)

APSI is given by Atheromatic <sup>TM</sup>, which gives the ability to classifier how well it separates between the classes. It will be computed based on the mean feature strength (MFS) of DL and ML systems. In our study, we used the TL paradigm also, but finding the MFS for TL is not relevant for APSI. Because it is a fusion of features from natural images and target labels. Equation 3.4 gives the formula for the computation of APSI with MFS <sub>AF</sub> of symptomatic and MFS <sub>AS</sub> of asymptomatic of each center. The APSI DL will give a strong separation index as shown in Table 3.6.

$$APSI_{K10}^{AImodel}(\%) = \left(\frac{|MFS_{AF} - MFS_{AS}|}{MFS_{AF}}\right) *100$$
 (3.4)

Table 3.6: APSI of AI methods used in this study.

AI Method	Center	Symptomatic	Asymptomatic	APSI
Al Michiga	Center	$(MFS_{AF})$	$(MFS_{AS})$	(%)
DL		18.62	9.95	46.56
ML with LBP	London	2.76	2.40	13.04
ML with HOG		0.055	0.051	7.27
ML with Hu-Moments		0.045	0.035	22.22
DL		22.16	17.86	19.4
ML with LBP	Lisbon	2.81	2.26	19.57
ML with HOG	Lisoon	0.057	0.06	5.00
ML with Hu-Moments		0.062	0.06	3.23
DL		23.81	10.99	53.84
ML with LBP	Mixed	2.92	2.18	25.34
ML with HOG	iviixed	0.057	0.062	8.77
ML with Hu-Moments		0.063	0.069	9.52

#### 3.8 Discussion

#### 3.8.1 Mixed Cohort

In our study, we used mixed LisD and LonD cohorts called MixD (Mixed Cohort). We used a mixed cohort to validate the cross-sectional study between the two centers. We achieved all the characteristics three datasets are the same except the Bispectrum of LisD, which is due to the intensity distribution of the Lisbon cohort varies with the London cohort. So, a mixed cohort helps to identify the cause of the disturbance.

A mixed dataset consisting of London patients who are British ethnicity, and Lisbon they are more Spanish ethnicity. However, they are from different geographical locations. There is no influence of this difference in our study. Because we are not doing any measurements of IMT and not considered ethnicity. There is no influence of geological location on ML, and DL performance. Because we are considered only delineated plaque sections in this study.

#### 3.8.2 Strengths/Weakness/Extensions

This study is the first of its kind, we achieved an optimization point in the conventional neural network for the classification of CUS scans on different geological locations. Our study was able to characterize the symptomatic plaque using MFS, Bispectrum, and histogram.

The limitation of this study was the availability of supercomputers is not sufficient in developing countries. However, we can achieve the supercomputer slot for executing our massive slot. We used an augmented dataset for all three centers due to the number of samples is moderate. If the original dataset is large enough to perform the advanced AI techniques, then we can deduce more reliable performance. We are further working on advanced deep learning techniques with modified conventional convolutional neural networks which gives a more comprehensive analysis of the classification and characterization.

#### 3.9 Conclusion

This study can classify and characterizes the CUS scans using AI methods on different geological locations. we achieved 95.66% (AUC:0.956), 96.90% (AUC:0.969) and 88.1% (AUC:0.89) accuracy in DL, 91.56% (AUC:0.915), 96.38% (AUC:0.963) and 95.73% (AUC:0.96) accuracy in TL, 91.25% (AUC:0.912), 91.3% (AUC:0.913), and 84.46% (AUC:0.844) in ML on London, Lisbon, and Mixed cohort respectively. Our study shows an improvement of 10.41% (London), and 3.32% (Lisbon) higher than existing work. Our study shows the characterization using MFS of symptomatic plaque is higher than

# 3.9 Conclusion

asymptomatic plaque by 46.56%, 19.40%, and 53.84% higher than asymptomatic in all three centers.

# **CHAPTER 4**

# OPTIMIZED MODEL FOR CLASSIFICATION AND CHARACTERIZATION OF THE STROKE USING TRANSFER LEARNING

#### 4.1 Introduction

Several popular models exist in TL, and each model offers its own merits and demerits. For example, few models are focused on fast optimization while some are aimed at hyperparameter reduction. Few others apply the TL paradigm in edge devices, such as NVIDIA Jetson (www.nvidia.com), or Raspberry Pi (from Rasberry Pi Foundation, UK) [53]. Few applications of TL have been developed in medical imaging such as the classification of Wilson disease [17], COVID pneumonia [143], brain tumor [144], etc which has shown superior performance over DL. In this study, we choose ten types of TL architectures, where each *one* of these carries advantages such as (a) intense neural network, (b) modified kernel sizes, (c) solving vanishing gradient problems, and (d) feedforward nature to the features [52]. Therefore, we hypothesize that the performance of TL is superior to or comparable to DL.

The architecture of the online AI model is shown in Figure 4.1. It contains five blocks (i) image acquisition, (ii) pre-processing, (iii) AI-based models, and (iv-v) performance evaluation and validation. The image acquisition block is used for scanning the internal carotid artery. These scans are normalized and manually delineated in the pre-processing block to obtain the plaque region of interest (ROI). Even though we do not show the augmentation during the online processing, it is embedded in the pre-processing block that helps in increasing the cohort size. The online AI model block helps to determine the unknown label of the symptomatic *vs.* asymptomatic plaques. This is accomplished by transforming the test plaque image and using the offline training TL model or DL models.

In our online paradigm, since there are 11 models, we run each test patient's plaque using 11 (10 TL + 1 DL) different AI models for predicting 11 kinds of labels. We determine the performance of these 11 architectures and are followed by the ranking of their performance.

We proposed an optimized TL model for carotid ultrasound-based plaque tissue classification (Atheromatic<sup>TM</sup> 2.0<sub>TL</sub>, AtheroPoint<sup>TM</sup>, Roseville, CA, USA). Because the features using this system are computed using a deep learning paradigm, we hypothesize that the performance of TL is superior and/or comparable to DL. Lastly, we have also designed a computer-aided diagnostics (CAD) system for computing the heatmaps using an AI-based approach.

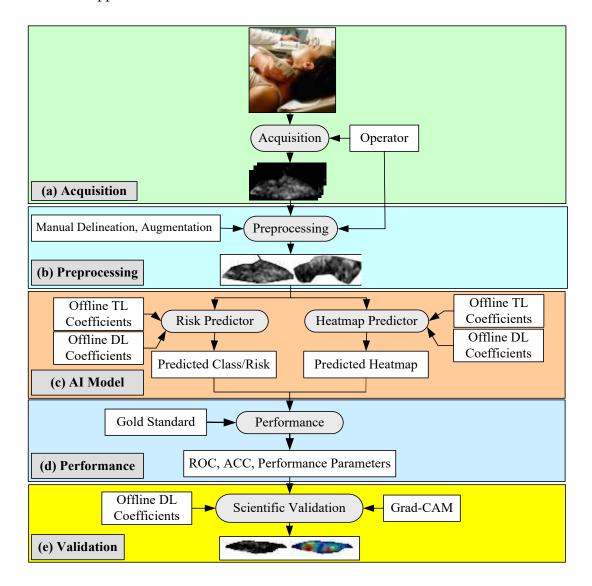


Figure 4.1: Online AI architecture of the Atheromatic™ 2.0TL study (TL: Transfer Learning, DL: Deep Learning, and Grad-CAM: Gradient-weighted Class Activation Mapping).

# **4.2** Methodology for Classification of Stroke Using Transfer Learning Methods

The choice of the TL architecture for lesion tissue characterization was motivated by (a) the diversity of the TL models and (b) the depth of the neural network models. Thus, we took two architectures from the VGG group (VGG-16 & 19), two architectures from the DenseNet architectures (DenseNet121, 169), and two architectures from the ResNet architectures (ResNet50, 101). All these models had a depth of neural networks extending to 169 layers while ensuring diversity. Note that some of the architectures such as MobileNet and XceptionNet are the most current, state-of-the-art, and popular TL architecture that demonstrates faster optimization (see Figure 4.2).

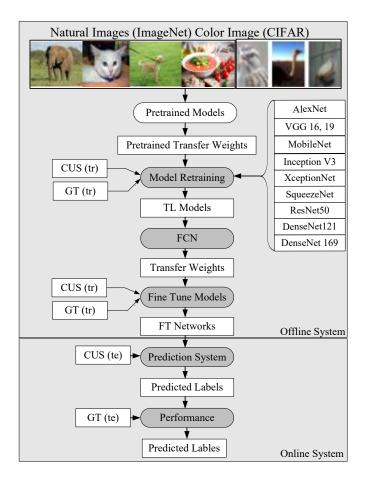


Figure 4.2: Global TL architecture using 10 different TL models ((i-ii) Visual Geometric Group-16, 19 (VGG16, 19); (iii) Inception V3 (IV3); (iv-v) DenseNet121, 169; (vi) XceptionNet; (vii) ResNet50; (viii) MobileNet; and (ix) AlexNet, and (x) SqueezeNet), te, stands for testing and tr stands for training, FN: fine-tune networks.

#### 4.2.1 Deep learning architecture

In our study, we benchmarked TL architectures with two DL architectures. One is

# 4.2 Methodology for Classification of Stroke Using Transfer Learning Methods

conventional CNN and another one is Modified Unet architecture. Although the UNet network is very popular for segmentation in medical image analysis, we used modified UNet architecture called Modified Unet for classification purposes. In the proposed Modified Unet architecture, we used separable convolution neural networks to reduce the overfitting and the number of parameters required for training. Figure 4.3 shows the Modified Unet architecture. Table 4.1 gives the detailed number of training parameters for Modified Unet.

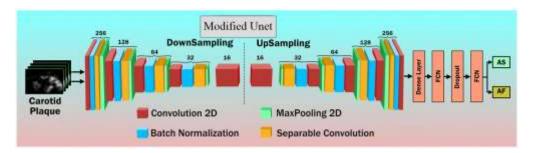


Figure 4.3: Modified Unet architecture.

Layer Type	Shape	#Param
Convolution 2D	128x128x32	896
Batch normalization	128x128x32	128
Separable Convolution 2D	128x128x64	2400
Batch normalization	128x128x64	256
MaxPooling 2D	64x64xx64	0
Separable Convolution 2D	64x64x128	8896
Batch normalization	64x64x128	512
MaxPooling 2D	32x32x128	0
Separable Convolution 2D	32x32x256	34176
Batch normalization	32x32x256	1024
MaxPooling 2D	16x16x256	0
Separable Convolution 2D	16x16x64	18752
Batch normalization	16x16x64	256
MaxPooling 2D	8x8xx64	0
Separable Convolution 2D	8x8x128	8896
Batch normalization	8x8x128	512
MaxPooling 2D	4x4x128	0
Separable Convolution 2D	4x4x256	34176
Batch normalization	4x4x256	1024
MaxPooling 2D	2x2x256	0
Flatten	1024	0
Dense	1024	1049600
Dropout	0.5	0
Dense	512	524800
Dropout	0.5	0
Dense (softmax)	2	1026

Total Trainable Parameters	1,687,330
----------------------------	-----------

#### 4.2.2 Experimental Protocol

Our study used 12 AI models (10 TL and 2 DL) with six augmentation folds and 1000 epochs using the K10 cross-validation protocol. It totaling to ~720,000 (720K) runs for finding the optimization point of each AI model. The mean accuracy of each model is calculated by using the following section.

#### Accuracy bar charts for each cohort corresponding to all AI models

If  $\eta(m,k)$  represents the accuracy of an AI model "m" using cross-validation combination "k" out of total combinations K, then the mean accuracy for all the combinations for the model "m", represented by  $\bar{\eta}(m)$  can be mathematically given by equation 4.1. Note that we considered the K10 protocol in our paradigm so K=K10=10.

$$\bar{\eta}(m) = \frac{1}{\kappa} \sum_{k=1}^{K} \eta(m,k) \tag{4.1}$$

# 4.3 Results

This section discussed *three* sets of experimentations for comparison of TL vs. DL to prove the hypothesis. The first experiment is the 3D optimization of the ten TL architectures by varying the augmentation folds. The second experiment is the 3D optimization of the Modified Unet architecture by varying the same fold. The third experiment was the benchmarking of TL architectures with Modified Unet and CNN by calculating the AUC.

#### 4.3.1 3D Optimization of TL Architectures and benchmarking against CNN

In this experiment, we used all the TL architectures for finding the optimized TL by varying the augmentation folds. There are 10 TL architectures, six augmentation folds, a K10 cross-validation protocol, and 1000 epochs. To train the model by empirically selecting each model's flatten point at a loss *vs.* accuracy, thus there were  $12x6x10x1000 \sim 720K$  runs. We used a total of 7,20,000 runs to obtain the optimization point. This is a reasonably large number of computations and needs high computation power. Thus, we used the Nvidia DGX V100 supercomputer at Bennett University, Gr. Noida. Figure 4.4 shows the performance of ten AI architectures, and the red arrow indicates the optimization point for each AI model when ran over 6 augmentations. The corresponding values are represented in Table 4.2. Using equation 4.1, we calculate the mean accuracy of the AI models.

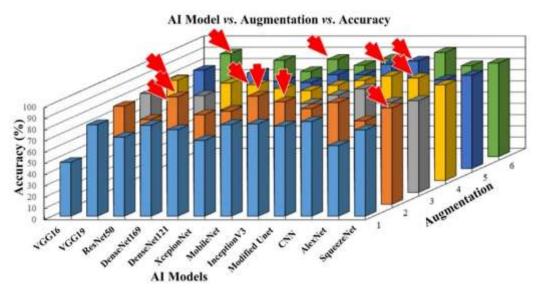


Figure 4.4: 3D bar chart representation of the AI model accuracy vs. augmentation folds.

As seen in Figure 4.4, MobileNet, and DenseNet 169 shows better accuracy than other TL architectures. They showed **96.19%** and **95.64%** of accuracy, respectively. Aug 2x is the optimization point for both models. Table 4.3 shows the comparison between ten types of TL which include VGG16, VGG19, DenseNet121, XceptionNet, MobileNet, AlexNet, InceptionV3, and SqueezeNet with seven types of DL. The ten types of TL and seven types of DL include CNN5, CNN7, CNN9, CNN11, CNN13, CNN15, and Modified Unet, respectively. Note that CNN5 to CNN15 was taken from our previous study [129].

Table 4.2: Accuracies of 10 TL and 2 DL models for 6 augmentations.

AI Model	Balanced	Aug 2x	Aug 3x	Aug 4x	Aug 5x	Aug 6x
VGG16	48	47.5	47.97	66.72	79.12	70.87
VGG19	81.5	87.33	88.07	89.08	87.5	91.56
ResNet50	70.4	75.4	78.2	70.5	68.7	66.5
DenseNet169	80.9	95.64	86.14	86.57	85.06	85.66
DenseNet121	76.99	79.69	73.29	85.17	77.33	75.81
Xception Net	67.49	82.74	79.99	81.87	76.49	86.55
MobileNet	81.49	96.19	72.82	79.99	83.59	81.24
InceptionV3	82.18	91.24	79	84.69	83.33	86.88
Modified Unet	80.32	85.09	86.50	88.93	92.77	84.95
CNN [129]	84.24	90.6	92.12	92.99	95.66	92.66
AlexNet	62.84	74.29	80.21	91.09	78.81	80.91
SqueezeNet	74.65	83.20	79.23	83.12	81.33	82.00

In Modified Unet architecture, there are 22 layers, while there is a varying number of

layers in CNN architecture ranging from five to 15. It is important to note that all CNNs except the CNN5 have accuracies above 92.27%. The overall mean and standard deviations of the DL accuracies were 90.86±3.15%. The innovation of the current study was the design and development of TLs. They are benchmarking against DL. In Table 4.3, the mean and standard deviation of ten TLs was 89.35±2.54%. Thus, the mean accuracy of TL systems is comparable to the mean accuracy of DL systems and in the range of ~1%. MobileNet has the highest accuracy among all the TL systems (96.19%), while CNN11 has the highest accuracy among all the DL systems (95.66%). Further, it is essential to note that the mean accuracy variations are less than equal to 3% within the limits of good design and operating conditions (typically, regulatory approved systems have variations less than 5%).

#### 4.3.2 3D Optimization of Modified Unet

In this set of experiments, we used the popular UNet architecture model for classification. Figure 4.3 represents the Modified Unet architecture inspired by UNet. We optimized the Modified Unet by varying the augmentation folds. Here, we also used the K10 CV protocol for training and testing. We choose 1000 epochs empirically. Therefore, the total number of runs for optimizing the Modified Unet leads to 60,000 (1 Modified Unet x 6 Aug folds x 10 combinations x 1000 epochs). We used the same set of hardware resources (used in the previous section) for this experiment. Table 4.3 represents the average accuracy at the augmentation folds. Modified Unet is optimized at aug 5x with an accuracy of 92.77 percent.

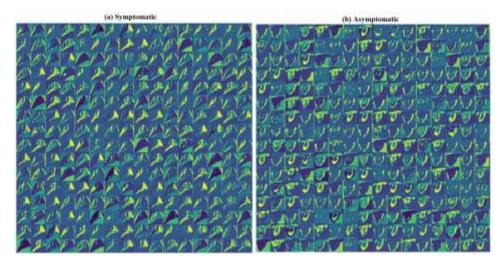
TL Type	TL Acc. (%)	DL Type	DL Acc. (%)
VGG16	79.12	CNN5	70.32
VGG19	91.56	CNN7	94.24
DenseNet169	95.64	CNN9	95.41
DenseNet121	85.17	CNN11	95.66*
Xception Net	86.55	CNN13	92.27
MobileNet	96.19*	CNN15	95.40
InceptionV3	91.24		
AlexNet	91.09	Modified	92.77
SqueezeNet	83.20	Unet	92.11
ResNet50	78.20		
Best TL	96.19	Best DL	95.66
Absolute Diff	0.53		

Table 4.3: TL systems vs. DL Systems.

#### 4.3.3 Visualization of the Modified Unet

<sup>\*</sup>Highest accuracy.

We visualized the intermediate layers of the Modified Unet to understand the learning ability of the model over CUS. Figure 4.5 represents the mean visualization of the training samples of symptomatic and asymptomatic classes from all the filters at the end layer before vectorization. Turquoise color represents the learned features, yellow represents the high-level features, and green represents the low-level features.



**Figure 4.5:** (a) Visualization of the intermediate layers of Modified Unet on symptomatic class (b) asymptomatic class.

#### 4.4 Performance Evaluation

This section aims to evaluate the samples required for the study using standard power analysis. Since we are using 12 AI models (10 TL, 2 DL), it is necessary to rank the model by considering all the performance parameters for finding the best performing AI model among the 12 AI models. In addition to that, we compared the performance of all 12 AI models with area-under-the-curve (AUC) using the receiver operating characteristic curve (ROC).

# 4.4.1 Power Analysis

We did a standardized protocol (power analysis) for analyzing the number of samples required at a certain threshold of the error margin. We considered a 95% confidence interval with a 5% margin of error and a data proportion of 0.5. We use Equation 4.2 below to compute the number of samples.

$$n = \left[ \left( z^* \right)^2 \times \left( \frac{\hat{p}(1-\hat{p})}{MoE^2} \right) \right] \tag{4.2}$$

Here, n is the number of samples (sample size),  $z^*$  is the z score (1.96) from the z-table, MoE is a margin of error and  $\hat{p}$  represents the data proportion. In our study, we had a total

#### 4.4 Performance Evaluation

of 2400 images. Using the power analysis, the total samples required for the study was 384. So, the number of samples used in this study was 84% higher than the required samples.

#### 4.4.2 Ranking of AI models

After getting the absolute values of 12 AI models' performance metrics, we sorted the AI models in increasing order, and then we compare each value with the highest possible value in the attribute. We considered five marks. If the percentage is more significant than 95%, four marks. If it is greater than 90 and less than 95, three marks. If it's more significant than 80% and less than 90%, two marks if it's more significant than 75%, one mark if it is greater than 50%, less than 50% is considered zero. The resultant rank table of the AI models is shown in Table 4.4. We color-coded each AI model from red to green. Each model is color-coded in this band. If the model is performing low, it is represented as red. If it is performing well represented as green.

Rank Model  $\mathbf{o}$ F F1 DS D TT Me **AUC** AS % A Se Sp VGG19 78.18 78.18 MobileNet CNN11[129] 76.36 AlexNet 63.60 Inception 61.82 DenseNet169 61.82 58.18 XceptionNet Modified Unet 54.85 VGG16 54.55 SqueezeNet 50.90 DenseNet 121 47.27 ResNet50 41.80 

Table 4.4: Ranking table of the AI models.

#### 4.4.3 Benchmarking the performance with existing

We computed Area-under-the-curve (AUC) for all the proposed AI models and compared the performance with our previous existing work [129], consisting of the CNN model with an accuracy of 95.66% and AUC: 0.956 AUC. Figure 4.6 represents the ROC comparison

<sup>\*\*</sup>O: Optimization, A: Accuracy, F: False Positive Rate, F1: F1 Score, Se: Sensitivity, Sp: Specificity, DS: Data Size, D: DOR, TT: Training Time, Me: Memory, AUC: Area-Under-the-Curve, and AS: Absolute Score.

of 10 AI methods. Among all the architectures, MobileNet showed the highest AUC value as 0.961 (*p*-value<0.0001) and it shows higher performance than CNN [129].

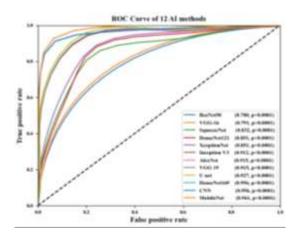


Figure 4.6: ROC comparison of 12 AI models (10 TL and 2 DL).

#### 4.5 Scientific validation vs. clinical validation

In this section, we discussed the validation of the hypothesis. Scientific validation was done by heatmap analysis using TL-based "Grad Cam" technique and clinical validation was proved using correlation analysis of biomarker with AI.

# 4.5.1 Scientific validation using Heatmaps

We applied a novel visualization technique called Gradient weighted Class Activation *Map* ("*Grad Cam*") for identifying the diseased areas in the plaque cut sections using VGG16 transfer learning architecture. Grad-CAM produces heatmaps based on the weights generated during the training. Here we take feature maps of the final layer. It gives the essential regions of the target and heatmaps highlight these regions. Figure 4.7 and Figure 4.8 represents the heatmaps of the nine patients of symptomatic and asymptomatic class. The dark red color region represents the diseased region in symptomatic plaque, whereas it represents the higher calcium area in asymptomatic plaque.

The Grad-Cam works on the training weights generated during the training phase. The DL model captures the important regions of the target label. We compared the heatmaps with original images of both symptomatic and asymptomatic images. We observed that heatmaps exhibit darker regions surrounded by grayscale regions. Whereas in asymptomatic regions, DL observes grayscale regions. Figure 4.8, a1, a2, b1, and c1 are the important regions observed by DL of symptomatic images, and d1, e1, e2, e3, f1, f2, f3 are the observed important regions of the asymptomatic images by the DL model. This

comparison proves our hypothesis that symptomatic plaques are hypoechoic, dark, and asymptomatic plaques are bright and hyperechoic.

#### 4.5.2 Correlation Analysis

We correlated all the biomarkers for the detection of the risk with AI. Table 4.5 represents the correlation coefficient of all the biomarkers. Among all the biomarkers, GSM vs. FD shows a better *p*-value. We computed the correlation coefficient using MedCalc. We computed the Euclidean distance (ED) between the centers of the two clusters (sym and asym). Table 4.6 represents the ED between two clusters symptomatic *vs.* asymptomatic. AI shows constant variation among all the techniques, whereas GSM with FD and HOS shows maximum distance. Figure 4.9 represents the correlation of AI (Modified Unet), GSM, FD, and HOS, and the black dot represents the center of each class. The clusters of symptomatic and asymptomatic are represented with red and violet colors respectively. The black dot represents the center of the cluster and the eclipse on the cluster represents the high-density area. Figure 43 (b), (d), and (e) represents the strong correlation between the biomarkers, (a), (c) represents moderate correlation, and (f) represent the weak correlation.

**Table 4.5:** Correlation analysis.

Comparison	Sym	ptomatic	Asymptomatic		Abs.
	CC	<i>p</i> -Value	CC	<i>p</i> -Value	Difference
FD vs. HOS	0.07221	0.0149	0.156	0.0017	1.160366
FD vs. GSM	-0.241	< 0.0001	-0.383	< 0.0001	0.589212
GSM vs. HOS	0.0725	0.0147	-0.0630	0.0208	1.868966
Modified Unet vs.	0.0017	0.009	-0.0437	0.0031	26.70588
GSM	0.0017	0.009	-0.0437	0.0031	20.70388
Modified Unet vs.	-0.0234	0.006	-0.0394	0.0042	0.683761
HOS	-0.0234	0.000	-0.0334	0.0042	0.083701
Modified Unet vs.	0.0623	0.0021	0.01347	0.0079	0.783788
FD	0.0023	0.0021	0.01547	0.0079	0.783788

Table 4.6: Euclidean distance between biomarker pairs.

Comparison	<b>Euclidean Distance</b>
Modified Unet vs FD	9.82
Modified Unet vs GSM	9.83
Modified Unet vs HOS	8.83
FD vs GSM	24.20
GSM vs HOS	24.19
FD vs HOS	2.18

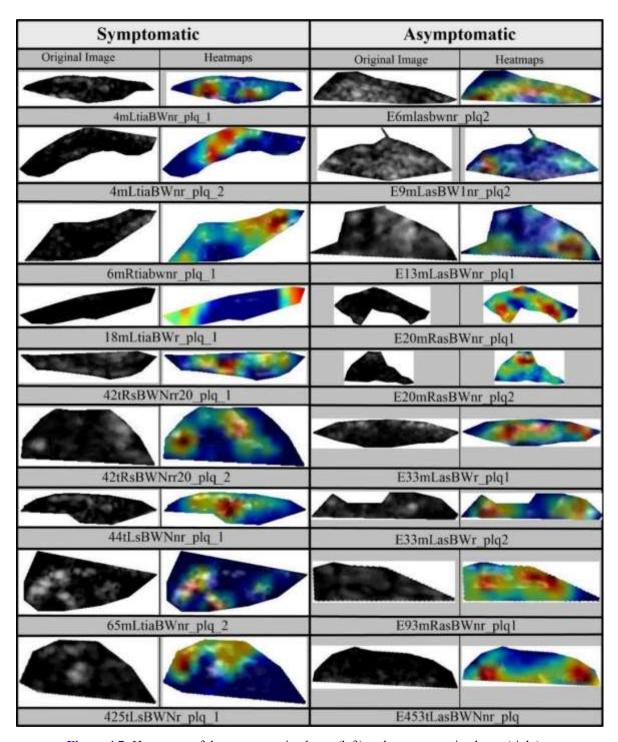
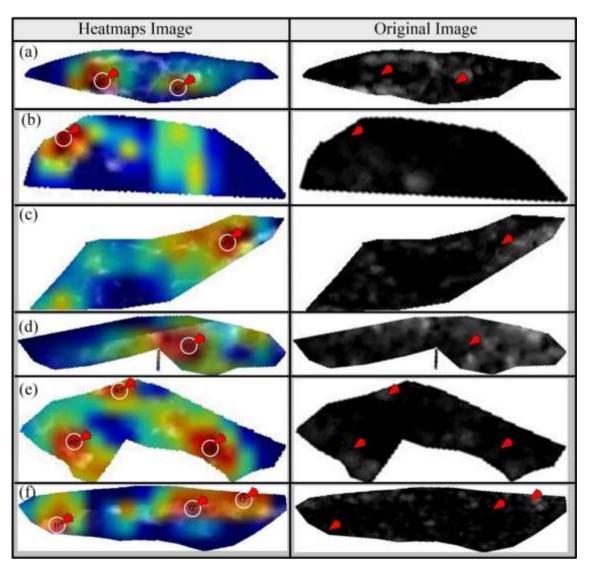


Figure 4.7: Heat maps of the symptomatic plaque (left) and asymptomatic plaque (right).



**Figure 4.8:** (a), (b), and (c) are the symptomatic image heatmaps vs. original images, (d), (e), and (f) are asymptomatic image heatmaps vs. original images (Red color arrow represents the important regions).

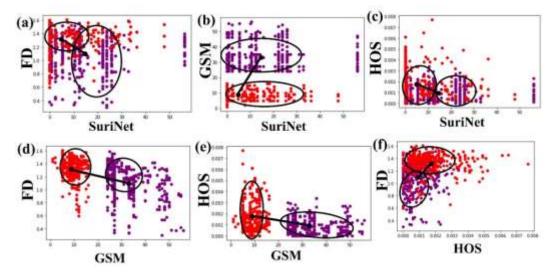


Figure 4.9: Correlation of AI (Modified Unet) and the three biomarkers - FD, GSM, and HOS.

# 4.6 Discussion

TL architectures use the pre-trained weights for retraining the model for target label prediction. But the TL architecture training time depended on the size of the pretrained weights and hardware resources. Various TL models discussed in Table 4.3 had their advantage over the other model, which is explained in Table 4.8 and 4.9.

# 4.6.1 Advantages of TL Models

The TL models have been designed has similarities and differences between them. These are explained in Table 4.7, along with the key findings of every TL model.

Table 4.7: Benchmarking Table.

	C1	C2	C3	<b>C4</b>	C5	C6
SN#	Authors, Year	Features Selected	Classifier Type	Dataset	AI Type	ACC. (%) AUC (p- value)
R1	Christodoulou <i>et al.</i> , (2003) [71]	Texture features	SOM KNN	230 (-)	ML	73.18, 68.88, 0.753, 0.738
R2	Mougiakakou <i>et al.</i> , (2006) [116]	FOS* and Texture Features	NN with BP and GA	108 (UK)	ML	99.18, 94.48, 0.918
R3	Seabra <i>et al.</i> , 2010 [35]	Five Features	Adaboost using LOPO*	146 Patients	ML	99.2
R4	Christodoulou <i>et al.</i> , 2010 [33]	Shape Features, Morphology Features, Histogram Features, Correlogram features	SOM KNN	274 Patients	ML	72.6, 73.0
R5	Acharya <i>et al.</i> , (2011) [115]	Texture Features	SVM with RBF Adaboost	346 (Cyprus)	ML	82.48, 81.78, 0.818, 0.810 p<0.0001
R6	Kyriacou <i>et al.</i> , 2012 [114]	Texture features with second-order statistics spatial gray level dependence matrices	probabilistic neural networks and SVM	1121 Patients	ML	77, 76
R7	Acharya <i>et al.</i> , (2012) [26]	Texture Features	SVM	346 (Cyprus)	ML	83.8 p<0.0001
R8	Acharya <i>et al.</i> , (2012) [67]	DWT* Features	SVM	346 (Cyprus)	ML	83.78 p<0.0001
R9	Gastounioti <i>et al.</i> , (2014) [117]	FDR+ Features	SVM	56 US Image	ML	88.08, 0.90

R10	Molinari <i>et al.</i> , 2018 [37]	Bidimensional empirical mode decomposition and entropy features	SVM With RBF	1173 Patients	ML	91.43 p<0.0001
R11	Skandha <i>et al.</i> , 2020 [129]	Automatic Features	Optimized CNN	2000 Images (346 Patients)	DL	95.66 p<0.0001
R12	Skandha <i>et al.</i> , 2020 [145]	Automatic Features	CNN with 13 Layers	2311 Images (346 Patients)	DL	89 <i>p</i> <0.0001
R13	Proposed	Automatic Features	10 TL Architectures VGG16 VGG19 DenseNet169 DenseNet121 XceptionNet MobileNet InceptionV3 AlexNet SqueezeNet ResNet50	346 Patients (Augmented from balanced to 6x)	DL	96.18 0.961 p<0.0001
R14	Proposed	Automatic Features	Modified Unet	346 Patients (Augmented from balanced to 6x)	DL	92.7 0.927 <i>p</i> <0.0001

Table 4.8: Comparison of TL models.

SN#	Author, Year	Name of the Network	Dataset	Purpose	Pretrained weight size (MB)	Type of Layers
1	Krizhevsky <i>et al.</i> , 2012 [55]	AlexNet	ImageNet	Classification	244	Convolution, Max Pooling, FCN
2	Simonyan <i>et al.</i> , 2015 [47]	VGG -16, 19	ImageNet	Object Recognition	528, 549	Convolution, Max Pooling, FCN
3	Szegedy <i>et al.</i> , 2015 [50]	InceptionV3	ImageNet	Object Recognition	92	Convolution, Max Pooling, Inception,FCN
4	He et al., 2016 [51]	ResNet 50, 101 , and 152	ImageNet, CIFAR	Fast optimization for extremely deep neural networks	98,171, 232	Convolution, Ave. Pooling, Residual, FCN
5	Howard <i>et</i> al.,	MobileNet	ImageNet	Classification &	16	Convolution, Depthwise

	2017 [53]			Segmentation		Convolution,
				in mobiles.		AveragePooling,
						FCN
				Modified		
				depthwise		Convolution,
	Chollet et		ImageNet,	separable		SeparableConvolution,
6	al.,	XceptionNet	JFT	convolution.	88	MaxPooling,
	2017 [54]		JI I	Advancement		Global AvgPooling,
				of		FCN
				InceptionV3		
				Gradient		Convolution,
	II	DenseNet		problem		*
7	Huang et	121, 169,	CIEAD	substantially	22 57 90	MaxPooling,
/	al., 2018	201, and	CIFAR	reduces the	33, 57, 80	Transition,
	[52]	264		number of		Dense, FCN,
				parameters		Global AvgPooling
				Reducing the		Convolution
	T am dala ak			number of		Convolution,
8	Landola et	ConsonaNid	I a.a.N.:4	parameters,	4.0	FireModule,
8	al., 2017	SqueezeNet	ImageNet	efficient	4.8	MaxPooling,
	[56]			working on		FCN, Global Avg
				edge devices		Pooling

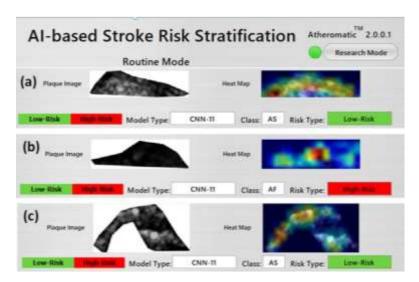
**Table 4.9:** Similarities and differences between the TL models.

Architecture	<b>Key Findings</b>	Similarities	Differences
AlexNet	First deep neural network using convolution It is developed to reduce the number of parameters	<ul> <li>All the models are pre-trained on ImageNet</li> <li>All the Models use convolution</li> </ul>	<ul> <li>MobileNet is focused on solving the computer vision problems in edge devices</li> <li>Densenet is trained</li> </ul>
SqueezeNet	required for AlexNet with the same accuracy. Effectively used for edge devices	<ul> <li>Every model uses a softmax activation function in the output layer and a</li> <li>Rel u activation</li> </ul>	and tested on the CIFAR dataset where the remaining models use ImageNet  XceptionNet only
VGG	Reducing the number of parameters in convolution and training time	• Every model loads the pretrained	uses the JFT dataset for pre-training.  • Except for Xception and MobileNet, all the other model uses
InceptionV3	Effective object detection for solving variable	weights from the cloud/offline.	standardized convolution

	size objects using	•	Every model uses a	•	Except for IV3, and
	kernels of		network-based TL		Xception, all other
	different sizes in		paradigm.		model uses depth-
	each layer.		1 0		wise kernels.
	Solving the				
	vanishing				
	gradient problem				
ResNet	in the deep neural				
	network using				
	skip (shortcut)				
	connections				
	The first model				
	was developed				
	for supporting				
MobileNet	Tensor flow in				
Modifienet	edge devices				
	using light-				
	weighted Tensor				
	Flow.				
	Fast optimization				
	and reducing the				
XceptionNet	trainable				
Aceptioninet	parameters in				
	IV3 using depth-				
	wise convolution				
	Increasing the				
	feed-forward				
	nature in the				
	neural networks				
DenseNet	using dense				
	layers by				
	concatenating the				
	features from its				
	previous layers				

#### 4.6.2 GUI Design

AtheroPoint<sup>TM</sup> had developed the Atheromatic<sup>TM</sup> 2.0<sub>TL</sub> system a computer-aided diagnostic system for stroke risk stratification. Figure 4.10 represents the screenshot of the CAD system. This CAD system will give the plaque risk and heatmaps generated by the Grad-Cam with the help of TL/DL models. In the CAD system, the heatmap would be predicted on the test image once the training model is selected.



