

AUTOMATED QUALITY ENHANCEMENT, MODELLING AND MANAGEMENT OF DIAGNOSTIC SCAN IMAGES WITH AI TECHNIQUES

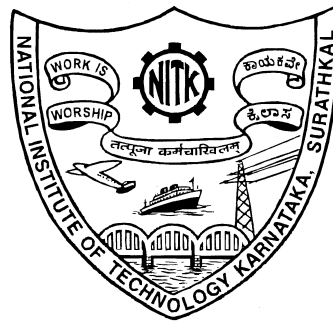
THESIS

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

KARTHIK K.



DEPARTMENT OF INFORMATION TECHNOLOGY
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALURU - 575 025

MAY 2022

DECLARATION

I hereby declare that the Research Thesis entitled “**AUTOMATED QUALITY ENHANCEMENT, MODELLING AND MANAGEMENT OF DIAGNOSTIC SCAN IMAGES WITH AI TECHNIQUES**” which is being submitted to **National Institute of Technology Karnataka, Surathkal** in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy in Information Technology** is a bonafide report of the research work carried out by me. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

Place : NITK - Surathkal
Date : 03rd May 2022



KARTHIK K.
Reg.No.: 177149IT500
Department of IT,
NITK Surathkal.

CERTIFICATE

This is to certify that the Research Thesis entitled, “**AUTOMATED QUALITY ENHANCEMENT, MODELLING AND MANAGEMENT OF DIAGNOSTIC SCAN IMAGES WITH AI TECHNIQUES**” submitted by **KARTHIK K. (Reg. No. 177149IT500)**, as the record of research work carried out by him, is accepted as the Research Thesis submission in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy**.

Place : NITK - Surathkal
Date : 03rd May 2022


Dr. Sowmya Kamath S

Research Guide
Assistant Professor Grade I
Department of IT
NITK Surathkal.


Dr. Jaidhar C D
(HOD, IT)

Chairman - DRPC
Department of IT
NITK Surathkal.

Chairman - DUGC / DPGC / DRPC
Department of Information Technology
NITK - Surathkal, Srinivasnagar P. O.,
Mangalore - 575 025, INDIA

Dedicated to

*My God, Guru, Wife, Father's Memory and all my
family members who were always there for me by giving
all the inspiration and support I need.*

Acknowledgements

Firstly, I would like to thank my esteemed supervisor Dr. Sowmya Kamath S., for her invaluable guidance, support and tutelage during the course of my Ph.D. degree. Her immense knowledge and guidance steered me in the right direction which helped me in all the time of my research, writing thesis and also in my daily life. It's an honor to work under an eminent guide for my Ph.D. study.

I want to express my sincere gratitude to my RPAC panel: Dr. Aparna P., Dept. of ECE, and Dr. Biju R. Mohan, Dept. of IT, who were involved in the validation for this research, also by their insightful comments and encouragement throughout this research work.

I would like to extend my deepest gratitude to Dr. Shashidhar G. Koolagudi, Dept. of CSE for his encouragement in all the time of my life. My gratitude also extends to the Science and Engineering Research Board, Department of Science and Technology, Government of India for its fellowship and financial support through the Early Career Research Grant to undertake my research studies.

I would also like to extend my thanks to Dr. Surendra U. Kamath, Head of Department of Orthopaedics, KMC Hospital, Manipal University and Dr. Jayakumar Jeganathan, Department of General Medicine, KMC Hospital, Mangaluru for giving their knowledgeable inputs that helped to fine tune and strengthen the works during this research work.

I am also grateful to all faculties and staff of the Department of Information Technology, NITK, for all the support I have got when I need it. I again thank the institution for providing me the necessary platform and opportunity for my research work.

The completion of my research work would not have been possible without the support and nurturing of my wife Mrs. Krithika. She has been incredibly supportive to me throughout this entire process and has made countless sacrifices to help me get to this point.

I like to extend my special thanks to my friends – Gokul S. Krishnan, Mahesh Chikkanna., Shreenivasa K., Veena Mayya, Ashwin T. S., Sanjay S. Bankapur,

Reshma U., Shashank S., Sanket Salvi S., Ranjit P. K., and others who all had supported me at all times and also for sharing their experiences of hard times, motivating me for moving ahead when I required it. Also, I thank my fellow lab mates and fellow research scholars for the stimulating discussions, working together in many events and for all the fun we had in the last four years.

Finally, I would like to thank my family members for supporting me spiritually throughout writing this thesis and my life in general. Last but not least, I thank all others who have helped or supported me in one way or the other in accomplishing the completion of my research work.

KARTHIK K.

Abstract

Diagnostic scanning is extensively used for investigation of internal organ-related ailments and managing patient care. With the proliferation of imaging-based diagnostic procedures in healthcare, patient-specific scan images constitute huge volumes of data, thus creating a need for automated healthcare information management systems (HIMS) to facilitate their efficient organization and management, and for supporting clinical decision support applications. Medical images often require varied processing for enabling effective representation and modeling for building higher-level decision-support applications. One of the critical gaps in automated systems is limited attention to certain standards in meeting the quality of the scanned images. Compounding this problem is the availability of multi-vendor, non-standard scan resolution machines and also ill-trained medical technicians. Automatically making computers understand the content of an image and offering a reasonable description in natural language has gained more attention recently in computer vision and natural language processing research communities. The caption prediction task in the medical domain is thus very relevant, as it aims to generate textual descriptions of the images, which can be used to improve indexing mechanisms in HIMS.

The focus of the research work presented in this thesis is on building an effective framework for medical image representation, modeling and management, for enabling advanced clinical applications like similarity based diagnostics, decision support, etc. In clinical diagnosis, diagnostic images that are obtained from the scanning devices serve as preliminary evidence for further investigation in the process of delivering quality healthcare. However, often the medical image may contain fault artifacts introduced due to noise, blur and faulty equipment. The reason for this may be low-quality or older scanning devices, the test environment or technician's lack of training etc, however, the net result is that the process of fast and accurate diagnosis is hampered. Towards this, automated image quality improvement approaches are adapted and benchmarked for the task of medical image quality enhancement through super-resolution.

To design approaches for leveraging the enhanced medical images for further analysis and modeling for supporting applications like categorization, retrieval and automated captioning using machine learning and deep learning techniques, the concept of Content-based Medical Image Retrieval (CBMIR) systems is incorporated. The CBMIR system designed can model heterogeneous views, body orientation, etc for supporting similar image retrieval for diagnosis. In diagnostic medical images, the patient body orientation or view of the scanning posture like anterior or frontal view, posterior or back view and the lateral or side views, also known as left lateral or right lateral can be used during scanning. However, computer-aided diagnosis systems often do not provide this piece of header information of the image. Hence, image orientation identification is essential for qualitative and quantitative analysis in diagnostic applications. If such patient body orientations are not recorded or are documented using an incorrect label, automated system indexing may be inconsistent, and may also result in improper interpretation by computers and radiologists. Thus, a learnable neural model for accurately identifying the view positions of different organs of the body is proposed and designed.

For a radiologist to delineate the imaging study's findings/observations as a textual report is a manual, time consuming and tedious task, further exacerbated by the volume of generated images. Automated methods for radiographic image examination for identifying abnormalities and generating reliable radiology report are thus a critical requirement in clinical workflow management applications. The features extracted using neural network architectures are used to automatically generate the diagnosis medical report for scanned images, thus providing a way to build a robust medical imaging application for quality diagnosis. The promising achieved results underscore the performance of the approaches designed in this research and reveal much scope for adaptation in the healthcare field for improving the quality of healthcare delivery and management.

KEYWORDS: *Image Quality Assessment, Image Super-Resolution, Content Based Medical Image Retrieval, Natural Language Processing, Artificial Intelligence*

Contents

Abstract	ii
Abbreviations	xvi
Nomenclature	xxi
 Part I - Introduction and Background	
1 Introduction	1
1.1 Diagnostic Image Management	4
1.2 Medical Image Representation	6
1.3 Dealing with variance in medical images	7
1.4 Automating Diagnostic Image Management	8
1.5 Prevalent Challenges and Issues	9
1.5.1 Thesis Contributions	11
1.6 Summary	12
1.7 Thesis Organization	12
 2 Literature Review	 15
2.1 Background	15
2.2 Related Work	16
2.2.1 Medical Image Quality Management	16
2.2.2 Medical Image Modeling and Representation	21
2.2.3 Dealing with variance in Medical Images	27
2.2.4 Automating Medical Image Understanding	31
2.3 Outcome of Literature Review	33
2.4 Summary	35
 3 Problem Description	 37
3.1 Background	37

3.2	Scope of the Work	37
3.2.1	Problem Statement	38
3.2.2	Research Objectives	38
3.3	Brief Overview of Proposed Methodology	40
3.3.1	Medical Image Quality Enhancement	40
3.3.2	Medical Image Modeling and Representation	40
3.3.3	Dealing with Variance - View Classification	42
3.3.4	Generating Medical Image Descriptions	42
3.4	Research Contributions	43
3.5	Summary	44

Part II - Medical Image Quality Enhancement

4	Medical Image Quality Enhancement	47
4.1	Introduction	47
4.1.1	Problem Definition	48
4.2	Radiography Image Quality Improvement	49
4.2.1	Experimental Results and Discussion	53
4.3	Neural Super-Resolution Models for Quality Enhancement	55
4.4	Experimental Evaluation and Results	58
4.5	Summary	62

Part III - Medical Image Modeling and Representation

5	Medical Image Modeling and Representation	67
5.1	Introduction	67
5.1.1	Problem Definition	67
5.2	Hybrid Feature Modeling for Content-Based Medical Image Retrieval	68
5.2.1	Noise Removal & Contrast Enhancement	68
5.2.2	Feature Modeling	70
5.2.3	Pair-wise Similarity & Class Label Prediction	73
5.2.4	Content-based Image Retrieval	74
5.3	Experimental Results and Discussion	74
5.4	Swarm Intelligence based BoVW Model for CBIR	79
5.4.1	Preprocessing and Feature Generation	79
5.4.2	Visual Vocabulary Construction and Training	81
5.4.3	Image Indexing and Retrieval	83

5.4.4	PSO based Retrieval Optimization	83
5.5	Experimental Results and Discussion	86
5.5.1	Classification and Retrieval Results	87
5.6	Deep Neural Models for Effective CBMIR	92
5.6.1	Content-Based Image Retrieval Task	95
5.7	Experimental Results and Discussion	97
5.7.1	Classification Task	97
5.7.2	Retrieval Task	100
5.7.3	Benchmarking against State-of-the-art Works	102
5.8	Summary	104

Part IV - Dealing with Variance in Body Orientation Views

6	Medical Image View Classification	109
6.1	Introduction	109
6.1.1	Problem Definition	110
6.2	Body Orientation: Multi-View Classification	110
6.2.1	Experimental Results and Discussion	112
6.3	Deep Neural Models for X-ray View Orientation Classification . . .	113
6.4	Experimental Results and Discussion	118
6.5	Summary	120

Part V - Automatic Generation of Medical Image Descriptions

7	Generating Medical Image Description	125
7.1	Introduction	125
7.1.1	Problem Definition	126
7.2	Abnormality Detection and Classification of Plain Radiographs . . .	126
7.2.1	Abnormality Classification	127
7.2.2	Abnormal Region Detection	129
7.2.3	Automatic Diagnostic Text Report Generation	130
7.3	Experimental Evaluation and Results	132
7.4	Automated Multi-task Diagnostic Scan Management	139
7.4.1	Scan Quality Enhancement	139
7.4.1.1	Image Quality Assessment	142
7.4.2	Orientation Classification	142
7.4.3	Diagnostic Report Generation	143

7.5	Experiments Results and Discussion	145
7.5.1	Ablation Study	149
7.6	Learning COVID-19 Disease Representations from Multimodal Data	151
7.6.1	Chest X-ray based Screening of COVID-19	151
7.6.2	Automatic Diagnostic Report Generation	151
7.7	Experimental Results & Discussion	153
7.7.1	Chest X-ray Classification for COVID-19 Diagnosis.	153
7.7.2	Automated chest X-ray text report generation task.	154
7.8	Summary	155
8	Conclusion & Future Work	159
8.1	Future Work	161
	Publications based on Research Work	162
	References	164

List of Figures

1.1	Using the Virtual Grid for Radiography	4
3.1	Overall workflow of the proposed framework	39
3.2	Medical Image Quality Enhancement Process	41
3.3	Medical Image Modeling and Representation process	42
3.4	Orientation Identification process	42
3.5	Medical Image Caption Generation process	43
4.1	Proposed Radiography Image Quality Improvement Approach.	49
4.2	Sample images and corresponding histograms	51
4.3	Comparative Evaluation of UM, CLAHE, Bicubic, VDSR and SR-CNN for X-ray image enhancement.	54
4.4	Illustration of X-ray Image Enhancement with quality metrics.	54
4.5	Deep CNN model for Automated Scan Quality Enhancement	56
4.6	Sample images from IXI dataset	57
4.7	PSNR performance of various models	61
5.1	Workflow of the Proposed CBMIR Model	69
5.2	Radiographic image enhancement process	70
5.3	HCSF extraction from X-ray images	72
5.4	Sample IRMA dataset images and their IRMA codes	75
5.5	AUROC values for <i>one versus all</i> classes using Cosine kNN classifier	76
5.6	Observed retrieval results (classes with high accuracy)	78
5.7	Observed retrieval results (classes with average accuracy)	79
5.8	Proposed BoVW+PSO approach for CBMIR	80
5.9	SURF feature generation process	81
5.10	Histogram of visual word occurrences with different vocabulary sizes	82
5.11	Sample images from IRMA dataset.	86
5.12	Observed performance for varying vocabulary sizes	88
5.13	Feature Point Representation using PSO	89
5.14	Classification Results for different classes	89