**Figure 4.10:** GUI screenshot of the Atheromatic<sup>™</sup> 2.0<sub>TL</sub> system.

#### 4.6.3 Strengths/Weakness/Extensions

We evaluated the optimization point of the TL models against various augmentation folds and compared the performance of TL against the DL models such as Modified Unet and CNN. TL model showed an improvement in symptomatic vs. asymptomatic plaque classification accuracy. Furthermore, our Atheromatic<sup>TM</sup>  $2.0_{TL}$  system predicts the risk of plaque and vulnerability using the color heatmaps on test scans.

Even though the power sample suggests that we have enough samples for the training, the main limitation of this study was the moderate cohort size. In addition to the cohort size, another limitation of this study is the limited availability of the hardware resources such as supercomputers availability, especially, in third world developing countries.

Our study had a manual delineation of ICA data sets. In the future, there could be a need to design an automated ICA segmentation system [127]. Another option is to enhance the CNN by using a better DCNN model with a modified rectified linear unit (ReLU) activation function that ensures "differentiable at zero" [48]. There are dense networks such as DenseNet121, DenseNet169, and DenseNet201, which could be tried and compared [87]. Further, one can combine hybrid deep learning models for PTC [146]. Finally, the proposed AI models can be extended to a big data framework by including other risk factors.

#### 4.7 Conclusion

The proposed study is the first of its kind to characterize and classify the carotid plaque

#### 4.7 Conclusion

using an optimized transfer learning approach and Modified Unet (a class of Atheromatic<sup>TM</sup> 2.0<sub>TL</sub>). Eleven AI models were implemented, and the best AUC was **0.961** (*p*<**0.0001**) from MobileNet and **0.927** (*p*<**0.0001**) from Modified Unet. We validated the performance using grayscale median, fractal dimension, higher-order spectra, and spatial heatmaps. TL showed equal and comparable performance with deep learning. Atheromatic<sup>TM</sup> 2.0<sub>TL</sub> model showed a performance improvement of **12.9**% over Atheromatic<sup>TM</sup> 1.0<sub>ML</sub> (AtheroPoint, CA, USA) compared to the previous machine learning-based paradigm. The system was validated with the widely accepted dataset.

#### **CHAPTER 5**

### CLASSIFICATION AND CHARACTERIZATION OF THE STROKE USING HYBRID DEEP LEARNING MODELS

#### 5.1 Introduction

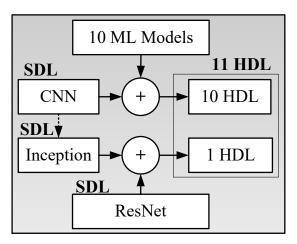
Hybrid Deep Learning (HDL) was recently introduced as a fusion technique that offers the benefit of ML, DL, and TL. Recent studies by Jena *et al.* [146], Abuhmed *et al.* [147], Dang *et al.* [148], and Suri *et al.* [149] have demonstrated the usage of hybrid techniques, such as DL with ML, or TL with DL, and TL with ML by showing better performance to fulfills the weakness of TL or solo DL (SDL) models [149-151]. We, therefore, hypothesize that the performance of HDL models is superior to TL or SDL.

In the proposed study for the design of Atheromatic<sup>TM</sup> 2.0<sub>HDL</sub>, we use two kinds of HDL innovations; the *first* innovation is the fusion of SDL which consists of the fusion of Inception Network [43] with Residual Network [152]. In the *second* set of innovations, we fuse SDL (say convolution neural network (CNN)) with ML in which ML has a fixed place and acts like a cascaded model to SDL. The second innovation offers the advantage and flexibility of changing or replacing the type of ML model for exhaustive training and choosing the best SDL-ML combination. Thus, we took the liberty of applying 10 types of ML models in a sequence with the DL model during tandem connection. The *third* innovation is to examine the three kinds of loss functions such as cross-entropy loss (CEL), hinge loss (HL), or mean-squared-error loss (MSEL) and evaluate the best DL-ML combination. These overall paradigms of HDL were tried on multicenter data sets to understand its behavior with the change in grayscale plaque characteristics, which can be part of the *fourth* innovation in our comprehensive analysis.

In summary, this is the first pilot study of its kind that uses 11 HDL models, five TL models, and 1 DL model for in-depth analysis. We further benchmark the HDL

framework with *five* TL models such as VGG16, ResNet50, DenseNet121, MobileNet, and AlexNet, and *one* SDL model, leading to a total of 17 AI models.

Based on the hypothesis that HDL should do better than SDL or TL, this study originated with a set of two key innovations: (i) <u>one</u> HDL is designed using a fusion of SDL-based Inception Network [43], SDL-based Residual Network [152], and (ii) a cascade of <u>ten</u> ML models with one SDL thread (i.e., CNN). Figure 5.1 represents the proposed eleven (10+1) HDL models. Since the weights of the AI models can be transferred for an efficient DL design, we propose <u>five</u> types of DL models under the class of TL. Lastly, we compare the set of 16 HDL against an independent SDL (without any cascaded ML, i.e., CNN alone) which totals 17 AI models. Thus, one needs to design the architectures to understand their internal structure to facilitate the PTC. Note that these AI models are implemented under three categories of loss function such as CEL, HL, and MESL. Due to the limited sample size for the PTC design, we use augmentation protocol using AI's flexibility.



**Figure 5.1:** Representation of 11 HDL models (i) fusion of Inception and ResNet (1 HDL model) and (ii) fusion of 10 ML models with CNN, leading to 10 HDL models.

#### 5.2 Dataset

Two sets of demographics were collected from two geographical areas of the world. The first was collected from St. Mary's Hospital, Imperial College, London, UK, symbolized as LonD. The patients with contralateral cerebral hemispheric/retinal symptoms, as well as those who had a transient ischemic attack (TIA), were included only if they had been asymptomatic for more than six months. Note that this data was taken from our previous study (Nicolaides *et al.* [125]), which consisted of 196 *symptomatic* and 150 *asymptomatic* cases. The study had a mean age of 69.9±7.8 years of which, 61% were male who had ICA

stenosis of 50% to 99%. Note that these patients did not have any previous symptoms or neurological abnormalities. The second data was collected from Instituto cardiovascular de Lisboa, Lisbon, Portugal, symbolized as LisD. This was approved by the ethical committee and the data taken from our previous study (Sanches *et al.* [126]). It consists of 50 *symptomatic* and 110 *asymptomatic* scans with a mean age of patients 67.5±0.77 years.

#### 5.2.1 Ultrasound Image Acquisition from Multicentre Study

The ground truth was obtained from the Pathology of the plaque and was considered for the classification of the carotid plaques in symptomatic vs. asymptomatic plaques and ground truth labeling [91, 103, 126]. For LonD, all the scans were acquired using an ATL HDI-3000 duplex scanner, Advanced Technology Laboratories, Seattle, WA, the USA with a linear broadband width 4-7 MHz transducer. The mean resolution was 20 pixels per millimeter. Region-of-Interest (ROI) of the scans was delineated and normalized using a medical practitioner with a mouse and "Plaque Texture Analysis Software". For LisD, a Philips HDI 5000 machine with a 5-12 MHz broadband linear array transducer was used to obtain these scans. The mean resolution was 25 pixels per millimeter. ROI was delineated and normalized using MATLAB by the medical practitioner. We combined LonD and LisD to form a new cohort called a mixed cohort labeled MixD. It contained 246 (196+50) symptomatic, 260 (150+110) asymptomatic scans. Figure 5.2 shows the representative images of these cohorts.

#### **5.2.2** Augmentation Protocol

Both cohorts are moderate in size. Thus, we have used augmentation techniques for increasing the cohorts into 2x, 3x, 4x, 5x, and 6x folds as used in our previous work [48, 127-129]. Initially, we took the unbalanced cohort and then balanced the classes labeled as balanced fold (1x) and then augmented up to six folds (6x). We used "Augmentor" for these purposes consisting of geometrical operations such as random flipping and rotation of 180 degrees.

#### 5.3 Methodology

Since there are 17 AI models, we divided the architecture layout into four parts: Part A presents the HDL model that uses the combination of Inception Network and Residual Network models. Part B presents the HDL models that use the fusion of CNN with 10 ML models in sequence. Part C presents the generalized TL architecture design, while part D is a solo DL model.

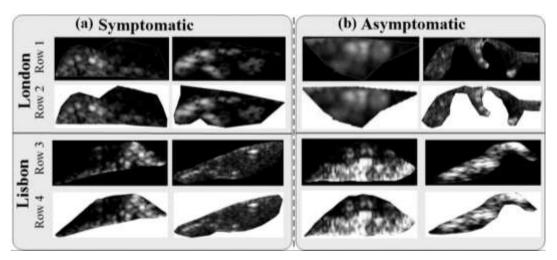


Figure 5.2: Sample (a) symptomatic (b) asymptomatic the US cut sections of the carotid plaque.

#### 5.3.1 Part A: One Hybrid Deep Learning Architecture using two SDL

The first HDL model uses two SDL that implements a fusion of Inception Network [43] with Residual Network [152], symbolized as HDL-IRN (referred to as HDL1, as shown in Figure 5.3).

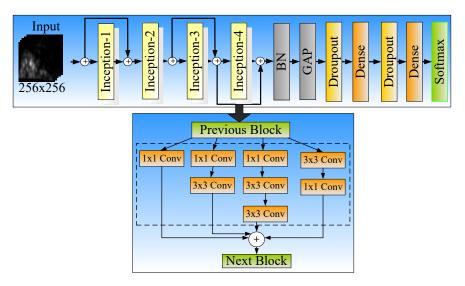


Figure 5.3: HDL-IRN model using two SDL (Inception+ResNet).

The proposed model contains five inception blocks (IB) in which each IB contains four parallel layers. The first column of the layer contains a 1x1 convolution layer (CL), the second column of layers contains one 1x1 CL and 3x3 CL, and the third column of layers contains one 1x1 CL and two 3x3 CL, and the fourth column contains 3x3 CL and 1x1 CL. The output of four-column layers is fused using the residual connection (RC) represented as a circle with the "+" symbol in Figure 5.3. These IBs are connected with the output of the previous block using RC, as represented in Figure 5.3. At the end of the last IB, we add a batch normalization (BN) layer to normalize the output from the residual connection. The

output of BN is passed through the global average pooling (GAP) instead of FCN to have a better correlation between feature maps and classes [72]. Following the GAP layer, a pair of dropout layers with a 50% dropout rate and a dense layer is added for complex feature learning. We used K10 cross-validation (training: testing: 9:1) for accuracy estimation and optimization. We use 100 epochs with stochastic gradient descent (SGD) optimizer [153], with a learning rate of 0.01.

#### 5.3.2 Part B: Ten Hybrid Deep Learning Architectures

In the second proposed HDL model, we bank on the simplicity of the ML model to design the HDL framework by cascading the ML to SDL in the end. Such cascading in the AI framework has been attempted before, either to improve the AI performance [154] or to calibrate the rough output of stage one into a smooth output [85]. Using the spirit of our previous design where we cascade ML with DL models for cardiovascular application, the fusion offers multiple advantages other than simplicity, this includes (i) optimal feature extraction using the CNN model in the SDL framework and (ii) handling the complex baton of SDL into a simpler ML framework for an efficient HDL design output. Thus, the nomenclature of HDL design is based on the selected ML configuration. Because there are 10 types of ML models, the system, as a result, has 10 types of HDL models labeled as HDL2 to HDL11. This is because HDL1 nomenclature was used by the HDL-IRN model. Similarly, Figure 5.4 shows the 10 HDL architectures where ML is represented by the 10 spokes of the wheel. Initially, this model extracts the features from target labels using SDL which consists of CNN with four convolution layers (CL) and four average pooling layers (APL).

Finally, the 2D feature maps are converted into a 1D vector. These vectors are converted to point data i.e., comma-separated value (CSV) files. The feature vector point data is then split into training and tested using the K10 CV protocol. The training CSV file is sent to be used for ML-based training while the test CSV file is used for predicting the class labels. Since the CSV format was used, this makes the system fast in the HDL framework.

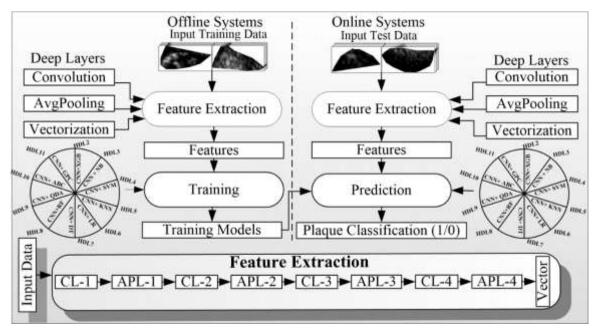


Figure 5.4: Cascading of ML models with SDL for the HDL2 to HDL11 model design architecture, a class of Atheromatic<sup>™</sup> 2.0HDL. The feature extraction process is discussed below showing four CL and four APL layers, leading to vector design in the end.

#### 5.3.3 Part C: Five Transfer Learning Models - A class of DL

In this study, we benchmark the performance of the HDL models against the popular *five* TL models such as AlexNet [55], ResNet50 [51], DenseNet121 [52], MobileNet [53], and VGG19 [47]. In all the five TL models, we froze the initial layers of the models (middle layer, left) with the pre-trained weights from "ImageNet" (middle layer, right). At the end of the model, we added a pair of FCN with a dropout layer (50%) for improving the complex learning ability of the model. Target labels (ground truth) were fed at the end of the model for retraining the models for prediction. Figure 5.5 represents the TL architecture.

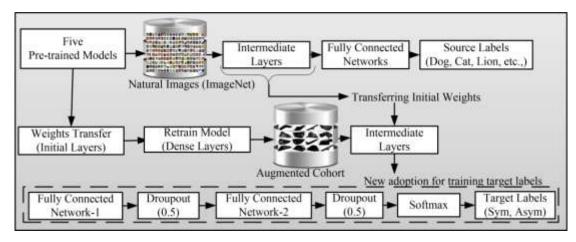


Figure 5.5: Generalized transfer learning architecture.

#### 5.3.4 Part D: One Solo Deep Learning Model

Figure 5.6 shows the SDL model in which the design consisted of an 11-layered CNN model. This consisted of four layers of convolution, four layers of average pooling, and two dense layers along with one dropout layer. Note that the fatten layer is not part of the 11 layered models, however, it is used for converting the 2D feature maps into the 1D vector. The function of the softmax layer is to predict the class labels [129, 155].

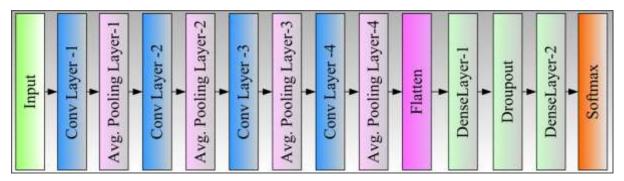


Figure 5.6: Solo deep learning architecture.

#### 5.3.5 Loss Functions

We used three kinds of loss functions for the HDL models, as shown in equations 5.1 to 5.3.

Cross-Entropy Loss = - 
$$[(y_i \times \log a_i) + (1-y_i) \times \log(1-a_i)]$$
 (5.1)

where,  $y_i$  is the class label for input and  $a_i$  is the predicted probability of the class being  $y_i$ .

Hinge Loss = 
$$\max(0, 1-y^i(x^i-b))$$
 (5.2)

where  $y^i$  and  $x^i$  refer to the  $i^{th}$  instance in the training set and b refer to the bias term.

Mean Squared Error Loss=
$$\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)$$
 (5.3)

where N is the total number of data samples,  $y_i$ : observed class  $\hat{y}_i$ : is predicted, class.

#### 5.4 Experimental Protocol

#### 5.4.1 Accuracy Bar Charts for each cohort corresponding to all HDLs

If  $\eta(h,c, CEL,k)$  represents the accuracy of an HDL type "h" for a cohort "c" using cross-validation combination "k" (out of total combinations K), while considering the CEL loss function, the accuracy of each combination was computed using

#### **5.4 Experimental Protocol**

 $\eta$  (%)=[ $\frac{(TP+TN)}{(TP+FP+FN+TN)}$ ]×100, where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative. These values are computed by comparing the predicted label against the target label. then the mean accuracy for all the combinations, represented by  $\bar{\eta}(h, c, CEL)$  can be mathematically given by Equation 5.4. Note that we considered the K10 protocol in our paradigm so K=K10=10.

$$\bar{\eta}(h,c,CEL) = \frac{1}{K} \sum_{k=1}^{K} \eta(h,c,CEL,k)$$
(5.4)

Using the same strategy, and considering the remaining two types of cohorts, LonD and MixD, we obtain equations 5.5 and 5.7, respectively.

$$\bar{\eta}(h,\text{LonD,CEL}) = \frac{1}{K} \sum_{k=1}^{1K} \eta(h,\text{LonD,CEL},k)$$
 (5.5)

$$\bar{\eta}(h, \text{LisD,CEL}) = \frac{1}{K} \sum_{k=1}^{K} \eta(h, \text{LisD,CEL}, k)$$
 (5.6)

$$\bar{\eta}(h,MixD,CEL) = \frac{1}{\kappa} \sum_{k=1}^{K} \eta(h,MixD,CEL,k)$$
 (5.7)

#### 5.4.2 Mean Feature Strength for each cohort using a CEL loss function

If MFS (LonD, CEL, h) is the feature strength for an HDL combination h using loss function CEL while considering the cohort LonD, and if  $H_d$  represents the total number of HDL combinations starting from 2 to 11, i.e., 10., then the mean values of the feature strength  $\overline{\text{MFS}}(\text{LonD}, \text{CEL}, h)$  can be mathematically computed as per Equation 5.4. Applying a similar strategy to the other cohorts such as LisD and MixD, we can compute the feature strength  $\overline{\text{MFS}}(\text{LisD}, \text{CEL}, h)$  and  $\overline{\text{MFS}}(\text{MixD}, \text{CEL}, h)$ , respectively. This is shown in Equations 5.8 and 5.10.

$$\overline{MFS}(LonD,CEL,h) = \frac{1}{H_d} \sum_{h=2}^{H} MFS(LonD,CEL,h)$$
 (5.8)

$$\overline{MFS}(LisD,CEL,h) = \frac{1}{H_d} \sum_{h=2}^{H} MFS(LisD,CEL,h)$$
 (5.9)

$$\overline{MFS}(MixD,CEL,h) = \frac{1}{H_d} \sum_{h=2}^{H} MFS(MixD,CEL,h)$$
 (5.10)

Note that since the CEL loss function was the best among the three kinds of the loss function, we therefore only considered CEL in our analysis. However, it is straightforward to compute the mean feature strengths for other loss functions.

#### 5.4.3 Cross-Validation Protocol

In this study, we have 17 AI models, three cohorts, six augmentation folds, and three loss functions. We used the K10 cross-validation protocol for the computation of accuracy.

Each fold contains 90% of training samples from the augmentation fold, and 10% of testing samples, such that the test samples are unique in each fold. The mean of the accuracy is computed by considering all the folds. The above process is repeated by changing the loss function and running for all the 11 HDL models. A similar procedure was adopted for all the augmentation folds for each of the cohorts and corresponding accuracies are recorded.

#### 5.4.4 Computational efficiency using supercomputer framework

The best way to compare these 17 (11 HDL, 5 TL, and 1 DL) AI models is to evaluate the performance of the models at their optimal point. There are two categories in which the 17 models can fall. In the first category, we take DL and HDL models (12 models) for complexity computations, In the second category, we take the remaining TL models. The fundamental difference between these two categories is the number of loss functions considered. In DL/HDL, we have three kinds of loss function (CEL, HL, MSEL), while in TL, there is only one type (CEL) of the loss function. This is because TL is pretrained and does not require changing the loss function.

Keeping this assumption in mind, the total # of computations is based on the number of attributes. This includes the number of classifiers, type of folds, types of loss functions, number of epochs, number of augmentations, and the total number of cohorts. For DL/HDL combination, these attributes take the values 12, 10, 3, 100, 6, and 3. Multiplying these values of attributes leads to 648,000 ~ 700K computations. On the contrary, the values for the TL paradigm are 5, 10, 1, 100, 6, and 3, leading to 90,000~90K computations. Thus, the number of computations in HDL/DL models is nearly seven times the TL computations. As a result, there is a need for a GPU-based supercomputer framework.

#### 5.5 Results

#### 5.5.1 Results of HDL Optimization on Three Cohorts

Since the number of computations is large, it is mandatory to run such kinds of systems in a GPU or supercomputer settings. In the past [156], we have used this kind of framework for DL-based systems, Thus, we follow the same hardware configurations (NVidia DGX V-100, with a series of 7 GPUs each having 16 GB, Bennett University, Gr. Noida, India). To achieve the optimization in all three cohorts, we sequentially run these cohorts corresponding to all 11 HDL models. Using equations 5.5, 5.6, and 5.7, the accuracy plots can be seen in Figure 5.7 (a, b, c) using the three kinds of loss functions (CEL, HL, MSE). CEL and MSE showed HDL7 being the best model with an accuracy (in %) of 99.60±1.15,

99.40±1.25, and 99.50±1.37, corresponding to LonD, LisD, and MixD with CEL, and 99.10+2.56, 99.20+6.58, 98.00+3.65 corresponding to LonD, LisD, and MixD with MSE. On the contrary, for the HL, the optimized values were 99.50±2.58, 99.66±0.75, and 99.20±9.87, respectively, corresponding to LonD, LisD, and MixD. Note that each HDL accuracy (bar) in Figure 5.8 (a, b, c) are the optimized values corresponding to the best augmentation data.

To compare the <u>three</u> states of unbalanced *vs.* balanced *vs.* optimized scenarios, we, therefore, choose HDL7 in the CEL framework to show the <u>six</u> types of performance metrics by using the <u>three</u> types of cohorts shown in Table 5.1. Note that these results are for the augmentation x5 which has the following sample distribution (LonD: 2000, LisD: 1500, MixD: 2600). Note that the results on HDL are superior to DL systems [LonD: 99.60±1.15 *vs.* 88.00±1.02, LisD: 99.40±1.25 *vs.* 86.85±8.78, MixD: 99.50±1.37 *vs.* 85.87±7.48], which further validates the hypothesis.

## 5.5.2 Results of 17 AI models using three cohorts for best loss function and best augmentation

For relative performance for all the 17 AI models, we selected the best loss function and best augmentation data for each of the AI models and then arranged these models in increasing order of accuracy for all three cohorts (see Figure 5.8). Since DT has a strong ability to classify, it when combined with best features via CNN, leads to a stronger combination of CNN+DT, so-called HDL7. Thus, HDL7 (CNN+DT) turned out to be the best model. The corresponding values can be seen in Table 5.2. Among all the models, HDL7 and HDL2 show better performance compared to TL and DL, which further validates our hypothesis. HDL models exhibit superior performance compared to TL by 8.78% on LonD, 3.61% on LisD, and 3.05% on MixD cohorts.

Similarly, HDL models show superior performance compared to DL with 13.18% in LonD, 14.79% in LisD, and 15.87% in MixD cohorts, respectively. Due to the fusion of CNN features with ML models, they begin to exhibit higher performance. Jena *et al.* [146], also proposed that the fusion of SDL with ML gives the best performance from the ML classifier.

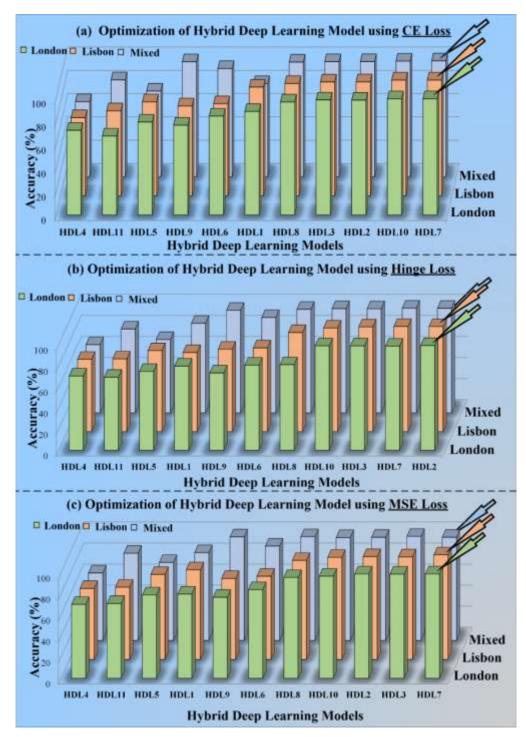
**Table 5.1:** Accuracies of three cohorts under unbalanced vs. balanced vs. optimized conditions for HDL7 (SDL+DT) using CEL and K10 protocol.

SN#	Cohorts	Unbalanced	Balanced	Optimized	
Accuracy Analysis (%)					
1	LonD	92.80	94.50	99.60	
2	LisD	78.50	82.50	99.40	
3	MixD	92.54	98.30	99.50	
	\$	Sensitivity Anal	ysis (%)		
4	LonD	0.956	0.961	0.999	
5	LisD	0.833	0.880	0.999	
6	MixD	0.954	0.983	0.999	
	Specificity Analysis (%)				
7	LonD	0.891	0.923	0.999	
8	LisD	0.743	0.772	0.999	
9	MixD	0.892	0.985	0.999	
	F1 Score Analysis (%)				
10	LonD	0.936	0.952	0.996	
11	LisD	0.770	0.846	0.995	
12	MixD	0.933	0.983	0.994	
AUC Analysis ( <i>p</i> -value<0.0001) [0-1]					
13	LonD	0.930	0.945	0.996	
14	LisD	0.790	0.825	0.994	
15	MixD	0.925	0.983	0.995	

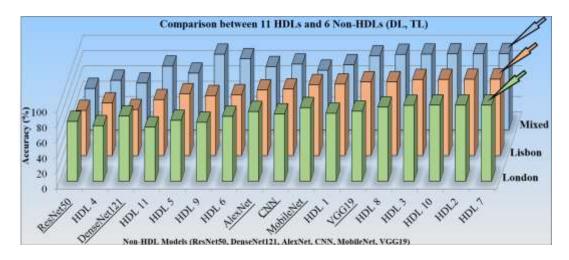
Table 5.2: Accuracy comparison between 11 HDLs and 6 Non-HDLs (1 DL + 5 TL).

AI	London	Lisbon	Mixed
Models	(%)	(%)	(%)
ResNet50*	78.20 <u>+</u> 5.65	58.88 <u>+</u> 2.32	54.03 <u>+</u> 5.87
HDL4	72.60 <u>+</u> 2.48	68.66 <u>+</u> 1.23	64.68 <u>+</u> 1.41
DenseNet121*	85.17 <u>+</u> 9.87	59.99 <u>+</u> 5.45	61.11 <u>+</u> 5.87
HDL11	71.00 <u>+</u> 6.41	72.99 <u>+</u> 2.01	83.19 <u>+</u> 2.54
HDL5	79.60 <u>+</u> 1.42	80.66 <u>+</u> 3.02	74.04 <u>+</u> 1.23
HDL9	77.00 <u>+</u> 3.23	78.22 <u>+</u> 3.13	98.71 <u>+</u> 3.20
HDL6	84.82 <u>+</u> 2.33	79.47 <u>+</u> 4.14	92.87 <u>+</u> 2.24
AlexNet*	91.09 <u>+</u> 6.58	85.66 <u>+</u> 5.14	82.58 <u>+</u> 1.24
CNN*	88.00 <u>+</u> 1.02	86.85 <u>+</u> 8.78	85.87 <u>+</u> 7.48
MobileNet*	96.19 <u>+</u> 8.42	92.60 <u>+</u> 3.23	77.81 <u>+</u> 6.41
HDL1	88.75 <u>+</u> 3.65	93.33 <u>+</u> 1.23	85.00 <u>+</u> 3.22
VGG19	91.56 <u>+</u> 4.21	96.30 <u>+</u> 2.98	96.55 <u>+</u> 1.14
HDL8	96.90 <u>+</u> 2.12	96.43 <u>+</u> 8.89	98.54 <u>+</u> 2.03
HDL3	99.15 <u>+</u> 2.03	99.33 <u>+</u> 6.58	98.83 <u>+</u> 3.56
HDL10	99.60 <u>+</u> 1.56	99.40 <u>+</u> 4.14	99.30 <u>+</u> 1.02
HDL2	99.50 <u>+</u> 3.23	99.66 <u>+</u> 0.23	99.20 <u>+</u> 1.22
HDL7	99.60 <u>+</u> 7.58	99.78 <u>+</u> 1.02	99.50 <u>+</u> 1.47

<sup>\*</sup>represents the non-HDL models (1 DL and 5 TL).



**Figure 5.7:** Optimization of 11 HDL models using three kinds of loss function on three kinds of cohorts (LonD, LisD, and MixD). (a) CE, (b) HL, (c) MSE. HDL7 is best for CEL and MSE, while HDL2 is best for HL.



**Figure 5.8:** Comparison between 17 AI models over multicenter cohorts using the best loss function (CEL vs. HL vs. MSEL) and best augmentation data for each of the HDL 17 categories.

The objective of the performance evaluation is to evaluate how well the 17 AI models performed on three different kinds of cohorts (LonD, LisD, and MixD) under different *loss* conditions (CEL, HL, and MSEL) and different *augmentation* states (unbalanced, balanced, and optimized). This evaluation can be accomplished using the receiving operating characteristics (ROC). The second aspect of PE is to understand how the HDL can segregate the symptomatic plaque *vs.* asymptomatic plaque and quantify this segregation using the *atheromatic plaque separation index* as a ratio. The third aspect of this study is to perform the power analysis by finding the sample size *vs.* the minimum required for the study. Lastly, in the fourth part of the PE, we show the usage of different statistical tests such as (i) T-test, (ii) paired T-test, (iii) diagnostics odd's ratio (DOR), and (iv) Cohen, and Kappa statistics for validation of hypothesis.

#### 5.6 Performance Evaluation

#### 5.6.1 ROC Analysis

We studied the AUC-ROC analysis using the CEL function on all 17 AI models. Figure 5.10 shows the AUC-ROC plots with significant values in increasing order. Our observation showed that HDL7 has higher performance compared to the rest of the models. Moreover, HDL shows superior performance to TL by 3.41% in the LonD cohort, 3.11% in LisD, 3.21% in MixD, and against DL by 11.64% in LonD, 12.57% in LisD, 13.76% in MixD, respectively. This proved our hypothesis that "the performance of HDL is superior to TL and DL".

#### 5.6.2 Power Study

It is important to have enough samples for the protocol, we thus used the standardized power study as shown in equation 5.11. In this equation,  $Z^*$  represents the standard z-table, and  $\hat{p}$  represents the data proportion, taken as 50% between the two classes. MoE represents the margin of error, which was assumed to be 5% with a 95% confidence interval. Using the above variables, our calculation showed that the number of samples required for the study was 384. Considering the augmentation 2-fold for the cohorts LonD, LisD, and MixD, the samples computed were 800, 600, and 1020, respectively. This is much higher than the basic need of 384, ensuring an increase in the sample size by 52%, 36%, and 62%, respectively.

$$N = \left[ \left( z^* \right)^2 * \frac{\hat{p}(1-\hat{p})}{M_0 E^2} \right]$$
 (5.11)

#### 5.6.3 Atheromatic Plaque Separation Index

It tells how best the classifier distinguishes between the two classes. It is computed based on the mean feature strength (MFS) of the classifiers and is also considered a separate index of the classifier. Since the process of classification is considered using an Atheromatic system, we further call it as Atheromatic Plaque Separation Index (APSI). It is mathematically computed as the absolute difference between the MFS of the symptomatic class and the asymptomatic class divided by the MFS of the symptomatic class. Table 5.3 shows the APSI of all the AI models except TL models since they are using ImageNet features initially. Our observations showed that MFS<sub>Sym</sub>>MFS<sub>Asym</sub>. It is also validated in our previous studies as well [156]. See Figure 5.9 for MFS.

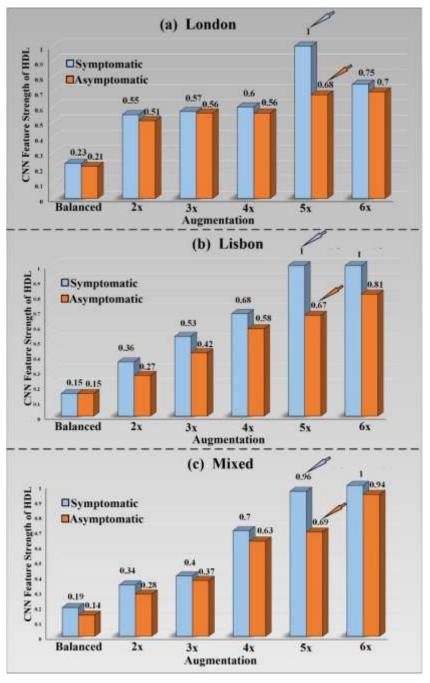
 Table 5.3: Atheromatic Plaque Separation Index for 12 HDL models.

Center	AI Model	MFS <sub>Sym</sub>	<b>MFS</b> <sub>ASym</sub>	APSI
	HDL1 (HDL-IRN)	0.86	0.57	50.87
LonD	HDL2 to HDL11	1.00	0.68	47.05
	SDL	1.00	0.78	28.20
LisD	HDL1 (HDL-IRN)	0.78	0.52	50.00
	HDL2 to HDL11	1.00	0.67	49.25
	SDL	1.00	0.79	26.58
MixD	HDL1 (HDL-IRN)	0.92	0.56	64.28
	HDL2 to HDL11	0.96	0.69	39.13
	SDL	0.82	0.79	3.70

#### 5.6.4 Statistical Test

We performed a statistical test on the proposed classifier to validate our hypothesis. We

performed a T-test, paired T-test, DOR, Cohen, and Kappa statistics for the validation of the hypothesis. It was observed that the order of the DOR performance shows HDL>TL>DL for LonD, HDL>DL>TL for LisD, and HDL>TL>DL for MixD, respectively. Similarly, for Cohen and Kappa analysis the behaviour was consistent and took the following order HDL>TL>DL for all the three types of cohorts. In the Mann-Whitney test HDL, the *p*-value is consistent and ~equal to 0.085, whereas TL was ~0.088 and DL ~0.21.



**Figure 5.9:** Mean feature strength (taken over HDL2-HDL11) for 5 augmentation states corresponding to (a) LonD, (b) LisD, and (c) MixD cohorts (Blue: symptomatic, Orange: asymptomatic).

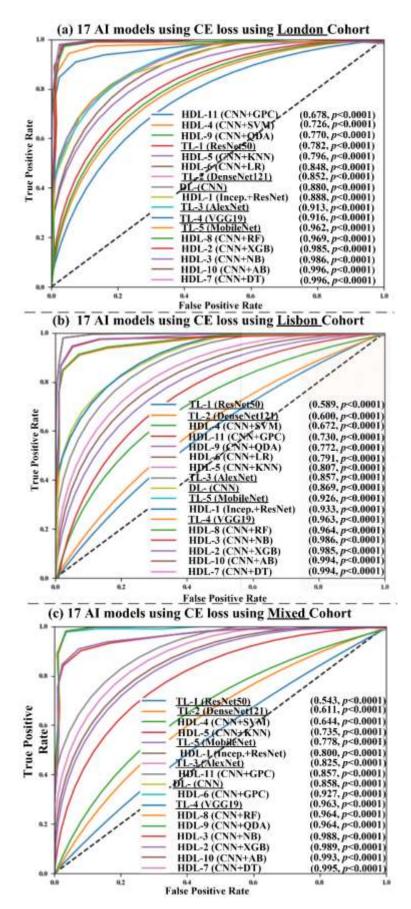


Figure 5.10: ROC analysis for 17 AI models using the three cohorts: (a) LonD, (b) LisD, and (c) MixD.

#### 5.7 Discussion

The study is the first of its kind that uses 17 AI models which include 11 kinds of HDL models, 5 kinds of TL, and 1 kind of DL model for PTC that classifies symptomatic *vs.* asymptomatic carotid plaques automatically. The system, Atheromatic<sup>TM</sup> 2.0<sub>HDL</sub> was optimized concerning augmentation protocol while using three kinds of loss functions (CEL, HL, MSEL) in a ten-fold CV framework. These AI models demonstrated effectiveness to understand the complex patterns of the grayscale carotid plaques for PTC. Our analysis was able to characterize and distinguish (a) symptomatic plaques that are chaotic, echo lucent, contain low collagen, and low calcium, as well as (b) the asymptomatic plaques that are less chaotic, echogenic, high collagen, and high calcium. The hypothesis was validated using mean feature strength, demonstrating that HDL models are superior to TL and DL models. We showed that LonD, HDL (CNN+DT) had the highest accuracy of 99.6% with an AUC of 0.996 (p<0.0001). Similar observations were seen for LisD and MixD databases. The highest performance was seen for augmentation x5 having a data size of 2000 samples for LonD, 1500 samples for LisD, and 2600 samples for MixD.

#### 5.7.1 A short on hardware resources

In this study, we required a high-performance machine due to the number of computations. Our observations showed that when we run the HDL models on the local machine having limited GPU power (2GB), it increased in training time. However, it was less than the traditional DL models. The training time of DL models using the K10 cross-validation paradigm took 23 hrs. for completing one augmentation fold, unlike, in HDL1, which took 14 hrs. For other HDL2 to HDL11 models, the training time was 2.5 hrs. This was because of the conversion of grayscale to point data (CSV format).

#### 5.7.2 A short note on HDL vs. DL and HDL vs. TL

In recent years, the HDL models in AI have started to solve complex problems in healthcare bioinformatics [147, 148, 150]. By removing the fully connected network and replacing it with ML-based models, the HDL-based architectures showed an improvement in performance by displaying a reduction in training times, and further facilitating the process of handling the complex learning of features. Thus, the overall performance of the HDL models improved in comparison to TL and solo DL models.

Our study results were consistent with our previous study [156], which adapted the

multicenter approach, in which we efficiently characterized the carotid ultrasound scans as symptomatic vs. asymptomatic. It is important to note that the difference in feature strengths in HDL was higher compared to DL models [156], which further validated our protocol. The accuracy of HDL improved by 13.44% compared to the DL model [156], and there was an improvement of 3.49% against TL in the MixD data set. Lastly, the HDL model showed a superior performance against DL while computing DOR, Cohen, and Kappa analysis [156], in all three cohorts while using all the three-loss functions. We also demonstrated lower training time in HDL vs. DL.

The performance of TL came out to be comparable with HDL in the London cohort and for the rest of the cohorts HDL is better than TL. The mean absolute difference came out to be 0.40%, 9.21%, and 15.56% for London, Lisbon, and Mixed cohorts, respectively.

#### 5.8 Conclusion

This multicenter study is the first of its kind that used 11 HDL models, six TL models, and one SDL model for plaque tissue characterization of the carotid ultrasound scans into symptomatic and asymptomatic. We validated the hypothesis using three loss functions and evaluated the performance of the HDL models on three kinds of states (unbalanced, balanced, and optimized). Further, we showed the system design in a multicenter paradigm by taking three kinds of cohorts. We used the mean feature strength for the characterization of the HDL models. Our results show that the behavior of the HDL models is consistent with our previous study using DL models, however, our HDL models are superior in performance. In all the three cohorts, HDL>TL>DL having the accuracies 99.78>96.5>88 (in %), and AUC of 0.996, 0.965, 0.88 (p<0.0001), respectively. The HDL showed 6.4% and 3.2% better performance compared to SDL and TL, respectively. We validated the performance of the HDL models with diagnostics odds ratio and Cohen and Kappa statistics. The online system runs in < 2 seconds.

#### **CHAPTER 6**

# LIGHT WEIGHT CNN MODEL FOR IDENTIFICATION OF LESION IN LUNG USING CT SCANS

#### 6.1 Introduction

According to the World Health Organization (WHO), cancer is the second biggest cause of mortality in the world, accounting for one out of every six fatalities (WHO) [8, 9]. Among all the cancer types, lung cancer is most common in men and women [10]. The detection and diagnosis of lung cancer in the early stages increase the mortality rate [11]. Among all the non-invasive techniques, computed tomography(CT) has reached many milestones in finding malignancy detection [157]. However, the size of the malignancy cell varies from 3mm to 30mm, and malignance regions less than equal to 3mm are hard to detect.

Early detection of cancer in CT scans requires skilled experts and advanced techniques in computer vision to help the radiologist for better predictions. This study focused on the early detection of lung cancer using a well-established and openly available dataset, "LIDC-IDRI". A radiologist analyzes these CT Scans for identifying lesions in the form of pulmonary nodules within the CT data [8].

The complexity of the detection of lung cancer depends on the number of nodules in the lung. Moreover, more nodules present in the lungs may increase the chances of lung cancer [158]. However, the radiologist's efficiency depends on intra-observer variability [159], and early detection of the nodule is also hard to find with naked eyes due to its size and shape[160]. In addition, some nodules are surrounded by anatomical structures such as vessels and pulmonary walls. However, the higher number of intra-parenchymal and Ground-glass opacification (GGO) type nodules present a higher chance to turn into lung cancers. From past studies, it has been observed that malignancy has a strong correlation

with intensity distribution and relative positions [161]. In recent years, we have also worked on the characterization of the lesions on COVID [87, 162], stroke [128, 129, 156], and Wilson disease [48].

Recent advances in artificial intelligence (AI) techniques take significant steps to detect the early detection of lung nodules. Moreover, among all the AI techniques, machine learning varies with efficiency by changing the exciting features. However, Deep learning (DL) overcame this gap and came up with automatic feature extraction and detection [98]. In our study, we proposed a conventional convolution neural network (cCNN), modified convolution neural network (mCNN), transfer learning (TL) technique (i.e., VGG19), and model-based on extreme learning machine. The performance of all the proposed models has been compared with learning techniques.

We hypothesized that malignance lesions are brighter than healthy lesions, and malign lesions are softer than healthy lesions. Therefore, the proposed AI systems classify and characterize the malignance lesion also. We summarize our work with the following novelties in this study (1) This is the first time applying modified CNN to the detection of lung nodules, (2) comparison of the performance of optimized cCNN, mCNN, VGG19, extreme learning (EL), and ML with a combination of features (Haralick, Histogram, HOG, LBP), (3) characterize the malignant and healthy lesions using novel means feature strength (MFS), (4) offers a tool to characterize the malignant lesion using traditional models Bispectrum and fractal dimension. (5) Benchmarking with previously published works.

#### 6.1.1 Multi-Nodule Classification

After achieving more advanced technologies for diagnosing the disease even though cancer is a big threat to society. This is the second most common disease, and the different government has paid enormous attention to minimizing this disease. Moreover, many research community has been working to detect cancer in the early stage, because as time passes the severity of cancer disease increases. Moreover, if cancer is not detected in the early stage, then it is invincible to drugs and radiation. There are two methods to detect cancer 1) Non-invasive and 2) Invasive. Non-invasive techniques have been raising their importance and detailed discussion is available in various literature. Non-invasive techniques are 1) X-rays 2) CT scans 3) PET Scans 4) MRI Scans, etc.

Among all non-invasive techniques, computed tomography (CT) has reached many

milestones in finding malignancy detection. These CT scans are analyzed by a radiologist for identifying lesions in the form of pulmonary nodules within the CT data [7]. The amount of this work is enormous and requires a lot of patience to analyze and detect the regions since the lung nodule diameter is the size of 10 mm or even less. A CT scan acquired after scanning of full lung contains more than 90,000 voxels [4]. Lung nodules with diameters between 3 to 10 mm occupy only 0.085% to 0.01% of the lung volume. So, it is a challenge for a radiologist to detect the malignance area within less amount of time [4]. The complexity increases when the number of nodules increases. On the other hand, if many nodules are present in the lung, it may increase the chance of lung cancer [7].

Radiologists' efficiency of interpretation depends on how it is required and reconstructed. In general, the shape of the lung nodules is spherical, due to its shape, it is too hard to find when it is surrounded by anatomical structures such as vessels, and pulmonary walls. Intra-parenchymal [8], and GGO [9], type nodules have a higher chance to turn into lung cancers. Since malignancy has a strong correlation with intensity distribution and relative positions.

In the proposed work, our main focus is on the classification of Multi nodule, single nodule, and healthy low-dose CT scan images. For the classification purpose, a convolution neural network model has been proposed. In this model, there are four convolution and maxpolling layers. After processing the data from the convolution and max-pooling layer, it is processed at a fully connected layer. The experimental results show the efficacy of the proposed algorithm over traditional machine learning approaches and famous CNN models under various parameters i.e., accuracy, precision, recall, and Cohen Kappa.

#### **6.2** Background Literature

For the literature survey, we considered three significant repositories: IEEE, Springer, and Knimbus ( repository of all the well-established databases such as "JSTOR", "EBSCO", "Oxford", "Harvard", and "Elsevier" and used the following keyword for searching the articles "lung nodule detection", "lung cancer detection using AI", "Pulmonary nodules detection using deep learning", "lung cancer detection using deep learning", "lung nodule classification using machine learning", "lung nodule classification and characterization", "lung nodule characterization".

Researchers have worked on detecting lung cancer and classifying the lung

nodule(s) using statistical techniques and AI techniques in the last two decades. Han *et al.* [28] extracted 2D and 3D Haralick texture features from the LIDC-IDRI dataset and fed these features into a support vector machine (SVM) for classifying the CT lung nodule. Moreover, the authors have achieved the area-under-the-curve 0.9441on 3D haralick and 0.9372 on 2D haralick texture features. Tran *et al.* [163] have proposed a CAD system to classify pulmonary nodules into non-nodule vs. nodules on LIDC-IDRI and Luna16 cohorts. The authors have used 15-layer 2D CNN for automatic feature extraction and classification. They have improved the accuracy by using focal loss and achieved an accuracy of 97.2%.

Maldonado *et al.* [164] proposed a classifier called BORDERS (benign versus aggressive nodule evaluation using radiomic stratification)and fed with eight features such as location, shape, size, texture, and surface. The authors validated the classifier with well-established models (Brock) and achieved an AUC of 0.90 with a CI of 0.85 to 0.94. Jacobs *et al.* [165] compared the state of art CAD systems to detect lung nodules in the LIDC-IDRI cohort. They have used two CAD systems 1) a commercial CAD system and 2) an educational CAD system. They have achieved the best sensitivity with 82% in the commercial CAD system. Nibali *et al.* [152] proposed a CAD system to detect malignancy nodules from cropped CT images of LIDC-IDRI. These images are fed to the deep residual networks built on the ResNet model and achieved an accuracy of 91.07%.

#### 6.3 Dataset

#### 6.3.1 Patient Demographics

The publicly available dataset LIDC-IDRI [166] has been used for the experimentation. LIDC-IDRI contains the 1018 helical CT scans of the lung, and these Nodules' size lies between 3 to 30 mm. Moreover, each slice is examined and annotated by an expert radiologist. In this study, we considered 905 nodules, out of which 422 were malignant and 483 benign nodules. However, we have discarded such a slice with a higher malignancy rating, i.e., three or more out of five. As the proposed study focuses on the early detection of lung cancer, the higher rating scans are more suspicious of malignancy. Sample images of the LIDC-IDRI dataset are depicted in Figure 6.1.

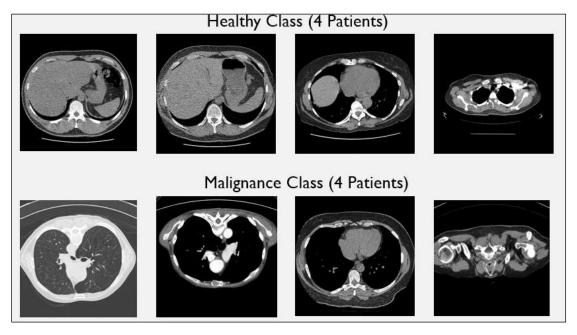


Figure 6.1: Sample of LIDC-IDRI dataset [34].

#### 6.3.2 Augmentation

Deep learning applications require a considerable amount of data for better performance. The LIDC-IDRI cohort contains an average number of data samples, and it has a class imbalance problem. In this study, we did not consider some samples (nodules size>3mm). After balancing the dataset, we augmented2, 3, 4, 5, and 6 number folds. The details of augmented data are presented in Table 6.1. For the augmentation purpose, four geometrical transformations such as a) rotation from -5 degrees to 10 degrees, b) random flipping, c) rotating 270 degrees, and d) skew have been applied. After completing the augmentation process, the data has been passed into different deep learning models for classification purposes.

Benign Malignant Aug x Balanced 500 500 1000 Aug2x 1000 Aug3x 1500 1500 Aug4x 2000 2000 2500 2500 Aug5x

**Table 6.1:** Details of the augmentation.

#### 6.3.3 Dataset for Multi-Nodule Detection

Aug6x

The suggested study employed a publicly accessible dataset called Early Lung Cancer

3000

3000

Action Program (ELCAP) [18], which included Low Dose CT images. This dataset comprises 50 CT scan pictures of patients ranging in age from 40 to 85 years old, including second-hand smokers and those at risk from occupational exposure. Moreover, many cases are stage I Lung cancers [19] and every scan is labeled and annotated by professional radiologists. All scans are taken by single breath-hold with 1.25 mm slice thickness and spacing between slice is 0.5 mm because here most nodule size is between 2 mm to 5 mm [14]. In our proposed model, 42 cases have been used for training, 4 cases for validation, and 4 cases for testing. Sample of Dataset given in Figure 6.2.

In ELCAP Dataset all the slices are in DICOM (Digital Imaging and Communications in Medicine) format. All the images are radiographic images and the unit for CT Scan images is Hounsefield Unit (HU). So, all images are converted from HU units to digital pixels using the following equation 6.1. After the conversion of the images from HU to pixel values, there are 12,000 healthy images, 3,000 images with single nodule, and 150 images with multi nodules. As the number of samples in each class is not similar so, to make the number of samples almost equal in each class, data augmentation has been applied. Dataset is divided into three different parts namely train, test, and validation training, validation, and testing split with k-fold cross-validation of 80:10:10 manually in such a way that each split should take at least 1 to 2 augmented images with original images.

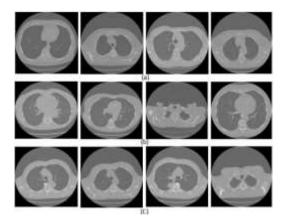


Figure 6.2: Sample of Images of ELCAP Dataset (a) Single Nodule (b) Multi Nodule (c) Healthy.

Hounsefield Unit = pixelvalue 
$$*$$
 slope + Intercept (6.1)

#### 6.4 Proposed Model for Lung Lesion Classification and

#### Characterization

In Figure 6.3, the global architecture of the DL frameworks is presented. Seven deep conventional convolutional neural networks (DCNN)have been considered for optimization over augmented data. These seven models varied in the number of layers 5, 7, 9, 11, 13, 15, and 19. In all these models, the number of combinations of convolution (CL) and average pooling (AP) varies, but the number of dense layers and dropout layers are constant (2+1 Layers). All the AI models have been applied to all the augmentation folds with K10 cross-validation (90% training and 10% testing) to obtain an optimization point.

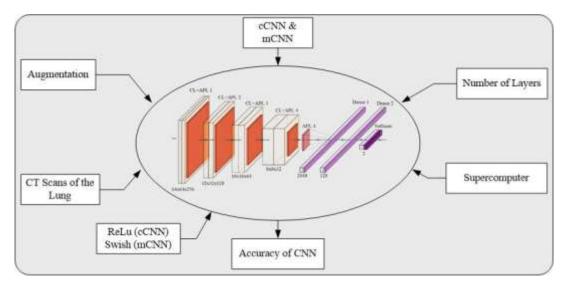


Figure 6.3: The global architecture of deep learning architecture.

We have x combinations of CL and AP layers in all DL architectures, followed by a flattened layer that converts 2D Feature maps to 1D Feature maps. Then we have fully linked networks (so-called Dense Layer) with dropout to reduce the possibilities of overfitting, and the ReLu activation function has been utilized to offer nonlinearity in the DL models. The Sigmoid activation function was employed in the output layer since there are only two classes in the dataset. Finally, the loss in convolution neural networks was computed using categorical cross-entropy. Equation 6.2 shows the formula for categorical cross-entropy.

Loss = 
$$-[(y_i \times \log a_i) + (1 - y_i) \times \log(1 - a_i)$$
 (6.2)

Where i is the class label for input and  $a_i$  is the predicted probability of class being  $y_i$ .

#### 6.4.1 Modified CNN

In this architecture, we have used the modified version of Sigmoid to provide the nonlinearity in the model. This activation function is the first-order derivative of the sigmoid function known as the "Swish" activation function (refer to equation 6.3).

$$S_a(x) = x^* \frac{1}{1 + e^{-x}} \tag{6.3}$$

Here  $S_a(x)$  is the Swish activation function, and x is the input. In the proposed mCNN, we applied the Swish activation function on all the layers except the output layer. Moreover, after applying the Swish activation function, there is a significant improvement in the performance of the mCNN.

#### 6.4.2 Transfer Learning Architecture

For the experimental work, we have also used the optimized VGG-19 [47] architecture. Our study used mCNN, and cCNN up to 19 layers, so it is helpful to compare all the DL models' performances. Initially, VGG19 is trained on the ImageNet dataset, and we freeze initial layer weights with ImageNet and fed the augmented LIDC-IDRI cohort at the end layers (i.e., from FCN). We added two dropout layers at the end of the model for controlling the overfitting. The architecture of the TL model is depicted in Figure 6.4.

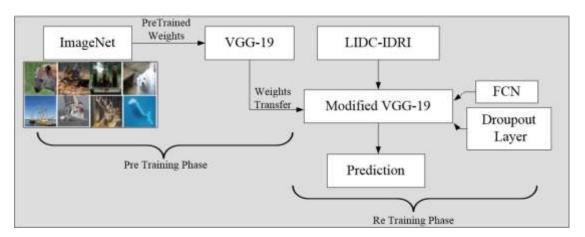


Figure 6.4: Global TL architecture.

#### 6.4.3 Machine Learning Architecture

We compared the proposed DL system to ML models and input four texture characteristics into the ML system for classification: haralick, histogram, histogram of gradients, and local binary patterns texture features. At the optimal augmentation, we chose the best mix of features and fed them to typical ML classifiers including support vector machine (SVM), logistic regression (LR), random forest (RF), k-nearest neighbor, and decision tree (DT).

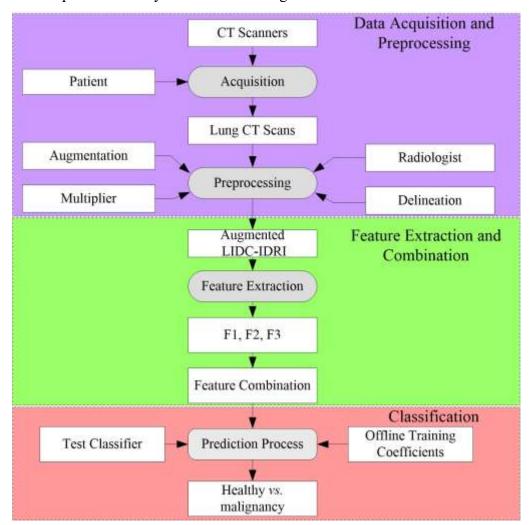


Figure 6.5 depicts the ML system's overall design.

Figure 6.5: ML system architecture.

#### 6.4.4 Proposed Model for Multi Nodule Detection

A convolution neural network has been constructed for cancer detection in the suggested study. Four convolution layers, four max polling layers, and two flatten layers were employed in the suggested model. The architecture of the proposed model is depicted in Figure 6.6. Table 6.2 gives the hyper parameters used in the proposed model.

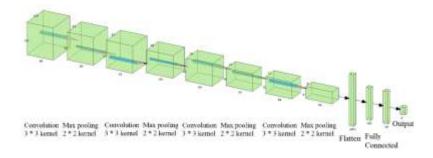


Figure 6.6: Proposed CNN Architecture.

Hyper-parameter	Description	
No. of convolution layer	4	
No. of max-polling layer	4	
Dropout rate	3	
Learning rate	0.001	
Momentum	0.999	
Epochs	2000	
Batch size	32	
Network weight initialization	Glorot uniform	
Activation	Relu	

**Table 6.2:** Hyper-parameter tuning.

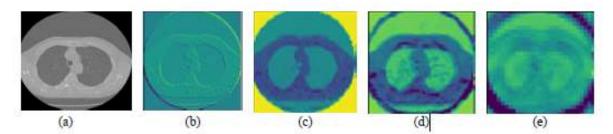
For the calculation of loss, a categorical cross-entropy function has been used and the formula for the same is represented in equation 6.4.

$$E = -(y\log(p) + (1 - y)\log(1 - p))$$
(6.4)

Here p-predicted probability, y- binary indicator for observation classification. After performing the convolution operation in the model data has been passed to a fully connected layer. At output layer score for each class has been calculated using the softmax function shown in equation 6.5.

$$F(x) = \frac{e^{x_i}}{\sum_{j=0}^{k} e^{x_j}} \quad \text{where } i, j = 0, 1, 2, 3$$
 (6.5)

Where  $x_i$  is the score calculated by the model and k is the number of classes. Figure 6.7 describes the information stored in the intermediate layers and the loss, and accuracy curve of the proposed model is presented in Figures 6.8 and 6.9.



**Figure 6.7:** (a) Original image and Activations after (b) First convolution layer (c) Second convolution layer (d) Third convolution layer and (e) fourth convolution layer.

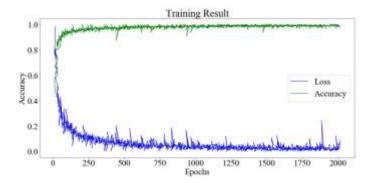


Figure 6.8: Training accuracy graph with epochs.

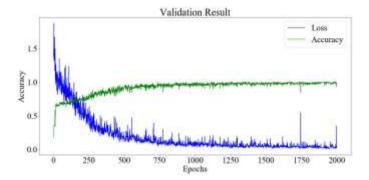


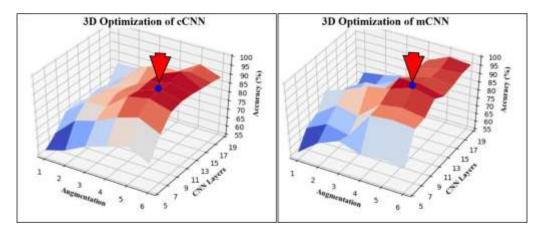
Figure 6.9: Validation accuracy graph with epochs.

#### 6.5 Results

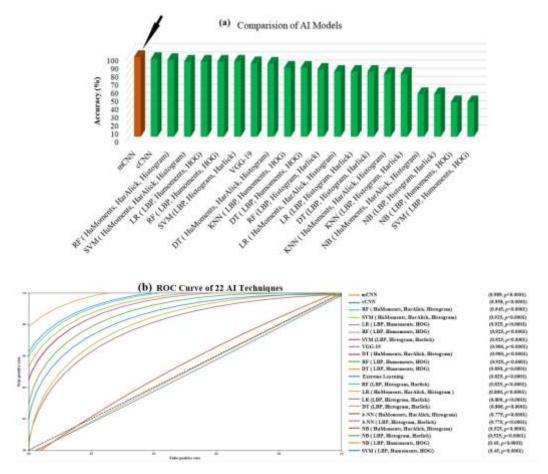
This section offers information on the experimental setup and outcomes analysis. All the experiments have been carried out using the NVIDIA DGX v100 machine, with 40600 CUDA cores, 5120 tensor cores, 128 GB RAM, and 1000 TFLOPS speed. We use K10 cross-validation to send all of the enhanced folds to the suggested AI models, which is 90% training and 10% testing, and we use 100 epochs to train the model experimentally. We have tested our models on various hyperparameters, and the best set of parameters has been chosen based on low validation errors.

mCNN, cCNN contains a range of layers from 5 to 19; we fed all the augmented folds to each model. Similarly, we have applied the same CV protocol to TL architectures. So, we ran18000 epochs (18x100x10) to get the optimization point of each model. The optimization results of mCNN and cCNN are presented in Figure 6.10, the optimization point is represented using a blue color dot.

Figure 6.11 represents the performance comparison of AI models at an optimization point in terms of accuracy. Among all the techniques, mCNN shows better performance than all the other proposed AI models. The accuracy of all AI models is presented in Table 6.3.



**Figure 6.10:** 3D surface plots of optimization of cCNN and mCNN. The blue color dot represents the optimization point.



**Figure 6.11:** (a) Comparison of proposed AI models at an optimization point (Aug5x). (b) ROC Curves 22 AI Techniques with AUC.

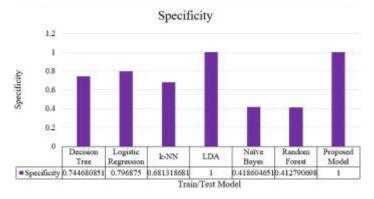
Table 6.3: Accuracy comparison of AI models.

AI Model	Accuracy (in %)
mCNN	98.75
cCNN	95.8

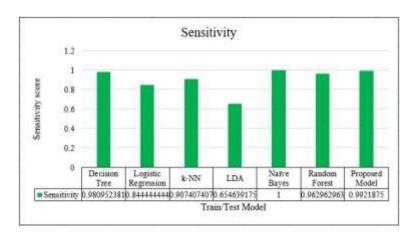
RF ( HuMoments, HarAlick, Histogram)	94.5
SVM ( HuMoments, HarAlick, Histogram)	92.5
LR (LBP, Humoments, HOG)	92.5
RF (LBP, Humoments, HOG)	92.5
SVM (LBP, Histogram, Harlick)	92.5
VGG-19	90.45
DT ( HuMoments, HarAlick, Histogram)	90
KNN ( LBP, Humoments, HOG)	85
DT (LBP, Humoments, HOG)	85
Extreme Learning	82.7
RF (LBP, Histogram, Harlick)	82.5
LR ( HuMoments, HarAlick, Histogram)	80
LR (LBP, Histogram, Harlick)	80
DT (LBP, Histogram, Harlick)	80
KNN ( HuMoments, HarAlick, Histogram)	77.5
KNN (LBP, Histogram, Harlick)	77.5
NB ( HuMoments, HarAlick, Histogram)	52.5
NB (LBP, Histogram, Harlick)	52.5
NB (LBP, Humoments, HOG)	42.5
SVM ( LBP, Humoments, HOG)	42.5

#### 6.5.1 Results for Multi-Nodule Lesion Classification

In Table 6.4, the accuracy comparison of the proposed model and the typical machine learning model is computed, as well as the accuracy comparison of alternative methods in terms of accuracy and F1 score. Figures 6.12 and 6.13 show performance comparisons of several machine learning algorithms (e.g., SVM, k-NN, Decision Tree, etc.) using different assessment metrics, such as specificity and sensitivity.



**Figure 6.12:** Specificity for traditional machine learning algorithms i.e., decision tree, logistic regression, k-NN, LDA, Naive Bayes, Random Forest, and Proposed model.



**Figure 6.13:** Sensitivity for traditional machine learning algorithms i.e., decision tree, logistic regression, k-NN, LDA, Naive Bayes, Random Forest, and Proposed model.

Table 6.4: Performance Comparison with Other Machine Learning.

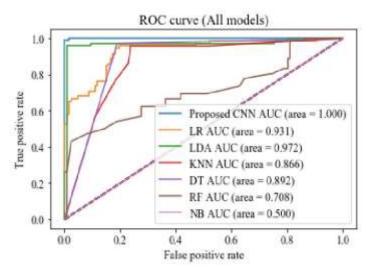
Features	Model	Accuracy	F1Score
	Decision Tree	0.583	0.515
Haralick	Logistic Regression	0.935	0.951
Hu-moments HSV	k-NN	0.95	0.961
histogram	LDA	0.653	0.784
LBP	Nave Bayes	0.497	0.351
	Random Forest	0.492	0.348
	Decision Tree	0.869	0.887
Haralick	Logistic Regression	0.829	0.87
Hu-moments	k-NN	0.804	0.834
HSV	LDA	0.663	0.791
histogram	Nave Bayes	0.497	0.35
	Random Forest	0.487	0.337
	Decision Tree	0.492	0.365
11017	Logistic Regression	0.854	0.89
HSV	k-NN	0.809	0.848
histogram Hu-moments	LDA	0.894	0.923
Tu-moments	Nave Bayes	0.497	0.351
	Random Forest	0.804	0.828
	Decision Tree	0.437	0.309
	Logistic Regression	0.884	0.916
Haralick,	k-NN	0.814	0.845
Hu-moments	LDA	0.714	0.816
	Nave Bayes	0.497	0.351
	Random Forest	0.472	0.314
Proposed Model		0.995	0.996

The probability curve is called ROC (Receiver operating characteristics), and AUC (Area under the curve) is a measure of separability between true and negative classes. It is based on binary classification; therefore, one class must be positive, and the others must be

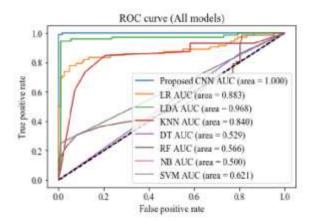
negative. If the AUC is near 1, it has good separability, meaning it can categorize 1s as 1s and 0s as 0s. AUC that is close to 0 classifies 1s as 0s and 0s as 1s. As a result, the higher the AUC score, the better the model. Figures 6.14 and 6.15 show the ROC-AUC curves for all of the methods with various characteristics, and it can be shown that the suggested model outperforms the others.

Cohen Kappa is a measure of rater trustworthiness. It's a step forward from % agreement between two classification outcomes. Furthermore, its value goes from 0 to 1, with a higher number indicating greater dependability. Because a classifier for illness detection has been built in the proposed study, high reliability is sought, and the offered methods obtained the highest score under the Cohen Kappa measure (see Figure 6.16).

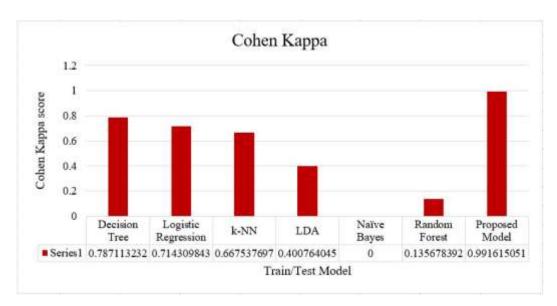
To compare the results, two well-known CNN models, namely VGG16 [20] and Inception V3 [21] were also on the schedule. Figures 6.17 and 6.18 show the architecture of VGG16 and Inception V3. Both models, however, were created for 1000 classes and are modified at the output layer, as in our case. There are only three courses, which is an issue. VGG16 and Inception V3 are also compared to a suggested CNN model in terms of storage space, inference time, and accuracy. Figures 6.19, 6.20, and 6.21 show the final findings.



**Figure 6.14:** ROC-AUC for healthy lung images with 3 features i.e. HSV, Humoments, and Haralick features as input in machine learning algorithms.



**Figure 6.15:** ROC-AUC for healthy lung images with 3 features i.e. HSV, Humoments, and Haralick features as input in machine learning algorithms.



**Figure 6.16:** Cohen Kappa for traditional machine learning algorithms i.e., decision tree, logistic regression, k-NN, LDA, Naive Bayes, Random Forest, and Proposed model.

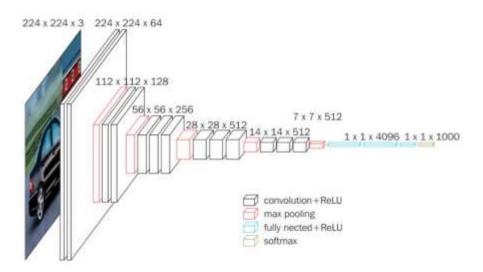


Figure 6.17: The architecture of VGG16.

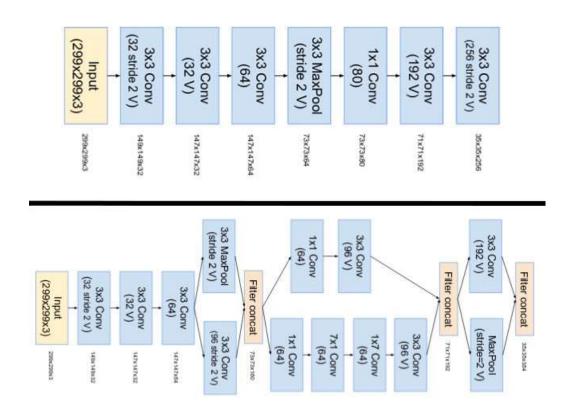


Figure 6.18: The architecture of Inception V3.

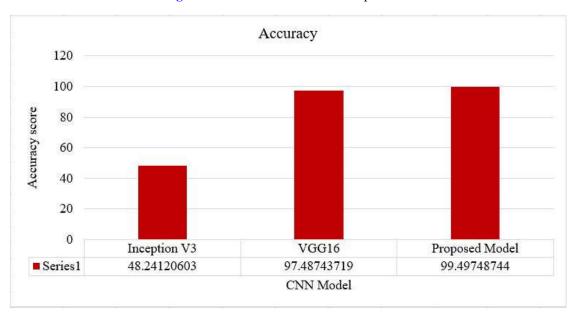


Figure 6.19: Accuracy comparison among Inception V3, VGG16, and proposed CNN model.

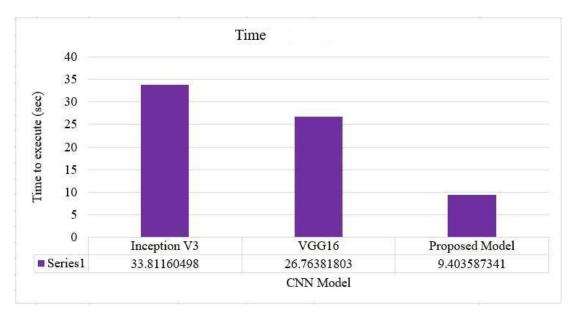


Figure 6.20: Inference time comparison among Inception V3, VGG16, and proposed CNN model.

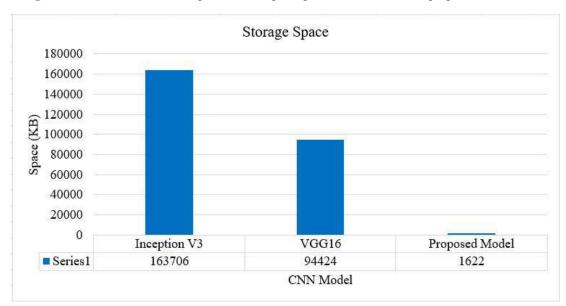


Figure 6.21: Space comparison among Inception V3, VGG16, and proposed CNN model.

Because ML methods rely on hand-crafted characteristics, the suggested methodology outperforms existing machine learning (ML) approaches. In the case of a neural network, however, the model extracts the key information by itself. Furthermore, the well-known CNN models VGG16 and Inception V3 perform poorly since both have numerous layers and are prone to overfitting. Because too many parameters were learned in VGG 16 and Inception V3, the reaction time in these models is excessively long. In contrast to VGG 16 and Inception V3, the layers in the proposed model are constrained, reducing the risk of overfitting. Furthermore, since fewer parameters have been educated to do the requested job, the reaction time is reduced.

# 6.5.2 Characterization of the Lung Lesion

In this section, we studied the histogram analysis of healthy and malignant nodules. We found an increase in the area by 2.1% in malignant nodules compared to health nodules. Figure 6.22 represents the histogram analysis of healthy and malignant. Validation of histogram is carried out by fractal dimension.

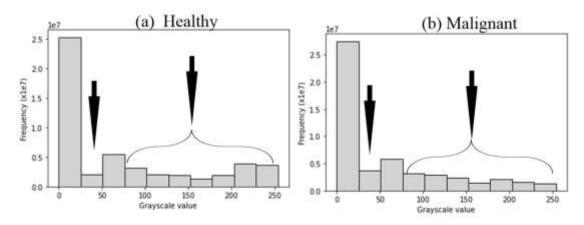


Figure 6.22: Histogram of (a) healthy and (b) malignant lung nodules.

# Hypothesis

we hypothesized that malignant lesions are softer than healthy lesions. We validated this hypothesis using Mandelbrot's fractal dimension (FD) equation, represented in equation 6.6. If the FD of malignancy is less than the healthy lesions, then the malignancy lesion is softer than healthy lesions. Because FD is a measure of chaotic nature in the images. We plot the FD for all the trained images at the optimization point, shown in Figure 6.23.

$$D = \frac{\log(N)}{\log(r)} \tag{6.6}$$

D is the dimensionless quantity of self-similar objects, N is the number of boxes that cover the pattern, and r is the magnification.

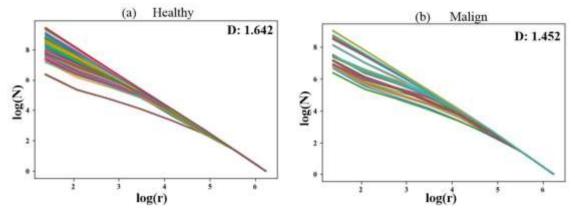


Figure 6.23: Fractal dimension of (a) healthy and (b) malignant.