5.15	Image Retrieval results without using Filter approach and PSO . . .	90
5.16	Image Retrieval results using Filter approach and PSO	90
5.17	Proposed CNN model for Radiograph classification	92
5.18	Dataset image sample and specifics	98
5.19	Classification performance of proposed CNN model for various classes	101
5.20	Observed retrieval results for Best-match classes	102
5.21	Observed retrieval results for Average-match classes	102
6.1	Predicted orientation label on images of <i>Cervical Spine</i> class	112
6.2	Predicted orientation label on images of <i>Neuro Cranium</i> class . . .	113
6.3	Proposed Approach for View/Body Orientation Classification . . .	114
6.4	Architecture of the proposed ViewNet model	117
6.5	Sample dataset images showing the IRMA class code and code de- scription.	118
7.1	Abnormality Classification and Report Generation process.	127
7.2	Architecture of the proposed MSDNet model	128
7.3	Automatic Report Generation Process for Chest X-ray Images. . . .	131
7.4	Sample images of MURA dataset in Hand, Forearm, Wrist, Finger, Shoulder, Humerus and Elbow classes	133
7.5	AUROC performance of the proposed Ensemble Model.	136
7.6	Illustration of the abnormal area detection process for sample im- ages from the <i>Shoulder</i> class.	137
7.7	Sample model generated report, with the ground-truth data.	137
7.8	Proposed Automated Multi-task Diagnostic Scan Management. . .	140
7.9	Architecture of proposed SRResNet.	140
7.10	RB and RRDB in SRGAN.	141
7.11	ViewNet Model for Body Orientation Classification.	143
7.12	Architecture of the Automated Diagnostic Report Generation Model	144
7.13	ESRGAN performance evaluation	146
7.14	Example of BLEU match with n-gram approach.	148
7.15	Visual Comparison representing the outcome of each component in ESRGAN	150

List of Tables

2.1	Summary of Existing works in Medical Image Quality Management	19
2.2	Summary of Medical Image Modeling and Representation.	26
2.3	Summary of View Orientation Classification based works.	30
2.4	Summary of Existing works on Medical Image Description Generation.	32
4.1	Quality Evaluation Metric Scores across different Models	61
4.2	Benchmarking the proposed ResNet SRCNN model against State-of-the-art.	62
5.1	Summary of Feature Modeling processes	73
5.2	Observed results for the Standard Euclidean Pairwise distance	75
5.3	Benchmarking the proposed approach with existing works	76
5.4	Classification accuracy for different kNN variants	76
5.5	Evaluation of retrieval with precision@ k for $k=3, 5, 10$	77
5.6	Evaluation of retrieval performance for all 116 classes	78
5.7	Observed performance on Training and Test sets	87
5.8	Evaluation of retrieval with precision@ k for $k = 5, 10, 15, 20$	91
5.9	Benchmarking the proposed approach against State-of-the-art models	91
5.10	Performance of proposed CNN model w.r.t distance measures	99
5.11	Classification results of Base AlexNet Model in comparison to various distance measures.	100
5.12	Evaluation of retrieval with precision@ k for $k=3, 5, 10$. (Average-match classes)	103
5.13	Benchmarking the proposed model against state-of-the-art approaches, using Error Score and Retrieval accuracy metrics.	104
6.1	Classification Model parameters	117
6.2	Observed classification performance w.r.t different CNN Models (<i>before class label refinement</i>)	120
6.3	Observed classification performance w.r.t different CNN Models (<i>After class label refinement</i>)	120

7.1	Sample images from the Indiana Chest X-ray dataset, along with the associated indications, findings and impressions	133
7.2	Classification Performance w.r.t different classes for the proposed MSDNet model	135
7.3	Classification Accuracy of various models	135
7.4	Proposed model's performance w.r.t Kappa Score against State-of-the-art models	138
7.5	Benchmarking proposed model against state-of-the-art models using standard metrics.	138
7.6	Performance of ESRGAN for the <i>Frontal</i> and <i>Lateral</i> classes	146
7.7	Orientation classification performance with different CNN Models .	147
7.8	Comparison of Report generation performance with and without BN layer.	150
7.9	Classification model parameters.	152
7.10	Performance evaluation of the chest X-ray image classification task	154
7.11	Performance evaluation of chest X-ray report generation	154

Abbreviations

AHE	Adaptive Histogram Equalization
AI	Artificial Intelligence
AMBE	Absolute Mean Brightness Error
ANN	Artificial Neural Network
AP	Antero-Posterior
AUC	Area Under the ROC Curve
BLEU	Bilingual Evaluation Understudy
BoVF	Bag of Visual Features
BoW	Bag of Words
BPHE	Brightness Preserving Histogram Equalization
CAD	Computer Aided Diagnosis
CBIR	Content-Based Image Retrieval
CBMIR	Content-Based Medical Image Retrieval
CBSF	Contour-Based Shape Feature
CDSS	Clinical Decision Support Systems
CI	Confidence Intervals
CLAHE	Contrast limited adaptive histogram equalization
CNN	Convolutional Neural Network
CT	Computed Tomography
CV	Computer Vision
DCNN	Deep Convolutional Neural Network
DCP	Dark Channel Prior
DICOM	Digital Imaging and Communications in Medicine
DNN	Deep Neural Network
DoG	Difference of Gaussian
FFT	Fast Fourier Transform
FN	False Negatives
FP	False Positives

FPR	False Positive Rate
GCE	Grad-Contrast Enhancement
HE	Histogram Equalization
HIMS	Healthcare Information Management Systems
HOG	Histogram of Oriented Gradients
HR	High Resolution
HSV	Hue, Saturation, and Value
HVS	Human Visual System
IP	Image Profile
LBP	Local Binary Patterns
LR	Low Resolution
LSTM	Long Short Term Memory
MAP	Mean Average Precision
MDI	Medical Diagnostic Imaging
MedIR	Medical Image Retrieval
MISR	Multi-image Super Resolution
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
MSER	Maximally Stable Extremal Regions
MS-SSIM	Multi-scale Structural Similarity Measure
NICU	Neonatal Intensive Care Unit
p@K	Precision@k
PA	Posterior-Anterior
PACS	Picture Archiving and Communication Systems
PAS	Predictive Analytics Systems
PCA	Principal Component Analysis
PET	Positron-emission Tomography
PGHD	Patient Generated Health Data
PHOG	Pyramid of Histograms of Orientation Gradients
PSNR	Peak Signal to Noise Ratio
PSO	Particle Swam Optimization
QBE	Query By Example
ReLU	Rectified Linear Unit
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
SFL-CT	Sharp Frequency Localization-Contourlet Transform
SGDM	Stochastic Gradient Descent Momentum

SIFT	Scale-Invariant Feature Transform
SISR	Single Image Super Resolution
SR	Super Resolution
SRCNN	Super-Resolution Convolutional Neural Network
SSIM	Structural Similarity Index Measure
SURF	Speed Up Robust Feature
SVM	Support Vector Machines
UMLS	Unified Medical Language System
VIF	Visual Information Fidelity
WebMIRS	Web-based Medical Information Retrieval System
WTE	Wavelet Transform Enhancement

Nomenclature

Notation	Meaning	Chapter No.
f_1, f_2	Patch Size of LR and HR images	4
n_1, n_2	HD vector	4
I, \hat{I}	Ground truth and reconstructed image	4
M	Numbers of rows of an image	4
N	Numbers of columns of an image	4
X_{ij}, Y_{ij}	Intensity of original and reconstructed image	4
C_1, C_2	Constants	4
σ_x, σ_y	Contrast comparison functions	4
μ	Mean	4
σ	Standard deviation	4
I	Input tensor	4
O	Output tensor	4
p	Pooling sub-tensor	4
l_M, c_j, s_j	luminance, contrast and structure comparison functions	4
α, β, γ	Weights set to 1	4
I	Maximum intensity of a grayscale image	4
C	Reference image	4
F	Distorted image	4
E	Image that the Human Visual System perceives	4
I^T	Transpose of Image I	5
Q	Number of query/test images	5
v_i	Current velocity of the particle	5
y_i	Best position of the particle	5
w	Inertia weight	5
c_1, c_2	Acceleration constants	5
λ	Regularization strength	5
n_d	Digits, where, $n_d \in \{3,4\}$	5
B_j^{ik}	Number of possible labels at position i	5
M	Number of training examples	7
K	Number of classes	7
w_k	Weight for class k	7
y_m^k	Target label for training example m for class k	7
x_m	Input for training example m	7

Notation	Meaning	Chapter No.
h_θ	Represents a model with neural network weights θ	7
ℓ	Iteration number	7
θ	Parameter vector	7
γ	Contribution of previous gradient step to current iteration	7
\hat{y}_c	Model's prediction of a class	7
$Pr(a)$	Actual observed agreement	7
$Pr(c)$	Chance agreement	7
L_1	Evaluates 1-norm between recovered and ground-truth image	7
$C(i)$	Number of i -gram tuples in candidate document	7

PART I

Introduction and Background

Chapter 1

Introduction

Medical imaging is a multidisciplinary field with an intersection of varied technologies like computer-aided modeling/design and mathematics applied to the field of medicine. Over the past two decades, there have been significant advancements in medical imaging technologies for disease diagnosis, through various imaging modalities like X-Ray, MRI (Magnetic Resonance Imaging), CT Scans (Computed Tomography), PET (Positron Emission Tomography) etc. These procedures have become the *de-facto* standard for facilitating the diagnosis of many common diseases in modern healthcare.

Healthcare quality and patient safety are deeply connected as healthcare quality is a broad term that encompasses many aspects of patient care. Quality healthcare is care that is safe, effective, patient-centered, timely, efficient, and equitable¹. Quality of healthcare is an important aspect in the promotion of health and well-being of people around the world. According to published reports², the US healthcare system has extensively invested in the use of Information Technology in healthcare services, with a budget that exceeds the world average income by more than eight times. India, the world's second most densely populated country, has made vast strides towards the implementation of nation-wide healthcare systems for its people. The Indian Constitution makes the provision of healthcare in India the responsibility of the state governments, and makes every state responsible for "raising the level of nutrition and the standard of living of its people and the improvement of public health as among its primary duties."

The National Health Policy was endorsed by the Parliament of India in 1983 and has been updated twice since, in 2002 and 2017. In 2017, four main updates have been proposed that stress the need to focus on the emergence of the

¹Crossing the Quality Chasm: A New Health System for the 21st Century. Online: <https://www.nap.edu/read/10027/chapter/1#iii>

²WHO Global Health Expenditure Database. Online: <http://apps.who.int/nha/database>

robust healthcare industry, growing incidences of unsustainable expenditure due to healthcare costs and rising economic growth enabling enhanced fiscal capacity (Sekher, 2013). In practice, the private healthcare sector is a major healthcare provider in India, and most healthcare expenses are paid directly out of pocket by patients and their families, rather than through health insurance (Berman *et al.*, 2010). Government health policy has thus far largely encouraged private-sector expansion in conjunction with well designed but limited public health programmes (Britnell, 2015).

Ayushman Bharat, an ambitious government-funded health insurance project launched by the Government of India in 2018, aims to cover the lower 50% of the country's population and offer them free treatment even at private hospitals (Zodpey and Farooqui, 2018). The two interrelated components of Ayushman Bharat are: 1) Health and Wellness Centres (HWCs) and 2) National Health Protection Scheme like Pradhan Mantri Jan Arogya Yojana (PM-JAY). Health and Wellness Centres are envisioned as a foundation of the health system to provide comprehensive primary care, free essential drugs and diagnostic services. National Health Protection Scheme is envisaged to provide financial risk protection to poor and vulnerable families arising out of secondary and tertiary care hospitalization (AyushmanBharat, 2018) to the tune of five lakh rupees per family per year which will enable the realization of the aspiration for UHC. This digital transformation story of healthcare in India has seen a technology-led program that aims to connect 500 million citizens to a nationwide network of hospitals via a single comprehensive process – from registration to cashless hospitalization and discharge. Typically, this connectivity would be weighed down by reams of documentation, a slew of sub-processes, and multiple physical interventions at every stage. Now, this re-imagined digital healthcare solution brings together central and state medical facilities to enrol citizens, empanel hospitals, process medical claims and generate auto approvals.

Modern healthcare has emerged as a data-driven ecosystem, one that is slowly realizing the utility and value of collecting varied health related data. Healthcare data comprises of a wide variety – claim data, electronic health records, administrative data, medical records, clinical trial, drug responses, research data, and so on. The emergence of eHealth and mHealth (mobile health) have expanded the definition of health data by creating new opportunities for patient-generated health data (PGHD) (Shapiro *et al.*, 2012). Digital health prescribes a patient-centric healthcare system in which patients are empowered by the availability of their health data, so that they can manage their own health and wellness with

assistive/wearable technologies³. Availability of such valuable streaming data enables providers to deliver personalized treatment, model patient-specific health risks and plan for adverse events.

Medical Diagnostic Imaging (MDI) is critical to modern healthcare, due to the availability of varied procedures for constructing the visual representations for internal organs of the human body for diagnosis and clinical interpretation. Diagnostic procedures like X-rays, CT and PET scans are extensively used in modern medical treatment and delivery, the net result being the generation of large volumes of data adding to the burden of hospital data management systems. Diagnostic scan images characterize the health assessment of the human body on various levels, such as microscopic, macroscopic, etc. The role of medical imaging in clinical diagnosis, treatment planning and procedures cannot be overstated. Due to the availability of advanced, state-of-the-art software and hardware, there has been rapid advancement in the field of medical imaging technology. The global medical imaging market was valued at \$25.7 billion in 2016 and is forecast to grow at a modest 5.4% between 2017 and 2024, with a projected global sales of \$43.3 billion in 2021 ([Report, 2017, 2021](#)).

The proliferation of medical image data from hospitals, documented in digital forms is a significant and valuable resource for improving diagnosis. According to published reports ([Report, 2019, 2021](#)), the global diagnostic imaging market is projected to reach USD 33.5 billion by 2024 from USD 25.7 billion in 2019, at a CAGR of 5.5% from 2019 to 2024. Based on modality, the diagnostic imaging services market is segmented into MRI, ultrasound, CT, X-ray, nuclear imaging, and mammography. In 2018, the X-ray segment accounted for the largest share of the market. The key factors driving the growth of this market include the rising geriatric population, lower cost of X-rays as compared to other imaging modalities, favorable returns on investments and technological advancements in X-ray imaging systems. Thus, diagnostic imaging data repositories are now a major source of knowledge for detailed analysis related to diseases like cancer, tumors, fractures, *etc.*, enabled through intelligent Healthcare Information Management Systems (HIMS). Efficient collection and management of diagnostic imaging and related data can be utilized for supporting intelligent decision-support applications, which is increasingly becoming a critical requirement. Extraction of clinically relevant information or knowledge from medical images for building CDSSs and predictive analytics systems have thus garnered much research interest in the past decade.

³<https://www.fda.gov/medical-devices/digital-health-center-excellence>

1.1 Diagnostic Image Management

The digitization of X-ray radiography devices and advancements leading to mobile X-ray devices, has resulted in portable examinations at the bedside of hospitalized patients and in the NICU (Neonatal Intensive Care Unit). Grid (a metallic filter made of lead strips), which removes scattered X-rays, is widely used in clinical settings since it improves image contrast. If X-rays penetrate to the grid at an oblique angle; an image of uneven density may result. However, in portable examinations where the grid is often slanted due to misalignment of the bed and other factors, uneven densities (Refer Fig.1.1a) can result, making it difficult to read the image. Medical technicians sometimes perform portable examinations without a grid although they understand that it may cause deterioration of image quality due to scattered X-rays (shown in Fig.1.1b). Virtual Grid is an image processing technology that converts deteriorated image quality due to scattered X-rays (Fig.1.1b) to an improved quality image (Fig.1.1c) by reducing the effect of scattered X-rays ([Kawamura *et al.*, 2015](#)).