We validated our hypothesis with a higher-order spectrum (HOS) and took the optimized augmented cohort rotated from 0 to 180 degrees to get the best angle for Bispectrum. We got the best angle at 60 degrees and found that the Bispectrum of malignancy is less than the Bispectrum of healthy. Figure 6.24 represents the intensity distribution of the malignance and healthy lesions. The red arrow represents the peak of the spectrum.

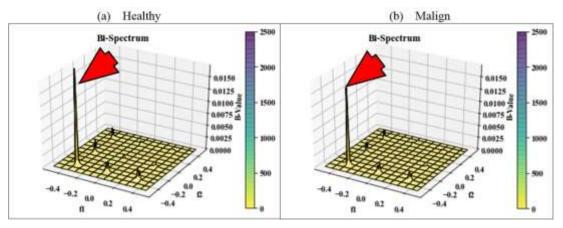
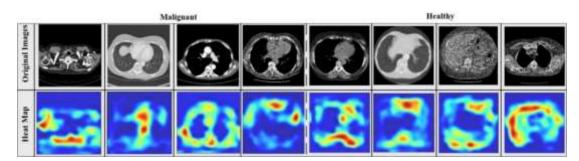


Figure 6.24: Higher-order spectra of (a) healthy and (b) malignant. The red arrow is pointing to the peak of the spectrum.

#### 6.5.3 Heat maps

We have also generated the Heat maps using gradient weighted class activation mapping (Grad-CAM) [167]. It represents the ability of the classifier to identify the region of interest in the input image. We have used VGG-19 model weights for generating the heat maps, and Figure 6.25 represents the Heat maps of the healthy class and malignant class. The red color represents the region of interest of the classifier.



**Figure 6.25:** Heat Maps of the malignant and healthy class generated from the VGG-19 Grad-Cam Technique.

#### **6.6 Performance Evaluation**

# 6.6.1 Power Study

We validated our study using power analysis, and the formula of the power study is presented in equation 6.7. Here we choose a 95% confidence interval with a 5% margin of error, and the value of data proportion is 0.5.

$$n = \left[ \left( z^* \right)^2 \times \left( \frac{\hat{p}(1-\hat{p})}{MoE^2} \right) \right] \tag{6.7}$$

Here n is the number of samples required,  $z^*$ : standard z-table,  $\hat{p}$ : Data proportion, MoE: margin of error.

After performing the power study, it has been concluded that the minimum number of required samples is 384 in the dataset. However, in our dataset, the number of samples is 1184, 67.56% higher than the required samples.

# 6.6.2 Diagnostics of Odds Ratio(DOR)

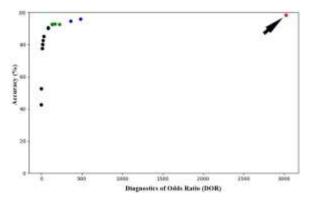
DOR is a practical diagnostic test on medical imaging, a ratio of target disorder from subjects without it. It takes any value between 0 to infinity, and higher the DOR means better performance. We can calculate DOR from sensitivity (S<sub>e</sub>) and specificity (S<sub>p</sub>) using equation 6.5. DOR of used AI methods is presented in Table 6.5. Moreover, the DOR vs. Accuracy comparison of all 22 AI models is presented in Figure 6.26.

$$DOR = \frac{(S_e * S_p)}{(1 - S_e) * (1 - S_p)}$$
(6.5)

Table 6.5: DOR	vs. Accuracy	(Descending	order of DOR).
I word out Dore	vs. riccaracy	(Descending	oraci or Dorej.

SN#	Model	DOR	Accuracy
1	mCNN	3024.99	98.24
2	cCNN	485.99	95.8
3	RF ( HuMoments, HarAlick, Histogram)	364.5	94.5
4	SVM ( HuMoments, HarAlick, Histogram)	135.2	92.5
5	LR (LBP, Humoments, HOG)	135.2	92.5
6	RF ( LBP, Humoments, HOG)	135.2	92.5
7	SVM (LBP, Histogram, Harlick)	225.33	92.5
8	VGG-19	86.4	90.45
9	DT ( HuMoments, HarAlick, Histogram)	86.4	90
10	KNN ( LBP, Humoments, HOG)	32.5	85
11	DT ( LBP, Humoments, HOG)	32.57	85

12	RF (LBP, Histogram, Harlick)	23	82.5
13	Extreme Learning Machine	23	82.7
14	LR ( HuMoments, HarAlick, Histogram)	16.56	80
15	LR (LBP, Histogram, Harlick)	17.25	80
16	DT (LBP, Histogram, Harlick)	18.29	80
17	KNN ( HuMoments, HarAlick, Histogram)	13.5	77.5
18	KNN (LBP, Histogram, Harlick)	13.5	77.5
19	NB ( HuMoments, HarAlick, Histogram)	1.22	52.5
20	NB (LBP, Histogram, Harlick)	1.22	52.5
21	NB ( LBP, Humoments, HOG)	0.52	42.5
22	SVM ( LBP, Humoments, HOG)	0.52	42.5



**Figure 6.26:** DOR vs. Accuracy comparison of 22 AI techniques. The red color arrow represents the highest DOR (mCNN).

# 6.6.3 Ranking

In this study, we computed the performance of 22 AI models, and all the AI models are arranged according to a gradient scheme for ranking the models. The ranking of every model is presented in Table 6.7. Ten attributes have been selected (refer to Table 6.6), and all attributes have been given maximum grades with a value of 5.

SN	Attribute	High Grade (4-5)	Medium Grade (3-2)	Low Grade (1-0)	
1	Optimization	High Aug (>5)	Avg Aug (<5 and >=3)	Low Aug (<3)	
2	Accuracy	> 95	>85 to <95	<85	
3	False Positive Rate	<0.1	>0.1 and. <0.2	>0.2	
4	F1 Score	>0.9	>0.8 and <0.9	<0.8	
5	Sensitivity	>0.9	>0.8 and <0.9	<0.8	
6	Specificity	>0.9	>0.8 and <0.9	<0.8	

**Table 6.6:** Attributes and range values.

7	Data Size	>= 2000	>1500 and <2000	<=1000
8	DOR	>500	>100 and <500	<100
9	Training Time	< 1 Hrs	>1 Hrs and <5	>5 Hrs
10	AUC	> 0.95	>0.85 to <0.95	< 0.85

After collecting all the values, we have graded the performance of the AI model according to Table 6.4. For providing the rank to individual models, we have computed the absolute sum of all the grades and divided it with the absolute sum of the graded sum with the maximum grade, i.e., 50 (10 attributes x 5). Out of all the AI models, the performance of mCNN is comparatively better. However, NB with HuMoments, Haralick, and Histogram combination features shows undervalued performance. The performance of the AI model is color-coded from green to red. better performer coded with green and undervalued performed colored with red.

**Table 6.7:** Gradient scheme for ranking the AI models.

Model	0*	A	F	F1	Se	Sp	DS	D*	TT	AUC	AS	%
mCNN	5	5	5	5	5	5	5	5	3	5	48	96
cCNN	5	4	5	5	5	5	5	3	2	4	43	86
SVM (LBP, Histogram, Harlick)	5	3	5	5	4	5	5	3	5	3	43	86
RF ( HuMoments, HarAlick, Histogram)	5	3	5	5	5	4	5	3	3	3	41	82
RF (LBP, Humoments, HOG)	5	3	4	4	4	4	5	2	5	3	39	78
SVM ( HuMoments, HarAlick, Histogram)	5	3	4	4	4	4	5	2	4	3	38	76
LR (LBP, Humoments, HOG)	5	3	4	4	4	4	5	2	4	3	38	76
VGG-19	5	2	4	4	4	4	5	1	2	2	33	66
DT ( HuMoments, HarAlick, Histogram)	3	2	4	4	4	4	4	1	1	2	29	58
KNN ( LBP, Humoments, HOG)	3	2	3	3	2	3	4	0	5	2	27	54
DT ( LBP, Humoments, HOG)	3	2	2	3	3	2	4	0	5	2	26	52
Extreme Learning	3	2	2	3	3	3	4	0	4	2	26	52
LR ( HuMoments, HarAlick, Histogram)	5	1	1	2	3	0	5	0	4	1	22	44
RF (LBP, Histogram, Harlick)	3	1	2	2	2	1	4	0	5	1	21	42
DT (LBP, Histogram, Harlick)	2	1	3	1	1	3	3	0	5	1	20	40
LR (LBP, Histogram, Harlick)	3	1	1	1	2	0	4	0	5	1	18	36
KNN (LBP, Histogram, Harlick)	2	0	2	1	1	1	3	0	5	0	15	30
KNN ( HuMoments,	2	0	2	1	1	1	3	0	4	0	14	28

HarAlick, Histogram)												
NB (LBP, Histogram, Harlick)	3	0	0	0	0	0	4	0	5	0	12	24
NB ( LBP, Humoments, HOG)	3	0	0	0	0	0	4	0	5	0	12	24
SVM ( LBP, Humoments, HOG)	3	0	0	0	0	0	4	0	5	0	12	24
NB ( HuMoments, HarAlick, Histogram)	2	0	0	0	0	0	3	0	4	0	9	18

<sup>\*\*</sup> O\*: Optimization, A: Accuracy, F: False Positive Rate, F1: F1 Score, Se: Sensitivity, Sp: Specificity, DS: Data Size, DOR: DOR, TT: Training Time, AUC: AUC

### 6.6.4 Mann-Whitney Test and Parried T-test

Our study involved four kinds of AI techniques, so we need to calculate the correlation between them. Therefore, to compare the two groups without assuming that they are typically distributed, and Mann-Whitney test gives such types of comparisons. Along with this test, we have considered paired sample t-test, which gives the mean difference between the two sets of observations.

Mann-Whitney test comparison among the models has been presented in Table 6.8. Similarly, we have also checked the correlation between features extracted for ML techniques presented in Table 6.9; Table 6.10 gives the paired T-test observations.

**Table 6.8:** Mann-Whitney test between AI models.

AI Comb#	z	p	U
mCNN-cCNN	0.674	0.5001	50.50
mCNN-ML	1.343	0.1794	36.00
cCNN-ML	1.764	0.0778	30.00

Table 6.9: Mann-Whitney test between ML features.

ML Comb#	Z	p	U
FC1-FC2	0.356	0.475	16.50
FC1-FC3	1.324	0.501	8.95
FC2-FC3	1.565	0.156	8.85

FC1: HuMoments, Haralick, Histogram; FC2: LBP, HuMoments, HOG; FC3: LBP, Histogram, Haralick.

**Table 6.10:** Paired sample t-test between AI models.

Comb#	t	p	DF
mCNN-cCNN	-0.0706	0.016	10
mCNN-ML	-1.443	0.0018	9
cCNN-ML	-0.801	0.113	9

# 6.7 Discussion

# 6.7.1 Benchmarking

The existing works on lungs in AI for the LIDC dataset are shown in Table 6.11. It contains five columns and ten rows. The first column represents the author and the year of publication of the work. The second and third columns show the features selected and the AI techniques used by the corresponding authors. The fourth column represents the architecture used by corresponding authors on the LIDC-IDRI dataset. Finally, the fifth column represents the performance of the model under accuracy, AUC, and p-value.

Table 6.11: Benchmarking with existing work.

	C1	C2	C3	C4	C1	C5	C6
SN #	Authors, Year	Features Selected	Technique used	Application	Dataset	ML vs. DL vs. TL	Acc. (%) AUC with p-value
R1	Han et al. 2013 [28]	Texture Features 2D and 3D Haralicks	Support Vector Machine (SVM)	Classification of benign vs. malign	LIDC- IDRI	ML	0.9441 (3D Haralick), 0.9372 (2D Haralick)
R2	Jacobs <i>et al.</i> 2015 [165]	Contrast, Morphologic al Features	Commercial CAD systems (Visia, Herakles), Academic CAD systems (ISICAD)	Classification of benign vs. malign	LIDC- IDRI	ML	82 (Sen)
R3	Shen et al.	Automatic Discriminati	Multiscale CNN,	Classification of benign vs.	LIDC- IDRI	DL	86, 85.82,

	2015 [168]	ve features, HOG, LBP	SVM, RF.	malign			85.60
R4	Xie et al. 2017 [169]	Texture, shape, and CNN	Adaboosted back propagation neural network (BPNN)	Classification of benign vs. malign	LIDC- IDRI Divided into three parts using maligna ncy rate	DL	0.9665, 0.9445, 0.8124,
R5	Nibali <i>et al</i> . 2017 [152]	Automatic	ResNet-18	Classification of benign vs. malign	LIDC- IDRI	TL	89.90 0.9459
R6	Deyet al. 2018 [170]	Automatic	3D CNN, 3D Multi- output CNN, 3D DenseNet, 3D Multi- output DenseNet	Classification of benign vs. malign	LIDC- IDRI patches	DL	84.35, 85.84, 88.42, 90.40, 0.9064, 0.8205, 0.9451, 0.9548,
R7	Deyet al. 2018 [171]	PSO optimized CNN features	CNN with PSO	False-positive reduction in the classification of benign vs. malign	LIDC- IDRI	DL	97.62, 0.955
R8	Nóbrega et al. 2018 [172]	Xception, ResNet50, MobileNet, VGG16, VGG19, InceptionV3, InceptionRes NetV2, NASNetMo bile, NASNetL, DNet169, DNet201 extracted features	Bayes, MLP, Nearest Neighbors, Random Forest, SVM linear, SVM RBF	Classification of benign vs. malign	LIDC- IDRI	ML	88.41, 0.9319 (SVM RBF+ ResNet50)
R9	Al-Shabi et al. 2019 [173]	Local and Global Features	Deep Local-Global Network (Combinatio n of ResidualBlo cks, Non-	Classification of benign vs. malign	LIDC- IDRI	DL	88.46, 95.62

			Local Blocks)				
R10	Proposed Method	Automatic Features	Optimized Modified CNN, Conventiona 1 CNN	Classification of benign vs. malign	LIDC- IDRI	DL	98.75,0.988 95.80, 0.96 92.7, 0.93 (p<0.0001)

#### 6.7.2 Hidden Layer visualization

Our study used three kinds of DL architectures cCNN, mCNN, and TL. TL framework works on the weights transfer from a model trained on another set of data samples, so we excluded the visualization of the TL models. Instead, we have compared the final features learned from the mCNN, as shown in Figure 6.3. Information presented in Figure 6.12 is the output of each filter at the last convolution, i.e., the layer before the vectorization.



Figure 6.27: Intermediate layer visualization top (mCNN)DL architectures.

#### 6.7.3 Memorization vs. Generalization

The generalization of the proposed model (mCNN11) shows the ability of the model. We have computed the generalization of the mCNN11 by varying the dataset with training percentages from 10 to 90%. We used the K10 cross-validation protocol for this purpose. Figure 6.28 represents the memorization vs. generalization of the mCNN model, and it shows that 60% of the cohort is required for memorizing the patterns, and, after that model generalizes the performance.

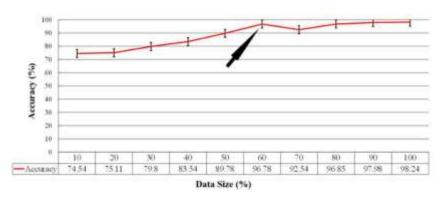


Figure 6.28: Memorization vs. generalization of mCNN11 at Augment Fold 5.

#### 6.7.4 Effect of Training Time on Data Size

We have compared the training time required for the mCNN on various cross-validation protocols. Figure 6.29 represents the training time vs. CV protocol. All the experiments

were performed in Nvidia DGX- v100 (supercomputer) at Bennett University, Gr. Noida, India. In addition, we compared the training time with a local machine I5 tenth generation processor with 24 GB Ram and 6 GB Nvidia 1660Ti. From the results, it has been observed that the training time for the CV protocol of K10 takes more time than other CV protocols.

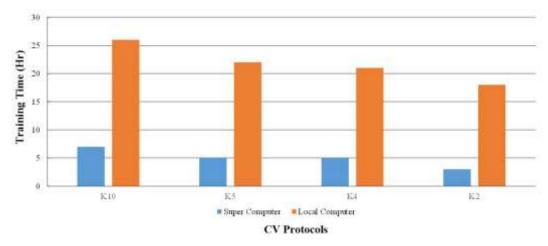


Figure 6.29: The training time of mCNN11 at augmented fold 5 with varying CV protocols.

#### 6.8 Conclusion

This study is the first of its kind. It uses four kinds of AI paradigms, namely machine learning, deep learning, transfer learning, and extreme learning on lung nodule classification and characterization. All the paradigms were optimized with varying augmentation folds. Moreover, characterization of lung nodules using AI-based MFS has taken place, and it is validated with signal processing-based Bispectrum and statistical analysis-based histogram analysis named fractal dimension analysis. Thus, it validates our hypothesis on the lung nodules. The proposed mCNN has achieved the best accuracy of  $98.75\pm1.38$  with an area-under-the-curve of 0.987 (p<0.0001). Experimentally, it has been proven that the accuracy of the proposed model is 2.5% more than the existing work, 2.5%, and 4.3% higher than the machine learning models. We have also ranked the performance of all the twenty-two models, and it has been observed that the proposed mCNN has achieved the highest rank among all the models.

The goal of multimodule detection is to detect single or multiple nodules in low-dose CT pictures, reducing the likelihood of a radiologist making a mistake. Furthermore, in the Early Lung Cancer Action Program, a CNN-based model with a 99.5 percent accuracy has been presented for prediction (ELCAP). In comparison to the proposed approach, traditional machine learning and pre-trained CNN methods are useless. The results were

compared to traditional machine learning algorithms such as SVM, k-NN, Decision Tree, LDA, Naive Bayes, and others using sensitivity, specificity, and Cohen Kappa as performance measures. The suggested model has four convolution layers and a maximum polling layer. As a consequence, it's far smaller than models like Inception V3 and VGG16 that have been pre-trained. Furthermore, since the proposed model contains fewer parameters, it takes 75 percent less time to infer than VGG16 and roughly 66 percent less time than Inception V3. Our long-term goal is to improve illness prediction accuracy on 3D CT scans while reducing storage capacity, so that this model may be used in hardware devices.

# CHAPTER 7

# LIGHT CNN MODEL FOR IDENTIFICATION OF LESION IN WILSON DISEASE USING MRI SCANS

#### 7.1 Introduction

Wilson's disease (WD) is one of the rears and it appears in one in 30,000 to 40,000 people worldwide as per the report published by National Organization for Rare Disease (NORD). Moreover, the main cause of WD is the deposition of excessive copper in the lungs and brains [12]. As per the study published in the year 2005, there are 9000 people affected by WD in the United States (US) [13]. There are many methods such as blood tests, and histology to diagnose WD but these approaches are very slow and unreliable [14, 15].

We can also diagnose the WD disease by finding the ATP7B genes in the genetic analysis [16]. Moreover, most of the methods used for WD identification were non-imaging and time-consuming processes [17]. However, early diagnosis of WD reduces the severity of the disease and increases the mortality rate [174]. Therefore, there is a need to build an image-based WD diagnosis such as MRI analysis. MRI analysis shows promising results for diagnosing WD as it shows the white matter hyperintensity (WMH) [175]. Moreover, MRI is an imaging tool for WD, but it suffers from inter-observer variability, and WMH's subtle nature between WD and control causes complications in diagnosis. Therefore, there is a need to develop an artificial intelligence-based computer-aided diagnosis system (CAD) to identify the WD.

Researchers have been applying advanced AI techniques to solve complex computer vision problems in the medical domain for the last two decades [37, 43, 176-178]. Moreover, Machine learning (ML) is a subset of AI, it shows efficiency in classification problems but it suffers from reliability issues because it takes hand-crafted features [151]. The performance of the ML model varies with the selected features. This gap was covered by deep learning (DL) techniques using automatic feature extraction. DL shows promising results in complex problems in radiology [98].

However, DL shows efficient computer vision problems, but it mainly depends on the hardware resources. Third, developing countries like India will not get high-performance computers for research scholars. On the other hand, the amount of training time required for DL models is also comparable. So, to avoid all these issues in AI, another technique is evolved: Transfer Learning (TL). It reduces the training time and dependence on hardware resources [17, 83, 179]. In the TL framework, we will use the pre-trained weights of the model (so-called pre-trained models) to retrain target variables. Due to weight transfer, the model training time is reduced and does not require any sophisticated hardware resources.

We should consider three essential characteristics in AI models: speed, accuracy, and reliability. ML techniques are speed and accuracy, DL techniques are accurate and reliable, whereas TL techniques are speed and reliability. However, it shows the same performance as DL. While characterizing the medical images, if any two characteristics are better and the third characteristic is comparable, that model can be used for characterization. So, we hypothesized that the training speed of TL is better than DL and WMH is higher in control images.

This study aims to characterize and classify the WD tissues into control and disease. WMH intensity is higher in WD rather than in control. We were validating the hypothesis using signal processing techniques. In this study, we used five TL architectures optimized on augmentation folds starting from 1 to 6x for detecting the best performance. We implemented all the architectures using the K10 cross-validation protocol (i.e., 90% training and 10% testing). We characterized the WD using novel means feature strength (MFS) and validated it with Bispectrum. We benchmarked with the existing systems and novel architecture called Modified Unet.

# 7.2 Background Literature

Since most of the diagnostics methods of WD were non-image, so AI is not given an impression on WD classification. Few of the studies are using AI on this topic. Most of the existing works are non-AI but we have to consider those studies. The last two decades of studies have been completed, where biomarkers such as serum or urine have been used to diagnose or segregate patients with WD mostly depending on their threshold ranges [12, 14, 180]. These procedures included 24-hour urine and serum laboratory testing to identify WD. Peter Ferenci *et al.* [180] proposed another diagnosis method using Kayser-Fischer

rings, neurologic symptoms, and low ceruloplasmin for the detection of WD.

There is another kind of method that exists for the detection of WD i.e laboratory-based test (blood test) and gene code detection test. Vrabelova *et al.* [181] identified WD disease using H1069Q mutation of ATP7B in European origins. Rosencrantz and Schilsky *et al.* [15] detected WD disease using mutation analysis of ATP7B, neurological disorders, Kayser-Fleischer rings, elevated urine and hepatic copper, and histological changes in the liver.

All the detection techniques earlier were non-imaging techniques, and they are time-consuming. Due to the recent advancement in MRI. We can visualize the WMH in the brain [182]. WMH was used to detect several diseases like stroke [183], wills configuration [184], and mood disorder [185]. Several studies showed that WMH is used for the detection of WD. Kim *et al.* [186] uses T2-weighted MRI images of pediatric patients and made them into three groups based on initial MRI findings and observed WMH in the caudate nucleus, globus pallidus, thalamus, midbrain, and pons.

Recent advances in AI bring detection of WD using AI architectures using fMRI. Hu *et al.* [187] did a study of WD patients with frequency-dependent channels in low-frequency function amplitude. As a result, resting-state functional magnetic resonance imaging was used to acquire MRI images. Kaden *et al.* [188] used a support vector machine and parametrized generalized learning vector quantization was used to classify WD illness (PGLVQ). They were 87.5 percent and 90.1 percent accurate, respectively.

Jing et al. [189] studied the classification of the WD disease using 30 patients' MRI scans. They used the independent component analysis technique (ICA) to extract the features, then used SVM to calculate the WD scores based on these features, achieving an AUC of 0.94. These features are fed to aberrant functional networks and obtained an accuracy of 89.4%. Agarwal et al. [48] used the MRI scans of the WD patients for classification and characterization using "improved convolutional neural networks" optimized on the augmentation folds. They achieved an accuracy of 98.28% with an AUC of 0.99. They can characterize the WD using AI-based FMS and Bispectrum.

Among all these studies it is observed that WD classification and characterization using AI have few research works. Our study is the first of its kind using the transfer learning paradigm for the characterization and classification of WD.

# 7.3 Dataset and Pre-processing

#### 7.3.1 Dataset

Between 2011 and 2015, a cohort of 46 patients' T2W-TSE MRI scans was evaluated (average age: 40.7311.3 years, equal M/F ratio) (permission was received from the Institutional Ethics Committee, Azienda Ospedaliero Universitaria (AOU), Cagliari, ITALY). The imaging investigations were carried out utilizing a 1.5 Tesla superconducting magnet with a head coil (Philips, Best, The Netherlands) and a defined technique. Diffusion-weighted imaging (DWI) was done on each participant using a single-shot spinecho with two diffusion-sensitivity settings of 0 and 1000 s/mm2 along the transverse axis. Axial and sagittal 2D FLAIR images (10000/140/2200 msec for TR/TE/TI; matrix: 512 x 512; FOV: 240 x 240 mm2; slice thickness: 5 mm) were obtained as part of our general brain methodology. Axial spin-echo T1-weighted pictures (500–600/15/2 for TR/TE/excitations) and fast spin-echo T2-weighted images (2200–3200/80–120/1,2 for TR/TE/excitations; turbo factor, 2) were also acquired with the same section thickness, in addition to FLAIR and DWI sequences.

#### 7.3.2 Data Augmentation

Our radiologist team manually categorized the raw MRI data, which was then prepared for further processing. We had an uneven number of photos in both groups since the cohort comprised 37 controls and 9 WD patients. Because each patient's MRI examination contained 12-13 slices, 458 control pictures, and 115 WD images were produced. The augmentation procedure utilizing a python "Augmentor" API was implemented in the WD class for maximum performance with an imbalanced dataset, resulting in 343 additional WD pictures. Because deep CNN (DCNN) requires a high number of pictures for effective training and performance, we raised the number of photos in both classes from 458 to 2x, 3x, 4x, and 5x, and then trained and tested the system to see which enhanced results produce ideal results. We followed the accepted technique of rotating the picture by -10 to 10 degrees randomly throughout the augmentation phase to prevent unrealistic brain MRI images. This would make it impossible to use techniques such as horizontal or vertical flipping or turning at bigger degrees.

#### 7.3.3 Pre-processing: Skull and background removal

Pre-processing is an important part of the classification process since it aids in the extraction of the ROI from MRI images. There are two crucial processes in pre-processing:

# 7.4 Methodology

To prepare for the segmented ROI, the skull area was removed, and (ii) the black backdrop was removed. The popular tolls BrainSuite [35] and VolBrain [36] were used for picture segmentation and background removal. The DICOM images were converted to nii files (nii file type is principally connected with NIfTI-1 Data Format by Neuroimaging Informatics Technology Initiative) and the grayscale pictures of the brain with the skull were obtained using the BrainSuite program. The VolBrain aids in the creation of a brain mask that may be used to eliminate the skull from MRI grayscale pictures. The data was pre-processed using BrainSuite and VolBrain, and the resulting pictures were then segmented to eliminate the backdrop. Figure 7.1 shows an example pair of pictures from a patient with WD and a control. When compared to control pictures, the WD segmented brain images displayed brighter regions (greater WMH) in the convoluted zones of the brain (as seen within yellow dotted rectangles Figure 7.1). A sample of six patients with segmented MRI images from the controlled and sick groups is shown in Figures 7.2 and 7.3.

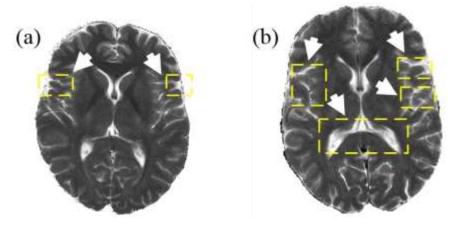


Figure 7.1: Control (left) and WD (right). Both images show the skull with the removal of the background.

# 7.4 Methodology

# 7.4.1 Global Architecture: Transfer Learning

We have developed a TL global architecture for the characterization and classification of WD. Figure 7.4 represents the global architecture of the TL framework. This study is the first of its kind to use the TL paradigm for the classification of WD. In this study, we have used four TL architectures named as AlexNet [55], ResNet [51], InceptionV3 [50], and DenseNet121 [52] which achieved the best performance in ImageNet challenge.

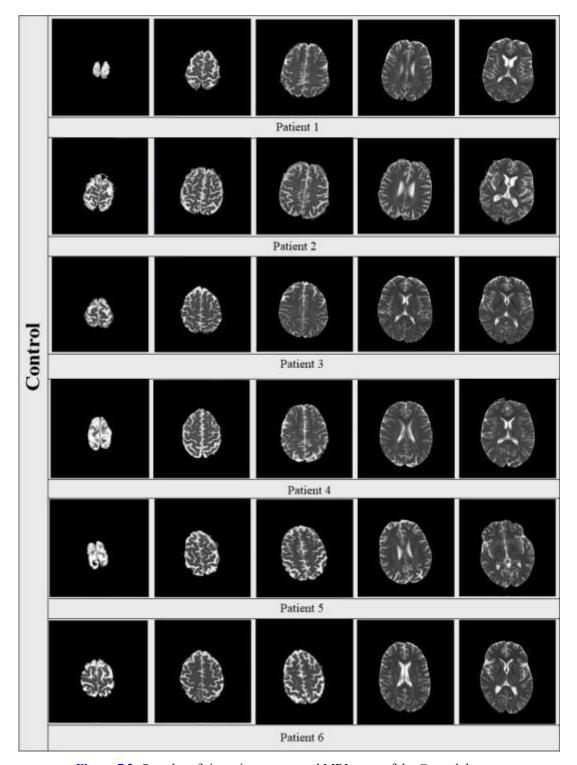


Figure 7.2: Samples of six patient segmented MRI scans of the Control dataset.

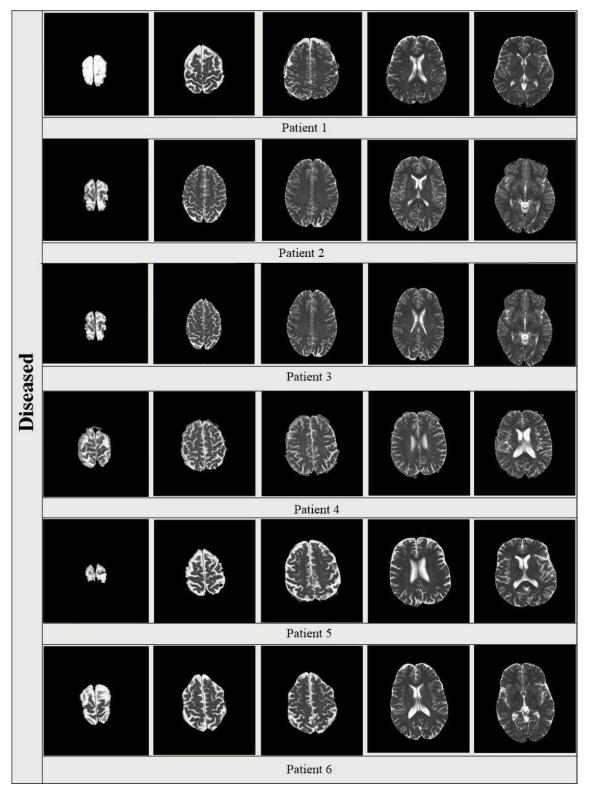


Figure 7.3: Samples of six patient segmented MRI scans of the diseased dataset.

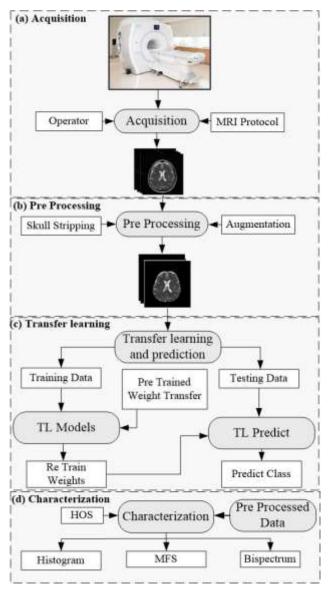


Figure 7.4: Global Architecture of our study.

#### 7.4.2 AlexNet

AlexNet was the first intelligent model designed for the classification of ImageNet using a deep convolution neural network. This architecture was proposed by Alex Krizhevsky *et al.* [55] in the year 2012 for classifying 1.2 million high-resolution images with 1000 classes at the ImageNet challenge. AlexNet architecture won the first prize in the ImageNet challenge. This architecture uses a combination of convolution and max-pooling with an input image size of 224x224 and achieves a top-5 test error rate of 15.3%. The architecture of AlexNet is pictorially represented in Figure 7.5, which as follows

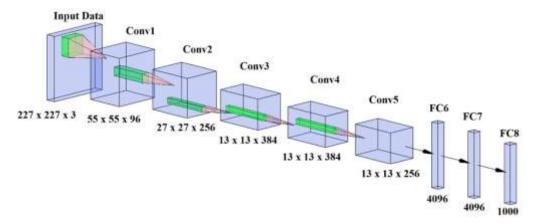


Figure 7.5: AlexNet Architecture.

In the TL paradigm, we took the pre-trained weights of the AlexNet model and sue these weights at the initial and intermediate layers. We fed the WD cohort at the Dense layer (FC6) for retraining the AlexNet for WD classification. Moreover, an extra dropout layer has been added between FC6 and FC7 with a 50% rate and the number of nodes in the output layer has been changed from 1000 to 2 as our cohort has only two classes.

#### 7.4.3 ResNet50

Kaiming He *et al.* [51] built the ResNet architecture and achieved a 3.57% error in the top 5%. This model also won the title in the challenge and the authors proposed a wide range of different ResNet architectures containing 20 layers to 1202 layers.

In our study, we have selected the ResNet50 for experimental purposes. In ResNet50, there are 50 layers with 25 residual blocks and one fully connected network (FCN) followed by a Softmax layer. We implemented this model under the TL paradigm by freezing all the layer's weights with pre-trained weights of ResNet on ImageNet. We modified the end layers by adding dropout layers after the first FCN, and then added one more FCN followed by a softmax layer. We retrained the model from the first FCN using the WD cohort. Here we also changed the number of nodes at the end layer from 1000 to 2 as the WD cohort has only two classes. The architecture of ResNet is presented in Figure 7.6.

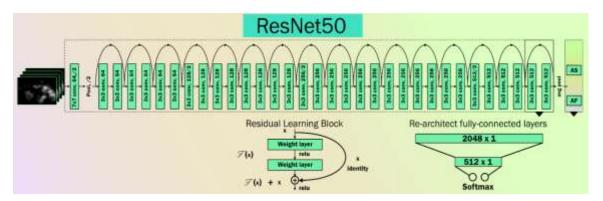


Figure 7.6: Resnet50 architecture.

#### 7.4.4 **DenseNet161**

DenseNet was another popular architecture that won the best paper award at CVPR in 2017. Gao Huang *et al.* [52] proposed densely connected convolutional neural networks for solving the vanishing gradient problem and better feature propagation with fewer training parameters. This architecture contains dense blocks which concatenate the features of the current block and previous block. Unlike ResNet, in DenseNet, all the features are fused not summating with the previous blocks. The proposed range of architectures with varying feature maps from 32 to 48 with varying depths from 121 to 161. This model achieves a 6.15% error rate in the top 5% in ImageNet 2017.

In our study, we choose DenseNet161 which consists of 161 layers and implemented it in the TL framework. To make the DenseNet fits in the TL framework, we freeze all the initial to intermediate layers with Imagenet-2017 weights. After then, we altered the end layer architecture in DenseNet 161 by adding the dropout and dense layer. The modified DenseNet architecture is presented in Figure 7.7, which is as follows:

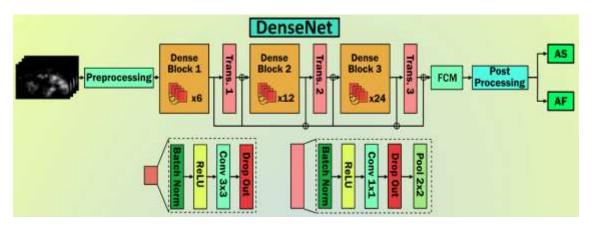


Figure 7.7: DenseNet161 architecture.

# 7.4.5 XceptionNet

XceptionNet is also one of the popular architectures developed by google for image classification and it is based on InceptionNet [54]. Moreover, XceptionNet introduces a modified depth-wise separable convolution technique for ImageNet classification and achieves an error rate of 0.790 in top-1 accuracy and 0.945 in top-5 accuracy.

In our study, we have used XceptionNet in the TL paradigm by freezing the initial layers and modifying the end layers by retraining the model using the WD cohort for classification. The modified architecture of XceptionNet has been depicted in Figure 7.8.