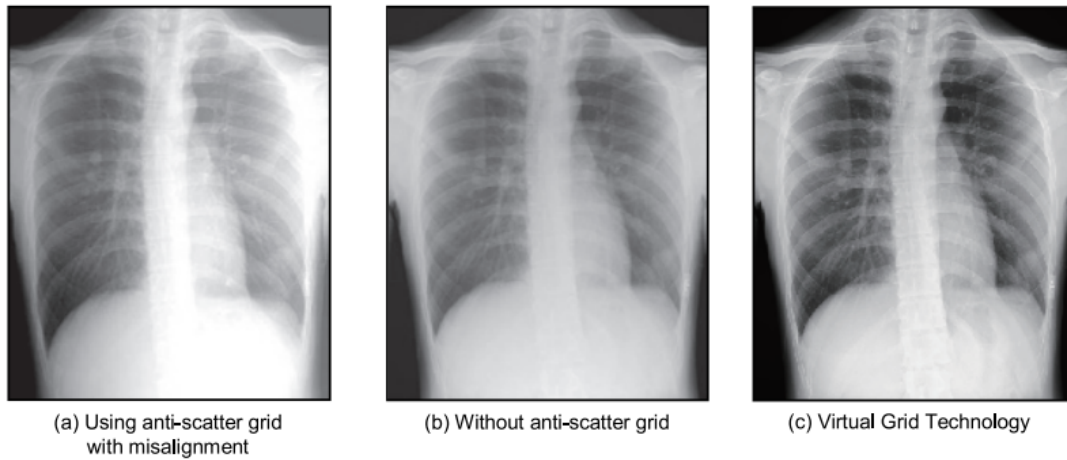


Figure 1.1: Using the Virtual Grid for Radiography ([Kawamura *et al.*, 2015](#))

Some quality control parameters adopted for achieving high-quality digital radiography include -

1. Inclusion of validated imaging protocols so that the consistency of image quality and radiation dose can be established and maintained between rooms and between sites;
2. Use of appropriate compression techniques on image data to facilitate transmission or storage, without loss of clinically significant information;
3. Archiving of data to maintain accurate patient medical records in a form that may be retrieved in a timely fashion;

4. Application of image processing for better display of acquired information;
5. Promotion of clinical efficiency and continuous quality improvement.

The earliest clinical insights are generally obtained via different modalities of medical images such as X-rays scans, CT and MRI among others (Binh and Tuyet, 2015). Hence, acquiring good quality diagnostic images is essential for analyzing and determining disease onset, criticality and progression. Due to the technical restrictions and various economic and physical conditions, there exist several challenges in obtaining a good quality diagnostic image, like, low resolution (LR) images, under-exposure or over-exposure, occurrence of artifacts introduced by faulty or older scanning equipment etc, which can often render diagnostic scans unsuitable for further analysis (Summers, 2012). Due to this, methods that improve the spatial resolution of medical images are gaining increasing importance in clinical workflow management systems.

To create a high resolution (HR) medical image, numerous image enhancement algorithms have been proposed. Image super-resolution (SR) by Ha *et al.* (2018) is an approach of restoring high-resolution images from images of lower resolution. Super Resolution can be categorized as Single Image Super Resolution (SISR) and Multi-image Super Resolution (MISR) based on a total of low-resolution images taken as input. SISR can be defined as a method where one low-resolution input image is utilized to restore high-resolution image details. MISR is primarily a reconstruction-based algorithm that takes multiple variants of LR images and attempts to combine them for recovering the details of the HR image. Further, enhancing medical images (Miffin, 2007) can help medical experts for evaluating diagnosis accurately with more details in pathology research. As a result, medical image enhancement process can substantially enhance the accuracy of computer-aided automatic detection (Ramakrishna *et al.*, 2009).

Image super-resolution is an active research field, however, many medical diagnostic image modalities do not lend themselves well to it. This is due to the inherent structure and manifold information contained in the medical scans. Thus, there is a need for super-resolution techniques to enhance the poor quality images through computational means. This minimizes the overall burden in analysis of the disease for physicians, saving their time and also aiding in accurate diagnosis, specifically in cases where the disease is in its early stages. This also reduces the need for rescanning, wastage of medical resources, cost, effort, time of patients, medical personnel and hospital administration.

1.2 Medical Image Representation

The proliferation of medical image data from hospitals, documented in digital forms, is a significant asset for diagnostic medical informatics. Such diagnostic tools contribute varied types of medical images like X-rays, CT and PET scans, generated in huge volumes continuously, given a large number of patients seeking medical attention; thus, efficient management becomes critical. For improving the organization and storage of medical scan images, several standards have been proposed and standardized. The DICOM standard (Digital Imaging and Communications in Medicine) (Mildenberger *et al.*, 2002) can be used for storing patient information and their scan images for communication. System like PACS (Picture Archiving and Communication Systems) (Choplin, 1992) was created for improving access to stored DICOM files so that higher-level applications like decision making can be supported. However, most such systems have many restrictions, like the dependency on the text-based search for images, requiring keyword matching capabilities. Thus, image retrieval often suffers from limited accuracy due to ambiguous/sparse textual descriptions of images, non-existent/low-quality image annotations (e.g., test description). Effective techniques called Content-Based Image Retrieval (CBIR) that use pure visual cues to retrieve relevant images from huge image collections were developed to overcome these limitations.

Medical image classification and retrieval is a challenging and active research area with applications in similarity-based automated diagnosis, decision support and medical data management. Keyword-based querying has been traditionally used for Medical Image Retrieval (MedIR), which is typically supported by text-based annotations that are created and stored along with each image, and used during the retrieval process. The PACS (Lehmann *et al.*, 2003a) is a significant effort to overcome these challenges with a focus on storing and retrieving medical images effectively. Despite this, a notable drawback of PACS is that the techniques used were dependent on keywords and related text annotations stored together with an image for retrieval. However, good quality textual annotations are hard to come by, due to the prohibitive manual effort and time constraints. When available, these annotations are often ambiguous due to the unstructured nature of natural language or incomplete, thus adversely affecting retrieval results. Advancements in image management systems have led to the development of CBIR systems, which have been adapted for medical image management.

In contrast to keyword-based image querying systems, a CBIR system aims to capture the latent features from an image without requiring any external infor-

mation (*e.g.* text metadata associated with images). In CBIR systems, potential matching images are identified as per their actual visual content overlap with a given query image and ranked as per their similarity. CBIR makes use of image-level features like texture, shape and color for finding the most significant images during retrieval concerning a given query image. Thus, most CBIR systems are dependent on low-level features like color, shape and texture to obtain relevance rankings. An adaption of CBIR concepts and their application to medical images has resulted in Content-Based Medical Image Retrieval (CBMIR) Systems.

Several prototype systems that assist users in efficient image retrieval have been designed and deployed in the healthcare domain. One of them, ASSERT (Shyu *et al.*, 1999) was specifically developed for lung CT images, and relies both on both text annotations and image-level features for identifying any anomalous regions in images to detect the possibility of disease. Pathfinder (Wang, 2000) and the I-Browse system (Tang *et al.*, 1999) focus on automatic labeling of histopathological images like tissue slides for facilitating retrieval. Similarly, Web-based Medical Information Retrieval System (WebMIRS) (Antani *et al.*, 2002) is a system that focuses on the indexing and retrieval of human lumbar and cervical spine scans only. The IRMA system⁴ mainly focuses on the classification and retrieval of images into anatomical areas, modalities and viewpoints. However, due to the limited feature modeling and also a dependency on text annotations, the adaptability of these systems to multi-modal medical image data is limited. Despite significant improvements in medical image retrieval, conventional retrieval models that support querying based on text/keywords fail to capture the latent visual features in an image, emphasizing the need for effective adaptations of CBIR systems for medical images. As medical images are inherently multi-dimensional and complex in the varied information that can be inferred from them, designing effective feature extraction mechanisms can help to improve overall retrieval accuracy.

1.3 Dealing with variance in medical images

Radiological procedures like X-rays have evolved which is a crucial diagnostic imaging tool for identifying abnormalities in different body parts, which may require insights derived from various views/body orientations of the patient. Often, *frontal view* and *lateral view* are used in such cases. For Computer Aided Diagnosis (CAD), internal and external shapes are very important in identifying the abnormality. While scanning, *i.e.*, during the diagnostic image capturing process,

⁴<https://www.kaggle.com/raddar/irma-xray-dataset>

the scanning equipment is focused on the injured part of the body, and scans are typically performed in different positions to aid effective diagnosis. There are other organs of the body which are taken at different views for proper diagnosis. Therefore, the medical images of the same organ that are taken at varied angles require proper categorization which needs to be trained differently, according to the image view.

Currently, the projection view/ image orientation of radiographs are labeled manually by radiologists and technicians. Manual corrections for wrongly labelled views makes it impractical in PACS and digital imaging systems, as it involves cost and time of human resources. Instead of manually labeling such multi-oriented images, it can be accomplished automatically by intelligent algorithms that are trained to understand the patterns with large-scale images. Methods that can assess this automatically and provide the necessary information regarding the view of the organ at which the scan is taken can be beneficial.

In solving the challenges for view classification, most of the approaches were developed by using traditional hand-crafted features, such as Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT) features, Histogram of Oriented Gradients (HOG) (Xue *et al.*, 2015). There are very limited works when it comes to neural network based orientation classification (Takeuchi *et al.*, 2019). Hence, a method for automatically recognizing the projection view and the patient-relative orientation of different organs of medical images would be very useful because it reduces the incidence of mislabeled or unlabeled images, and saves time in reorienting images and helps improve the image management process. The outcome of such orientation classification model should be accurate, continuous, operating in the real time in the clinical workflow CAD systems.

1.4 Automating Diagnostic Image Management

Automatically making computers understand the content of an image and offering a reasonable description in natural language has gained importance due to the challenges of large volume and streaming nature. In clinical practice, medical specialists and researchers usually write diagnosis reports to record microscopic findings from images, so automatic captioning on medical images will benefit healthcare providers with valuable insights and reduce their burden across the overall clinical workflow. The medical image captioning challenge (de Herrera *et al.*, 2018) aims to advance methodological development in mapping visual information from medical images to condensed textual descriptions. The caption task can be seen

as a part of the medical image classification task (Villegas *et al.*, 2015; Seco *et al.*, 2016) and can be a significant addition to HIMS software.

Image processing and Computer Vision (CV) based techniques have been applied for designing applications for surgical and imaging interventions. Such systems extend clinical decision making capabilities to the healthcare professionals, by automating certain tasks related to diagnosis, or by forecasting the severity of several abnormalities and radiology reports. Incorporating Artificial Intelligence (AI) in these systems to support learning behaviour so that systems can detect abnormalities at the earliest disease onset in a wide variety of diagnostic media like radiology, CT scans, MRIs etc are of critical importance. The radiologist can utilize these insights for enabling and optimizing the quality of diagnosis. The marked region can help the physicians to focus on early and effective treatment recommendations. Further, automated retrieval of radiological diagnosis reports will minimize the manual work involved in reporting observations from the radiological images, while also alleviating the cognitive burden of the radiologists due to the huge load of cases that they typically handle each day.

1.5 Prevalent Challenges and Issues

Despite vast strides achieved in the way medical diagnosis is performed, there exist several prevalent issues that hinder the scale at which the utility of generated volume of clinical knowledge is being exploited. Several challenges were observed that need to be addressed, for full-scale adoption of intelligent automated systems in healthcare delivery and diagnosis. Some of these observations are listed below.

1. **Scarcity of well-annotated medical imaging datasets:** To find a dataset that provides well-documented and labeled data for designing large-scale medical image categorization tasks is significantly more difficult in the medical domain when compared to other domains. Crowdsourcing of images is one of the solution to address this challenge, where people can be asked to share images based on their agreement. One such example is, where COVID positive patients were asked to share their X-ray images, leading to curation of large repositories of Chest X-ray scans. Another way to deal with this is making the data well-documented and open access, i.e., organisations make the anonymized datasets openly available for researchers to carry out their research. Some of the publicly available datasets are CheXphoto, MURA, ImageCLEF and others.

2. **Quality of medical scans:** The acquired medical scans are often poor in quality when compared to natural scene images, which are more clear and of higher resolution when compared to the latter. Often in healthcare delivery scenarios, utmost quality control is difficult, and scanning errors are introduced because of faulty equipment, lab environmental conditions, limited availability of well-trained lab technicians, patient non-cooperation (in case of kids, accident victims, terminally ill patients etc). These factors sometimes affect the quality of the acquired images (Saunders Jr *et al.*, 2007; Boita *et al.*, 2021). Such differences in training and test data can lead to disparity in the quality of training and performance of learning models. Therefore, it is essential to have curated data that provide high-quality and consistent scan images. This is a difficult requirement for most hospitals, due to the hectic pace of patient-centric activities, sheer volume and generation frequency of medical scan data. A solution is seen in the design and development of effective image quality enhancement pipelines, and their incorporation in Medical Image Management Systems, for automatic curation of high-quality data.
3. **Generalisation and variance:** Medical images contain manifold information, and are multimodal in nature. Clinical diagnosis processes are often dependent on multiple scan views with respect to varied patient positions for effectively assessing the prognosis of the patient. Medical personnel utilize such multi-modal data to gain insights into the symptoms that are indicative of a particular medical condition. Dealing with such additional knowledge on a single patient is important, and is essentially performed manually in current scenario.
4. **Expertise:** Radiologists are highly trained to infer clinical observations from scan images, for providing insights to the referring specialists in the process of disease diagnosis. Radiologists typically deal with 200-300 scan images a day even in small-scale hospitals, and analysis is performed manually, resulting in high cognitive burden. Despite this, though radiologist contributed knowledge is crucial for diagnoses, it is rarely stored for future access and learning, in practice.
5. **Data annotation quality:** Most medical image datasets exhibit a significant lack of high-quality annotations and labeling, making any large-scale learning task difficult. Another modality of data that is seldom utilized in

the expert-generated diagnosis reports. Disease-specific knowledge contained in expert-generated content like radiology reports, nursing reports, doctors' notes etc are rarely stored and thus, a crucial body of valuable data is lost. Any available reports are often very brief, containing precise medical terminology or abbreviations, which could be very useful if correctly processed for enabling future diagnoses tasks.