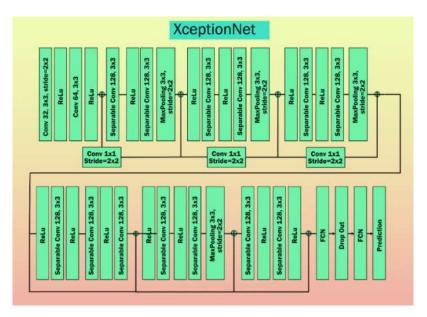


Figure 7.8: The proposed architecture of XceptionNet.

# 7.4.6 InceptionV3

Christian Szegedy *et al.* [50] have developed the InceptionV3 for solving the computer vision problems such as label smoothing and propagating the label information deep down the neural network. The author proposed a series of architectures under the inception family. Among which InceptionV3 (IV3) achieves the lowest error rate 4.2% among the Top 1%.

We considered IV3 under the TL paradigm by transferring the pre-trained weights of IV3 (initially trained on ImageNet) from initial to intermediate layers. We modified the end layers i.e from FCN to the Output layer and added two dropout layers between the FCNs with a 50% drop rate with 2 nodes in the output layer. We fed the WD cohort at FCN for retraining the IV3 model for target labels. The architecture of InceptionV3 is presented in Figure 7.9, which is as follows.

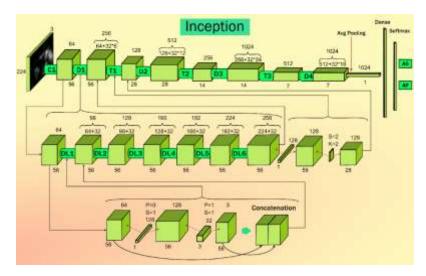


Figure 7.9: The modified architecture of InceptionV3.

#### 7.4.7 Modified Unet

The performance of proposed TL architectures has been compared with a novel DL architecture named Modified Unet. It uses the separable convolution neural network with batch and normalization combinations. After each combination of these sets of layers, we have used the max-pooling layer. Initially, we used a small number of filters i.e., 32, and then increased it to 256. The shape of the architecture is like a teacup which is presented in Figure 7.10.

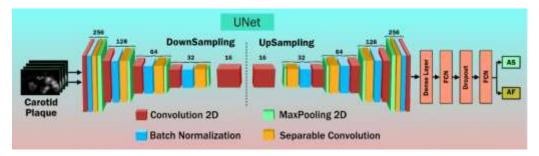


Figure 7.10: Modified Unet architecture.

# 7.5 Results and Characterization

#### 7.5.1 TL Optimization

We have optimized the performance of the TL framework using Six TL architectures with augmentation folds from 1x to 6x and achieved the best performance (in terms of accuracy) in IV3. Figure 7.11 represents the 3D optimization of the TL and Modified Unet and corresponding values are presented in Table 2. We have used the K10 cross-validation protocol (90% training, 10% testing) for all the AI architectures and achieved the best

accuracy from IV3 with 97.82±1.52% whereas the mean accuracy of 91.83±2.23%. Results are shown in Table 7.1.

We compared the TL performance with Modified Unet using the K10 cross-validation protocol and Modified Unet achieved the best accuracy of 98.43±1.45% with a mean accuracy of 90.96±4.63%. Modified Unet shows an improvement of 0.62% in the best accuracy combination with comparable mean accuracy. Moreover, we have computed the Area-under-the-curve (AUC) of all the AI architectures which are presented in Figure 7.12.

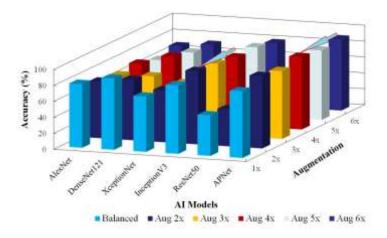


Figure 7.11: Optimization of TL and Modified Unet with Augmentation.

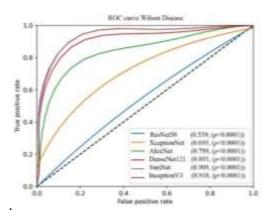


Figure 7.12: ROC analysis of six AI models.

Table 7.1: 3D-Optimization of TL architectures and Modified Unet.

Model	Aug1x	Aug 2x	Aug 3x	Aug 4x	Aug5x	Aug6x	Mean
AlexNet	79.98	71.10	67.21	69.81	62.49	68.05	71.94
DenseNet121	89.36	75.76	68.39	81.41	74.47	72.39	78.74
Xception Net	69.30	65.33	62.1	61.87	57.81	52.72	64.67
Inception V3	85.98	91.83*	89.40	85.41	86.23	79.51	88.16

ResNet 50	50.86	46.35	48.24	53.90	50.34	51.55	49.84
Modified UNet	83.57	90.79	85.23	90.96*	87.49	89.58	87.64
Mean	76.45	73.53	70.10	73.89	67.81	67.52	

#### 7.5.2 Characterization

We hypothesized that the White Matter Hyperintensity (WMH) of WD is higher than the control, to validate the proposed hypothesis we have used two strategies 1) AI-based mean feature strength, and 2) signal processing based on higher-order spectra (HOS) called Bispectrum.

# A) Mean Feature Strength (MFS)

AI-based MFS [129] is a novel technique for computing the strength of the features learned by the DL models at every layer. MFS will be calculated from DL models only because TL models take ImageNet features at initial and intermediate layers, so it is not a good option to compute the strength of the features. Therefore, the trained weight of Modified Unet has been taken for computation of MFS. Here we pass all the training images of the best accuracy combination and computed the feature strength for each image at every layer. After computation of the feature strength, the mean of all the feature strengths of all the training images at every layer has been computed. Figure 7.13 represents the MFS of the Modified Unet and also shows that the MFS of WD is higher than the control by 21.57% so, we conclude that MFS (WD) > MFS (Control).

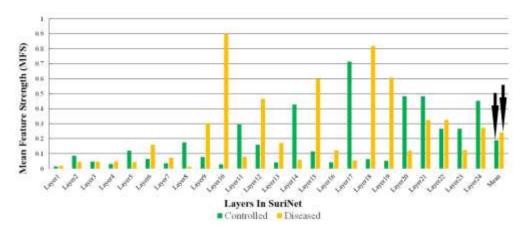


Figure 7.13: Mean Feature Strength of Modified Unet.

#### B) Higher-Order Spectrum

For validating the MFS, we choose a non-AI-based technique i.e., signal processing-based

HOS by computing Bispectrum of the same combination used for computing the MFS. Here also we computed the mean B value of every image by rotating at a random angle from 0 to 180 degrees. Figure 7.14 (a) represent the mean B value of the training images at every angle from 0 to 180 degree with a step size of 15 degrees. Figure 7.14 (b) represents the HOS of the WD at the best possible angle of 60 degrees. HOS also exhibits similar results to MFS and BiSpectrum of WD is higher than the control by 9.24%.

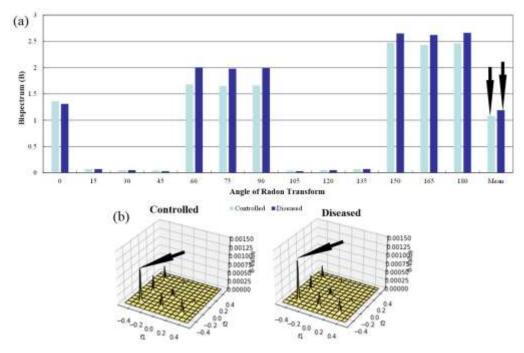


Figure 7.14: (a) Mean Bispectrum analysis (b) Higher-order spectra of the Wilson disease.

#### 7.6 Discussion

#### 7.6.1 Benchmarking

We benchmarked our current TL paradigm with existing work which is represented in Table 7.2. We found that except for the present study remaining existing works are not done any classification work. The current study is the first of its kind on TL optimization in comparison with Modified Unet for the classification of WD. The benchmarking table contains the author's name with year and its application with modality, techniques used for the application, and accuracy with area-under-the-curve (AUC) shown in the C1 to C7 columns in the below table.

Table 7.2: Benchmarking with existing systems.

	C1	C3	C4	C5	C6	C7
SN	Author, Reference, Year	Application	Modality	Technique	AI	ACC (%) (AUC)
R1	Kaden <i>et al.</i> [188] (2015)	WD	MRI	SVM and PGLVQ	ML	90.1
R2	Suk et al. [190] (2017)	ALZa	MRI (ADNI)	DBN+NN	DL	<b>85.91</b> (0.91)
R3	Abiwinanda <i>et al.</i> [191] (2018)	BT&	MRI	CNN	DL	84.19
R4	Zhang <i>et al.</i> [192] (2019)	MCI <sup>b</sup>	fMRI	SSGSR	ML	<b>88.50</b> (0.965)
R5	Abrol <i>et al.</i> [193] (2018)	MCI <sup>b</sup> vs. ALZ <sup>a</sup>	MRI	ResNet	DL	<b>82.7</b> (0.89)
R6	Jing et al. [189] (2019)	WD	fMRI	SVM and ICA	ML	<b>89.4</b> (0.94)
R7	Richhariya <i>et al.</i> [194] (2020)	MCI <sup>b</sup>	MRI	SVM	ML	90
R9	Liu et al. [195] (2020)	MCI <sup>b</sup>	MRI	SVM	ML	<b>88.5</b> (0.90)
R10	Saba et al. [17] (2020)	CNN	MRI	VGG-19	DL	95.46 (0.954)
R11	Agarwal <i>et al.</i> [48] (2020)	iDCNN	MRI	CNN	DL	98.28 (0.990)
R12	Proposed Work	CNN	MRI	Optimized Inception V3	TL	Optimized Comb: 97.82 91.83 (0.918)
R13	Proposed Work	CNN	MRI	Modified UNet	DL	Optimized Comb: 98.43 90.96 (0.91)

<sup>a</sup>ALZ: Alzheimer, <sup>&</sup>BT: Brain Tumour, <sup>b</sup>MCI: Mild Cognitive Impairment

# 7.6.2 A short note on WD characterization

We have done the characterization of WD using the AI method i.e., MFS, and the non-AI method i.e., Bispectrum. The characterization based on the AI method depends on the input data. However, we have the cohort, which is initially unbalanced, and it is having a higher controlled class. However, Bispectrum is based on pixel intensity. Both the characterization techniques are valid for our datasets and a large amount of data is required to strengthen the validation under these paradigms.

#### 7.6.3 Power Analysis

We have also performed the power analysis for finding the sample size required for this study and used equation 7.1 for the calculation of sample size.

Sample Size=
$$\frac{2 \times (Z_{\alpha} + Z_{1-\beta})^{2} \times \sigma^{2}}{\Delta^{2}}$$
 (7.1)

Here we choose  $Z_{\alpha}$ =3.2905 for type 1 error having a value of 1%, and  $Z_{1-\beta}$ =1.6449 for type II error having a value of 1%. Here,  $\sigma$  (standard deviation) =2.53 and  $\Delta$  (mean difference) =0.627. By substituting these values in equation 7.1, 793 samples are required for the proposed study. Moreover, the number of samples in the cohort was 916 under the balanced category, which is 13.42% higher than required.

#### 7.7 Conclusion

This study is the first of its kind to use the TL framework for the classification and characterization of the WD. We have achieved the best-optimized combination in IV3 with 97.82±1.52% accuracy. Moreover, we have also compared the performance of TL with a novel DL architecture called "Modified Unet". Modified Unet achieves the best accuracy of 98.43±1.45% from Augmentation 4 with K10 Fold. After classification, we are further able to do the characterization of WD using AI-based MFS and signal-based Bispectrum and concluded that WD was higher than the control by 21.57%. Using Bispectrum we obtained that WD is higher than the control by 9.24%. We validated our systems using widely accepted heart and lung data.

# **CHAPTER 8**

# COMPRESSION OF THE OPTIMIZED CNN MODEL USING GENETIC ALGORITHM

#### 8.1 Introduction

In the last decade, deep learning has been extensively explored due to its vast range of applications in various domains, such as agriculture [196], health care [188], transportation [197], and security surveillance [198]. There are two main approaches adapted in deep learning: (i) developing models from scratch and tuning hyper-parameters to get better results and optimization, and (ii) using pre-trained models via transfer learning. In the first approach, the user defines the architecture of the deep learning model and manually tunes the hyper-parameters for optimization. This process is computationally exhaustive, as the user has to design several models by changing the hyper-parameters [86].

However, in the second approach, several pre-trained architectures are available that can be utilized in different applications such as classification [56, 82]. Pre-trained architectures have provided encouraging results since they are trained on large image datasets (referred to as ImageNet) [55]. Moreover, the user can utilize these architectures by retraining them on a customized dataset for a specific number of epochs. This process is also known as transfer learning. However, all such pre-trained architectures have many convolution layers, which may increase the storage space and inference time required.

Edge computing devices such as mobile phones, Rasberry pi, and Jetson nano are used in various applications (i.e., human activity recognition, smart homes, and precision agriculture) [196]. Such devices are also used in making decisions based on collected data. However, deploying a pre-trained model on edge devices is problematic due to the limited storage and processing capability of these devices.

Genetic algorithms have been established before for the optimization of machine

learning algorithms. We hypothesize that this family of algorithms, which comes under evolutionary methods can be used for the optimization of the machine learning paradigms. Therefore, the current study describes a novel genetic algorithm (GA)-based approach that compresses the convolution neural network (CNN) without compromising the model's performance. The main advantages of GA over existing techniques are its simplicity, better compression, and acceleration results obtained from predefined datasets. Further, GA can be well adapted in situations where the problem complexity using brute force is exponential [196]. In GA-based approach found that the near-optimum solution for chromosome number (N) will be  $2^N$  (where N is the number of the chromosome). The compressing size of the model using the GA-based approach reduces the size reasonably and this compressed model can be deployed on edge computing devices with improved inference time. The main contributions of the present work are as follows:

- A novel genetic GA-based approach for compression and acceleration of the trained CNN.
- A novel method is developed representing the chromosome form of the jagged array
- A multi-objective fitness function is developed that is dependent on the number of hidden & convolution filters in convolution layers units in the fully connected network (FCN) and the accuracy of the model.
- A novel method for performing crossover between the chromosomes (represented in the form of a matrix) is developed.
- Termination criteria are set in such a way that the compression algorithm is stopped immediately once the difference between previous and current accuracy exceeds 1%.
- On data sets such as MNIST, CIFAR-10, and CIFAR-100, popular pre-trained architectures (such as AlexNet, VGG16, ResNet50, and SqueezeNet) were trained, followed by an application of the suggested compression and acceleration technique.
- The proposed approach is validated first by developing a CNN model and training it
  on an LIDC dataset. After training and testing the model, it was compressed using
  the proposed GA-based approach. The compressed model only required 1/10<sup>th</sup> of the
  space and provide a 25% improvement in inference time without compromising the
  accuracy of the model.

#### 8.1.1 Introduction to Genetic Algorithm

GA is one of the most popular and oldest meta-heuristic techniques and has been widely

used to solve various optimization problems [199]. The flowchart of GA is presented in Figure 8.1. The process of GA is started after a generation of chromosomes that represent a valid solution to the optimization problem. At the start of the GA process, chromosomes are randomly generated, some of which are selected (based on their fitness values) to generate the initial population. Many techniques are available for selecting chromosomes, such as the roulette wheel and tournament selection methods [200], any of which can be used to complete the selection process. After the selection operation is finished, the next step in GA is the crossover operation.

During the crossover operation, two-parent chromosomes are randomly selected. These chromosomes then share information to generate child chromosomes, whose "fitness" values are then evaluated. After the crossover operation, the next step is mutation. The mutation is a genetic operator that is used to maintain a population's genetic variety from one generation to the next. For performing mutations, a variety of approaches have been devised [201]. After the crossover and mutation operations, the fitness values of the parent and child chromosomes are computed; the best two chromosomes are then added to the initial population, and the other two are discarded. All the intermediate steps are repeated either for a certain number of epochs or until the termination criteria are satisfied [202].

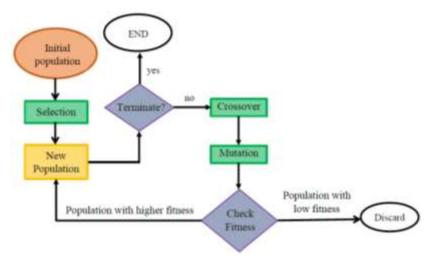


Figure 8.1: The flowchart of GA.

#### 8.1.2 Introduction to pre-trained models

In the current work, a novel GA-based approach is developed to compress and accelerate CNN models. First, the proposed methodology was applied to pre-trained models. Then, the proposed theory was applied to a customized CNN to classify lung cancer CT images.

The remainder of this section discusses the utilized pre-trained models such as AlexNet, VGG16, ResNet50, and SqueezeNet giving the reader a better understanding of these well-known CNN architectures. These pre-trained models are well known and have different varied architectural styles in terms of the depth of the model and varied convolution operations. In addition to that, these models offer better availability to interface with evolutionary methods.

# A) AlexNet

Alex Krizhevsky developed AlexNet in 2012 [55], to solve the 1000-class classification problem, often referred to as the ImageNet large-scale visual recognition challenge (ILSVRC) [55]. Krizhevsky's work won the title in the same year. The architecture of AlexNet is presented in Figure 8.2.

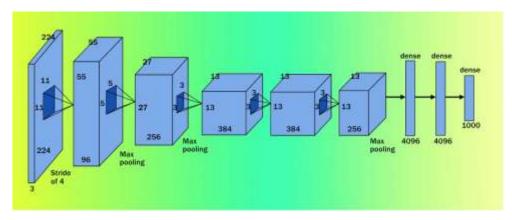


Figure 8.2: AlexNet architecture [203].

AlexNet's architecture consists of five convolution layers with three fully connected layers [144]. In each convolution layer, the RELU activation function is applied to provide the model with non-linearity. Moreover, a dropout layer is placed between the fully connected layers to reduce overfitting. The model works well for colored images with a size of 227×227. Moreover, the basic architecture of AlexNet has trained on the ImageNet [55] dataset with 1000 classes, achieving remarkable accuracy with 62.3 million trainable parameters.

#### **B)** VGG16

K. Simonyan and A. Zisserman proposed VGG16 in 2014, which is one of the most well-known CNN architectures for image classification [82]. VGG16 receives input in the form of images with a size of 224×224×3 (where '3' represents the number of channels). The architecture of VGG16 is depicted in Figure 8.3. The VGG16 has 16 layers (13 convolution

layers and three fully connected layers). In the convolution layers, the RELU activation function was applied for non-linearity, with the number of filters ranging from 64 to 512. The model was also trained on the ImageNet dataset and achieved a top-5 accuracy of 92.5%, with 533 MB of storage space.

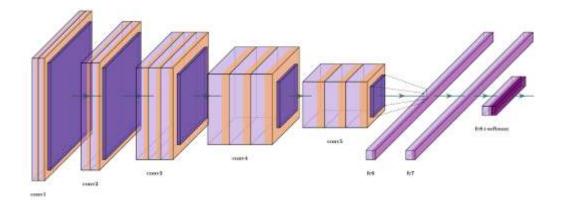


Figure 8.3: Architecture of VGG16 [204].

#### C) ResNet50

The ResNet architecture comprises five stages with convolution and identity blocks [204]. Each identity block, like each convolution block, has three convolution layers. The ResNet50 model has 23 million trainable parameters. The ResNet50 model is depicted in Figure 8.4.

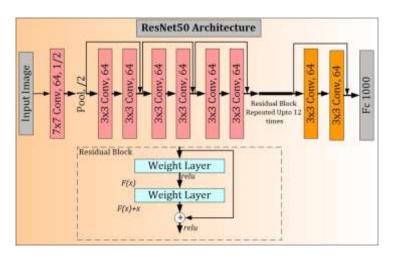


Figure 8.4: Architecture of ResNet50.

## D) SqueezeNet

The architecture of SqueezeNet is presented in Figure 8.5. SqueezeNet is a compressed version of AlexNet that has 50 times fewer parameters and 0.5 MB of storage space. In SqueezeNet, there are eight fire modules, which are responsible for reducing the floating-point operations in AlexNet [204]. Figure 8.6 is a visual representation of a fire module.

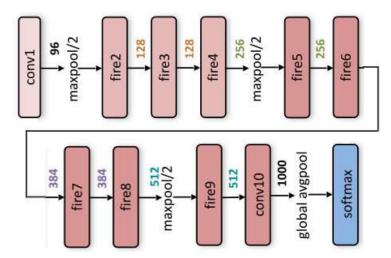


Figure 8.5: The architecture of SqueezeNet.

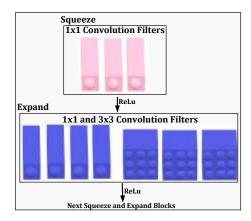


Figure 8.6: The architecture of the fire module.

# 8.2 Background Literature

Since the introduction of NVIDIA's GPU and its applications [205], Since it has considerably increased computer speed, the field of deep learning has gotten a lot of interest. Further, the importance of hardware has increased due to pre-processing steps such as filtering or noise reduction [206], which causes the cost and complexity to increase in conventional methods or knowledge-based methods. Attempts were made to speed the deep learning-based training models using supercomputer applications such as arterial imaging leading to cardiovascular risk stratification imaging [128, 129, 204] and brain imaging [48]. Initially, researchers attempted to construct deep learning models, and they were successful in obtaining exceptional accuracy in both solo and hybrid scenarios [162] and unseen AI paradigms [207] or transfer learning frameworks [204]. However, accuracy is not sufficient to judge the performance of a model and it could be even biased [207-210]. Further, these deep learning models possesses a threat to edge devices as they cannot be deployed on them, since they are low in memory and power, hence called low-powered or low-energy

devices. Here, we discuss research works by the author (s) that provided their method (s) for reducing the size of deep learning models and improving their inference time.

Gong *et al.* [73] proposed dense layers provide approximately 90% of the storage contribution in CNN models, according to researchers. As a result, the authors devised techniques for lowering the number of nodes in the dense layer based on (a) matrix factorization and (b) vector quantization. Matrix factorization decreases the picture size without sacrificing data, but vector quantization is a lossy reduction. The authors discovered that they could reduce the model's size by 16-24% without sacrificing 1% of its accuracy. Cheng *et al.* [211] discussed a CNN compression approach based on pruning. During pruning, redundant weights were discarded, allowing the size of the model to be reduced significantly. Redundant weights are those that are having low priority and do not affect the performance of the model. This type of pruning requires efficient hardware. On AlexNet, the best compression ratio was 4.94 (compression ratio was defined as the ratio of original model parameters divided by compressed model parameters).

For compressing the CNN model, Han et al. [77] devised a three-stage technique. The authors began by identifying all weights with values less than a predetermined threshold. The weight-sharing process was then started using a K-means clustering technique. Finally, to save storage space, the authors used the Huffman coding method. The authors were able to compress a 1000-class data set 35 and 49 times using the AlexNet and VGG16 models, respectively. During model training, Luo et al. [75] presented "Thinet," an effective model compression framework that removed a set of filters whose absolute sum of weights was smaller than a threshold value. When applied to ResNet50, this method lowered parameters by 66 percent while increasing accuracy by 4.46 percent. Anwar et al. [76] devised a pruning-based model compression approach that would speed up the procedure. Evolutionary particle filtering and activation sum voting are two strategies outlined by the authors for selecting the best candidates for pruning. The authors showed that by utilizing CIFAR-10, they were able to prune 70% of the network while only sacrificing 1% of the accuracy. For pre-trained models like GoogleNet and ResNet, He et al. [212] suggested three techniques to prune pre-trained CNNs (single layer pruning, full network pruning, and multi-branch pruning).

Meanwhile, Han et al. [213] devised a network pruning procedure in which all

convolution and dense layer weights below a certain threshold are set to zero. Li *et al.* [78] suggested a compression method that includes pruning convolution layer filters with little influence on model correctness. The authors used the L1-norm to rank the filters, and low-ranking filters were deleted from each layer. Liu *et al.* [79] developed a network slimming strategy that uses sparsity-based regularisation to prune irrelevant channels during training by computing the channel sparsity and its scaling factor around zero. Choudhary *et al.* [80] gave a review of prior research on model compression and acceleration strategies, as well as their experiments on AlexNet utilizing MNIST and CIFAR-10 datasets. From the structural, algorithmic, and implementation viewpoints, Zhang *et al.* [81] investigated numerous strategies for CNN compression and acceleration. Layer decomposition and network pruning may be used to decrease duplication in CNN at the structural level, according to the authors. In addition, the fast Fourier transform (FFT) was employed to accelerate the model at the algorithm level.

Chen et al. [214] presented four strategies for model compression: Pruning, quantization, and compression, 2) matrix factorization and filtering to speed up matrix multiplication, 3) knowledge distillation and transfer learning, and 4) hybrid models that combine two or more of the other three models. Cheng et al. [74] presented four more model compression techniques: 1) model pruning through quantization and binarization, 2) low-rank matrix factorization and filtering to speed up matrix multiplication, 3) knowledge distillation and transfer learning, and 4) hybrid models that combine any two of the other models. Another four model compression methods were described by Cheng et al. [74] 1) model pruning using quantization and binarization, 2) low-rank matrix factorization using DCT, and wavelet transforms, 3) transferred/compact filters, and 4) training a compact network by distilling knowledge from a larger network. On the MNIST dataset, Yang et al. [215] showed the use of a multi-objective fitness function for LeNet compression. By reducing the number of parameters for breast CT scan pictures, Samala et al. [216] were able to compress AlexNet using GA by 34%. On a plant leaf image dataset, Aggarwal et al. [217] used meta-heuristic techniques to compress common segmentation architectures, particularly SegNet and UNet, achieving compression of 25x and 15x, respectively.

# 8.3 Methodology

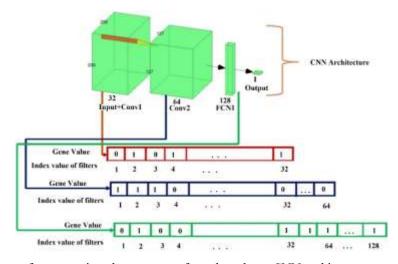
This section describes the proposed GA-based method for compressing and accelerating the CNN model, including appropriate examples. The GA process started with chromosome

representation, followed by the generation of the initial population. After that, other intermediate operations of GA, such as selection, crossover, and mutation, are explained using examples.

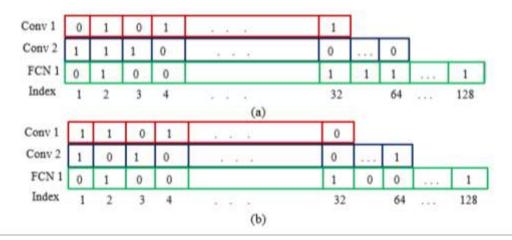
#### **8.3.1** Chromosome representation

We represent chromosomes in the form of a jagged array (i.e., a matrix with different lengths of columns), and the number of columns in each row has a varying length. The first row of the matrix represents the first convolution layer, while the last fully connected layer is represented by the last row of the matrix. Meanwhile, the columns represent the number of filters present in corresponding layers. The matrix is randomly filled with binary values (i.e., 0 or 1). As shown in the CNN architecture presented in Figure 8.7, there are two convolution layers and one fully connected layer with 32, 64, and 128 hidden units, followed by the output layer. The input layer is combined with the first convolution layer and takes a 256×256 color image as the input. Figure 8.7 also illustrates the process by which chromosomes are generated from a specific architecture.

Two randomly generated chromosomes are presented in Figures 8.8 (a) and (b). We used algorithm 1 to generate random chromosomes. In the present work, 100 chromosomes were generated; the accuracies achieved by these chromosomes are presented in Figure 8.9 to understand that the randomly generated have diversified nature.



**Figure 8.7:** Process for generating chromosomes for a three-layer CNN architecture comprising two CNN and FCN layers with 32, 64, and 128 hidden units, respectively.



**Figure 8.8:** Two randomly generated chromosomes (a) and (b) for the CNN architecture are presented in Figure 8.7

#### 8.3.2 Fitness Function

Our objective in the present work is to compress and accelerate the CNN architecture by minimizing the number of hidden units without compromising accuracy. To this end, we developed a fitness function that considers both parameters i.e., the number of hidden units, and accuracy.

#### Objective 1: Minimising the number of nodes in the model

Our goal in the current work is to develop a CNN model that is faster (in terms of inference time) and consumes less storage space than previous models. The main cause for large storage space and high inference times in CNN models is the massive number of hidden units in the CNN architecture. A large number of hidden units in a model cause it to incur many floating-point operations, which increase the computational time. Similarly, this situation increases the storage space, as the model must store information about hidden units in terms of weight and bias. Therefore, we can minimize the storage and inference time by minimizing the number of hidden units. As such, our first objective is represented as follows (equation 8.1)

# Algorithm 1 Generation of chromosome

# Input

8:

9:

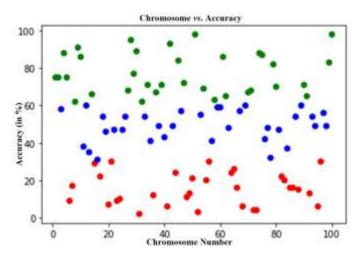
- 1. K ▷ Represents the convolution or FCN Layers in CNN architecture
- 2.  $F[p] > Represents the number of hidden units at <math>p^{th}$  CNN or FCN layer
- 3. N ▶ Denotes the number of chromosomes

Output : ChroMat[][][]
$$\triangleright$$
 List of chromosomes1: procedure GENCHROMOSOME(K,F[p], N)2:  $A \leftarrow \{\};$  $\triangleright$  To make A a empty list3:  $p=1$  $\triangleright$  Initialization of p4: for  $i \leftarrow 1$  to N do $\triangleright$  Loop to generate N chromosomes5: for  $j \leftarrow 1$  to K do $\triangleright$  Selection of  $j^{th}$  layer6: for  $k \leftarrow 1$  to  $F[p]$  do $\triangleright$  Selection of  $k^{th}$  hidden unit7:  $A[j][k] = \text{random}() \% 2$  $\triangleright$  Assignment of binary values

ChroMat  $[i][j][k] = A [j][k] \Rightarrow Assigning the binary value <math>k^{th}$ 

hidden unit at j $^{th}$  hidden layer in i $^{th}$  chromosome

p=p+1



**Figure 8.9:** The fitness values of randomly generated chromosomes (green color dots indicate the accuracy of chromosome greater than 60%, blue color dots indicate chromosome accuracy greater than 40%, and less than 40% represented as red color).

Objective 1=minimization of 
$$(\frac{\delta}{\gamma})$$
 (8.1)

Where  $\gamma$  represents the number of hidden units in the network and  $\delta$  is the selected number of hidden units after applying the proposed methodology.

## Objective 2: Maximising accuracy

We have selected increasing accuracy as another objective because minimizing the number of hidden units could remove nodes that have important information; thus, the performance of the model could suffer. By considering accuracy in the fitness function, we hope to maintain the original model's accuracy in the compressed model. The second objective is represented by equation 8.2. Accuracy is calculated using equation 8.3.

$$\eta = \frac{\sum Fc}{\sum n}$$
 (8.3)

Here ' $\eta$ ' is an accuracy, 'Fc' represents full corrected, and 'n' is the number of samples. As there are two objectives, a weight (w) has been assigned to establish a trade-off between these them. The fitness function used in the proposed compression method is presented in equations 8.3 and 8.4.

Fitness Value=maximize 
$$\left[\left(1-\frac{\delta}{\gamma}\right)\times w+(1-w)\times\eta\right]$$
 (8.3)

such that:

$$0 \le w \le 1; \delta \le \gamma \tag{8.4}$$

The constraints of the fitness function are provided in equation 8.4, in which w lies between 0 and 1. Selecting a value closer to 1 gives more weight to accuracy. For the current experimental work, the value of w has been set to 0.1. Moreover, the value of  $\delta$  is always less than or equal to  $\gamma$ , which represents that the number of hidden units selected is always less than or equal to the number of hidden units in the uncompressed model.

#### 8.3.3 Selection

Selection is one of the intermediate operations in GA. Using a selection operation, we identified which chromosomes should be passed on to the next operation (e.g., crossover and mutation). There are various methods for selection (i.e., tournament, roulette wheel, stochastic universal sampling, rank selection). We have used the roulette wheel selection method in the proposed work.

#### 8.3.4 Crossover

Crossover is another intermediate step in GA. In this step, two-parent chromosomes share information and generate two child chromosomes. In the proposed work, we used a one-point crossover operation. The point at which the crossover operation is conducted was

chosen randomly with a constraint in place dictating that the index value of the crossover point cannot be less than the minimum number of hidden units in any layer. An example of a crossover operation is presented in Figure 8.10.

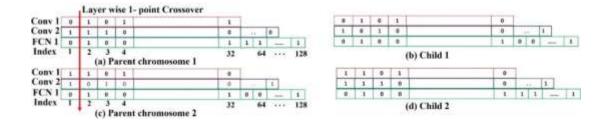


Figure 8.10: Crossover operation.

The experiments were also performed using recombination, by which a random number was generated for each gene position and compared with a probability value (0.6). If the random number was greater than the probability value, then the parent 1 gene was used at that position; otherwise, the parent 2 gene was used. The results were almost the same when this type of crossover was employed.

#### 8.3.5 Mutation

Genes were altered at random in this stage. The program chose a gene at random and reset it to 0 if the gene value was 1. This indicates that the node corresponding to that gene did not participate in the training when the model was being run. One of the current project's goals is to use mutation to minimize the number of nodes in the model. To meet this objective, 5% of the genes were randomly flipped from 1 to 0. Thus, the mutation process generated an entirely new type of chromosome after each iteration, which helped us search a very vast search space to find a global optimum instead of a local optimum that suits only a specific type of chromosome. The various parameters used in the Genetic Algorithm are given in Table 8.1.

SNo	Parameter	Description
1	Gene value	Binary (0-1)
2	No. of initials chromosomes	100
3	Selection criteria	Roullete wheel
4	Crossover	1-point/ Recombination
5	Mutation	Flipping 1-bit
6	Termination Criteria	Change in fitness < 0.000001

Table 8.1: Parameters used for Genetic Algorithm.

The GA-based strategy suggested in this study outperformed previous compression

algorithms significantly. Furthermore, we regarded precision to be a critical feature in GA's fitness function, hence chromosomes with poor fitness values were removed. Furthermore, inverted mutations were applied, which transformed gene values from 1 to 0, reducing the number of hidden units to a minimum. The fitness values were reviewed after mutation, if the fitness score was lower than the prior best, it was also deleted. The goal of both the crossover and mutation processes was to reduce the number of concealed units while keeping the performance assessment measure intact.

#### 8.4 Results

Python was used to run the tests on an NVIDIA DGX v100 computer, on UBUNTU 20.0.1 operating system, this computer has 40600 CUDA cores, 5120 tensor cores, 128 GB RAM, and 1000 TFLOPS performance. VGG16, SquezNet, ResNet50, AlexNet, and customized CNN architectures were used in the experiments. We used MNIST [218], CIFAR-10 [219], CIFAR-100 [219], and LIDC-IDRI [28] datasets.

#### 8.4.1 Datasets

The MNIST dataset contains 70,000 black-and-white photographs of handwritten digits in ten classifications. The CIFAR-10 dataset additionally includes 60,000 color photos of items divided into 10 classifications. The CIFAR-100 dataset, meanwhile, has 100 classifications and 60,000 color photos of objects. Furthermore, appropriate padding has been used to fit the input size of the various pre-trained models. We have also applied the proposed approach to compress different CNN architectures trained on a publicly available cancer dataset (i.e., LIDC-IDRI). The LIDC-IDRI dataset contains 1018 helical CT scans of the lung with nodule sizes between 3 and 30 mm. Each slice was examined and annotated by four expert radiologists.

In this study, we considered 905 helical CT scans, of which 422 were malignant and 483 were benign. We discarded slices with higher malignancy ratings (three or more out of five). This is because this work focused on the early detection of lung cancer; as such, scans with higher ratings were discarded as they are prone to malignancy. Sample images of LIDC-IDRI datasets are presented in Figure 8.11.

## 8.4.2 Storage space comparison

On the MNIST and CIFAR-10 datasets, the initial model sizes of AlexNet, VGG16, ResNet50, and Squeezenet were 122.08 MB, 63.73 MB, 104.88 MB, and 2.98 MB, respectively. AlexNet, VGG16, ResNet50, and Squeezenet have model sizes of 122.44 MB,

64.09 MB, 105.25 MB, and 3.16 MB, respectively, on the CIFAR-100 dataset. The GA-based technique was used for comparison purposes, to minimize changes in accuracy. On the trained model, we ran the suggested compression strategy for 10 iterations and reported the accuracy for each iteration. Figure 8.12 gives the change of accuracy concerning epochs.

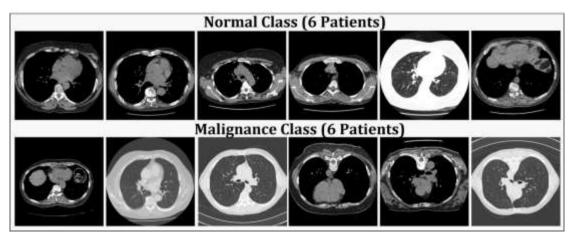
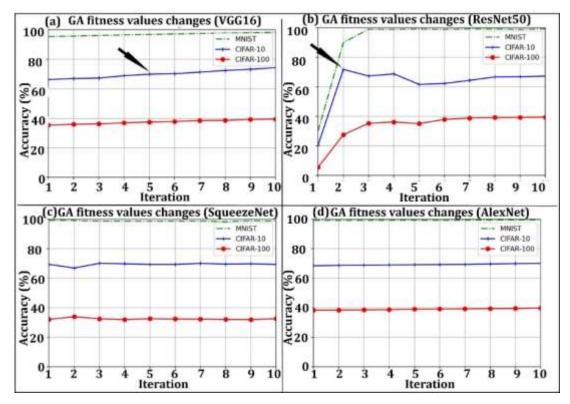
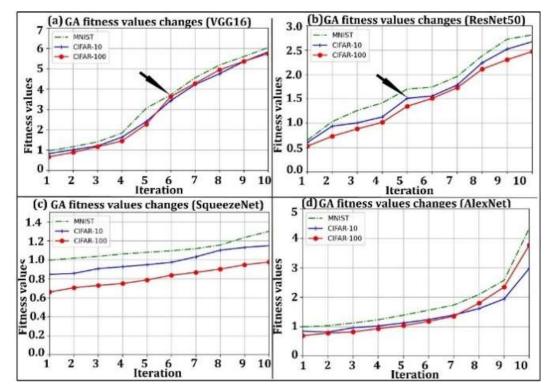


Figure 8.11: Sample images of LIDC- IDRI datasets for four parents.

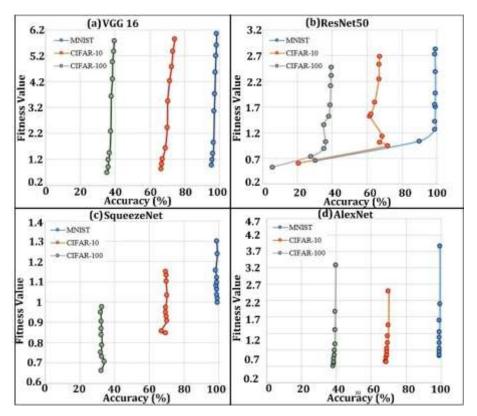
Figure 8.13 & 8.14 (a) and (b) clearly shows that accuracy improved as the model was compressed. Due to the excessive number of hidden units, the model may enter an overfitting state; therefore, after reducing the number of hidden units, the models are in their optimal states. Figure 8.13 & 8.14 (c) and (d) illustrate that accuracy did not change throughout the compression process. The main reason for this is that the fitness function depends on the accuracy of the model; if the accuracy drops below a certain threshold, the proposed model discarded that chromosome (solution).



**Figure 8.12:** Variations in accuracy (%) values while executing GA for 10 epochs over different datasets (i.e., CIFAR-100, CIFAR-10 and MNIST) for (a) VGG16 (b) ResNet50 (c) SqueezeNet (d) AlexNet.



**Figure 8.13**: Variations in fitness values while executing GA for 10 epochs over different datasets (i.e., CIFAR-100, CIFAR-10 and MNIST) for (a) VGG16 (b) ResNet50 (c) SqueezeNet (d) AlexNet.



**Figure 8.14:** Variations in fitness values vs. Accuracy while executing GA for 10 epochs over different datasets with weight(w) is equal to 0.99 (i.e., CIFAR-100, CIFAR-10 and MNIST) for (a) VGG16 (b) ResNet50 (c) SqueezeNet (d) AlexNet.

Previously, many researchers have applied different techniques (e.g., pruning, slimming) to compress CNN models. The performance statistics of the proposed algorithm and other compression techniques are provided in Tables 8.2, 8.3, and 8.4. Meanwhile, Table 8.5 presents the changes in accuracy that occurred during compression in all four models using the proposed approach. Equation 8.5 was used to compute the percentage decrease in space.

$$Dr(\%) = \frac{K-M}{K} \times 100$$
 (8.5)

Here, Dr(%) represents the percentage of decrease, K represents the original size of the CNN model and M represents the size of the compressed model. The experiments were also performed with crossover in the form of recombination. The compression results for AlexNet, SqueezeNet, ResNet50, and VGG16 on different datasets are compared in Table 8.5. Since the results were nearly the same, it was presumed that the two methods could be used interchangeably.

Table 8.2: Comparison of statistics with different models in terms of size on MNIST dataset.

Sn#	Model Name	Size (in MB) after	Percent
SII#	Wodel Name	compression	decrease
1	Pyramid Structure on AlexNet [220]	73.253	40%
2	Magnitude based pruning on AlexNet [80, 213]	47.138	61.39%
3	L1-Norm pruning on AlexNet [78, 80]	34.295	71.90%
4	N/W slimming on AlexNet [79, 80]	23.421	80.81%
	Proposed models		
5	AlexNet+GA	15.112	87.62%
6	SqueezeNet+GA	1.848	38.11%
7	ResNet50+GA	22.587	78.46%
8	VGG16+GA	5.421	91.49%

## 8.4.3 Inference time comparison

The suggested technique not only compresses the model but also greatly decreases the inference time. The compressed model has fewer floating-point operations, which is only feasible by lowering the number of filters and nodes, resulting in a faster inference time. The performance statistics in terms of inference time are shown in Tables 8.5, 8.6, and 8.7. Figure 8.15 shows the activation values for the VGG16 model when applied to the MNIST dataset. Because the ReLU activation function had converted the negative weights to 0, several of the filters didn't contain any information.

Table 8.3: Comparison of statistics with different models in terms of size on the CIFAR-10 dataset.

Sn#	Model Name	Size (in MB) after compression	Percent decrease			
1	Pyramid Structure on AlexNet [220]	73.253	40%			
2	Magnitude based pruning on AlexNet [80, 213]	46.777	61.68%			
3	L1-Norm pruning on AlexNet [78, 80]	34.295	71.90%			
4	N/W slimming on AlexNet [79, 80]	23.255	80.95%			
5	Channel pruning on ResNet50 [221]	26.327	74.9%			
6	Pruning Filters on ResNet50 [78]	90.519	13.7%			
7	NISP on ResNet50 [222]	59.524	43.25%			
8	CNN-FCF on ResNet50 [223]	31.739	69.74%			
9	Channel pruning on VGG16 [221]	10.19	84%			
10	Pruning Filters on VGG16 [78]	22.943	64%			
	Proposed models					
11	AlexNet+GA	23.233	80.97%			
12	SqueezeNet+GA	1.838	38.44%			

13	ResNet50+GA	22.399	78.64%
14	VGG16+GA	5.879	90.77%

Table 8.4: Comparison of statistics with different models in terms of size on the CIFAR-100 dataset.

Sn#	Model Name	Size (in MB) after compression	Percent decrease
1	Channel pruning on ResNet50 [221]	37.890	64%
2	Channel pruning on VGG16 [221]	22.56	64.8%
	Proposed mod	dels	
3	AlexNet+GA	16.891	86.20%
4	SqueezeNet+GA	1.942	38.66%
5	ResNet50+GA	22.978	78.16%
6	VGG16+GA	5.631	91.21%

## 8.4.4 Compression of pre-trained architectures trained on the LIDC dataset

To validate the proposed method and its efficacy, we used it to compress pre-trained architectures trained on the LIDC-IDRI dataset. In the current experimental work, we trained four pre-trained models (i.e., VGG16, SqueezeNet, ResNet50, and AlexNet) on a lung dataset. The numbers of hidden units before and after compression for all four pre-trained models are presented in Figure 8.15.

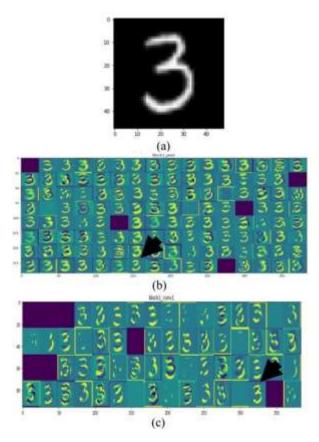
Table 8.5: Comparison of accuracy change on compression.

C #	Madalasassa	Original	Final Accuracy	Percent		
Sn#	Model name	Accuracy (%)	(%)	Compression		
		MINIST da	taset			
1	AlexNet+GA	99.32	99.14	87.62%		
2	SqueezeNet+GA	99.40	98.81	38.11%		
3	ResNet50+GA	30.20	99.33	78.46%		
4	VGG16+GA	97.33	99.44	91.49%		
	CIFAR-10 dataset					
5	AlexNet+GA	70.58	69.31	80.97%		
6	SqueezeNet+GA	69.49	69.44	38.44%		
7	ResNet50+GA	20.64	66.72	78.64%		
8	VGG16+GA	66.03	79.67	90.77%		
		CIFAR-100 d	lataset			
9	AlexNet+GA	40.02	36.34	86.20%		
10	SqueezeNet+GA	32.19	32.36	38.66%		
11	ResNet50+GA	5.59	38.89	78.16%		
12	VGG16+GA	40.04	40.62	91.21%		

## 8.4.5 Compression CNN trained on Medical Dataset (LIDC-IDRI)

We classified lung cancer by developing different CNN models by varying the hyper-

parameters presented in Table 8.9. After performing several experiments, we observed that the CNN model with nine layers (including convolution, average pooling, drop out and fully connected layers) was the best fit for the LIDC-IDRI dataset.



**Figure 8.15:** (a) Input image, (b) activation before compression, and (c) activation after compression in the first set of the convolution layer in VGG16 on the MNIST dataset.

In this model, the batch size value is 64 and the number of hidden units in different convolutions layers are 32, 16, and 8. We also included a fully connected layer with 128 hidden units. Furthermore, every convolution layer is followed by the average polling layer with two drop-out layers to prevent overfitting. We also used a Relu activation function in hidden layers to introduce non-linearity into the model with a drop-out value of 0.4.

We applied the proposed algorithm to compress the nine-layer CNN network on the lung dataset by considering a specific value of w (= 0.99). We observed that the space required to store the compressed model was approximately 1/10 of that required to store the original model. Moreover, we observed a 25% improvement in inference time after compression. The layer-wise filter values, along with other statistics, are presented in Figures 8.16, 8.17, and Table 8.10.

Table 8.5: Performance comparison in terms of inference time after compression on MNIST dataset.

Sn# Model name		Original time (In sec)	Final time (In sec)	Improvement in time
		Proposed mod	` ′	III tille
			C1	
1	AlexNet+GA	37.37	24.45	34.57%
2	SqueezeNet+GA	373.38	338.78	9.26%
3	ResNet50+GA	1668	437.21	73.78%
4	VGG16+GA	113.47	22.12	80.50%

**Table 8.6:** Performance comparison in terms of inference time after compression on the CIFAR-10 dataset.

Sn#	Model name	Original time	Final time	Improvement
311#	wiouer name	(In sec)	(In sec)	in time
1	Adaptive batch size speedup on	256.59	218.97	14.66%
1	RedNet20 [56]	230.39	218.97	14.00/0
2	Pruning Filters on ResNet50 [31]	1.31	0.99	24.42%
3	Channel pruning on VGG16 [56]	NG#	NG#	56.20%
	Propos	sed models		
4	AlexNet+GA	39.13	25.23	35.52%
5	SqueezeNet+GA	374.02	339.78	9.15%
6	ResNet50+GA	1669.12	437.89	73.76%
7	VGG16+GA	114	23.56	79.33%

#NG: Not Given

**Table 8.7:** Performance comparison in terms of inference time after compression on the CIFAR-100 dataset.

Sn#	Model name	Original time	Final time	Improvement
3II#	wiodei name	(in a sec)	(in a sec)	in time
1	Channel pruning on VGG16 [56]	NG#	NG#	45.30%
2	Pruning Filters on VGG16 [31]	NG#	NG#	34.20%
3	Channel pruning on VGG16 [56]	1.23	0.73	40.70%
4	ADABATCH [59]	66.24	44.34	33.01%
	Propo	sed models		
5	AlexNet+GA	35.11	22.21	36.74%
6	SqueezeNet+GA	373.89	339.12	9.29%
7	ResNet50+GA	1669.59	437.82	73.76%
8	VGG16+GA	113.84	22.78	79.98%

#NG: Not Given

**Table 8.8:** Comparison of static with different models and datasets for comparison using 1-point crossover and recombination crossover.

			Size (in MB) after	Size (in MB)
Sn#	Model Name	Dataset	compression	after
			(1-point crossover)	compression

				(recombination)
1	AlexNet+GA	MNIST	15.112	16.133
2	AlexNet+GA	CIFAR-10	23.223	22.184
3	AlexNet+GA	CIFAR-100	16.891	15.343
4	SqueezeNet+GA	MNIST	1.848	1.845
5	SqueezeNet+GA	CIFAR-10	1.838	1.854
6	SqueezeNet+GA	CIFAR-100	1.942	1.912
7	ResNet50+GA	MNIST	22.587	22.671
8	ResNet50+GA	CIFAR-10	22.399	22.581
9	ResNet50+GA	CIFAR-100	22.978	22.994
10	VGG16+GA	MNIST	5.421	5.601
11	VGG16+GA	CIFAR-10	5.879	5.913
12	VGG16+GA	CIFAR-100	5.631	6.071

**Table 8.9:** Set of hyper parameters.

Sn#	Hyper parameter	Range
1	Batch size	16-64
2	Learning rate	$10^{-5} - 10^{-3}$
3	Number of layers	5-13
4	Hidden units	8-64
5	Activation function	[Relu , Sigmoid, Tanh]
6	Drop out	10% - 70%

**Table 8.10:** Comparison of 9-layer customized CNN model before and after compression.

Sn#	Parameter	Value		
		Before compression	After compression	
1	Accuracy (in %)	99.14	99.14	
2	Storage space (in MB)	4148	459	
3	Inference time (in sec)	6.075	4.55	

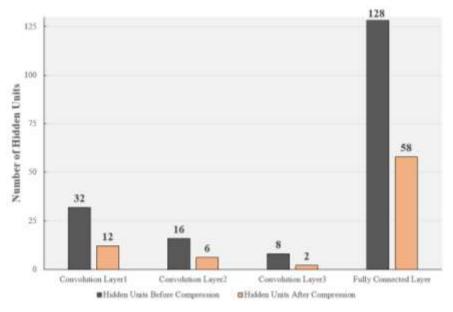
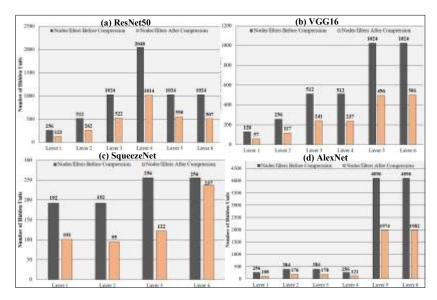


Figure 8.16: Hidden units before and after compression in a customized CNN on the LIDCIDRI dataset.



**Figure 8.17:** Hidden units before and after in (a) ResNet50, (b) VGG16, (c) SqueezeNet, and (d) AlexNet on the LIDC-IDRI dataset.

#### 8.5 Discussion

#### 8.5.1 Principal Finding

We demonstrated a strategy that uses optimized GA, which is used for compressing the deep convolution neural network (DCNN) models without compromising performance. Here, we used a novel method such as a jagged array for the representation of chromosomes and the same way of chromosome representation has been used to perform the other intermediate operations i.e., crossover and mutation. The GA identifies filters/nodes which do not significantly contribute to decision-making using CNN. After eliminating these useless hidden units and filters there is a significant improvement in the performance of the model in terms of storage and inference time. Reduction in the storage space for AlexNet was 87%, 86.55%, and 86.16% corresponding to the data sets MNIST, CIFAR-10, and CIFAR-100, respectively. Further, for VGG16, ResNet50, and SqueezeNet, the system was compressed by around 91%, 78%, and 38%, respectively. The inference time, our proposed strategy significantly improved by 35%, 9%, 73%, and 80%, respectively for the models AlexNet, SqueezeNet, ResNet50, and VGG16. In the LIDC-IDRI CT lung cohort, the compressed model's storage space and inference time showed a drastic improvement, in size by reducing by 4148KB to 459KB, whereas the inference time increases by 25%.

The convergence to a local optimum was tested by repeating the experiment with initial random chromosomes and the final compression results were found to be nearly the same. Thus, it shows that from any direction process of compression was eventually

reaching a similar compression solution which was near-optimal. Also, GA optimization processes are known to give not the exact optimal solution but near optimum.

#### 8.5.2 A special note on the Compression DL model

The utilization of the edge devices will reach almost 125 to 500 billion IoT devices by 2030 out of which 20% of devices would have cameras. The deep learning applications on edge devices are catching up slowly, due to the size and time constraints. So, there is a need for pruning the DL model's size. Compression of DL models offers the reduction of the size and inference time, which in turn helps in the utilization of the edge devices. Pruning offers to trim the redundant or not properly learned nodes in the DL architecture, so it reduces the trainable parameters in the models. If the trainable parameters reduce, the size of the model will be reduced. We ran the proposed system on different operating systems, Table 8.11 gives the inference time of the two testing images prediction.

**Environments Specification** VGG16 AlexNet SqueezeNet ResNet50 Intel XenonES-2698 v4, RAM NVIDIA 3.34 3.64 2.54 7.5 128 GB, Nvidia DGX128 GB Intel Core i5-9750H CPU, HP Desktop pc 3.54 4.15 3.49 11.14 RAM 16 GB, UHD Graphics Quad core Cortex-A72 (ARM Raspberry-pi 6.34 8.23 6.29 15.34 v8), 4GB SDRAM Apple A11 Bionic (10 nm),

8.19

10.32

7.94

18.82

Table 8.11: Inference time (in sec) of prediction on different OS.

*Note:* these values are for inference time of predicting two images.

Hexa-core 2.39 GHz, 3GB

RAM, 64 GB ROM

# 8.6 Conclusion

iPhone

This study described a novel GA-based method for compressing a trained CNN. All intermediate GA steps (i.e., chromosome representation, crossover, and mutation) were discussed, and examples were provided. Accuracy and the number of hidden units were utilized as the evaluation criteria for the GA-based approach. The fitness function of the approach caused the compression process to stop if too great of a change in the evaluation criteria occurred.

In the current context, CNN compression is crucial since hefty models cannot be stored on edge devices owing to their limited capacity. In this study, we used the suggested method on pre-trained models and used the results to verify the modified CNN architecture. For both types of CNNs investigated in this work, the suggested technique yielded

# 8.6 Conclusion

favorable results. We also demonstrated how to expand the suggested strategy to condense the object detection and segmentation model.

# **CHAPTER 9**

# **CONCLUSIONS AND FUTURE WORKS**

#### 9.1 Conclusion

The main objective of the thesis was to classify and characterize the lesions in non-invasive medical imaging such as US, MRI, and CT by applying deep learning techniques. There are many studies available in the literature review on ML and few studies are available on DL. Although CNN was applied for classification but most of the studies in the medical field for characterization of the lesions.

In Chapter 2, we proposed a 3D optimized lightweight CNN model used for the identification of the lesion at the carotid arteries. We have two types of lesions such as symptomatic and asymptomatic collected from the st. mary's hospital, London. We augmented the dataset by balancing it and then we make the folds up to 6x and we ran our proposed models using K10 cross-validation (90% training and 10% testing) for all the experimentations, we achieved an accuracy of 95.66%, and 0.956 (p-value<0.0001) respectively. The proposed model shows an improvement of 7.01% over previous ML models and a 12.05% improvement from the previous studies. We confirmed our hypothesis that symptomatic plaque is heterogeneous, dark (echolucent), and patchy by computing two novel strategies, one is based on DL named as mean feature strength and the other one is Bispectrum based on higher-order spectra. Further, our system performance was evaluated and validated with DOR, power analysis, memorization vs. generalization, ROC analysis, and Kappa analysis. These models are also validated using the most widely used benchmarking data (Facial data, cats, and dogs). Our proposed model was also tested on the local machines for prediction time analysis, we achieved a prediction time of less than 2 sec.

In Chapter 3, We used Multicenter data collected at two geographical locations namely London and Lisbon. We confirmed our hypothesis, using two datasets and validated

it with another dataset by mixing both cohorts. When implementing mean feature strength and Bispectrum, six types of artificial intelligence models were tested, which include four machine learning (ML) systems, one transfer learning (TL) system, and one deep learning (DL) system. Based on the suggested model, all three data sets, namely London, Lisbon, and Combined (London + Lisbon), demonstrated consistent and stable findings. Using the K10 partition technique, the mean accuracies and mean AUCs had the same order DL>TL>ML, with values of 0.938, 0.946, and 0.889 (p<0.0001), respectively. On lesion characterization in all three data sets, the mean feature strength for the symptomatic plaque was higher than asymptomatic plaque by 46.56%, 19.40%, and 53.84%, respectively, thus validating our assumptions. We benchmarked our proposed model with the existing methods, it shows an improvement of 10.41% for London and 3.32% for Lisbon. We evaluated and validated the performance of the models using DOR, power analysis, Cohen, and kappa analysis. All the results were consistent with the previous studies.

In Chapter 4, We developed a CAD system based on the transfer learning paradigm. We proposed Eleven AI models out of which ten are transfer learning models and one novel modified Unet model for classification and characterization. We achieved the best accuracy with MobileNet and its AUC of 0.961~(p<0.0001) and Modified Unet's AUC of 0.927~(p<0.0001). Grayscale median, fractal dimension, higher-order spectra, and spatial heatmaps were used to verify the performance. With deep learning, TL performed similarly to deep learning. We built a CAD system based on a deep learning system that generates the risk of the lesion and heatmaps. We validated our study with correlation analysis and compared it with the ground truth of the medical practitioners. Our proposed model exhibits showed a performance improvement of **12.9**% than previous existing systems.

In Chapter 5, We proposed a new kind of hybrid DL model on US multicenter paradigm by fusion of two TL models and another kind of model is a fusion of DL and ML. We tested the HDL models' performance on three different types of states and used three different loss functions to confirm the hypothesis (unbalanced, balanced, and optimized). We also demonstrated the system architecture utilizing three cohorts in a multicenter scenario. To characterize the HDL models, we employed the mean feature strength. The findings reveal that the HDL models perform much better than the DL models in our prior research. HDL>TL>DL (99.78>96.5>88 (in percent) and 0.996, 0.965, 0.88 (p0.0001), respectively, were the order of accuracy and AUC in all three cohorts. The performance of the HDL models was 6.4 percent and 3.2 percent higher than the SDL and TL models,

respectively. DOR, Cohen, and Kappa statistics were used to verify the HDL models' performance. It takes less than two seconds for the online system to function.

In Chapter 6, We used Lung CT scans for the classification and characterization of the benign nodule lesions and malignance. Unlike the previous study here the case of characterization is reverse i.e the strength of the healthy lesions is higher than the malignance. In lung nodule classification and characterization, we employed four AI paradigms: machine learning, deep learning, transfer learning, and extreme learning. With varied augmentation folds, all the paradigms were optimized. Furthermore, lung nodules have been characterized using AI-based MFS, which has been confirmed using signal processing-based Bispectrum and statistical analysis-based histogram analysis known as fractal dimension analysis. As a result, it supports our idea about lung nodules. With an area-under-the-curve of 0.987 (p<0.0001), the suggested mCNN attained the highest accuracy of 98.75±1.38. Experiments have shown that the suggested model is 2.5 percent more accurate than previous work, 2.5 percent is more accurate than machine learning models, and 4.3 percent more accurate than machine learning models. We also rated the performance of all twenty-two models, and the suggested mCNN received the highest score of all the models. We also implemented another CAD system for the classification of the multi nodule with sizes less than 3mm to 50 mm. In the Early Lung Cancer Action Program, a CNN-based model with a 99.5 percent accuracy has been presented for prediction purposes (ELCAP). Traditional machine learning and pre-trained CNN techniques have been demonstrated to be ineffective in comparison to the suggested methodology. The findings were compared to classic machine learning algorithms which exhibit superior performance to ML. Pre-trained models like Inception V3 and VGG16 are substantially larger than the suggested model. Furthermore, since the proposed model contains fewer parameters, it takes 75 percent less time to infer than VGG16 and 66 percent less time than Inception V3.

In Chapter 7, We shifted our paradigm to MRI medical scans of Wilson's disease. In this case, also the hypothesis reverses in comparison with the US. This is the first research of its sort to employ the TL framework to categorize and characterize the WD. With 97.82+1.52 percent accuracy, we obtained the best-optimized combination in IV3. The performance of TL was compared to that of a new DL architecture termed "Modified Unet." Augmentation 4-Fold obtains the highest accuracy of 98.43±1.45 percent. We were also able to describe the WD using signal-based Bispectrum and AI-based MFS. We discovered

that WD was 21.57 percent greater than the control. We found that WD was 9.24 percent greater than the control using Bispectrum. We used well-acknowledged cardiac and lung statistics to verify our systems.

In Chapter 8, We used a meta-heuristic approach for the compression of the proposed CNN models. A unique GA-based strategy for compressing a trained CNN was reported in this study. All intermediate GA phases (chromosome representation, crossover, and mutation) were covered, with examples. The GA-based strategy was evaluated using accuracy and the number of concealed units as the criterion. If there was too much of a change in the assessment criteria, the approach's fitness function forced the compression process to cease. In the current context, CNN compression is crucial since hefty models cannot be stored on edge devices owing to their limited capacity. In this study, we used the suggested method on pre-trained models and used the results to verify the modified CNN architecture. For both types of CNNs studied in this study, the suggested technique yielded positive results. We also demonstrated how to expand the suggested strategy to condense the object detection and segmentation model.

#### 9.2 Future Work

The findings of this research can be applied to other medical imaging techniques such as PET, X-Ray, Echo Cardiograms, and Nuclear Medicine. On other hand, we can extend this technique to invasive techniques such as mammograms, in-vivo, ex-vivo, and biopsy imaging. The cohorts we used were of a low to moderate size. Thus, there is a need for similar investigations using a large database within a big data framework. To make deep learning solutions compatible with these small edge devices, suggested CNN compression approaches may be used. The presented method may be expanded to compress additional well-known pre-trained models such as MobileNet and Inception-V3 while keeping the performance assessment measure intact. This study might be expanded in the future to include real-time field robots for autonomous identification in the area of radiology.

# **BIBLIOGRAPHY**

- [1] G. M. Cooper and R. E. Hausman, "The development and causes of cancer," *The cell: A Molecular Approach*, vol. 2, 2000.
- [2] P. Anand, A. B. Kunnumakara, C. Sundaram, K. B. Harikumar, S. T. Tharakan, O. S. Lai, *et al.*, "Cancer is a preventable disease that requires major lifestyle changes," *Pharmaceutical Research*, vol. 25, pp. 2097-2116, 2008.
- [3] P. Libby, "Inflammation and cardiovascular disease mechanisms," *The American Journal of Clinical Nutrition*, vol. 83, pp. 456S-460S, 2006.
- [4] W. C. Meijers and R. A. de Boer, "Common risk factors for heart failure and cancer," *Cardiovascular Research*, vol. 115, pp. 844-853, 2019.
- [5] P. Libby, P. M. Ridker, and A. Maseri, "Inflammation and atherosclerosis," *Circulation*, vol. 105, pp. 1135-1143, 2002.
- [6] P. Libby and D. I. Simon, "Inflammation and thrombosis: the clot thickens," ed: American Heart Association, 2001.
- [7] M. Ueda, "ATIS," *Drugs*, vol. 70, pp. 3-8, 2010.
- [8] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, pp. 7-34, 2019.
- [9] A. C. Society:, "Cancer facts & figures " 2016.
- [10] C. Bain, D. Feskanich, F. E. Speizer, M. Thun, E. Hertzmark, B. A. Rosner, *et al.*, "Lung cancer rates in men and women with comparable histories of smoking," *Journal of the National Cancer Institute*, vol. 96, pp. 826-834, 2004.
- [11] R. Battista and S. Grover, "Early detection of cancer: an overview," *Annual Review of Public Health*, vol. 9, pp. 21-45, 1988.
- [12] P. Dusek, T. Litwin, and A. Czlonkowska, "Wilson disease and other neurodegenerations with metal accumulations," *Neurologic Clinics*, vol. 33, pp. 175-204, 2015.
- [13] G. J. Brewer and F. K. Askari, "Wilson's disease: clinical management and therapy," *Journal of Hepatology*, vol. 42, pp. S13-S21, 2005.
- [14] V. Medici, L. Rossaro, and G. Sturniolo, "Wilson disease—a practical approach to diagnosis, treatment and follow-up," *Digestive and Liver Disease*, vol. 39, pp. 601-609, 2007.

- [15] R. Rosencrantz and M. Schilsky, "Wilson disease: pathogenesis and clinical considerations in diagnosis and treatment," in *Seminars in Liver Disease*, 2011, pp. 245-259.
- [16] L. Saba, A. Tiwari, M. Biswas, S. Gupta, E. Godia-Cuadrado, A. Chaturvedi, *et al.*, "Wilson's disease: A new perspective review on its genetics, diagnosis and treatment," *Frontiers in Bioscience (Elite edition)*, vol. 11, pp. 166-185, 2019.
- [17] L. Saba, M. Agarwal, S. Sanagala, S. Gupta, G. Sinha, A. Johri, *et al.*, "Brain MRI-based Wilson disease tissue classification: an optimised deep transfer learning approach," *Electronics Letters*, 2020.
- [18] O. Salvado, C. Hillenbrand, J. Suri, and D. L. Wilson, "MR coil sensitivity inhomogeneity correction for plaque characterization in carotid arteries," in *Medical Imaging 2004: Image Processing*, 2004, pp. 2114-2121.
- [19] C. R. Crawford and K. F. King, "Computed tomography scanning with simultaneous patient translation," *Medical Physics*, vol. 17, pp. 967-982, 1990.
- [20] L. Saba, G. M. Argioas, P. Lucatelli, F. Lavra, J. S. Suri, and M. Wintermark, "Variation of degree of stenosis quantification using different energy level with dual energy CT scanner," *Neuroradiology*, vol. 61, pp. 285-291, 2019.
- [21] H. Baradaran, C. R. Ng, A. Gupta, N. M. Noor, K. W. Al-Dasuqi, E. E. Mtui, et al., "Extracranial internal carotid artery calcium volume measurement using computer tomography," *International Angiology: A Journal of the International Union of Angiology*, vol. 36, pp. 445-461, 2017.
- [22] L. Saba, G. M. Argiolas, E. Raz, S. Sannia, J. S. Suri, P. Siotto, *et al.*, "Carotid artery dissection on non-contrast CT: Does color improve the diagnostic confidence?," *European Journal of Radiology*, vol. 83, pp. 2288-2293, 2014.
- [23] J. S. Suri, C. Kathuria, and F. Molinari, *Atherosclerosis disease management*: Springer Science & Business Media, 2010.
- [24] T. S. Hatsukami and C. Yuan, "MRI in the early identification and classification of high-risk atherosclerotic carotid plaques," *Imaging in Medicine*, vol. 2, p. 63, 2010.
- [25] H. E. Tinsley and D. J. Weiss, "Interrater reliability and agreement," in *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, ed: Elsevier, 2000, pp. 95-124.
- [26] U. R. Acharya, S. V. Sree, M. M. R. Krishnan, F. Molinari, L. Saba, S. Y. S. Ho, *et al.*, "Atherosclerotic risk stratification strategy for carotid arteries using texture-based features," *Ultrasound in Medicine & Biology*, vol. 38, pp. 899-915, 2012.
- [27] U. Acharya, S. V. Sree, M. Mookiah, L. Saba, H. Gao, G. Mallarini, et al., "Computed tomography carotid wall plaque characterization using a combination of discrete wavelet transform and texture features: A pilot study," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, vol. 227, pp. 643-654, 2013.

- [28] F. Han, G. Zhang, H. Wang, B. Song, H. Lu, D. Zhao, *et al.*, "A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database," in *2013 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, 2013, pp. 14-18.
- [29] C. I. Christodoulou, E. Kyriacou, M. S. Pattichis, C. S. Pattichis, and A. Nicolaides, "A comparative study of morphological and other texture features for the characterization of atherosclerotic carotid plaques," in *International Conference on Computer Analysis of Images and Patterns (CAIP)*, 2003, pp. 503-511.
- [30] S. Lennartz, A. Mager, N. Grosse Hokamp, S. Schäfer, D. Zopfs, D. Maintz, *et al.*, "Texture analysis of iodine maps and conventional images for k-nearest neighbor classification of benign and metastatic lung nodules," *Cancer Imaging*, vol. 21, pp. 1-10, 2021.
- [31] A. AlKubeyyer, M. M. Ben Ismail, O. Bchir, and M. Alkubeyyer, "Automatic detection of the meningioma tumor firmness in MRI images," *Journal of X-Ray Science and Technology*, vol. 28, pp. 659-682, 2020.
- [32] B. Fei, J. S. Suri, and D. L. Wilson, "Three-Dimensional Volume Registration of Carotid MR Images," *Studies in Health Technology and Informatics*, vol. 113, pp. 394-411, 2005.
- [33] C. Christodoulou, C. Pattichis, E. Kyriacou, and A. Nicolaides, "Image retrieval and classification of carotid plaque ultrasound images," *The Open Cardiovascular Imaging Journal*, vol. 2, 2010.
- [34] E. Kyriacou, P. Vogazianos, C. Christodoulou, C. Loizou, A. S. Panayides, S. Petroudi, *et al.*, "Prediction of the time period of stroke based on ultrasound image analysis of initially asymptomatic carotid plaques," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 334-337.
- [35] J. Seabra, L. M. Pedro, J. F. e Fernandes, and J. Sanches, "Ultrasonographic characterization and identification of symptomatic carotid plaques," in 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (EMBS), 2010, pp. 6110-6113.
- [36] J. C. Seabra, F. Ciompi, O. Pujol, J. Mauri, P. Radeva, and J. Sanches, "Rayleigh mixture model for plaque characterization in intravascular ultrasound," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 1314-1324, 2011.
- [37] F. Molinari, U. Raghavendra, A. Gudigar, K. M. Meiburger, and U. R. Acharya, "An efficient data mining framework for the characterization of symptomatic and asymptomatic carotid plaque using bidimensional empirical mode decomposition technique," *Medical & Biological Engineering & Computing*, vol. 56, pp. 1579-1593, 2018.
- [38] E. P. V. Le, N. Evans, J. Tarkin, M. Chowdhury, F. Zaccagna, H. Pavey, *et al.*, "105 Machine learning and carotid artery CT radiomics identify significant differences

- between culprit and non-culprit lesions in patients with stroke and transient ischaemic attack," ed: BMJ Publishing Group Ltd and British Cardiovascular Society, 2020.
- [39] M.-H. Roy-Cardinal, F. Destrempes, G. Soulez, and G. Cloutier, "Assessment of carotid artery plaque components with machine learning classification using homodyned-K parametric maps and elastograms," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, pp. 493-504, 2018.
- [40] C. He, Z. Li, J. Wang, Y. Huang, Y. Yin, and Z. Li, "Atherosclerotic Plaque Tissue Characterization: An OCT-Based Machine Learning Algorithm With ex vivo Validation," *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020.
- [41] S. Peng, Y. Liu, W. Lv, L. Liu, Q. Zhou, H. Yang, *et al.*, "Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study," *The Lancet Digital Health*, vol. 3, pp. e250-e259, 2021.
- [42] Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, *et al.*, "A deep learning system for differential diagnosis of skin diseases," *Nature Medicine*, vol. 26, pp. 900-908, 2020.
- [43] M. Biswas, V. Kuppili, D. R. Edla, H. S. Suri, L. Saba, R. T. Marinhoe, *et al.*, "Symtosis: A Liver Ultrasound Tissue Characterization and Risk Stratification in Optimized Deep Learning Paradigm," *Computer Methods and Programs in Biomedicine*, 2017.
- [44] W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, *et al.*, "Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study," *The Lancet Oncology*, vol. 21, pp. 233-241, 2020.
- [45] U. R. Acharya, F. Molinari, S. V. Sree, G. Swapna, L. Saba, S. Guerriero, *et al.*, "Ovarian tissue characterization in ultrasound: a review," *Technology in Cancer Research & Treatment*, vol. 14, pp. 251-261, 2015.
- [46] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," presented at the The International Conference on Learning Representations (ICLR), 2014.
- [48] M. Agarwal, L. Saba, S. K. Gupta, A. M. Johri, N. N. Khanna, S. Mavrogeni, *et al.*, "Wilson disease tissue classification and characterization using seven artificial intelligence models embedded with 3D optimization paradigm on a weak training brain magnetic resonance imaging datasets: a supercomputer application," *Medical & Biological Engineering & Computing*, pp. 1-23, 2021.
- [49] K. Purohit, A. Kesarwani, D. R. Kisku, and M. Dalui, "COVID-19 Detection on Chest X-Ray and CT Scan Images Using Multi-image Augmented Deep Learning Model," *BioRxiv*, p. 2020.07.15.205567, 2020.