6. **Multimodal feature learning:** Incorporating multiple types of clinical data, such as, medical images and clinical text, is a field which has received very little research attention. Automated systems that can combine the visual features extracted from the medical scans combined with effectively modeled expert knowledge contained in clinical text, have a significant potential for augmenting diagnosis accuracy and also cut down on the time required for diagnosis.

1.5.1 Thesis Contributions

Based on the understanding of the gaps identified in medical image modeling and representation of healthcare systems, the research problem addressed by the work presented in this thesis is defined as:

“To design and develop an effective framework for representation, modeling and management for diagnostic medical images for supporting advanced clinical decision support applications.”

The research work presented in this thesis elucidates the design of a framework for medical image modelling and representation for effective management of healthcare systems built using radiography data. The major contributions of the research work presented in this thesis are as follows:

- Improving clinical diagnosis performance with automated medical scan quality enhancement algorithms with deep neural image super-resolution models.
- A hybrid feature modeling approach, Swarm Optimization based Bag of Visual Words Model and a deep neural network model for Content-Based Medical Image Retrieval with multi-view classification.
- Deep neural ensemble models for abnormality detection and classification in plain radiographs.

- Design of automated view orientation classification techniques for X-ray images using deep neural networks.
- Deep neural models for automated multi-task diagnostic scan management, including automated generation medical image descriptions.

1.6 Summary

This chapter discusses issues and challenges in the healthcare delivery process, of which diagnostic imaging is a significant part. The issues relating to collection, representation, modeling and management of medical images are discussed and highlighted. A significant need for effective automated medical image enhancement methods for quality representation and improved diagnosis are observed. It was also observed that developing effective medical image representation techniques to capture the manifold information and region-of-interest from scan images to enable targeted retrieval and real-world medical diagnostics applications is critical. To help healthcare providers with valuable insights and to reduce their burden across the overall clinical workflow, techniques that can automatically analyze medical scans and generate natural language reports for them with reasonable accuracy are also the need of the day.

1.7 Thesis Organization

The rest of this thesis is organized as follows.

- In Chapter 2, an extensive literature review on the challenges in medical image management and observed gaps are explained.
- In Chapter 3, the research problem addressed is formally defined based on outcomes and gaps learned from the existing literature. The scope of this research and a brief description of the proposed methodologies are also provided in Chapter 3.
- Chapter 4 presents a detailed discussion on proposed approaches for dealing with medical image quality enhancement.
- In Chapter 5, approaches for medical image modeling and representation for enabling classification and retrieval are presented in detail.

- Chapter 6 proposes approaches for dealing with the problem of variance in medical scan views and models for orientation classification.
- Chapter 7 presents automatic approaches for medical image description generation.
- Chapter 8 presents concluding remarks about the research work carried out and possible directions of future research in the area.

Chapter 2

Literature Review

2.1 Background

In modern healthcare, medical imaging is a preferred diagnostic tool, due to its reliability and non-invasive nature. Available across multiple modalities, these services can aid in the process of accurate and decisive disease diagnosis, enabling curative action fast. The demand for advanced image analysis techniques stems from the recent proliferation of new biomedical imaging modalities. The number of scans currently performed in most hospital environments has increased exponentially placing unprecedented workloads on healthcare personnel associated with these services, performing tasks like capturing, analysis, interpretation and documentation. Alleviating such growing burden, by the introduction of automated systems for healthcare information management has received significant research interest over the past decade. Remarkable advances in large-scale and cost-effective availability of computational resources, data storage and the advent of learning based neural models have brought forth critical advancements in revolutionising healthcare delivery, paving the way to precision medicine applications. Intelligent systems for automatic medical image analysis, interpretation and decision making which can lead to improved diagnosis and a better understanding of disease progression.

Over the past decade, active research interest has been focused on the area of medical image representation and management with real-world implications in the medical image enhancement (Rui and Guoyu, 2017; Gao *et al.*, 2017; Liu *et al.*, 2017), medical image representation (Tizhoosh, 2015; Liu *et al.*, 2016; Zhu and Tizhoosh, 2016; Qayyum *et al.*, 2017), medical image categorization (Kao *et al.*, 2011; Xue *et al.*, 2015; Takeuchi *et al.*, 2019) and medical image modeling and interpretation (Stefan *et al.*, 2017; Su and Liu, 2018). Medical imaging informatics

has generated much interest among researchers and the healthcare community owing to the large number of practical applications, however, critical challenges still exist. In this chapter, a comprehensive review of the existing research in the area of medical image informatics for development of intelligent healthcare systems is presented. A detailed study of the merits and limitations of existing works is provided for insights into the gaps to be addressed.

2.2 Related Work

An extensive review of existing research in the domain of Medical Image Analytics for the design of intelligent medical image management systems was undertaken, and the various challenges in these areas were examined in depth. Our study concentrated upon the challenges identified during preliminary review (listed in Section 1.5), and existing works that address these issues are discussed in subsequent sections.

2.2.1 Medical Image Quality Management

The primary factors that typically affect the quality of captured medical scans are, noise, edge/contours and contrast. Gaussian noise and impulse noise are the two fundamental types that degrade the quality of a medical scan. Generally, median filtering is used to smoothen any impulse noise, but this does not improve the gray-contrast of an image. Histogram Equalization (HE) (Kim, 1997) is a popular method that could be applied to intensify the contrast of the given image; however, the new image that gets developed is often not of acceptable quality. Additionally, grayscale modalities often suffer from low contrast, making the minute details like hairline fractures, fissures etc challenging to identify even for trained medical professionals.

Other factors that affect digital radiographic images are low contrast, visual noise or X-ray scattering and blurring lead by the complexity and density of body tissues. Radiographic images are often found to need significant improvement in visual quality, including contrast and feature enhancements. Different techniques like Linear Contrast, HE (Histogram Equalization), CLAHE (Contrast limited adaptive histogram equalization) and BPHE (Brightness Preserving Histogram Equalization) were used by Ahmed *et al.* (2011) to enhance digital radiographic images. Georgieva *et al.* (2013) developed an enhancement method for X-ray images using CLAHE followed by morphological processing and noise reduction.

CLAHE improves the contrast of an image by reducing the noise in homogeneous areas. But, it was noticed that the artifacts increased when the block size consideration for the image enhancement was more than 16x16. To mitigate this, [Ren *et al.* \(2014\)](#) proposed a hybrid image contrast improvement method formed by the sharp frequency localization-contourlet transform (SFL-CT) and CLAHE. Also, a comprehensive pre-processing algorithm that was developed substantially enhanced the contrast, simultaneously reducing the artifacts.

A solution to this problem is the use of Super-Resolution (SR) techniques by dynamically enhancing the resolution, de-noising the medical images and applying Super-Resolution techniques such as patch based and orthogonal acquisition algorithms. [Huang *et al.* \(2016\)](#) proposed a two-stage filtering process and contrast enhancement for X-ray images. By using an adaptive median filter and bilateral filter, their method was able to suppress the mixed noise which contains Gaussian noise and impulsive noise, while preserving the important structures (e.g., edges) in the images. Afterward, the contrast of an image is enhanced by using gray-level morphology and CLAHE. However, the absolute mean brightness error (AMBE) was more than CLAHE. [Bhairannawar \(2018\)](#) effectively used the enhancement techniques using HSV Transform (Hue, Saturation, and Value) and Adaptive Histogram Equalization. Standard medical image dataset MEDPIX was used for this purpose and it was observed that the method performed better than the existing methods in terms of PSNR (Peak Signal-to-Noise Ratio). In order to improve the diagnosis efficiency and accuracy, medical X-ray image enhancement using dark channel enhanced method was performed by [Rui and Guoyu \(2017\)](#) for medical X-ray images. The DCP (Dark Channel Prior) method may lead to the noise amplification, this kind of granular noise impacts little on the medical diagnosis in most cases. However, this method combined with some denoising method could play a better performance on X-ray image enhancement.

Deep Learning models learn diverse patterns in data to automatically capture informative hierarchical representations, leveraging in achieving a pre-specified task. The enhanced characteristic of approximating capacity and hierarchical information flow property makes Artificial Neural Networks (ANN) the best tools for Deep learning. [Zhang and An \(2017\)](#) presented a deep learning and transfer learning-based super-resolution reconstruction method aims to reconstruct a high-resolution image from one single low-resolution image. Therefore, the proposed method can avoid collecting a high number of various medical images. They proposed a fast bicubic interpretation layer and SIFT feature-based transfer learning to speed up Deep Convolutional Neural Network (DCNN) to obtain sharper

outlines. Empirical experiments showed that the proposed method could achieve better performance than other conventional methods. Finally suggesting that this enhancement method is meaningful for clinical diagnosis, medical research and automatic image analysis. [Gao *et al.* \(2017\)](#) proposed a novel deep network model specifically for medical image super-resolution reconstruction. Their method considers the characteristics of medical image structure repetition and black border. Based on Super-Resolution Convolutional Neural Network (SRCNN) model, a convolution layer is added to carry out feature extraction to improve the feature performance, and overlapping pooling layers is adapted to highlight the important features. Further, a link layer was established between the second convolution layer and the reconstruction layer, which make local features and global features to complete the reconstruction together. The experimental results showed that average PSNR gained better results than the original SRCNN.

Super Resolution CNN (SRCNN) ([Dong *et al.*, 2015](#)) is a classical Super Resolution technique, which comprises a shallow network compared to other networks used in deep learning. It consists of three processes - patch extraction from the image and representation learning, non-linear mapping and image reconstruction. Low resolution (LR) images are initially upscaled to the appropriate required size using a technique known as bicubic interpolation ([Zhou *et al.*, 2017](#)) before passing them through the network. In the Neural network¹, the first layers perform standard convolution with Rectified Linear Unit (ReLU) units, which is passed through the non-linear mapping stage. Here, the mapping from a low-resolution vector to a high-resolution vector is performed using a sparse-coding mechanism after which, the vectors are used to reconstruct the image. The super-resolution results in medical images achieved better visual effect than other contrast algorithms. [Liu *et al.* \(2017\)](#) proposed a low-rank minimum variance estimation method. Especially, the proposed method first generates an initial HR image by nonlocal interpolation, then uses the low-rank minimum variance estimator to reconstruct it, and at last, iteratively applies the subsampling consistency constraint to further refine the reconstructed HR result. Additionally, there also exists the nonlocal self-similarity between the neighboring medical slices that can be used to further improve the performance of resolution enhancement.

[Zhao *et al.* \(2019\)](#) proposed a novel medical image enhancement method based on the modulation techniques. Luminance modulation was used to adjust the brightness and increase the contrast of the input image by shrinking the global

¹<https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks>

range of the input image. In addition, gradient modulation was used to enhance texture details of the image. The performance of the method showed a good improvement when compared with the other popular enhancement methods. [Zhu et al. \(2018\)](#) proposed a new image reconstruction method based on the artificial intelligence approach called AUTOMAP, which produces a high-quality image from less data reducing the radiation effects. The results identified that this approach improved the signal-to-noise ratio. As part of this objective, state-of-the-art super-resolution models were considered for medical scan image for quality enhancement and reconstruction.

Table 2.1: Summary of Existing works in Medical Image Quality Management

Work	Methodology	Remarks
Feng et al. (2008)	GTwo methods grad-contrast enhancement (GCE) combined with wavelet transform enhancement (WTE) were combined for contrast enhancement.	Beneficial for amplifying tiny areacharacters, such as tissues and fibrins, but sensitive to noise.
Saleem et al. (2012)	Fusion-based contrast enhancement technique.	Enhancing local and global contrasts while retaining the original image appearance. Over-enhancement of artifacts and no noise removal.
Georgieva et al. (2013)	Contrast limited adaptive histogram equalization.	CLAHE can enhance not only the contrast of the image, but it also reduces noise in homogeneous areas. However, artifacts are considerably amplified when tiles are more than 16*16.
Ren et al. (2014)	HA hybrid methodimage contrast enhancement based on sharp frequency localization-contourlet transform (SFL-CT) and CLAHE.	Greatly enhance the contrast and suppresses the artifacts simultaneously.

Work	Methodology	Remarks
Isaac and Kulkarni (2015)	Super-resolution techniques - patch based and orthogonal acquisition.	The resolution of low-resolution medical images can be satisfactorily increased to required levels, tested on HR images.
Huang <i>et al.</i> (2016)	Two-stage filtering process and contrast enhancement for X-ray images	able to suppress mixed noise while preserving important structures (e.g., edges) in images. Absolute mean brightness error was more than CLAHE.
Rui and Guoyu (2017)	Dark channel prior.	DCP is widely used for removing haze on images, but may lead to noise amplification. However, this kind of granular noise impacts little on the medical diagnosis in most cases.
Zhang and An (2017)	Deep Learning and Transfer Learning using SIFT feature-based Feature Technique.	Aims to reconstruct a high-resolution image from one single low-resolution image.
Gao <i>et al.</i> (2017)	Deep CNN	Establish a link layer between the second convolution layer and the reconstruction layer, which make local features and global features to complete the reconstruction together. Trained only on a small CT dataset.
Liu <i>et al.</i> (2017)	Nonlocal self-similarity and low-rank minimum variance estimator.	Low-rank priors have achieved great success in the field of image processing. Neighboring slices of images were taken for resolution computation.
Ahmed <i>et al.</i> (2011)	Four different histogram equalization algorithms were applied along with noise reduction techniques.	BPHE (brightness preserving histogram equalization) produced good contrast enhancement.

2.2.2 Medical Image Modeling and Representation

Modern medical diagnostic tools contribute varied types of medical images like X-rays, CT and PET scans, which are huge in volume and are also continuously generated. Due to this, manually creating adequate and sufficient textual annotations is quite a difficult and time-intensive task. When available, these annotations are often incomplete or ambiguous due to the unstructured nature of natural language, thus, adversely affecting retrieval results. The Picture Archival and Communication System (PACS) (Lehmann *et al.*, 2003a) is a significant effort to overcome these challenges to store, retrieve and transmit medical images effectively. However, a prime constraint is that it uses a method that depends on keywords and is connected with text notations stored with the image for retrieval. Later, this paved the way for advent and subsequent popularity of CBIR systems, especially for medical image management. The main objective of CBIR systems is capturing the latent features of an image dataset without depending on any information that is external to it (e.g., text meta-data associated with images). Most CBIR systems utilize features like color, shape and texture for generating a good relevance rank. However, a significant challenge faced here is the fact that most medical images are gray-scale. Hence, color cannot be considered as the most dominant feature. But, the image's texture and shape features play a crucial role that needs to be effectively captured.

Bag of Visual Features is adapted from the well-known Bag of Words method, commonly used in document classification and information retrieval. Two classifiers, simple Naive Bayes and linear Support Vector Machines (SVM), were used for classification and results showed that SVM produced meaningful results for high-dimensional data. Extracting image features for image classification and a bag of features (O'Hara and Draper, 2011) for retrieval focuses on improving the feature detection task for faster image retrieval. Different feature detectors like Gaussian Difference (GD) (Lowe, 1999), Scale Invariant Feature Transform (SIFT) (Lowe, 2004), Speed Up Robust Feature (SURF) (Bay *et al.*, 2006, 2008), Maximally Stable Extremal Regions (MSER) (Matas *et al.*, 2004) and Harris-Affine keypoint operators (Mikolajczyk *et al.*, 2005) have been benchmarked for the task of medical image modeling and representation. Wang *et al.* (2007) proposed a BoF method for medical image retrieval using AdaBoost for generating visual words. Vocabulary size was tuned in different sizes of K from 200 to 800 in step size of 100. With the value, k=700 had given the best accuracy for the boosted weighting method. Experiments were carried out using the SIFT keypoint detectors on the

three medical image datasets - ImageCLEFmed, 304 CT and Basal-Cell Carcinoma datasets. For enabling retrieval, a fusion rule was incorporated by reducing the vocabulary size $k=10$, and performance was measured by recall and mean average precision. [Avni et al. \(2010\)](#) presented a bag of visual words technique for X-ray image classification and retrieval. Three kinds of feature extraction methods were utilized and analyzed – local patch-based, normalized patch-based variance and SIFT features.

A hierarchical based classification using SVM classifier on each of the IRMA sub-codes was proposed by [Unay et al. \(2009\)](#). Accuracy was noted for each of the sub-code and the obtained results were 96.7, 85.6, 88.0 and 96.4%. Multiple visual features like GLCM, pixel values and edge-based canny detector were combined in [Mueen et al. \(2007\)](#) work, and they obtained an accuracy rate of 89% using ImageCLEF 2005 with 57 classes. [Tommasi et al. \(2008\)](#) extracted local and global features from images using pixel values and BoW. Then they combined these features at high, mid and low levels using multi-cue approaches, with 89.7% accuracy on ImageCLEF 2007. BoW was combined and fused with other feature extraction techniques like edge histogram, pixel value and LBP [Dimitrovski et al. \(2011\)](#), and this combined feature space helped the model achieve good classification performance. [Zare et al. \(2013\)](#) proposed an iteration based approach for automatic classification for medical X-ray images using the ImageCLEFmed dataset. In each iteration, low accuracy classes are again fed into the classification model, which is a Support Vector Machine (SVM) trained on radial basis function.

Over the past decade, extensive research has been undertaken to address the problem of medical image retrieval. [Zare et al. \(2011\)](#) proposed a combined low-level feature extraction method for classification and retrieval of medical X-ray images using shape and texture features. Their technique achieved good accuracy despite using a limited set of data. [Pourghassem and Daneshvar \(2013a\)](#) proposed a medical image retrieval framework based on a merging based classifier, using which similar images were ranked as per computed similarity values. Images in the result set are labeled as positive and negative by the user which is used to optimize the retrieval results further by employing a Random Forest classifier. During the feature extraction process, invariant moments are extracted from the main object in the binary image using Otsu's thresholding method, ([Otsu, 1979](#)) to capture the binary image's foreground and background space. Other features like the gray-level co-occurrence matrix and Fourier descriptors for texture & shape features are also used. The designed merging based classification helps in addressing two issues – it increases the interclass distance while reducing intraclass distance, and

also improves the accuracy of classification. However, body orientation changes were not addressed by the authors. A pattern similarity scheme for retrieval of medical images as per the PANDA framework was proposed by [Iakovidis *et al.* \(2009\)](#). This scheme involves the extraction of low-level features, which are then clustered according to the feature space to form relevant patterns. Clustering on the feature set is done with an expectation-maximization algorithm that uses an iterative method to decide the number of clusters by itself. Retrieval results were evaluated based on precision and recall measures, while the best retrieval was obtained using the k nearest distribution function. However, this approach suffers in performance for a larger dataset, due to lack of good image indexing schemes.

[Aggarwal *et al.* \(2013\)](#) implemented an independent CBIR framework for retrieval of lung images. An expanded dataset provided better variability in the retrieval set, on which, they used different distance metrics for nodule similarity assessment. A clustering method for medical image retrieval using dictionary learning was proposed by [Srinivas *et al.* \(2015\)](#) for grouping of large datasets. Two types of feature extraction methods were employed – initially, an image is divided into equal regions with concentric circles giving an invariant representation of the image which then finds the mean and variance at that circular region of the image. As a second step, an image is divided into four subparts; from each of these subparts, the mean and variance at circular regions are computed as major part of the object information is available at the center in medical images. Clusters are formed by applying K-means to these feature vectors and by using the K-SVD method, a dictionary is generated for each cluster. The performance of the methods was measured in terms of precision and recall with three different distance measures like Euclidean, Mahalanobis and Cross-correlation for cluster sizes of 3, 4 and 5. However, defining suitable attributes for medical images is a comparatively difficult task and incorporating visual attributes could potentially boost CBMIR performance.

Several distance measures are used in medical image classification and retrieval evaluations. Such distance measures can be of two types - global and local. A single value is obtained when a global distance measure is used, where as, in local distance measures, values per mesh vertex is determined. [Getto *et al.* \(2015\)](#) used the extended surface distance for 3D medical image segmentation, which is used to detect regions of bad segmentation quality, hidden in earlier scans. They reported that the use of the metric improved reliability, while reducing the asymmetry, provides more insights into the segmentation quality for medical experts' use. [Trapp *et al.* \(2013\)](#) proposed an effective object retrieval approach designed for

neuron structures of the organism, *Drosophila Melanogaster*, however it can be generalized for other species also. Domain experts reported that the retrieval results for neuronal structures are very good.

Deep CNNs have achieved good performance in image retrieval compared to traditional image processing based approaches. [Ahmad *et al.* \(2018\)](#) proposed a selective convolutional feature model that uses Fast Fourier Transform (FFT) to generate a sequence of bits. Initially, convolutional feature maps are obtained from a pretrained CNN, which are then converted to compact binary codes. The framework was tested with two large datasets of radiology and endoscopy images and experimental results showed that their method outperformed from other feature extraction and hashing schemes. [Liu *et al.* \(2016\)](#) used a CNN trained on Radon barcodes for retrieval of medical images from a large collection of 14,000 X-ray images. Once the training process is done, CNN codes are generated for image retrieval. Top 50 similar images for the query image that have the shortest Hamming distance are selected and Radon barcodes are calculated. Finally, the top 10 re-ordered results are presented to the user as the retrieval results. Refinement of the retrieval task, re-order of the result is not the best practice when it comes to medical imaging. [Zhu and Tizhoosh \(2016\)](#) used SVM classification on Radon barcodes for content-based image retrieval. Each dataset image is represented in a binary format along with the Radon barcodes and Radon transform is used for the extraction of Radon features. To categorize the latent information in the query images, a multi-class SVM classifier is trained on the extracted Radon features. Similar images are retrieved using the k-nearest neighbor's method, during the retrieval stage. A deep convolutional neural network for CBMIR ([Qayyum *et al.*, 2017](#)) was built on an intermodal dataset from which features are learned by the neural network, and then used for retrieval of medical images. The dataset used consists of only 24 classes, with very low-class imbalance, thus the authors reported a 99.7% accuracy rate for classification and 0.69 mean average precision for retrieval task. However, in larger medical datasets, an inherent class imbalance is common, thus, to overcome this designing an efficient retrieval algorithm is crucial for medical image management.

Training deep CNNs directly with high-resolution images requires significant compression of images at the input layer, resulting in loss of information which might be crucial for medical image abnormality detection. [Xi *et al.* \(2019\)](#) proposed an integrated approach for medical abnormality detection using deep CNN, where, pre-trained deep CNNs are first fine-tuned on image patches centered at medical abnormalities, later integrating them with class activation maps for build-

ing abnormality detectors. A deep patch classifier was tested on a single class mammogram dataset obtaining an overall classification accuracy of 92.53% compared to the traditional approach using manual features. [Madani *et al.* \(2018\)](#) proposed a learning algorithm of GAN that labels input images using a semi-supervised classifier for disease prediction based on chest X-ray images. The network was trained on small sized annotated images. Two sets of chest X-ray images from National Institute of Health (NIH), prostate, lung, colorectal, and ovarian (PLCO) cancer datasets and NIH Chest X-Ray collection from Indiana University was considered for the study. They used a semi-supervised GAN architecture with a loss function to assimilate both labeled and unlabeled real data. Here, the loss function is divided into three parts, having output layer of the discriminator with $K+1$ classes, $K=2$ for normal and abnormal classes and $K+1$ for distinguishing the generated images. Parallely, GAN performance was compared with CNN by varying the number of labeled images in the experiment. The results showed that with a fewer number of images (about 10) in each class, the semi-supervised model achieved an accuracy of 73.08%, whereas, for CNN it required more than 250 images.

[Camlica *et al.* \(2015\)](#) used a context-aware saliency algorithm to detect salient regions from the medical images, so that relevant information can be extracted. They reported an IRMA error (Eq. 5.28, Section 5.7.3) of 146.55, which is the lowest achieved error so far. However, the algorithm is extremely slow and saliency calculation is a time-consuming process that works only with offline maps generated during testing, making it impractical. [Khatami *et al.* \(2018b\)](#) proposed a search space based approach for retrieving the most similar images for a given test image. A two-step hierarchical shrinking search space was used with local binary patterns. Transfer learning via CNN is utilized in the first stage for shrinking the search space, followed by creation of a selection pool using Radon transform for further reduction resulted with an error score of 168.05. The authors also proposed a parallel deep approach based on convolutional neural networks with a local search using LBP, HOG and Radon features [Khatami *et al.* \(2018a\)](#), which achieved an error rate of 165.55 (mean value). However, they did not use high-level features when using parallel deep solutions and ensemble methods for decision making during the image retrieval task. [Avni *et al.* \(2009\)](#) proposed a multi-resolution patch-based dictionary approach by employing principal component analysis (PCA) on the densely sampled patches. Training on the bag-of-words, they used a support vector machine (SVM) classifier and reported an IRMA error of 169.5 on the IRMA dataset. [Müller *et al.* \(2009\)](#) combined two different image descriptors, *i.e.* LBP

and modSIFT (Tommasi and Orabona, 2010) for different SVM based classification approaches and reported an IRMA error of 178.93. Liu *et al.* (2016) utilized CNN architectures for classifying LBP and Radon transform codes to achieve an IRMA error of 224.13. Sze To *et al.* (2016) used deep autoencoders and Radon barcodes, which achieved an IRMA error of 344.08, while Sharma *et al.* (2016) used KNN for extracting features obtained from stacked autoencoders, but the IRMA error increased to 376.

Table 2.2: Summary of Medical Image Modeling and Representation.

Work	Methodology	Remarks
Arimura <i>et al.</i> (2002)	Template matching technique.	Accuracy of 94.7% achieved. Used only 1000 images and only two views of chest images.
Lehmann <i>et al.</i> (2003b)	Correlation function and distance measures.	Achieved 99.3% classification accuracy, however used only 1867 images with only two views.
Iakovidis <i>et al.</i> (2009)	block-based low-level feature extraction and feature space clustering.	Achieved a good AUC of 78% compared to earlier works, evaluated on 116 classes.
Zare <i>et al.</i> (2011)	Uses low level features extraction and SVM classifier.	70 classes had a classification accuracy greater than 80%, and also classes with more images had good accuracy.
Fesharaki and Pourghassem (2012),	Shape-based feature extraction techniques and used Bayesian rule classifiers.	Achieved good accuracy rate, but used only 28 classes.
Pourghassem and Daneshvar (2013b)	Merging based classification algorithm that measured weighted Euclidean distance for retrieval.	Similar classes were merged, oriented classes were combined, and then classification accuracy is reported.
Liu <i>et al.</i> (2016)	CNN and radon barcode.	Showed that error score decreased when image size increased. Total error score was average when compared to other works.

Work	Methodology	Remarks
Zhu and Tizhoosh (2016)	Combination of Radon projections and Support Vector Machine classifier.	Improvement in classification, but retrieval error was average when compared to other works.
Qayyum <i>et al.</i> (2017)	Deep CNN based model	Achieved good classification rate by using the central cropped image. Considered a small dataset with only 24 classes.

Based on the study of existing works, it was observed that researchers have tried to utilize the manifold information available in medical images for the classification task. However, some approaches overlooked the semantic gap *i.e.*, the dispute between the intention of the user and the images retrieved by the algorithm. Also, most other approaches experimented with small datasets for their experimental analysis, which makes them difficult to scale when applied to larger datasets. With these insights, it is deciphered that there is a significant requirement for scalable CBMIR models built of effective AI models, for enabling high accuracy and high precision medical image retrieval.

2.2.3 Dealing with variance in Medical Images

Classifying medical scans is an essential requirement for accurately indexing and categorizing the incoming image data in large-scale Hospital Information Management Systems (HIMS). Medical images can be varied or can belong to different classes even when they belong to the same diagnostic modality, as variety is introduced based on the particular body part/organ that the scan covers. Even if only scans covering one body part are considered (for e.g., chest X-rays), there is also variety in the way in which a particular scan is performed. Different orientations like anterior or frontal view, posterior or back view and the lateral or side views, also known as left lateral or right lateral can be used during scanning, as per expert inputs, so that anomalies are better captured. However, computer-aided diagnosis systems often do not capture this crucial information of the scan. The orientation identification for medical images is required for quality and quantitative analysis, in diagnostic applications. Scans of body parts and modalities where tissue orientation is also a significant need have to be addressed adequately,

and is a research gap that needs to be considered. Hence, there is a scope for view classification for other biological structures, that can be incorporated as an essential step during the indexing process applied to scanned radiograph images, thus aiding the overall management of HIMS.