- [50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818-2826.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.
- [52] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700-4708.
- [53] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [54] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2017, pp. 1251-1258.
- [55] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012, pp. 1097-1105.
- [56] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," arXiv preprint arXiv:1602.07360, 2016.
- [57] W. B. Winn, U. P. Schmiedl, D. D. Reichenbach, K. W. Beach, H. Nghiem, C. Dimas, *et al.*, "Detection and characterization of atherosclerotic fibrous caps with T2-weighted MR," *American Journal of Neuroradiology*, vol. 19, pp. 129-134, 1998.
- [58] K. Murata, N. Murata, B. Chu, H. Watase, D. S. Hippe, N. Balu, *et al.*, "Characterization of carotid atherosclerotic plaques using 3-dimensional MERGE magnetic resonance imaging and correlation with stroke risk factors," *Stroke*, vol. 51, pp. 475-480, 2020.
- [59] Q. Zhang, H. Qiao, J. Dou, B. Sui, X. Zhao, Z. Chen, *et al.*, "Plaque components segmentation in carotid artery on simultaneous non-contrast angiography and intraplaque hemorrhage imaging using machine learning," *Magnetic Resonance Imaging*, vol. 60, pp. 93-100, 2019.
- [60] G. Zhu, Y. Li, V. Ding, B. Jiang, R. L. Ball, F. Rodriguez, *et al.*, "Semiautomated characterization of carotid artery plaque features from computed tomography angiography to predict atherosclerotic cardiovascular disease risk score," *Journal of Computer Assisted Tomography*, vol. 43, pp. 452-459, 2019.
- [61] S. Arora, J. Chien, S.-C. Cheng, K. Chun, and M. Wintermark, "Optimal carotid artery coverage for carotid plaque CT-imaging in predicting ischemic stroke," *Journal of Neuroradiology*, vol. 37, pp. 98-103, 2010.

- [62] U. Acharya, S. Vinitha Sree, M. Mookiah, R. Yantri, F. Molinari, W. Zieleźnik, et al., "Diagnosis of Hashimoto's thyroiditis in ultrasound using tissue characterization and pixel classification," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in medicine, vol. 227, pp. 788-798, 2013.
- [63] U. R. Acharya, O. Faust, S. V. Sree, F. Molinari, L. Saba, A. Nicolaides, *et al.*, "An accurate and generalized approach to plaque characterization in 346 carotid ultrasound scans," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 1045-1053, 2012.
- [64] U. R. Acharya, S. V. Sree, M. M. R. Krishnan, L. Saba, F. Molinari, S. Guerriero, et al., "Ovarian tumor characterization using 3D ultrasound," *Technology in Cancer Research & Treatment*, vol. 11, pp. 543-552, 2012.
- [65] U. R. Acharya, S. V. Sree, F. Molinari, L. Saba, A. Nicolaides, and J. S. Suri, "An automated technique for carotid far wall classification using grayscale features and wall thickness variability," *Journal of Clinical Ultrasound*, vol. 43, pp. 302-311, 2015.
- [66] S. K. Banchhor, N. D. Londhe, T. Araki, L. Saba, P. Radeva, N. N. Khanna, *et al.*, "Calcium detection, its quantification, and grayscale morphology-based risk stratification using machine learning in multimodality big data coronary and carotid scans: a review," *Computers in Biology and Medicine*, vol. 101, pp. 184-198, 2018.
- [67] U. R. Acharya, O. Faust, S. V. Sree, F. Molinari, L. Saba, A. Nicolaides, *et al.*, "An accurate and generalized approach to plaque characterization in 346 carotid ultrasound scans," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 1045-1053, 2011.
- [68] J. C. Seabra, L. M. Pedro, J. F. e Fernandes, and J. M. Sanches, "A 3-D ultrasound-based framework to characterize the echo morphology of carotid plaques," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 1442-1453, 2009.
- [69] J. C. Seabra, J. Sanches, L. M. Pedro, and J. e Fernandes, "Carotid plaque 3d compound imaging and echo-morphology analysis: a bayesian approach," in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 2007, pp. 763-766.
- [70] D. Afonso, J. Seabra, J. S. Suri, and J. M. Sanches, "A CAD system for atherosclerotic plaque assessment," in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 2012, pp. 1008-1011.
- [71] C. I. Christodoulou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides, "Texture-based classification of atherosclerotic carotid plaques," *IEEE transactions on medical imaging*, vol. 22, pp. 902-912, 2003.
- [72] L. Saba, S. S. Sanagala, S. K. Gupta, V. K. Koppula, A. M. Johri, N. N. Khanna, *et al.*, "Multimodality carotid plaque tissue characterization and classification in the artificial intelligence paradigm: a narrative review for stroke application," *Annals*

- of Translational Medicine, vol. 9, 2021.
- [73] Y. Gong, L. Liu, M. Yang, and L. Bourdev, "Compressing deep convolutional networks using vector quantization," *arXiv* preprint arXiv:1412.6115, 2014.
- [74] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "Model compression and acceleration for deep neural networks: The principles, progress, and challenges," *IEEE Signal Processing Magazine*, vol. 35, pp. 126-136, 2018.
- [75] J.-H. Luo, J. Wu, and W. Lin, "Thinet: A filter level pruning method for deep neural network compression," in *Proceedings of the IEEE international conference on computer vision (ICCV)*, 2017, pp. 5058-5066.
- [76] S. Anwar, K. Hwang, and W. Sung, "Structured pruning of deep convolutional neural networks," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 13, pp. 1-18, 2017.
- [77] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," *arXiv* preprint arXiv:1510.00149, 2015.
- [78] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf, "Pruning filters for efficient convnets," *arXiv preprint arXiv:1608.08710*, 2016.
- [79] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, "Learning efficient convolutional networks through network slimming," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2736-2744.
- [80] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A comprehensive survey on model compression and acceleration," *Artificial Intelligence Review*, vol. 53, pp. 5113-5155, 2020.
- [81] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, "Recent advances in convolutional neural network acceleration," *Neurocomputing*, vol. 323, pp. 37-51, 2019.
- [82] M. Mateen, J. Wen, S. Song, and Z. Huang, "Fundus image classification using VGG-19 architecture with PCA and SVD," *Symmetry*, vol. 11, p. 1, 2019.
- [83] C. Lam, D. Yi, M. Guo, and T. Lindsey, "Automated Detection of Diabetic Retinopathy using Deep Learning," *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, vol. 2017, pp. 147-155, 2018.
- [84] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-rays," *Ingenierie et recherche biomedicale : IRBM = Biomedical engineering and research*, p. 10.1016/j.irbm.2020.07.001, 2020.
- [85] M. Biswas, V. Kuppili, T. Araki, D. R. Edla, E. C. Godia, L. Saba, *et al.*, "Deep learning strategy for accurate carotid intima-media thickness measurement: An

- ultrasound study on Japanese diabetic cohort," *Computers in Biology and Medicine*, vol. 98, pp. 100-117, May 12 2018.
- [86] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*: MIT press, 2016.
- [87] M. Agarwal, L. Saba, S. K. Gupta, A. Carriero, Z. Falaschi, A. Paschè, *et al.*, "A Novel Block Imaging Technique Using Nine Artificial Intelligence Models for COVID-19 Disease Classification, Characterization and Severity Measurement in Lung Computed Tomography Scans on an Italian Cohort," *Journal of Medical Systems*, vol. 45, pp. 1-30, 2021.
- [88] A. N. Nicolaides, S. K. Kakkos, E. Kyriacou, M. Griffin, M. Sabetai, D. J. Thomas, *et al.*, "Asymptomatic internal carotid artery stenosis and cerebrovascular risk stratification," *Journal of Vascular Surgery*, vol. 52, pp. 1486-1496.e5, 2010/12/01/2010.
- [89] E. J. Benjamin, P. Muntner, and M. S. Bittencourt, "Heart disease and stroke statistics-2019 update: a report from the American Heart Association," *Circulation*, vol. 139, pp. e56-e528, 2019.
- [90] S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, *et al.*, "Heart disease and stroke statistics-2020 update: a report from the American Heart Association," *Circulation*, vol. 141, pp. e139-e596, 2020.
- [91] S. K. Kakkos, M. B. Griffin, A. N. Nicolaides, E. Kyriacou, M. M. Sabetai, T. Tegos, *et al.*, "The size of juxtaluminal hypoechoic area in ultrasound images of asymptomatic carotid plaques predicts the occurrence of stroke," *Journal of Vascular Surgery*, vol. 57, pp. 609-618. e1, 2013.
- [92] L. Saba, H. Gao, E. Raz, S. V. Sree, L. Mannelli, N. Tallapally, *et al.*, "Semiautomated analysis of carotid artery wall thickness in MRI," *Journal of Magnetic Resonance Imaging*, vol. 39, pp. 1457-1467, 2014.
- [93] L. Saba and J. S. Suri, *Multi-Detector CT imaging: principles, head, neck, and vascular systems* vol. 1: CRC Press, 2013.
- [94] J. Seabra and J. Sanches, "Ultrasound Imaging: Advances and Applications," ed: New York: Springer, 2012.
- [95] N. D. Londhe and J. S. Suri, "Superharmonic imaging for medical ultrasound: a review," *Journal of medical systems*, vol. 40, p. 279, 2016.
- [96] M. A. Hussain, G. Saposnik, S. Raju, K. Salata, M. Mamdani, J. V. Tu, *et al.*, "Association between statin use and cardiovascular events after carotid artery revascularization," *Journal of the American Heart Association*, vol. 7, p. e009745, 2018.
- [97] L. Saba, N. Ikeda, M. Deidda, T. Araki, F. Molinari, K. M. Meiburger, *et al.*, "Association of automated carotid IMT measurement and HbA1c in Japanese patients with coronary artery disease," *Diabetes Research and Clinical Practice*, vol. 100, pp. 348-353, 2013.

- [98] L. Saba, M. Biswas, V. Kuppili, E. C. Godia, H. S. Suri, D. R. Edla, *et al.*, "The present and future of deep learning in radiology," *European Journal of Radiology*, 2019.
- [99] N. N. Khanna, A. D. Jamthikar, D. Gupta, M. Piga, L. Saba, C. Carcassi, *et al.*, "Rheumatoid arthritis: atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning–based tissue characterization," *Current atherosclerosis reports*, vol. 21, p. 7, 2019.
- [100] X. Huang, Y. Zhang, L. Meng, D. Abbott, M. Qian, K. K. Wong, *et al.*, "Evaluation of carotid plaque echogenicity based on the integral of the cumulative probability distribution using gray-scale ultrasound images," *Plos one*, vol. 12, p. e0185261, 2017.
- [101] K. Lekadir, A. Galimzianova, A. Betriu, M. del Mar Vila, L. Igual, D. L. Rubin, *et al.*, "A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 48-55, 2017.
- [102] A. N. Nicolaides, S. K. Kakkos, M. Griffin, G. Geroulakos, and E. Bashardi, "Ultrasound plaque characterisation, genetic markers and risks," *Pathophysiology of Haemostasis and Thrombosis*, vol. 32, p. 371, 2002.
- [103] A. N. Nicolaides, S. K. Kakkos, M. Griffin, M. Sabetai, S. Dhanjil, D. J. Thomas, *et al.*, "Effect of image normalization on carotid plaque classification and the risk of ipsilateral hemispheric ischemic events: results from the asymptomatic carotid stenosis and risk of stroke study," *Vascular*, vol. 13, pp. 211-221, 2005.
- [104] B. Liu, R. Tang, Y. Chen, J. Yu, H. Guo, and Y. Zhang, "Feature generation by convolutional neural network for click-through rate prediction," in *The World Wide Web Conference (WWW)*, 2019, pp. 1119-1129.
- [105] F. Yasar and F. Akgunlu, "Fractal dimension and lacunarity analysis of dental radiographs," *Dentomaxillofacial Radiology*, vol. 34, pp. 261-267, 2005.
- [106] C. P. Loizou, M. Pantziaris, C. S. Pattichis, and E. Kyriakou, "M-mode state based identification in ultrasound videos of the atherosclerotic carotid plaque," in 2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP), 2010, pp. 1-6.
- [107] C. P. Loizou, C. S. Pattichis, M. Pantziaris, E. Kyriacou, and A. Nicolaides, "Texture feature variability in ultrasound video of the atherosclerotic carotid plaque," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, pp. 1-9, 2017.
- [108] R. Doonan, A. Dawson, E. Kyriacou, A. Nicolaides, M. Corriveau, O. Steinmetz, et al., "Association of ultrasonic texture and echodensity features between sides in patients with bilateral carotid atherosclerosis," *European Journal of Vascular and Endovascular Surgery*, vol. 46, pp. 299-305, 2013.
- [109] R. U. Acharya, O. Faust, A. P. C. Alvin, S. V. Sree, F. Molinari, L. Saba, et al.,

- "Symptomatic vs. asymptomatic plaque classification in carotid ultrasound," *Journal of Medical Systems*, vol. 36, pp. 1861-1871, 2012.
- [110] U. R. Acharya, F. Molinari, L. Saba, A. Nicolaides, S. Shafique, and J. S. Suri, "Carotid ultrasound symptomatology using atherosclerotic plaque characterization: a class of Atheromatic systems," in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 2012, pp. 3199-3202.
- [111] U. R. Acharya, O. Faust, A. Alvin, G. Krishnamurthi, J. C. Seabra, J. Sanches, *et al.*, "Understanding symptomatology of atherosclerotic plaque by image-based tissue characterization," *Computer Methods and Programs in Biomedicine*, vol. 110, pp. 66-75, 2013.
- [112] M. Demirer, S. Candemir, M. T. Bigelow, S. M. Yu, V. Gupta, L. M. Prevedello, *et al.*, "A user interface for optimizing radiologist engagement in image data curation for artificial intelligence," *Radiology: Artificial Intelligence*, vol. 1, p. e180095, 2019.
- [113] J. Lee, D. Prabhu, C. Kolluru, Y. Gharaibeh, V. N. Zimin, L. A. Dallan, *et al.*, "Fully automated plaque characterization in intravascular OCT images using hybrid convolutional and lumen morphology features," *Scientific Reports*, vol. 10, pp. 1-13, 2020.
- [114] E. C. Kyriacou, S. Petroudi, C. S. Pattichis, M. S. Pattichis, M. Griffin, S. Kakkos, *et al.*, "Prediction of high-risk asymptomatic carotid plaques based on ultrasonic image features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 966-973, 2012.
- [115] U. R. Acharya, O. Faust, S. V. Sree, A. P. C. Alvin, G. Krishnamurthi, J. Sanches, et al., "Atheromatic<sup>TM</sup>: Symptomatic vs. asymptomatic classification of carotid ultrasound plaque using a combination of HOS, DWT & texture," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 2011, pp. 4489-4492.
- [116] S. G. Mougiakakou, S. Golemati, I. Gousias, A. N. Nicolaides, and K. S. Nikita, "Computer-aided diagnosis of carotid atherosclerosis based on ultrasound image statistics, laws' texture and neural networks," *Ultrasound in medicine & biology*, vol. 33, pp. 26-36, 2007.
- [117] A. Gastounioti, S. Makrodimitris, S. Golemati, N. P. Kadoglou, C. D. Liapis, and K. S. Nikita, "A novel computerized tool to stratify risk in carotid atherosclerosis using kinematic features of the arterial wall," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 1137-1145, 2014.
- [118] T. Araki, N. Ikeda, D. Shukla, P. K. Jain, N. D. Londhe, V. K. Shrivastava, *et al.*, "PCA-based polling strategy in machine learning framework for coronary artery disease risk assessment in intravascular ultrasound: a link between carotid and coronary grayscale plaque morphology," *Computer methods and programs in biomedicine*, vol. 128, pp. 137-158, 2016.

- [119] V. D. Tsakanikas, P. K. Siogkas, M. D. Mantzaris, V. T. Potsika, V. I. Kigka, T. P. Exarchos, *et al.*, "A deep learning oriented method for automated 3D reconstruction of carotid arterial trees from MR imaging," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 2408-2411.
- [120] J. Wu, J. Xin, X. Yang, J. Sun, D. Xu, N. Zheng, *et al.*, "Deep morphology aided diagnosis network for segmentation of carotid artery vessel wall and diagnosis of carotid atherosclerosis on black-blood vessel wall MRI," *Medical Physics*, vol. 46, pp. 5544-5561, 2019.
- [121] S. A. Sheth, V. Lopez-Rivera, A. Barman, J. C. Grotta, A. J. Yoo, S. Lee, *et al.*, "Machine Learning–Enabled Automated Determination of Acute Ischemic Core From Computed Tomography Angiography," *Stroke*, vol. 50, pp. 3093-3100, 2019.
- [122] L. Kats, M. Vered, A. Zlotogorski-Hurvitz, and I. Harpaz, "Atherosclerotic carotid plaques on panoramic imaging: an automatic detection using deep learning with small dataset," *arXiv* preprint arXiv:1808.08093, 2018.
- [123] N. H. Meshram, C. C. Mitchell, S. Wilbrand, R. J. Dempsey, and T. Varghese, "Deep Learning for Carotid Plaque Segmentation using a Dilated U-Net Architecture," *Ultrasonic Imaging*, vol. 42, pp. 221-230, 2020.
- [124] R. Zhou, A. Fenster, Y. Xia, J. D. Spence, and M. Ding, "Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images," *Medical Physics*, vol. 46, pp. 3180-3193, 2019.
- [125] A. Nicolaides, S. K. Kakkos, E. Kyriacou, M. Griffin, G. Geroulakos, and C. S. Pattichis, "Ultrasonic plaque characterization: Results from the asymptomatic carotid stenosis and risk of stroke (ACSRS) study," in *Ultrasound and Carotid Bifurcation Atherosclerosis*, ed: Springer, 2011, pp. 613-632.
- [126] U. R. Acharya, M. R. K. Mookiah, S. V. Sree, D. Afonso, J. Sanches, S. Shafique, *et al.*, "Atherosclerotic plaque tissue characterization in 2D ultrasound longitudinal carotid scans for automated classification: a paradigm for stroke risk assessment," *Medical & biological engineering & computing*, vol. 51, pp. 513-523, 2013.
- [127] P. K. Jain, N. Sharma, A. A. Giannopoulos, L. Saba, A. Nicolaides, and J. S. Suri, "Hybrid deep learning segmentation models for atherosclerotic plaque in internal carotid artery B-mode ultrasound," *Computers in Biology and Medicine*, vol. 136, p. 104721, 2021.
- [128] L. Saba, S. S. Sanagala, S. K. Gupta, V. K. Koppula, A. M. Johri, A. M. Sharma, *et al.*, "Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system," *The International Journal of Cardiovascular Imaging*, pp. 1-18.
- [129] S. S. Skandha, S. K. Gupta, L. Saba, V. K. Koppula, A. M. Johri, N. N. Khanna, *et al.*, "3-D optimized classification and characterization artificial intelligence

- paradigm for cardiovascular/stroke risk stratification using carotid ultrasound-based delineated plaque: Atheromatic<sup>TM</sup> 2.0," *Computers in Biology and Medicine*, vol. 125, p. 103958, 2020.
- [130] A. S. Hyman, "Asymptomatic heart disease: Observations made during the early recruiting period for Navy and Marine enlistments," *The American Journal of Medicine*, vol. 5, pp. 351-364, 1948.
- [131] A. El-Baz and J. S. Suri, *Big Data in Multimodal Medical Imaging*: CRC Press, 2019.
- [132] D. C. Haussen, J. A. Grossberg, S. Koch, A. Malik, D. Yavagal, B. Gory, et al., "Multicenter experience with stenting for symptomatic carotid web," *Interventional Neurology*, vol. 7, pp. 413-418, 2018.
- [133] M. Biswas, V. Kuppili, T. Araki, D. R. Edla, E. C. Godia, L. Saba, *et al.*, "Deep learning strategy for accurate carotid intima-media thickness measurement: an ultrasound study on Japanese diabetic cohort," *Computers in Biology and Medicine*, vol. 98, pp. 100-117, 2018.
- [134] G. Von Ingersleben, U. Schmiedl, T. Hatsukami, J. Nelson, D. Subramaniam, M. Ferguson, *et al.*, "Characterization of atherosclerotic plaques at the carotid bifurcation: correlation of high-resolution MR imaging with histologic analysis-preliminary study," *Radiographics*, vol. 17, pp. 1417-1423, 1997.
- [135] M. Wintermark, S. Arora, E. Tong, E. Vittinghoff, B. C. Lau, J. D. Chien, et al., "Carotid plaque computed tomography imaging in stroke and nonstroke patients," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 64, pp. 149-157, 2008.
- [136] B. Chu, B. A. P. Phan, N. Balu, C. Yuan, B. G. Brown, and X.-Q. Zhao, "Reproducibility of carotid atherosclerotic lesion type characterization using high resolution multicontrast weighted cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 8, pp. 793-799, 2006.
- [137] X. Zhao, D. S. Hippe, R. Li, G. M. Canton, B. Sui, Y. Song, *et al.*, "Prevalence and characteristics of carotid artery high-risk atherosclerotic plaques in Chinese patients with cerebrovascular symptoms: a Chinese atherosclerosis risk evaluation II study," *Journal of the American Heart Association*, vol. 6, p. e005831, 2017.
- [138] M. Wintermark, C. Glastonbury, E. Tong, B. C. Lau, S. Schaeffer, J. D. Chien, *et al.*, "Semi-automated computer assessment of the degree of carotid artery stenosis compares favorably to visual evaluation," *Journal of the Neurological Sciences*, vol. 269, pp. 74-79, 2008.
- [139] Y. Li, G. Zhu, V. Ding, Y. Huang, B. Jiang, R. L. Ball, *et al.*, "Assessing the Relationship between Atherosclerotic Cardiovascular Disease Risk Score and Carotid Artery Imaging Findings," *Journal of Neuroimaging*, vol. 29, pp. 119-125, 2019.
- [140] M. B. Griffin, E. Kyriacou, C. Pattichis, D. Bond, S. K. Kakkos, M. Sabetai, et al.,

- "Juxtaluminal hypoechoic area in ultrasonic images of carotid plaques and hemispheric symptoms," *Journal of Vascular Surgery*, vol. 52, pp. 69-76, 2010.
- [141] A. Nicolaides, M. Griffin, G. C. Makris, G. Geroulakos, D. Bond, E. Kyriacou, *et al.*, "Carotid Plaque Texture Analysis Using 3-Dimensional Volume Ultrasonic Imaging," in *Ultrasound and Carotid Bifurcation Atherosclerosis*, ed: Springer, 2011, pp. 299-324.
- [142] A. Nicolaides, M. Sabetai, S. Kakkos, S. Dhanjil, T. Tegos, J. Stevens, *et al.*, "The asymptomatic carotid stenosis and risk of stroke (ACSRS) study," *International Angiology: a Journal of the International Union of Angiology*, vol. 22, pp. 263-272, 2003.
- [143] H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. Khan, "Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms," *arXiv* preprint arXiv:2004.00038, 2020.
- [144] G. S. Tandel, A. Balestrieri, T. Jujaray, N. N. Khanna, L. Saba, and J. S. Suri, "Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm," *Computers in Biology and Medicine*, p. 103804, 2020.
- [145] S. S. S. Luca Saba, Suneet K. Gupta,, A. M. J. Vijaya K. Koppula, Aditya M. Sharma, Raghu Kolluri,, and A. N. Deepak L. Bhatt, Jasjit S. Suri, "Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system," *The International Journal of Cardiovascular Imagin*, vol. 37, pp. 1511--1528, 2021.
- [146] B. Jena, S. Saxena, G. K. Nayak, L. Saba, N. Sharma, and J. S. Suri, "Artificial Intelligence-based Hybrid Deep Learning Models for Image Classification: The First Narrative Review," *Computers in Biology and Medicine*, p. 104803, 2021.
- [147] T. Abuhmed, S. El-Sappagh, and J. M. Alonso, "Robust hybrid deep learning models for Alzheimer's progression detection," *Knowledge-Based Systems*, vol. 213, p. 106688, 2021.
- [148] H. V. Dang, H. Tran-Ngoc, T. V. Nguyen, T. Bui-Tien, G. De Roeck, and H. X. Nguyen, "Data-driven structural health monitoring using feature fusion and hybrid deep learning," *IEEE Transactions on Automation Science and Engineering*, 2020.
- [149] J. S. Suri, S. Agarwal, R. Pathak, V. Ketireddy, M. Columbu, L. Saba, *et al.*, "COVLIAS 1.0: Lung Segmentation in COVID-19 Computed Tomography Scans Using Hybrid Deep Learning Artificial Intelligence Models," *Diagnostics*, vol. 11, p. 1405, 2021.
- [150] T. S. Qaid, H. Mazaar, M. Y. H. Al-Shamri, M. S. Alqahtani, A. A. Raweh, and W. Alakwaa, "Hybrid Deep-Learning and Machine-Learning Models for Predicting COVID-19," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.
- [151] B. K. Singh, K. Verma, A. Thoke, and J. S. Suri, "Risk stratification of 2D ultrasound-based breast lesions using hybrid feature selection in machine learning paradigm," *Measurement*, vol. 105, pp. 146-157, 2017.

- [152] A. Nibali, Z. He, and D. Wollersheim, "Pulmonary nodule classification with deep residual networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, pp. 1799-1808, 2017.
- [153] S. K. Srivastava, S. K. Singh, and J. S. Suri, "A healthcare text classification system and its performance evaluation: A source of better intelligence by characterizing healthcare text," in *Cognitive Informatics, Computer Modelling, and Cognitive Science*, ed: Elsevier, 2020, pp. 319-369.
- [154] J. C. Than, L. Saba, N. M. Noor, O. M. Rijal, R. M. Kassim, A. Yunus, *et al.*, "Lung disease stratification using amalgamation of Riesz and Gabor transforms in machine learning framework," *Computers in Biology and Medicine*, vol. 89, pp. 197-211, 2017.
- [155] S. S. Sanagala, S. K. Gupta, V. K. Koppula, and M. Agarwal, "A Fast and Light Weight Deep Convolution Neural Network Model for Cancer Disease Identification in Human Lung (s)," in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019, pp. 1382-1387.
- [156] L. Saba, S. S. Sanagala, S. K. Gupta, V. K. Koppula, J. R. Laird, V. Viswanathan, et al., "A Multicenter study on Carotid Ultrasound Plaque Tissue Characterization and Classification using Six Deep Artificial Intelligence Models: a Stroke Application," *IEEE Transactions on Instrumentation and Measurement*, 2021.
- [157] W. Alakwaa, M. Nassef, and A. Badr, "Lung cancer detection and classification with 3D convolutional neural network (3D-CNN)," *Lung Cancer*, vol. 8, p. 409, 2017.
- [158] A. R. Larici, A. Farchione, P. Franchi, M. Ciliberto, G. Cicchetti, L. Calandriello, *et al.*, "Lung nodules: size still matters," *European Respiratory Review*, vol. 26, 2017.
- [159] A. McErlean, D. M. Panicek, E. C. Zabor, C. S. Moskowitz, R. Bitar, R. J. Motzer, *et al.*, "Intra-and interobserver variability in CT measurements in oncology," *Radiology*, vol. 269, pp. 451-459, 2013.
- [160] S. Birring and M. Peake, "Symptoms and the early diagnosis of lung cancer," ed: BMJ Publishing Group Ltd, 2005.
- [161] A. I. Baba and C. Câtoi, *Comparative oncology*: Publishing House of the Romanian Academy Bucharest, 2007.
- [162] L. Saba, M. Agarwal, A. Patrick, A. Puvvula, S. K. Gupta, A. Carriero, et al., "Six artificial intelligence paradigms for tissue characterisation and classification of non-COVID-19 pneumonia against COVID-19 pneumonia in computed tomography lungs," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1-12, 2021.
- [163] G. S. Tran, T. P. Nghiem, V. T. Nguyen, C. M. Luong, and J.-C. Burie, "Improving accuracy of lung nodule classification using deep learning with focal loss," *Journal of Healthcare Engineering*, vol. 2019, 2019.

- [164] F. Maldonado, C. Varghese, S. Rajagopalan, F. Duan, A. Balar, D. A. Lakhani, *et al.*, "Validation of the BRODERS classifier (Benign versus aggressive nODule Evaluation using Radiomic Stratification), a novel high-resolution computed tomography-based radiomic classifier for indeterminate pulmonary nodules," *European Respiratory Journal*, 2020.
- [165] C. Jacobs, E. M. van Rikxoort, K. Murphy, M. Prokop, C. M. Schaefer-Prokop, and B. van Ginneken, "Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database," *European Radiology*, vol. 26, pp. 2139-2147, 2016.
- [166] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, pp. 915-931, 2011.
- [167] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Gradcam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618-626.
- [168] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging (IPMI)*, 2015, pp. 588-599.
- [169] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT," *Information Fusion*, vol. 42, pp. 102-110, 2018.
- [170] R. Dey, Z. Lu, and Y. Hong, "Diagnostic classification of lung nodules using 3D neural networks," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 774-778.
- [171] G. L. F. da Silva, T. L. A. Valente, A. C. Silva, A. C. de Paiva, and M. Gattass, "Convolutional neural network-based PSO for lung nodule false positive reduction on CT images," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 109-118, 2018.
- [172] R. V. M. da Nóbrega, S. A. Peixoto, S. P. P. da Silva, and P. P. Rebouças Filho, "Lung nodule classification via deep transfer learning in CT lung images," in 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), 2018, pp. 244-249.
- [173] M. Al-Shabi, B. L. Lan, W. Y. Chan, K.-H. Ng, and M. Tan, "Lung nodule classification using deep local-global networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 1815-1819, 2019.
- [174] E. A. Roberts and M. L. Schilsky, "A practice guideline on Wilson disease," *Hepatology*, vol. 37, pp. 1475-1492, 2003.
- [175] M. Yousaf, M. Kumar, R. Ramakrishnaiah, R. Vanhemert, and E. Angtuaco,

- "Atypical MRI features involving the brain in Wilson's disease," *Radiology Case Reports*, vol. 4, p. 312, 2009.
- [176] S. K. Banchhor, N. D. Londhe, T. Araki, L. Saba, P. Radeva, J. R. Laird, *et al.*, "Well-balanced system for coronary calcium detection and volume measurement in a low resolution intravascular ultrasound videos," *Computers in Biology and Medicine*, vol. 84, pp. 168-181, 2017.
- [177] A. Jamthikar, D. Gupta, E. Cuadrado-Godia, A. Puvvula, N. N. Khanna, L. Saba, *et al.*, "Ultrasound-based stroke/cardiovascular risk stratification using Framingham Risk Score and ASCVD Risk Score based on "Integrated Vascular Age" instead of "Chronological Age": a multi-ethnic study of Asian Indian, Caucasian, and Japanese cohorts," *Cardiovascular Diagnosis and Therapy*, 2020.
- [178] T. Araki, N. Ikeda, N. Dey, S. Acharjee, F. Molinari, L. Saba, *et al.*, "Shape-based approach for coronary calcium lesion volume measurement on intravascular ultrasound imaging and its association with carotid intima-media thickness," *Journal of Ultrasound in Medicine*, vol. 34, pp. 469-482, 2015.
- [179] M. Singh, S. Bansal, S. Ahuja, R. K. Dubey, B. K. Panigrahi, and N. Dey, "Transfer learning based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data," 2020.
- [180] P. Ferenci, "Regional distribution of mutations of the ATP7B gene in patients with Wilson disease: impact on genetic testing," *Human Genetics*, vol. 120, pp. 151-159, 2006.
- [181] S. Vrabelova, O. Letocha, M. Borsky, and L. Kozak, "Mutation analysis of the ATP7B gene and genotype/phenotype correlation in 227 patients with Wilson disease," *Molecular Genetics and Metabolism*, vol. 86, pp. 277-285, 2005.
- [182] A. M. Brickman, K. L. Siedlecki, J. Muraskin, J. J. Manly, J. A. Luchsinger, L.-K. Yeung, *et al.*, "White matter hyperintensities and cognition: testing the reserve hypothesis," *Neurobiology of Aging*, vol. 32, pp. 1588-1598, 2011.
- [183] L. Saba, P. Lucatelli, M. Anzidei, M. di Martino, J. S. Suri, and R. Montisci, "Volumetric distribution of the white matter hyper-intensities in subject with mild to severe carotid artery stenosis: does the side play a role?," *Journal of Stroke and Cerebrovascular Diseases*, vol. 27, pp. 2059-2066, 2018.
- [184] L. Saba, R. Sanfilippo, M. Porcu, P. Lucatelli, R. Montisci, F. Zaccagna, *et al.*, "Relationship between white matter hyperintensities volume and the circle of Willis configurations in patients with carotid artery pathology," *European Journal of Radiology*, vol. 89, pp. 111-116, 2017.
- [185] M. Porcu, A. Balestrieri, P. Siotto, P. Lucatelli, M. Anzidei, J. S. Suri, *et al.*, "Clinical neuroimaging markers of response to treatment in mood disorders," *Neuroscience Letters*, vol. 669, pp. 43-54, 2018.
- [186] T. Kim, I. O. Kim, W. Kim, J. Cheon, S. Moon, J. Kwon, *et al.*, "MR imaging of the brain in Wilson disease of childhood: findings before and after treatment with

- clinical correlation," *American Journal of Neuroradiology*, vol. 27, pp. 1373-1378, 2006.
- [187] X. Hu, S. Chen, C.-B. Huang, Y. Qian, and Y. Yu, "Frequency-dependent changes in the amplitude of low-frequency fluctuations in patients with Wilson's disease: a resting-state fMRI study," *Metabolic Brain Disease*, vol. 32, pp. 685-692, 2017.
- [188] M. Kaden, M. Riedel, W. Hermann, and T. Villmann, "Border-sensitive learning in generalized learning vector quantization: an alternative to support vector machines," *Soft Computing*, vol. 19, pp. 2423-2434, 2015.
- [189] R. Jing, Y. Han, H. Cheng, Y. Han, K. Wang, D. Weintraub, *et al.*, "Altered large-scale functional brain networks in neurological Wilson's disease," *Brain Imaging and Behavior*, vol. 14, pp. 1445-1455, 2020.
- [190] H.-I. Suk, S.-W. Lee, D. Shen, and A. s. D. N. Initiative, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Medical Image Analysis*, vol. 37, pp. 101-113, 2017.
- [191] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in *World Congress on Medical Physics and Biomedical Engineering 2018*, 2019, pp. 183-189.
- [192] Y. Zhang, H. Zhang, X. Chen, M. Liu, X. Zhu, S.-W. Lee, *et al.*, "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," *Pattern Recognition*, vol. 88, pp. 421-430, 2019.
- [193] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, V. Calhoun, *et al.*, "Deep residual learning for neuroimaging: An application to predict progression to alzheimer's disease," *Journal of Neuroscience Methods*, p. 108701, 2020.
- [194] B. Richhariya, M. Tanveer, A. Rashid, and A. s. D. N. Initiative, "Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE)," *Biomedical Signal Processing and Control*, vol. 59, p. 101903, 2020.
- [195] J. Liu, Y. Pan, F.-X. Wu, and J. Wang, "Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification," *Neurocomputing*, 2020.
- [196] U. R. Acharya, M. R. K. Mookiah, S. V. Sree, R. Yanti, R. Martis, L. Saba, *et al.*, "Evolutionary algorithm-based classifier parameter tuning for automatic ovarian cancer tissue characterization and classification," *Ultraschall in der Medizin-European Journal of Ultrasound*, vol. 35, pp. 237-245, 2014.
- [197] M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, pp. 3152-3168, 2019.
- [198] G. Sreenu and M. S. Durai, "Intelligent video surveillance: a review through deep

- learning techniques for crowd analysis," Journal of Big Data, vol. 6, pp. 1-27, 2019.
- [199] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," 1988.
- [200] N. M. Razali and J. Geraghty, "Genetic algorithm performance with different selection strategies in solving TSP," in *Proceedings of the World Congress on Engineering*, 2011, pp. 1-6.
- [201] L. M. Schmitt, "Theory of genetic algorithms," *Theoretical Computer Science*, vol. 259, pp. 1-61, 2001.
- [202] M. Safe, J. Carballido, I. Ponzoni, and N. Brignole, "On stopping criteria for genetic algorithms," in *Brazilian Symposium on Artificial Intelligence*, 2004, pp. 405-413.
- [203] M. Agarwal, V. K. Bohat, M. D. Ansari, A. Sinha, S. K. Gupta, and D. Garg, "A convolution neural network based approach to detect the disease in corn crop," in 2019 IEEE 9th International Conference on Advanced Computing (IACC), 2019, pp. 176-181.
- [204] S. S. Sanagala, A. Nicolaides, S. K. Gupta, V. K. Koppula, L. Saba, S. Agarwal, *et al.*, "Ten Fast Transfer Learning Models for Carotid Ultrasound Plaque Tissue Characterization in Augmentation Framework Embedded with Heatmaps for Stroke Risk Stratification," *Diagnostics*, vol. 11, p. 2109, 2021.
- [205] R. Narayanan, P. Werahera, A. Barqawi, E. Crawford, K. Shinohara, A. Simoneau, *et al.*, "Adaptation of a 3D prostate cancer atlas for transrectal ultrasound guided target-specific biopsy," *Physics in Medicine & Biology*, vol. 53, p. N397, 2008.
- [206] P. Sudeep, P. Palanisamy, J. Rajan, H. Baradaran, L. Saba, A. Gupta, *et al.*, "Speckle reduction in medical ultrasound images using an unbiased non-local means method," *Biomedical Signal Processing and Control*, vol. 28, pp. 1-8, 2016.
- [207] S. S. Skandha, A. Nicolaides, S. K. Gupta, V. K. Koppula, L. Saba, A. M. Johri, *et al.*, "A hybrid deep learning paradigm for carotid plaque tissue characterization and its validation in multicenter cohorts using a supercomputer framework," *Computers in Biology and Medicine*, vol. 141, p. 105131, 2022.
- [208] J. Suri, S. Agarwal, S. K. Gupta, A. Puvvula, K. Viskovic, N. Suri, *et al.*, "Systematic Review of Artificial Intelligence in Acute Respiratory Distress Syndrome for COVID-19 Lung Patients: A Biomedical Imaging Perspective," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [209] S. Paul, M. Maindarkar, S. Saxena, L. Saba, M. Turk, M. Kalra, *et al.*, "Bias Investigation in Artificial Intelligence Systems for Early Detection of Parkinson's Disease: A Narrative Review," *Diagnostics*, vol. 12, p. 166, 2022.
- [210] S. Das, G. Nayak, L. Saba, M. Kalra, J. S. Suri, and S. Saxena, "An artificial intelligence framework and its bias for brain tumor segmentation: A narrative review," *Computers in Biology and Medicine*, p. 105273, 2022.