Very few works exist currently, that address this particular issue. [Arimura *et al.* \(2002\)](#) proposed a set of nine templates, one set for medium-sized patients consisting of 3 templates (1 PA and 2 lateral) and another set for small/large-sized patients consisting of 6 templates (2 PA and 4 lateral). The similarity of the chest image in lieu of one of these templates was determined using a template matching technique, considering a correlation value greater than 0.2. Their approach was a two-step process for identifying the orientations of the image; in the first step, two different views were determined for medium-sized patients. If it is unidentified then a check with the other set involving six templates is performed in the second step. A total of 1000 test images were used in their experiments, involving 500 PA and 500 lateral chest radiograph images. In the first step, 924 cases (92.4%) were correctly identified, while all other cases were identified in the second step. [Lehmann *et al.* \(2003b\)](#) determined chest radiographs' view by applying several distance measures and nearest-neighbor classification. Using tangent distance as the nearest-neighbor classification scheme, good accuracy was obtained for images of 32×32 pixels. [Boone *et al.* \(2003\)](#) proposed a feed-forward neural network to identify views in chest x-ray images, in which a series of chest images consisting of 999 lateral and 999 frontal were downsampled to a size of 16×16 during training. The network was able to identify the views of chest images with 98.8% on an average of six trials. [Kao *et al.* \(2006\)](#) developed a projection profile technique to identify frontal and lateral views for chest x-ray images. The projection profile was computed based on the computation of two indices, namely body symmetry index and background percentage index.

During chest X-ray screening, the orientation view information is a crucial aspect. [Santosh and Wendling \(2018\)](#) developed a novel method for classifying the chest X-ray image view as frontal and lateral. They incorporated a new technique called angular relational signature to extract features from the histogram. Multi-layer perceptron, random forest, and support vector machine were used to predict the classification accuracy attaining close to 100%. [Kao *et al.* \(2011\)](#) developed an automatic recognition of frontal PA and AP chest radiographs. Their work incorporated three features in identifying the chest radiographic views, i.e., the scapula's and clavicle's tilt angles and the extent of radiolucence in the lung. The method was evaluated with 1200 chest radiographs, consisting of 600 PA

and 600 AP images. The performance was measured with Receiver Operating Characteristic (ROC), which illustrated that the fusion of the above three features showed a high discriminant result. [Takeuchi *et al.* \(2019\)](#) developed an automated chest X-ray radiography classification for CheXpert dataset consisting of 65,240 patients images, labeled by an expert radiologist. The work explored different network architectures and found that the featured DenseNet121 passed into a decision tree classifier achieved an accuracy of 93%.

Scanned radiographs are stored frequently in the Picture Archive and Communication System (PACS) with unknown orientation label, making it ineffective for radiologist analysis. A solution to this problem was proposed by [Luo *et al.* \(2006\)](#) with an automated protocol for chest images. Desired regions of the chest features were extracted like - its size, rotation and translation and a trained classifier was used for identifying the directional view of the chest images. The alignment label was then distinguished considering the abdomen and the neck positions in the radiograph image. The experiment showed promising results of about 96%, with 6,680 images collected images from a hospital. [Shiraishi *et al.* \(2007\)](#) developed a computerized scheme for detection of lung nodules to improve the overall performance of CAD systems for posterior-anterior (PA) views using its lateral views of chest X-rays. Different pre-processing and ANNs were used in the overall workflow of the CAD scheme. The performance of the computerized scheme for lateral views was relatively low (60.7% sensitivity). However, the overall sensitivity (86.9%) was improved for PA views.

[Xue *et al.* \(2015\)](#) developed a hybrid feature model for chest X-rays categorizing the images into - frontal and lateral. The experiment was performed on two datasets - the NLM Indiana and the IRMA datasets, consisting of 8,000 images. Combined features of Image Profile (IP), Contour-Based Shape Feature (CBSF), and Pyramid of Histograms of Orientation Gradients (PHOG) with a 10-fold cross-validation achieved a good accuracy when used with CAD systems. However, from [Santosh *et al.* \(2016\)](#)'s work, it is observed that the algorithm was trained with frontal chest X-rays, where it won't classify the lateral chest images. Certain features are also essential when classifying both the views of an image. The primary reason is that the features (shape and texture) vary with both the frontal and lateral view images. Another novel technique had developed by [Santosh *et al.* \(2015\)](#) for identifying the rotated lungs in chest X-rays measures the rib-orientation using a generalized line histogram technique for quality control. On these observations, modeling the image variances is determined to be an important problem, and giving attention to the body view positioning by incorporating a multi-view

classification technique can contribute positively towards effective classification labeling.

Based on the view position of the radiography a CNN model was trained for 14 different types of thoracic diseases by Rubin *et al.* (2018). Different view positions like posteroanterior (PA), anteroposterior (AP) and lateral view positions of Chest X-ray images were included. It was observed that the overall performance of the model's DualNet classifier showed greater average AUC compared to the state-of-the-art individually trained classifiers. Bertrand *et al.* (2019) trained a simple DenseNet model separately for different types of chest X-ray diseases using PA or lateral images and found that the performance of the lateral view images showed better than the PA view for around eight different class labels. The conclusion drawn from this experimental analysis is that using lateral images helps in prediction tasks for certain types of the diseases, however a more extensive research and analysis is required.

Table 2.3: Summary of View Orientation Classification based works.

Work	Methodology	Remarks
Luo <i>et al.</i> (2006)	Size, rotation and translation invariant features with an automatic hanging protocol.	Resulted in 98.2% on projection view (without protocol, 62%), and 96.1% had correct orientation (without protocol, 75%)
Shiraishi <i>et al.</i> (2007)	Computerized scheme for detection of lung nodules	The overall sensitivity (86.9%) was improved for PA views using its lateral views.
Xue <i>et al.</i> (2015)	PHOG, CBSF using body size ratio and 10-fold CV	Achieved a good accuracy result of 99.2% for frontal and lateral chest images. Need to include other body organs, possibly with three orientation view labels.
Ittyachen <i>et al.</i> (2017)	A real case scenario	Highlights the importance of the lateral view positions of the Chest X-rays.
Rubin <i>et al.</i> (2018)	A CNN model	Overall performance of the model's DualNet classifier showed greater average AUC compared to the state-of-the-art methods.

Work	Methodology	Remarks
Kitamura et al. (2019)	Ensemble Models used (Inception V3, Resnet, and Xception CNNs)	Models used 3 views for each case and achieved an accuracy of 81%. Only one single class which included 298 normal and 298 fractured ankle studies were considered.
Bertrand et al. (2019)	A simple DenseNet model	Using lateral images helps in prediction tasks for certain types of the diseases, however a more extensive research and analysis is required.

2.2.4 Automating Medical Image Understanding

Leveraging latent clinical knowledge available in multiple clinical sources like medical images and text based clinical reports has been explored to a very limited extent. Recently, the ImageCLEF conference’s concept detection task focused on medical image caption prediction using medical concepts as sentence-level descriptions extracted from the Unified Medical Language System (UMLS) dataset. The goal of the task is to efficiently identify the relevant medical concepts from medical images as a predictor of figure captions. However, training a multilabel CNN on noisy datasets with a limited number of training samples is difficult due to a large number of parameters to be learned. [Stefan et al. \(2017\)](#) showed that a CNN pre-trained on single label image datasets, e.g., ImageNet, can be transferred to tackle the multi-label problem. [Harzig et al. \(2019\)](#) proposed a dual-word Long Short Term Memory (LSTM) sentence generation model, trained separately for abnormal and normal chest X-ray images. They reported that the dual-word LSTM helped increase the number of distinct sentences generated, however, it failed to address the findings or identification of abnormal regions in the image.

[Rajpurkar et al. \(2017b\)](#) worked on the Indiana university dataset consisting of chest x-rays with textual reports reporting observed abnormal conditions. The CheXNet model is a Dense Convolutional Network (DenseNet) adapted from ([Huang et al., 2017](#)) which is a 121-layer deep for detecting 14 categories of pathologies from the frontal-view of chest X-ray images. The performance of the model is calculated by taking the difference between the average F1 score of CheXNet and the average F1 score of the radiologists on the same set of samples using confidence intervals (CI). [Su and Liu \(2018\)](#) explored and implemented an encoder-decoder

framework to generate a caption for a given medical image on ImageCLEF Caption Prediction 2018 Task. Two types of CNN architectures were used in the model for comparison ResNet-152 and VGG-19. As a decoder, they used LSTM recurrent neural network. The task is more challenging, due to difficult medical terms. To address this, the model needs to be more intelligent with adequate reasoning ability, which may require more complex and hierarchical text modeling structure with the support of background knowledge. [Shin *et al.* \(2015\)](#) designed a text/image deep mining system applied to a large-scale PACS dataset, for extracting the semantic interactions from radiology reports. Given an image, the system interleaves between supervised and unsupervised learning on document, and sentence-level text collections, to generate semantic labels. When a scan image is fed into the system, semantic labels in radiology are predicted, and its associated keywords are also generated. The disease types are then detected as present or absent, for providing more specific interpretation to the scanned images. [Li *et al.* \(2018\)](#) proposed a novel Hybrid Retrieval-Generation Reinforced Agent (HRGR-Agent) to perform robust medical image report generation. The model generated robust reports on medical abnormal findings detection and best human preference, with good precision performance.

A multi-task learning framework for prediction of tags and the generation of reports was proposed by [Jing *et al.* \(2017\)](#), who incorporated CNNs with LSTM. The model is capable of not only generating high-level impressions, but also generating detailed descriptive findings. [Xue *et al.* \(2018\)](#) used a text-image embedding network integrated with multi-level attention models in an end-to-end CNN-RNN architecture for learning distinctive image and text representations [Wang *et al.* \(2018a\)](#). These models have shown promising results so far. Padchest by [Bustos *et al.* \(2020\)](#) is an initiative that involved trained physicians for manually annotating 27% of the UMLS CUI dataset samples. They then used a RNN attention method to label the remaining images, and also proposed a hierarchical taxonomy to categorize radiographic findings.

Table 2.4: Summary of Existing works on Medical Image Description Generation.

Work	Methodology	Remarks
Shin <i>et al.</i> (2015)	Deep CNN - text/image deep mining system	large-scale image/text analysis on a hospital's PACS.
Stefan <i>et al.</i> (2017)	ResNet - 152 pre-trained Neural Network.	Data augmentation techniques improved network generalization capabilities.

Work	Methodology	Remarks
Jing <i>et al.</i> (2017)	CNN to learn visual features, multi-label classifier to predict relevant tags.	Hierarchical LSTM network effectively captured long-range semantics and produced effective text reports.
Su and Liu (2018)	ResNet-152 and VGG-19 Neural Network.	Ranked second best result, however, did not address the complexity of medical term captioning. Hierarchical text modeling is required to handle medical terms.
Li <i>et al.</i> (2018)	HRGR-Agent based medical image report generation.	Bridges human prior knowledge and generative neural network via reinforcement learning, generating robust reports on abnormal findings.
Xue <i>et al.</i> (2018)	CNN with LSTM	Generated high-level impressions with detailed descriptive findings.
Wang <i>et al.</i> (2018a)	A text-image embedding network integrated with multi-level attention models.	End-to-end CNN-RNN architecture for learning distinctive image and text representations, improvement in the quality of generated reports needed.
Harzig <i>et al.</i> (2019)	Dual word LSTM sentence generation model	Dual word LSTM helped increase the number of distinct sentences generated.
Bustos <i>et al.</i> (2020)	A RNN attention method.	Hierarchical taxonomy used to categorize radiographic findings and diagnosis reports.

2.3 Outcome of Literature Review

The comprehensive survey of existing literature presented in the previous section helped identify several gaps in medical imaging in healthcare management. A key challenge to implementing quality improvement programs is to develop methods to collect knowledge related to quality control and to deliver that knowledge to

practitioners at the point of care (Johnson *et al.*, 2009). According to Hillman *et al.* (2004), “*quality is the extent to which the right procedure is done in the right way at the right time, and the correct interpretation is accurately and quickly communicated to the patient and referring physician*”. Initially, when a scan is taken of a particular organ of the body, the contrast and noise are the primary factors that affect the image quality, due to the nature of the scanning environment. Also, overexposure and underexposure can affect the quality of the scanned image. In addition, imaging is increasingly performed by radiology technicians, or radiologists with limited training in hospitals in rural and remote locations. Measuring and improving quality is essential to ensure optimum effectiveness of care and combat current trends leading to the commercialization of radiology services. In view of this, designing effective automated techniques for improving the quality of digital radiography was considered as part of this research work.

Over the past decades, the volume of medical scan images has grown exponentially as advanced medical technology has reached more people and the number of patients seeking medical attention keeps increasing. The problem of medical image classification and retrieval is also an area of active research. CBMIR evolved from CBIR, aims at retrieving similar images from the medical image repository for further diagnosis and improvement in the treatment based on additional background data. Conventional text-based search systems fail to capture the latent visual features in an image. Hand-engineered features are effective only in a small-scale datasets with limited number of categories. Thus, techniques that automatically capture latent features from patient scans for the development of intelligent applications that consume this data are crucial.

Scan images like X-rays, CT scans, etc., can encompass several internal organs; it is essential to devise automatic classification approaches to deal with the diversity. Existing classification and retrieval models are built for a specific category or class. However, the scans of the organs are taken at different views, to enable well-rounded assessment of patient prognosis. Effectively modeling this additional information for enabling automated diagnosis is another area which has received very limited research attention. This might be due to the non-availability of view-annotated datasets or due to the limited number of images in the datasets available for different organs. This is seen as another critical gap, which is considered in the proposed work.

Need for automatic annotation and description generation for medical images is another gap observed during our literature review, which is a critical requirement in developing advanced medical image management systems. Various approaches

are proposed in existing literature for dealing with the problems of automatic semantic tagging (Guillaumin *et al.*, 2009) and generating an automatic description for natural scene images (Elliott and Keller, 2013). However, there is limited work in the medical imaging domain, to the best of our knowledge. Interpreting and summarizing the insights gained from medical images such as radiology outputs is mostly a time-consuming, manual task that involves highly trained experts and often represents a bottleneck in clinical diagnosis pipelines. Consequently, there is a considerable need for automatic methods that describe the image contents. Automatically generating the medical insights available in each scanned image to assist medical personnel during diagnosis and treatment could be especially beneficial in reducing the significant burden in the overall workflow in patient care.

The demand for advanced image analysis techniques stems from the recent proliferation of new biomedical imaging modalities. The number of scans currently performed in most hospital environments has exploded, placing unprecedented workloads on personnel associated with their interpretation. At the same time, remarkable advances are being made in the field of deep neural networks. New algorithms are paving the way for automatic image interpretation, which can lead to improved diagnosis and a better understanding of disease progression. Biomedical imaging analysis techniques can be applied in many different areas to solve existing problems. The various requirements arising from resolving practical issues motivate and expedite the development of biomedical imaging analysis. With this motivation, an effective framework for representation and modeling for medical images is proposed, for supporting advanced decision support applications for improving healthcare delivery systems.

2.4 Summary

Various approaches and models that have been proposed as part of medical image representation and modeling are discussed in this chapter. The existing techniques in building an effective medical image model are grouped into four categories – Medical image quality enhancement, classification and retrieval of medical scan images, Orientation identification and automated medical image description generation. The extensive literature review revealed a definite requirement for approaches in developing super-resolution image enhancement methods, accurate classification, retrieval and orientation identification methods by introducing better feature modeling strategies. Adding to these models, automatic report generation methods for scan images give a better insight into this proposed works.

In Chapter 3, the research problem addressed in this thesis is formally defined, based on the identified research gaps in the existing literature. The proposed methodologies designed to address the observed research gaps are also discussed briefly, the details of which are presented in subsequent chapters of this thesis. Clinical decision support systems are a critical emerging area in the healthcare sector. Towards this, the work presented in this thesis focuses on making significant positive contributions to the ongoing research in this area.

Chapter 3

Problem Description

3.1 Background

In the previous chapter, an extensive review of existing approaches focusing on developing a practical framework for medical image modeling and representation for augmenting advanced healthcare systems was presented. The issues and requirements for designing an improved medical image modeling are also summarized. In this chapter, the identified research gaps are explicitly presented and defined as a problem statement. Additionally, the scope of the proposed research work presented in this thesis and a brief overview of the approaches designed for solving the formally defined problems are also discussed.

3.2 Scope of the Work

An extensive review of the existing research works in medical image modeling and representation of healthcare systems along with the research gaps identified are summarized in Chapter 2. From the review of existing approaches, it is clear that developing an effective framework for automated medical image modeling and representation methods using learning based models is a critical requirement. It is known that the effectiveness of the developed medical imaging models depends heavily on medical scan images which act as a primary source for any of the medical image representation models. With this purpose and as an aim of bridging the observed gaps, the research work presented in this thesis has contributed in five significant aspects, as listed below:

1. Design and development of automated image quality improvement approaches for medical image super-resolution.

2. Design and development of techniques for improved perceptual image quality management for enabling better diagnosis.
3. Design of hybrid feature modeling approaches for optimal representation of medical images, for classification and content-based retrieval tasks.
4. Design and development of efficient visual feature learning models for modeling variances in medical images, for addressing body view positioning challenges, through multi-view classification techniques.
5. Design of content-based medical report retrieval approaches, and automatic generation of diagnostic text reports for a given medical image.

3.2.1 Problem Statement

Based on the understanding of the gaps identified from the review of existing literature in medical image modeling and representation of healthcare systems, the research problem addressed by the work presented in this thesis is defined as:

“To design and develop an effective framework for representation, modeling and management for diagnostic medical images for supporting advanced clinical decision support applications.”

3.2.2 Research Objectives

Based on identified gaps and the problem statement, four research objectives have been defined, that are addressed in the research work presented in this thesis:

1. To design and develop automated image quality improvement techniques for enhancing the diagnostic scans.
2. To design and develop effective feature modeling and representation approaches for medical images.
3. To design efficient techniques for dealing with variance in scanned image views in practical scenarios.
4. To design and develop a system for automatically capture and describe the valuable insights in medical scan images for clinical decision support.

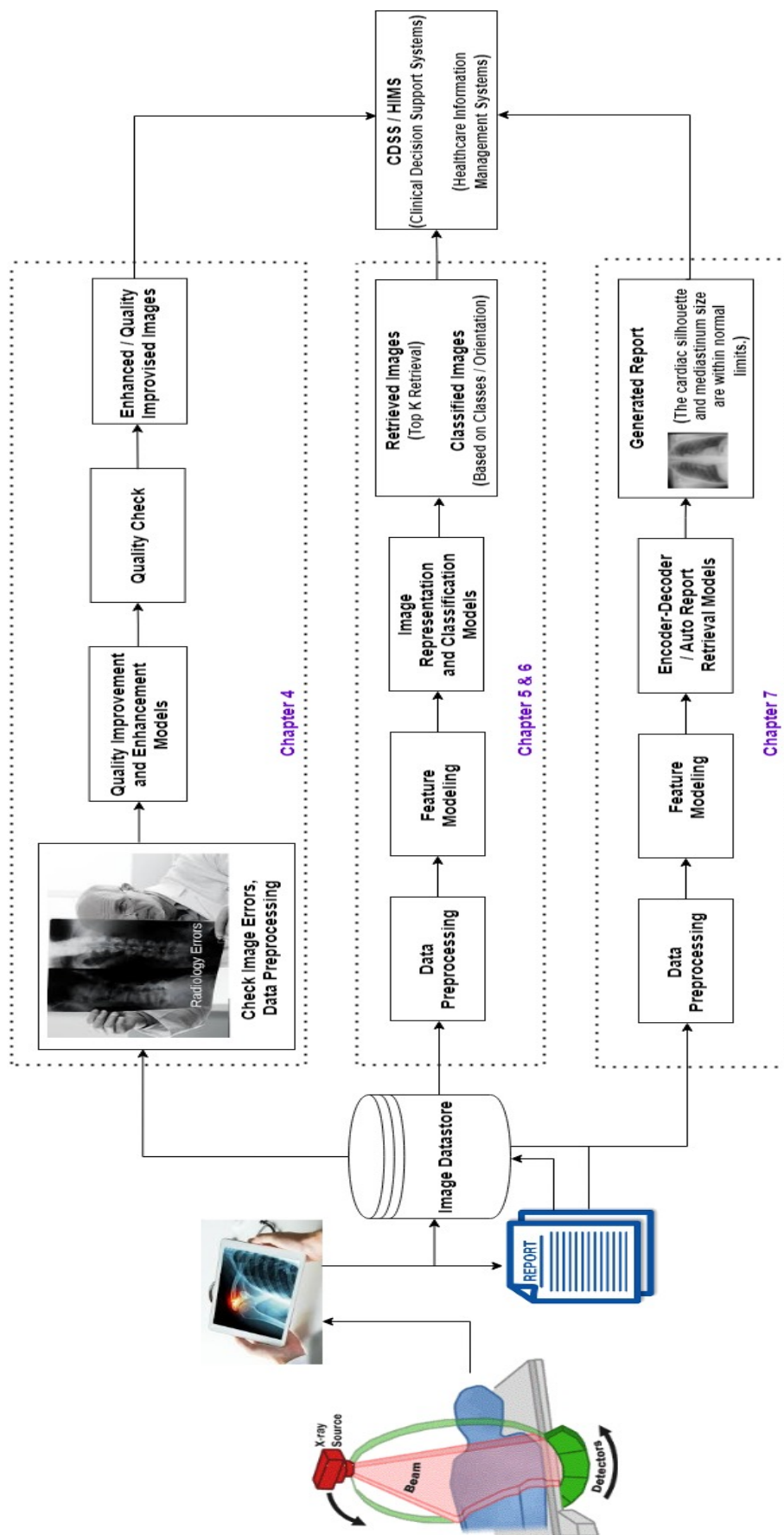


Figure 3.1: Overall workflow of the proposed framework for Quality Enhancement, Modelling and Management of Diagnostic Scans.

3.3 Brief Overview of Proposed Methodology

The overall system architecture of the proposed Integrated Medical Image Modeling and Representation framework for Healthcare Applications is depicted in Fig. 3.1. The diverse set of contributions made towards the defined research objectives with respect to the individual chapters, are presented in more detail in this thesis. A brief outline of the overall research work presented in this thesis is discussed in this chapter.

3.3.1 Medical Image Quality Enhancement

From the extensive research review, it was observed that most approaches in medical imaging mainly rely just on image data without having to do anything with the medical imaging machine configuration involved in image reconstruction. It is also identified that factors like scatter, blur and exposure levels adversely affects the quality of the image, despite significant advancements in imaging technology. Specifically in the case of X-Rays, beam intensity related scattering, metal-induced scattering in cases when the scanned body part has a metal implant, tissue density margin induced scattering and attenuation effect exist, which adversely affect the observations. In addition, faults that occur due to changes in projection angle and rotation axis, overexposed and underexposed images etc, also exist. This affects the resolution of the image itself, resulting in poor quality of X-rays.

Some existing mechanisms proposed for correcting such faults, like, Spatial Non-uniformity identification, Signal to Noise Ratio/Peak Signal to Noise Ratio analysis, Structural similarity indexing, Non-pixelating super-resolution, Gradient descent approaches etc, have been used for modeling the metrics to be controlled in the imaging process. The use of super-resolution techniques is explored for this task, the details of which are presented in Chapter 4. Fig. 3.2 depicts the general methodology defined for addressing this objective.

3.3.2 Medical Image Modeling and Representation

Most diagnostic scan images are generally monochromatic in nature (for e.g., X-ray, MRI). Hence, effective local and global-level analysis of the images is critical. Designing feature modeling and representation techniques to address this aspect is a critical task when designing any CBMIR system since higher-order intelligent applications are built on this data to provide clinical decision support to medical personnel. In view of this, four categories of features were identified-

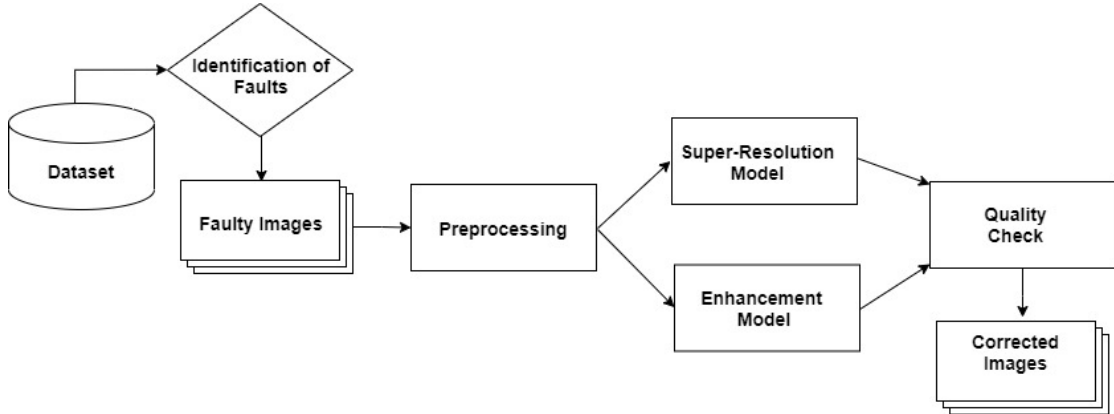


Figure 3.2: Medical Image Quality Enhancement Process

1. *General purpose features* are those which can be extracted from almost any type of image but sometimes not appropriate in case of all applications, e.g., *color* is unsuitable in case of grayscale medical images.
2. *Application-specific features* are specified for a particular problem that describes or capture its unique characteristics; these are semantic features intended to capture a specific meaning or context (Smeulders *et al.*, 2000).
3. *Global features* capture the overall characteristics of an image but fail to identify important visual characteristics from the image if those characteristics appear only in a relatively small part of an image.
4. *Local features* extract the characteristics from a small set of pixels (perhaps even one pixel), representing the details. (Datta *et al.*, 2008).

In most works, local features are used to a great extent since many medical images are not suitable candidates for extracting general-purpose features. However, the incorporation of local and global features still becomes an area of consideration for computer vision applications. To fulfill this objective, a hybrid feature representation model for medical images is proposed. Also, deep learning models are highly effective in the optimal representation of the manifold information contained in medical images. An ensemble deep neural model for classifying abnormal radiography images along-with designing a boundary detection algorithm for identifying the region of interest for facilitating anomaly detection was proposed. Fig. 3.3 depicts a generic workflow for classification and CBMIR task. Chapter 5 presents the detailed discussion regarding the contributions made with respect to this objective.

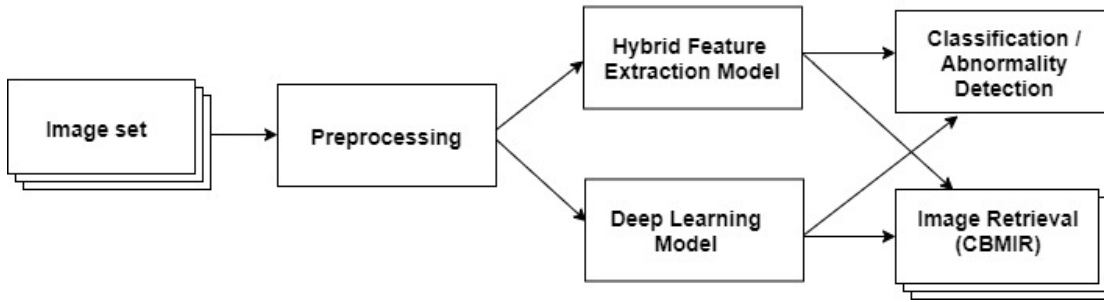


Figure 3.3: Medical Image Modeling and Representation process

3.3.3 Dealing with Variance - View Classification

Real-world medical scan repositories contain scans with inherent variance in modalities, orientation, views, etc, hence dealing with these kind of images is also crucial. At present, identifying the projection view and correcting image orientation of a radiograph are manually performed by radiologists and technicians. As observed from the literature review, very few works address the issues with reference to medical image variance and disparate views. Focusing on the other categorical view of medical scan images with orientation view is necessary for proper reporting and data management systems in clinical labs and hospitals by a radiologist. Hence, approaches that focus on representing variances in an optimal feature space, for supporting view classification are designed through a view orientation identification algorithm. A generic workflow of view orientation classification is shown in Fig. 3.4. Chapter 6 explains in detail the research works undertaken towards this objective.

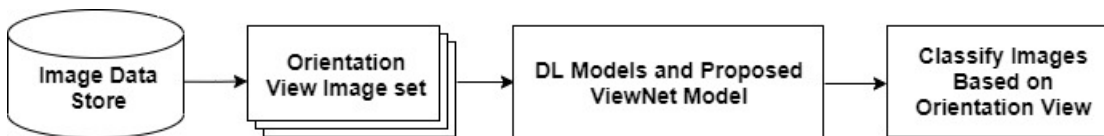


Figure 3.4: Orientation Identification process

3.3.4 Generating Medical Image Descriptions

For automatically generating descriptions of medical scan images, deep neural network models are adopted. Fig. 3.5 depicts the general methodology identified for this purpose. The associated challenges in medical image description generation lie in designing techniques for automatically mapping visual information from medical images towards effective natural language text generation. The semantic gap between captions (text) and the image has to be effectively bridged. Thus,

these two types of information will be utilized for automatically describing new medical images. A detailed explanation on the contributions of this thesis towards this objective is presented in Chapter 7.