- [211] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," *arXiv preprint arXiv:1710.09282*, 2017.
- [212] Y. He, X. Zhang, and J. Sun, "Channel pruning for accelerating very deep neural networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1389-1397.
- [213] S. Han, J. Pool, J. Tran, and W. J. Dally, "Learning both weights and connections for efficient neural networks," *arXiv* preprint arXiv:1506.02626, 2015.
- [214] C.-J. Chen, K.-C. Chen, and M.-c. Martin-Kuo, "Acceleration of neural network model execution on embedded systems," in *2018 International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*, 2018, pp. 1-3.
- [215] C. Yang, Z. An, C. Li, B. Diao, and Y. Xu, "Multi-objective pruning for cnns using genetic algorithm," in *International Conference on Artificial Neural Networks* (ICANN), 2019, pp. 299-305.
- [216] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, C. Richter, and K. Cha, "Evolutionary pruning of transfer learned deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis," *Physics in Medicine & Biology*, vol. 63, p. 095005, 2018.
- [217] M. Agarwal, S. K. Gupta, and K. Biswas, "A Compressed and Accelerated SegNet for Plant Leaf Disease Segmentation: A Differential Evolution Based Approach," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2021, pp. 272-284.
- [218] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278-2324, 1998.
- [219] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009.
- [220] I. Ullah and A. Petrosino, "About pyramid structure in convolutional neural networks," in 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 1318-1324.
- [221] Y. Hu, S. Sun, J. Li, X. Wang, and Q. Gu, "A novel channel pruning method for deep neural network compression," *arXiv preprint arXiv:1805.11394*, 2018.
- [222] R. Yu, A. Li, C.-F. Chen, J.-H. Lai, V. I. Morariu, X. Han, *et al.*, "Nisp: Pruning networks using neuron importance score propagation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 9194-9203.
- [223] T. Li, B. Wu, Y. Yang, Y. Fan, Y. Zhang, and W. Liu, "Compressing convolutional neural networks via factorized convolutional filters," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3977-3986.

#### **List of Publications**

#### **Published & Accepted**

- Skandha *et al.* "3-D optimized classification and characterization artificial intelligence paradigm for cardiovascular/stroke risk stratification using carotid ultrasound-based delineated plaque: Atheromatic<sup>TM</sup> 2.0." **Computers in Biology and Medicine** 125 (2020): 103958. (IF: 4.589) (Indexed in SCI/Scopus/WoS).
- 2 Skandha *et al.* "A Multicenter Study on Carotid Ultrasound Plaque Tissue Characterization and Classification Using Six Deep Artificial Intelligence Models: A Stroke Application." **IEEE Transactions on Instrumentation and Measurement** 70 (2021): 1-12. (IF: 4.47) (Indexed in SCI/Scopus/WoS).
- 3 Skandha *et al.* "Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system." **The International Journal of Cardiovascular Imaging** 37.5 (2021): 1511-1528. (IF: 2.357) (Indexed in SCI/Scopus/WoS).
- 4 Skandha *et al.* "Multimodality carotid plaque tissue characterization and classification in the artificial intelligence paradigm: A narrative review for stroke application." **Annals of Translational Medicine** 9, no. 14 (2021). (IF: 3.932) (Indexed in SCI/Scopus/WoS).
- Skandha *et al.*, "Ten Fast Transfer Learning Models for Carotid Ultrasound Plaque Tissue Characterization in Augmentation Framework Embedded with Heatmaps for Stroke Risk Stratification." **Diagnostics** 11, no. 11 (2021): 2109. (IF: 3.706) (Indexed in SCI/Scopus/WoS).
- 6 Skandha *et al.* A hybrid deep learning paradigm for carotid plaque tissue characterization and its validation in multicenter cohorts using a supercomputer framework, **Computers in Biology and Medicine**, Volume 141, 2022,105131, ISSN 0010-4825, (IF: 4.589). (Indexed in SCI/Scopus/WoS).
- Skandha *et al.* "A Fast and Light Weight Deep Convolution Neural Network Model for Cancer Disease Identification in Human Lung(s)," **2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)**, Boca Raton, FL, USA, 2019, pp. 1382-1387, Doi: 10.1109/ICMLA.2019.00225. (Core rank: B).

8 Siva Skandha et al. "A Novel Genetic Algorithm-based Approach for Compression and Acceleration of Deep Learning Convolution Neural Network: An Application in Computer Tomography Lung Cancer Data "in **Neural Computing and Application** (Submitted two revisions) (IF: 5.606). (Indexed in SCI/Scopus/WoS).

#### **Under Review**

- 9 Skandha et al. "Magnetic Resonance-based Wilson Disease Tissue Characterization in Artificial Intelligence Framework using Transfer Learning" Book chapter in **IOP press**.
- 10 Skandha et al. "Understanding COVID severity for COVID-19 Lung Characterization via Artificial Intelligence-Transfer Learning " Book chapter in **IOP press**.

# Synopsis on

# Developing lightweight CNN models for detecting lesions in the medical Imaging

Submitted By

Siva Skandha Sanagala

(Admission No. E18SOE841)

Under the Supervision of

Dr. Suneet Kr. Gupta and Dr K. Vijaya Kumar



Department of Computer Science Engineering Bennett University, (The Times Group)

Greater Noida, INDIA 201310

#### 1. Introduction

This studies mainly focus on detecting lesions in non-invasive medical imaging techniques such as ultrasound, computed tomography, and magnetic resonance imaging (MRI). For these three imaging techniques, we choose cardiovascular diseases (CVD), lung nodules, and Wilson disease. These three diseases are the leading causes of death worldwide and as per the World Health Organization (WHO) statistics, CVDs are the leading cause of death globally, with nearly 18 million deaths annually [1]. In the U.S. alone, approximately 700,000 people die each year—the equivalent of 1 in 4 people dying every 36 seconds [2, 3]. Cancer is the second leading cause of death [4], and it causes 9.6 million death [5].

#### 1.1. Introduction to lesions

Lesions may occur at any part of the body [6] and it is developed due to abnormal mass or swelling in the tissue. There are two types of lesions: benign lesions which is non-cancerous and malignant lesions are cancerous. Malignant lesions refer to the cells which grow out of control and invade other cells [6]. Infections, genetic inheritance, environmental factors, and poor lifestyles strengthen malignant lesions [7] and these factors also damage DNA and lead to cancer. However, all these DNA damages will be identified by the cells automatically and repaired. If the cells are damaged, they cannot repair themselves and it is so-called "cell of death." Cancer occurs when these damaged cells spread abnormally. In this study, we considered the lesions formed at the carotid, lung nodule, and Wilson disease.

CVD's leading cause is atherosclerosis which occurs due to lesions in the major arteries such as carotid [8, 9]. There are several risk factors for CVD and atherosclerosis, including hypertension, hyperlipidemia, diabetes, obesity, smoking, and a sedentary lifestyle [10]. The process of atherosclerosis is characterized by excessive plaque deposits in the walls of the arteries [11]. In the case of an unstable plaque, the associated fibrous cap breaks, causing a

form of thrombosis [12]. The thrombosis leads to embolism which in turn causes a blockage in blood flow resulting in a stroke or myocardial infarction [13].

#### 1.2. Imaging Modalities

Non-Invasive imaging techniques are raising the demand for the diagnosis of diseases. There are popular imaging techniques existed those are magnetic resonance imaging (MRI) [14], computer tomography (CT) [15], and ultrasound (US) [16]. MRI imaging uses powerful magnets for Imaging the tissues and CT uses x-ray radiations for Imaging, whereas sound signals have been used in ultrasound. Individual techniques have some other advantages i.e., MRI is useful for measuring the tissue thickness [17], whereas CT will give a better clinical view [15]. Both MRI and CT are emitting radiations, which causes side effects when the object is exposed in longer time. In other hand we had US, which portable, handy and radiation free (since it uses sound signals). Sample images of all these modalities are shown in Figure 1, Figure 2, and Figure 3. Figure 1 shows the US scans of the carotid scan (Cyprus). and CT scans of the Lung from Early Lung Cancer Action Program (ELCAP) dataset are shown in Figure 2; Figure 3 represents CT scans of the COVID affected lungs, and Figure 4 gives MRI scans of the Wilson disease effected brain.

## 1.3. Comparison of symptomatic vs. asymptomatic lesions

Lesions appear differently in different modalities [18]. However, their characteristics of morphological features are the same. Table 1 gives the carotid lesions in the US imaging.

Table 1. Comparison of symptomatic vs. asymptomatic lesions.

SN	Attribute	Symptomatic	Asymptomatic
1	Brightness	Dark	Bright
2	Echo	Echo lucent (dark)	Echogenic (bright)
3	Intensity	Hypogenic	Hyper genic
4	GSM	Low (20-40)	High (50-70)
5	Color	Patchy, UNSTABLE, more	Homogeneous, STABLE, less
		Lipids	Lipids

6	Type	Low Collagen	High Collagen
7	Calcium	Low Calcium	High Calcium
8	Deep Learning	High Level features are high	Low Level Features are high in
	Features	in the end layers	the initial layers
9	Fractal Di.	More Chaotic, More Patchy,	Less Chaotic, Less Patchy,
		higher Fractal Dimension	Lower Fractal Dimension
9	Riskiness	Riskier	Less Risky (Less Dangerous)
10	Lesion Type	lesion Type I and II	lesion Type III and IV
		type I: predominantly	type III: echogenic plaque with
		hypoechoic with thin	<50% hypoechoic areas
		echogenic rim	type IV: uniformly echogenic
		type II: echogenic plaque with	plaque
		>50% hypoechoic areas	

#### 1.4. Need of classification and characterization

Early detection of the lesions reduces the mortality rate in cardiovascular disease (CVD) and Cancer [6] and all the imaging modalities suffer from inter-observer variability [19]. On the other hand, if the number of images/scans is higher, much time is required to identify experts' malignancy tissue. Therefore, it is required to use computer vision techniques for identifying the region of interest (ROI) tissue.

## 1.5. Role of AI in medical imaging

Artificial Intelligence (AI) changes all the science and technology dynamics, especially in medical image processing. Machine learning (ML) is a subset of the AI. It shows promising results for lesion tissue classification and characterization [21-27]. ML model requires handcrafted features to be fed as input and performance may vary based on the selected features. To overcome this difficulty, AI offers a powerful paradigm called deep learning (DL) which offers the automatic feature extraction by convolutional neural networks (CNN) on image dataset [28-30].

# 1.6. Need of Deep learning in medical Imaging

DL offers flexible and reliable strategies for classification and characterization of the lesions. In deep learning models medical images (generally grayscale) and fed into the convolution layers followed by max-pooling/average pooling layers for automatic feature extraction. These extracted features are subsequently fed into a fully connected network for classification and risk assessment. DL models learn complex patterns from the input training images and then use these patterns to predict target labels. There are several models besides CNN that are popular in medical imaging. Some of these are the U-Net, Autoencoders, recurrent neural networks (RNNs), long short-term memory (LSTM), Region-based Convolutional Neural Networks (R-CNN), and Mask R-CNN. U-Net (a combination of encoders and decoders) plays an important role in the segmentation and classification processes [20].

Research on the use of DL in lesion classification and characterization is still emerging; other researchers have used DL to investigate the chest [21], coronary [22], liver [23], Intimamedia thickness (IMT) wall [24] of patients, as well as lumen characterization [25] and carotid risk measurement in diabetic patients [26] and Rheumatoid arthritis [27] in arthritic patients in particular.

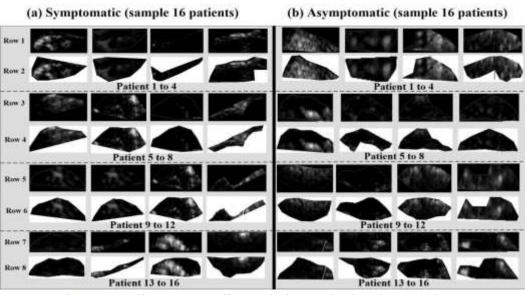


Figure 1. Full scan vs. Delineated plaques in the ultrasound.

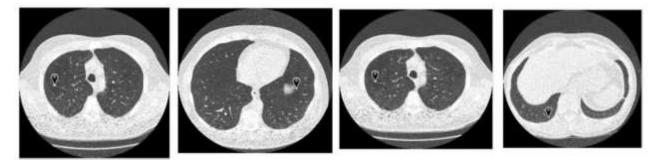


Figure 2. ELCAP dataset with annotation of the cancer cell (Source: ELCAP).

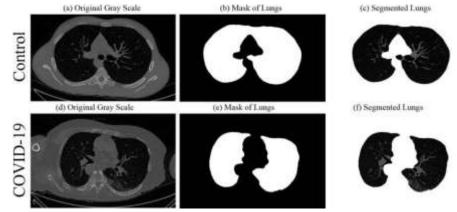


Figure 3. Sample control patient's (a) original grayscale image, (b) lung mask, (c) segmented lung; Sample COVID-19 patient's (d) original grayscale image, (e) lung mask, (f) segmented lung.

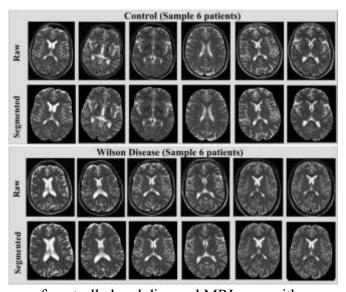


Figure 4. Sample images of controlled and diseased MRI scan with corresponding segmented images (skull stripping).

# 1.7. Compression of Convolution Neural Networks

Recent developments in AI and internet of things brings the application of DL into regular

human life using edge devices. The major setback of edge devices is the memory. The pretrained models such as Inception V3, MobileNet, ResNet50, AlexNet [28]. These architectures were developed for solving 1000 class object recognition problem, the basic issue is that the pre-trained models have large number of layers with huge number of nodes and consequently require huge memory space.

The utilization of the edge devices will reach almost 125 to 500 billion IoT devices by 2030 out of which 20% of devices would have cameras [29]. The deep learning applications on edges devices are catching up slowly, due to the size and time constraint. So, there is a need for pruning the DL model's size. Compression of DL models offers the reduction of the size and inference time, which in turn helps in the utilization of the edge devices. Pruning offers to trim the redundant or not properly learned nodes in the DL architecture, so it reduces the trainable parameters in the models. If the trainable parameters reduce, the size of the model will be reduced.

# 2. Research Gaps

- Research showed that deep learning models such as VGG16 [30], AlexNet [31], and Inception V3 [32] models perform the classification of the lesion in medical imaging. All these models contain a large number of layers and are faced with the highest false-positive rate. Depending on the imaging modalities the deep learning application is varying. So, there is a need of developing optimized models that work on all the imaging modalities with lightweight and faster prediction time.
- Most of the researchers applied deep learning techniques for the segmentation and classification of the lesion [33, 34]. The most important use of medical imaging is the characterization of the lesion [35]. This characterization of the lesion was done by the medical practitioners manually by calculating grey scale median (GSM), morphological textures, and calcium percentage. Some of the researchers are applied machine learning techniques for the detection of GSM levels [36]. However, these ML techniques suffer

from handcrafted images so there is a need for automatic feature extraction for unbiased characterization.

- The availability of the cohorts of different imaging techniques is limited and contains a smaller number of images. So, the augmentation of the images using geometrical operations such as rotation, flip, and tilt. There are no such attempts to take care of the effect of augmentation on the performance of the model. Along with the classification, we are presenting the risk of the disease.
- The deep learning models are computational expensive and also require lots of storage space due to it these models are compatible with devices which having limited computational power and storage space.

#### 3. Objective

- Our objective in this thesis is three-fold. We have first shown how effectively the light-weighted and optimized CNN models for classification of the lesion in stroke, lung, and Wilson disease with improved accuracy and highest diagnostic odds ratio.
- Secondly, characterizing the lesion using a novel deep learning technique called mean feature strength and validating with the existing statistical and image processing techniques.
- Creating a CAD system using multi-center trained deep learning models, hybrid CNN models, and pre-trained CNN models for disease characterization and generation of the heatmaps.
- Finally, we have utilized the meta-heuristic i.e., Genetic Algorithm (GA) for the
  compression of deep learning model to make the model compatible with devices i.e.,
  Rasberry Pi, Jetson Nano etc, which is having limited resources like processing power,
  storage.

#### 4. Scope of the research work

In this section, we discussed the research work carried by dividing the sections which describe the proposed optimized CNN models and characterization of the lesion using CNN and statistical methods in lung and stroke. Secondly, multicenter study on stroke using optimized CNN, the third application was on Wilson disease and stroke using pre-trained CNN models. The fourth application was on stroke using hybrid CNN models for classification and characterization of the multicenter cohort. Finally, we used a genetic algorithm for pruning the models.

Classification and characterization of the lesion in carotid and lung using optimized CNN We proposed several light-weighted CNN models for the classification of the lesions. We proposed CNN model by varying the number of layers from 5 to 22. We augmented the cohorts from balanced to 6x folds. We trained and tested each model on every fold using cross-validation protocols K10, K5, K4, K2, and TT. The optimized model exhibits superior performance in comparison with existing techniques and machine learning models using carotid lesions lung nodule lesions.

In the proposed CNN model, each convolution layer is followed by the average pooling layer. Since the cohort is a grayscale image and it is a cut section region. For complex pattern learning, we added a dropout layer with a 50% rate. "ReLu" activation function is used in every layer to provide nonlinearity in the model.

Carotid plaque: Carotid arteries are major arteries at the neck supplies blood and oxygen to the brain, neck, and face. Due to cholesterol or fat, calcium deposition at the arteries forms plaque or lesions. In due course of time, this plaque will rupture and block the arteries causes a stroke. This disease is called atherosclerosis. Two kinds of lesions are there for measuring the severity of the lesion they are symptomatic and asymptomatic. Symptomatic plaques contain high calcium. They appear dark in the imaging, on the other hand, asymptomatic lesion

contains low calcium and appears bright in the imaging.

Medical practitioners carefully observed the ultrasound scans and separated the lesion portion in the imaging it is referred to as "cut sections". Due to the moderate size of the cohort, we augmented the cohort using the geometrical transition specified above along with skew operation. Our proposed model shows top-1 accuracy of **95.66%** which is better performance than the existing work on the carotid.

Lung Nodule: According to the world health organization (WHO), cancer is the second leading cause of death globally; it causes one in six deaths, world health organization [37]. Among all the cancer types, lung cancer is most common in men and women [38]. The detection and diagnosis of lung cancer in the early stages increase the mortality rate [39]. Among all the non-invasive techniques, computed tomography (CT) has reached many milestones in finding malignancy detection [40, 41]. However, the size of the malignancy cell varies from 3mm to 30mm, and malignance regions less than are equal to 3mm are hard to detect. Early detection of cancer in CT scans requires skilled experts and advanced techniques in computer vision to help the radiologist for better predictions.

We proposed two kinds of CNN models one light weighted model contains seven layers and uses the "ReLu" activation function. Another model was using a modified sigmoid activation function called as "Swish activation" function. The model classifies the lung nodules with a size less than 50mm. It contains the multi-labels named single nodule, multi nodule, and health. Single nodule contains single nodule of size less than 30 mm and multi nodule contains the nodule size less than 50 mm. health scans do not contain any nodule. Our proposed models exhibit top-1 accuracy of 99.5% which is superior to already existing work.

#### Multicenter Study

A multicenter study is an important study in the medical field. It exhibits the different patterns and complexity of the problems. The problem that exists in this kind of study was the

availability of the cohort of the same disease under the same imaging conditions. We got carotid cohorts from two different geological locations with different ethnicity, and we created another cohort by mixing these cohorts, called the "Mixed cohort". Similarly, we had the free cohort of the lung nodule collected from the different geological locations.

We proposed CNN models on these cohorts for classification and characterization. We optimized each CNN model on these three cohorts with augmentation folds. Characterized each class using novel DL method "mean feature strength (MFS)". Which exhibits the class label behavior in each class at every layer. It is validated with image processing techniques and statistical methods such as higher-order spectrum (popularly known as Bispectrum) and histogram. These results were further validated with gold standards (GSM and plaque area) given by radiologists. Multicenter study on exhibits top-1 accuracy in carotid was 94.5%, in the lung was 98.75%, which is far superior to existing methods.

#### Classification and characterization using transfer learning.

In this approach, we used pre-trained CNN models in the thesis. We took 12 different kinds of TL models for classification and characterization of carotid, lung affected with covid, and Wilson disease. We augmented every cohort using the same geometrical operations and classified the carotid lesions into "symptomatic" and "asymptomatic", covid affected lungs into "covid" and "controlled", Wilson disease lesions into "disease" and "controlled". We generated the heatmaps of the lesion, which is intended the severity regions in the scans. These are developed from the pre-trained CNN models weights. We changed the end layers of the TL models by adding the dense layer and dropout layer to get complex learning. These heatmaps are further validated with the gold standards given by radiologists and popular biomarkers.

#### Classification and characterization using hybrid DL models

We fused the strength of ML models and DL models to form a hybrid DL model. These models are applied on the multicenter carotid for effective characterization of the lesions. Here we

optimized the model by varying the cohort's augmentation folds, centers, and loss functions. We studied the behavior of the models at unbalanced, balanced, and optimized cohort sizes.

Further, we characterized the lesions using MFS, HOS, and fractal dimensions. The proposed HDL models exhibit the highest diagnostics odds ratio (DOR). All the proposed AI models are ranked based on the grading mechanism obtained from averaging all the cycles of cross-validation. The proposed HDL models exhibit better ranking in comparison with TL and DL models. We achieved top-1 accuracy of 99.8%. These models are validated with popular cohorts such as ASSIRA, dogs, and cats.

#### Compression of the DL models using GA

In the proposed method we compressed the DL models proposed in our previous studies. We used genetic algorithm (GA) for optimization of the weights. GA is applied by converting the CNN layers into binary vectors. Further, fitness function in GA is applied based on (i) the minimization of hidden units and (ii) test accuracy.

We apply our GA-based strategy on different pre-trained architectures, namely, AlexNet, VGG16, SqueezeNet, and ResNet50 using different datasets, namely, MNIST, CIFAR-10, CIFAR-100, and LIDC-IDRI. The proposed compression model achieves better accuracy in the performance and occupies less storage. The system compresses by around 91%, 78% and 38%, respectively for VGG16, ResNet50 and SqueezeNet. AlexNet (by35%), SqueezeNet (9%), ResNet50 (73%), and VGG16 (80%) on the medical imaging.

#### 5. Conclusion

In this thesis, we focused mainly on three activities 1) classification of the lesions 2) characterization of the lesions 3) CAD systems for showing severity of the lesion and compression of the model. Experimentally all the proposed models/algorithms perform better than the existing system. We successfully characterized the lesions and its components using AI methods and validated with statistical methods. Further, all the proposed models are

validated with existing methods and well-known benchmarked datasets.

This research will be useful for the radiologist for characterizing and measuring the severity of the lesion. It also deals with lightweight models from heavy pre-trained models to use for deploying gin edge devices. Further, this work can be used by researchers for developing CAD systems and fine-tune real usages.

#### References

- 1. Prevention, C.f.D.C.a. 2020 10 June 2020]; Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
- 2. Mackay, J. and G.A. Mensah, *The atlas of heart disease and stroke*. 2004: World Health Organization.
- 3. Virani, S.S., et al., *Heart disease and stroke statistics-2020 update: a report from the American Heart Association.* Circulation, 2020. **141**(9): p. e139-e596.
- 4. Nomura, S., et al., Population health and regional variations of disease burden in Japan, 1990-2015: a systematic subnational analysis for the Global Burden of Disease Study 2015. Lancet, 2017. **390**(10101): p. 1521-1538.
- 5. D'Souza, N., N. Murthy, and R. Aras, *Projection of burden of cancer mortality for India,* 2011-2026. Asian Pac J Cancer Prev, 2013. **14**(7): p. 4387-92.
- 6. Cooper, G.M. and R.E. Hausman, *The development and causes of cancer*. The cell: A molecular approach, 2000. **2**.
- 7. Anand, P., et al., *Cancer is a preventable disease that requires major lifestyle changes.*Pharmaceutical research, 2008. **25**(9): p. 2097-2116.
- 8. Libby, P., *Inflammation and cardiovascular disease mechanisms*. The American journal of clinical nutrition, 2006. **83**(2): p. 456S-460S.
- 9. Suri, J.S., C. Kathuria, and F. Molinari, *Atherosclerosis disease management*. 2010: Springer Science & Business Media.
- 10. Meijers, W.C. and R.A. de Boer, *Common risk factors for heart failure and cancer*. Cardiovascular research, 2019. **115**(5): p. 844-853.
- 11. Libby, P., P.M. Ridker, and A. Maseri, *Inflammation and atherosclerosis*. Circulation, 2002. **105**(9): p. 1135-1143.
- 12. Libby, P. and D.I. Simon, *Inflammation and thrombosis: the clot thickens*. 2001, Am Heart Assoc.
- 13. Ueda, M., ATIS. Drugs, 2010. **70**(1): p. 3-8.

- 14. Council, N.R., Effects of Electromagnetic Fields on Organs and Tissues, in Assessment of the Possible Health Effects of Ground Wave Emergency Network. 1993, National Academies Press (US).
- 15. Saba, L., et al., Carotid artery dissection on non-contrast CT: Does color improve the diagnostic confidence? European journal of radiology, 2014. **83**(12): p. 2288-2293.
- 16. Saba, L., et al., Multi-modality atherosclerosis imaging and diagnosis. 2014: Springer.
- 17. Zhang, X., S. Zhu, and B. He, *Imaging electric properties of biological tissues by RF field mapping in MRI*. IEEE Transactions on medical Imaging, 2010. **29**(2): p. 474-481.
- 18. Halani, S., et al., *Ultrasound and infrared-based imaging modalities for diagnosis and management of cutaneous diseases.* Frontiers in Medicine, 2018. **5**: p. 115.
- 19. Arnold, J., et al., Carotid plaque characterization by duplex scanning: observer error may undermine current clinical trials. Stroke, 1999. **30**(1): p. 61-65.
- 20. 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*. 2015. Springer.
- 21. Kats, L., et al., Atherosclerotic carotid plaques on panoramic imaging: an automatic detection using deep learning with small dataset. arXiv preprint arXiv:1808.08093, 2018.
- 22. Abdolmanafi, A., et al., Characterization of coronary artery pathological formations from OCT imaging using deep learning. Biomedical Optics Express, 2018. **9**(10): p. 4936-4960.
- 23. Biswas, M., et al., *Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm.* Computer methods and programs in biomedicine, 2018. **155**: p. 165-177.
- 24. Savaş, S., et al., Classification of Carotid Artery Intima Media Thickness Ultrasound Images with Deep Learning. Journal of medical systems, 2019. **43**(8): p. 273.
- 25. Biswas, M., et al., *Deep learning fully convolution network for lumen characterization in diabetic patients using carotid ultrasound: a tool for stroke risk.* Medical & biological engineering & computing, 2019. **57**(2): p. 543-564.

- 26. Saba, L., et al., *Ultrasound-based carotid stenosis measurement and risk stratification in diabetic cohort: a deep learning paradigm.* Cardiovascular Diagnosis and Therapy, 2019. **9**(5): p. 439.
- 27. Jamthikar, A.D., et al., *Rheumatoid arthritis: its link to atherosclerosis imaging and cardiovascular risk assessment using machine-learning-based tissue characterization*, in *Vascular and Intravascular Imaging Trends, Analysis, and Challenges, Volume 2.* 2019, IOP Publishing. p. 4-1-4-29.
- 28. Sanagala, S.S., et al., Ten Fast Transfer Learning Models for Carotid Ultrasound Plaque Tissue Characterization in Augmentation Framework Embedded with Heatmaps for Stroke Risk Stratification. Diagnostics, 2021. 11(11): p. 2109.
- 29. Mohan, A., et al. *Internet of video things in 2030: A world with many cameras.* in 2017 *IEEE international symposium on circuits and systems (ISCAS).* 2017. IEEE.
- 30. Mateen, M., et al., Fundus image classification using VGG-19 architecture with PCA and SVD. Symmetry, 2019. **11**(1): p. 1.
- 31. Lam, C., et al., *Automated Detection of Diabetic Retinopathy using Deep Learning*. AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, 2018. **2017**: p. 147-155.
- 32. Narayan Das, N., et al., *Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-rays*. Ingenierie et recherche biomedicale: IRBM = Biomedical engineering and research, 2020: p. 10.1016/j.irbm.2020.07.001.
- 33. Biswas, M., et al., Deep learning strategy for accurate carotid intima-media thickness measurement: An ultrasound study on Japanese diabetic cohort. Comput Biol Med, 2018. 98: p. 100-117.
- 34. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.
- 35. Agarwal, M., et al., A Novel Block Imaging Technique Using Nine Artificial Intelligence Models for COVID-19 Disease Classification, Characterization and Severity Measurement in Lung Computed Tomography Scans on an Italian Cohort. Journal of Medical Systems, 2021. 45(3): p. 1-30.
- 36. Nicolaides, A.N., et al., Asymptomatic internal carotid artery stenosis and

- cerebrovascular risk stratification. Journal of Vascular Surgery, 2010. **52**(6): p. 1486-1496.e5.
- 37. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics*, 2019. CA: a cancer journal for clinicians, 2019. **69**(1): p. 7-34.
- 38. Bain, C., et al., Lung cancer rates in men and women with comparable histories of smoking. Journal of the National Cancer Institute, 2004. **96**(11): p. 826-834.
- 39. Battista, R. and S. Grover, *Early detection of cancer: an overview*. Annual Review of Public Health, 1988. **9**(1): p. 21-45.
- 40. Barnes, R., *Noninvasive diagnostic assessment of peripheral vascular disease.* Circulation, 1991. **83**(2).
- 41. Alakwaa, W., M. Nassef, and A. Badr, *Lung cancer detection and classification with 3D convolutional neural network (3D-CNN)*. Lung Cancer, 2017. **8**(8): p. 409.

#### **Publications**

#### **Published & Accepted**

- Skandha *et al.* "3-D optimized classification and characterization artificial intelligence paradigm for cardiovascular/stroke risk stratification using carotid ultrasound-based delineated plaque: Atheromatic<sup>TM</sup> 2.0." **Computers in Biology and Medicine** 125 (2020): 103958. (IF: 4.589) (Indexed in SCI/Scopus/WoS).
- 2 Skandha *et al.* "A Multicenter Study on Carotid Ultrasound Plaque Tissue Characterization and Classification Using Six Deep Artificial Intelligence Models: A Stroke Application." **IEEE Transactions on Instrumentation and Measurement** 70 (2021): 1-12. (IF: 4.47) (Indexed in SCI/Scopus/WoS).
- 3 Skandha *et al.* "Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system." **The International Journal of Cardiovascular Imaging** 37.5 (2021): 1511-1528. (IF: 2.357) (Indexed in SCI/Scopus/WoS).
- 4 Skandha *et al.* "Multimodality carotid plaque tissue characterization and classification in the artificial intelligence paradigm: A narrative review for stroke application." **Annals of Translational Medicine** 9, no. 14 (2021). (IF: 3.932) (Indexed in SCI/Scopus/WoS).
- Skandha *et al.*, "Ten Fast Transfer Learning Models for Carotid Ultrasound Plaque Tissue Characterization in Augmentation Framework Embedded with Heatmaps for Stroke Risk Stratification." **Diagnostics** 11, no. 11 (2021): 2109. (IF: 3.706) (Indexed in SCI/Scopus/WoS).
- 6 Skandha *et al.* A hybrid deep learning paradigm for carotid plaque tissue characterization and its validation in multicenter cohorts using a supercomputer framework, **Computers in Biology and Medicine**, Volume 141, 2022,105131, ISSN 0010-4825, (IF: 4.589). (Indexed in SCI/Scopus/WoS).
- Skandha *et al.* "A Fast and Light Weight Deep Convolution Neural Network Model for Cancer Disease Identification in Human Lung(s)," **2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)**, Boca Raton, FL, USA, 2019, pp. 1382-1387, Doi: 10.1109/ICMLA.2019.00225. (Core rank: B).

+

8 Siva Skandha et al. "A Novel Genetic Algorithm-based Approach for Compression and Acceleration of Deep Learning Convolution Neural Network: An Application in Computer Tomography Lung Cancer Data "in **Neural Computing and Application** (IF: 5.606). (Indexed in SCI/Scopus/WoS).

#### **Under Review**

- 9 Skandha et al. "Magnetic Resonance-based Wilson Disease Tissue Characterization in Artificial Intelligence Framework using Transfer Learning" Book chapter in **IOP** press.
- 10 Skandha et al. "Understanding COVID severity for COVID-19 Lung Characterization via Artificial Intelligence-Transfer Learning "Book chapter in **IOP press**.

# Profile of the Siva Skandha Sanagala



**Siva Skandha Sanagala**, completed his M.Tech from Bharath University, Chennai. He is into academics and research from 2009 onwards. Presently he is working in CMR College of Engineering & Technology as an assistant professor. He is currently pursuing PhD in computer science in the field of deep learning applications in medical imaging from Bennett University.