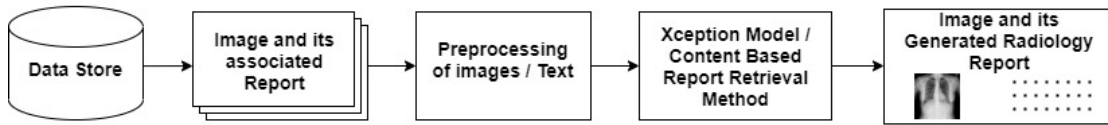


Figure 3.5: Medical Image Caption Generation process

3.4 Research Contributions

This thesis presents a framework for authorizing the design and development for medical image modeling and representation for an effective management in health-care systems is built using radiography data. The objectives are to design and develop automated image quality improvement techniques for enhancing the diagnostic scans, to develop effective feature modeling and representation approaches, along with an efficient techniques for dealing with variance in scanned image views in practical scenarios. Also, approaches for capturing and describing the valuable insights in medical scan images for clinical decision support are proposed for providing an insights into the patients' health outcomes, thus affording intelligent decision-making capabilities to medical/health personnel. With regards to the outcomes gathered from the literature review and the scope of work presented, the major contributions of the research work presented in the subsequent chapters of this thesis are as follows:

- Improving clinical diagnosis performance with automated medical scan quality enhancement algorithms with deep neural image super-resolution models.
- A hybrid feature modeling approach, Swarm Optimization based Bag of Visual Words Model and a deep neural network model for Content-Based Medical Image Retrieval with multi-view classification.
- Deep neural ensemble models for abnormality detection and classification in plain radiographs.
- Design of automated view orientation classification techniques for X-ray images using deep neural networks.
- Deep neural models for automated multi-task diagnostic scan management, including automated generation medical image descriptions.

3.5 Summary

This chapter presents the scope of the research work and the identified research gaps that are addressed in this research work. Based on these gaps, the research problem was formally defined and four objectives pertaining to specific gaps were formulated. The approaches designed for addressing the defined objectives are also discussed briefly, and are explained in detail in subsequent chapters of this thesis.

PART II

Automated Medical Image Quality Enhancement

Chapter 4

Medical Image Quality Enhancement

4.1 Introduction

In modern healthcare, diagnostic imaging is an essential component for diagnosing illness and delivering quality healthcare. The earliest clinical predictions are generally obtained via different modalities of medical images such as X-ray images, Computerized Tomography (CT), Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) among others ([Binh and Tuyet, 2015](#)). Equipment such as X-ray machines and CT/PET/MRI scanners are used in diagnosing disease by capturing a view of the human body's bone, soft tissue, and internal organs. The physical properties of the human body, such as radio-density, bone skeleton structure, etc are thus measured and later interpreted by clinical experts like radiologists for making a diagnosis based on the captured insights. Hence, acquiring good quality diagnostic images is essential for analyzing and determining disease occurrence and progression.

Protocols have been established in the medical fraternity to assess the process by which a medical scan image has been acquired. Often the end goal is to determine small, noticeable differences to indicate unexpected findings and thus successfully diagnosing a particular issue. Common issues observed during imaging process is low-resolution images, under-exposure or over-exposure, and introduction of unwanted artifacts in the image due to movement of the patient undergoing scanning. Hence, methods that improve the spatial resolution of medical images, new approaches to image reconstruction, efficient medical image enhancement methods are essential in clinical workflow management systems. These techniques help improve the visual perception of information to provide better visualization of the diagnosed image.

In this chapter, scan quality enhancement approaches for enhancing the im-

age's spatial resolution and methods for automatically assessing the quality of diagnostic scans are proposed to enable improved visualization of images for more detailed examinations. The objective here is to build a generalized pipeline to reconstruct medical images using different super-resolution techniques and achieve a high correlation with the human visual system. As the end-users of the proposed system would be medical practitioners, hence an accurate model according to human perception is an essential requirement.

4.1.1 Problem Definition

Over time, many medical image quality algorithms have been developed, to tackle the primary issues like noise, edge, contrast issues, poor quality images obtained from faulty or older scanning equipment. However, very few works have focused on addressing the problem of over and under-exposure issues in X-ray images that are often caused by the environmental conditions in which the scans are taken. These over and under-exposed images are not suitable for further diagnosis. High perceptual image quality is a crucial requirement to register the characteristics and details of the scan image in the medical domain. In such cases, super-resolution techniques can be used to enhance the scanned images through computational means. The focus of this work is automated image quality enhancement, with additional emphasis on reducing the computational cost and processing time.

The problem to be addressed here is defined as follows:

Given large number of medical scans, design techniques for automated X-ray Scan Quality Enhancement using super-resolution methods, for enhancing the spatial resolution of the image for improved visualization.

In this chapter, the solution to the issues identified are approaches in two different ways. Initially, the exposure level of each input image is analyzed using an exposure-level detection algorithm, and classified as under-exposed, over-exposed or normal, and normalized using image intensity equalization. Ultimately, the image quality is improved using five different image quality improvement algorithms. Next, reconstruction of medical images along with enhancement using different super-resolution techniques to achieve a high correlation with human visual perception is proposed. The proposed pipeline is automated and aims to provide full image reconstruction to aid the decision-making process after scanning, with significant saving in overheads like time, cost and re-visits of patients for additional scans.

4.2 Radiography Image Quality Improvement

In this study, approaches for automatically assessing the quality of diagnostic scan images as part of clinical workflow management are proposed. The exposure level of each input image is analyzed using an exposure detection algorithm, and classified as under-exposed, over-exposed or normal, and normalized using image intensity equalization. Ultimately, the image quality is improved using five different algorithms, and observations on their comparative performance using standard metrics are reported.

The overall workflow of the proposed X-ray image quality improvement method is depicted in Fig. 4.1. The main objective of this study is to identify, assess and fix faults in the images and improve the visualization so that physicians can appropriately examine the abnormalities. In the first phase, all image exposure levels are checked, based on which, the image is corrected using image intensity equalization methods. Various Super Resolution (SR) techniques are employed for improving the image quality for a better visualization, and performance is assessed using standard metrics.

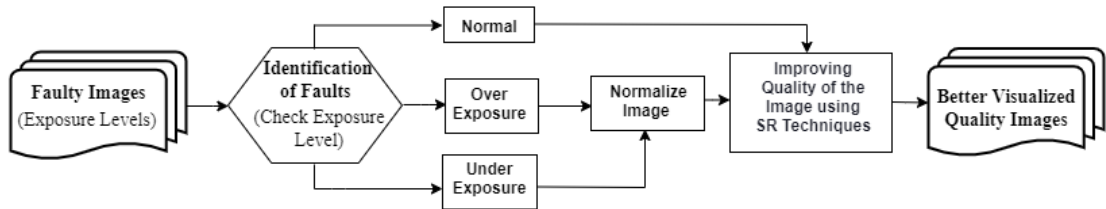


Figure 4.1: Proposed Radiography Image Quality Improvement Approach.

Algorithm 4.1 depicts the process of detecting the exposure level of an image. A sample of images with different exposure levels is shown in Fig. 4.2. Firstly, the image exposure level is checked by computing the histogram level of each image. Based on the exposure level, an image can be classified into three categories - over-exposed (Fig. 4.2(a)), under-exposed (Fig. 4.2(b)) and normal (Fig. 4.2(c)). To check the image exposure level, the threshold points considering the image histogram are used. If the histogram counts are evenly distributed on the scale of 0 to 255, then it is a normal image. If the bin values are more on a scale of 0 to 127, then it is an under-exposed image. Similarly, if the bin values are more on a scale of 128 to 255, then it is an over-exposed image.

If an image is found to be over or under-exposed, it is subjected to a normalization process using image intensity equalization, as shown in Algorithm 4.2. In the

Algorithm 4.1 Image Exposure Detection Algorithm

Input: A Sequence of Images.**Output:** Image Exposure Label with its histogram.

```

1: for each input image do
2:    $Count_I \leftarrow$  Compute the histogram of the image
3:    $Count_O = Count_U \leftarrow 0$ 
4:   for  $i = 1$  to  $\text{length}(Count_I)/2$  do
5:     if ( $Count_I(i) > 0$ ) then
6:        $Count_U = Count_U + 1$ ;
7:     end if
8:   end for
9:   for  $j = 128$  to  $\text{length}(Count_I)$  do
10:    if ( $Count_I(j) > 0$ ) then
11:       $Count_O = Count_O + 1$ ;
12:    end if
13:  end for
14:  if ( $Count_U > 100 \ \&\& \ Count_O > 100$ ) then
15:    Result  $\leftarrow$  'Normal'
16:  else if ( $Count_O > 100$ ) then
17:    Result  $\leftarrow$  'Over Exposed'
18:  else if ( $Count_U > 100$ ) then
19:    Result  $\leftarrow$  'Under Exposed'
20:  end if
21: end for

```

Algorithm 4.2 Image Exposure Correction Algorithm

Input: An Exposed Image**Output:** Normal Image.

```

1:  $Img \leftarrow$  Read the exposed image.  $\triangleright$  Under or Over Exposed image
2:  $N_{IMG} \leftarrow \text{uint8}(255 * \text{mat2gray}(Img))$ 
3:  $[HC, BL] \leftarrow \text{imhist}(Img)$   $\triangleright$  Calculates the histogram of the Image
4:  $\text{imshow}(Img)$   $\triangleright$  Display image
5:  $\text{imhist}(Img)$   $\triangleright$  Display histogram of the image

```

next step, the normalized images are enhanced using enhancement algorithms like Contrast Limited Adaptive Histogram Equalization (CLAHE) and Unsharp Masking (UM) using the Gaussian filtering technique. CLAHE is a histogram-based method used to improve contrast in images, which computes the histogram for the region around each pixel in the image, improving the local contrast and enhancing the edges in each region. Adaptive Histogram Equalization (AHE) over amplifies the noise in the image; CLAHE prevents this by limiting the amplification.

To apply CLAHE to the images, they are first convert to grayscale and then normalized. This approach is similar to N-CLAHE, but log normalization is not

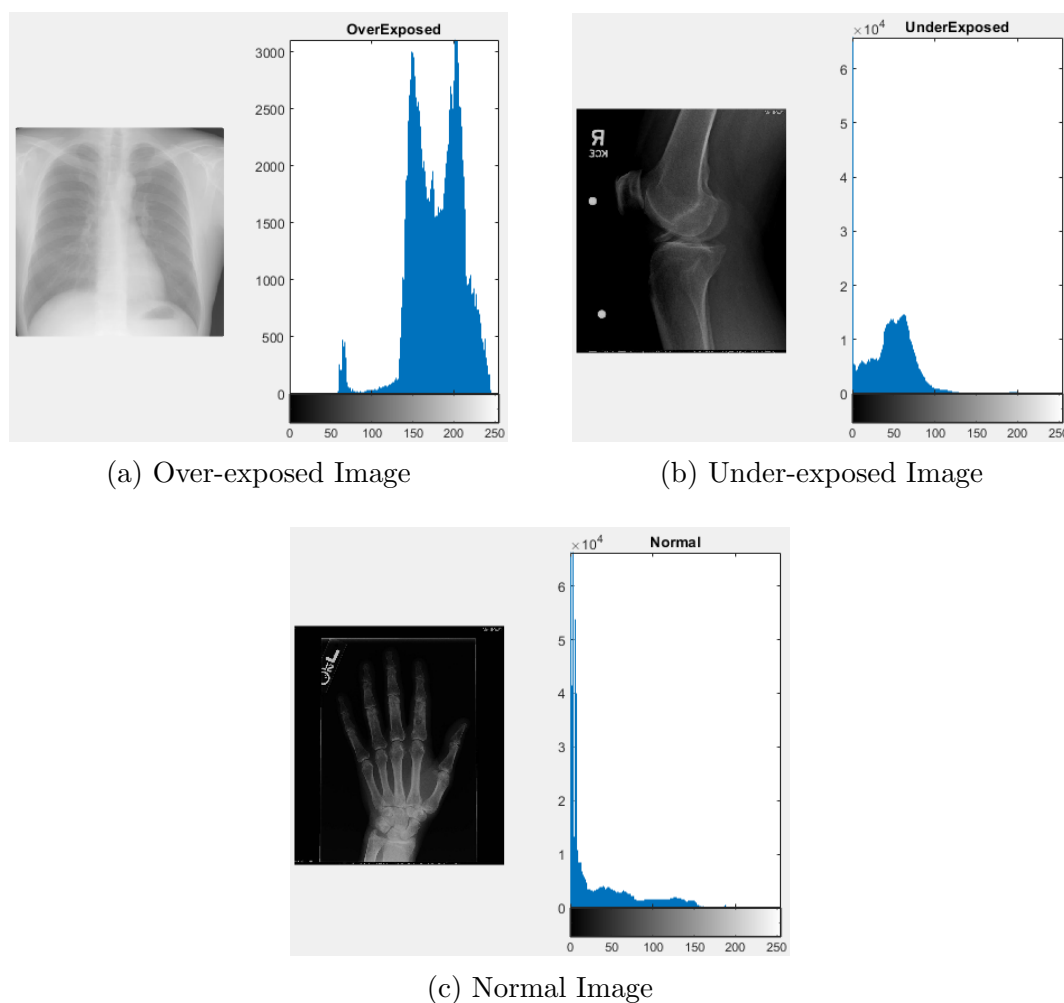


Figure 4.2: Some sample images with different exposure levels and their corresponding histograms.

used. The implementation of CLAHE requires three inputs - *window size* (the size of the rectangular region around the pixel to be processed), *clip limit* (maximum number of occurrences of the pixel in the histogram) and *iterations* (Number of clipping iterations). After this step, the image is padded by reflecting the pixels in the borders. Then, for each pixel in the image, the clipped histogram is calculated for the region around it, i.e., the maximum number of occurrences a pixel can have is defined. If the occurrence is greater than the clip limit, then the exceeding area is cut and redistributed to all other pixels. To improve the technique, this process can be repeated a certain number of times until the desired contrast image is obtained. With this clipped histogram, the probability of each pixel is calculated and the CDF (Cumulative Distribution Function) is computed using the cumulative sum of the ordered pixels. Then, each value of the function is multiplied by 255, to limit the image's values to $[0, 255]$. After calculating the CDF, all pixels will

have a transformation value, which is now applied to the pixel in the center of the region. For certain images, the image becomes very noisy when the clip limit chosen is very high. This may be because, when the limit is very high, no clipping is performed, and the CLAHE algorithm is essentially similar to AHE algorithm.

Unsharp masking is a linear filter that is capable of amplifying high frequencies of an image. The first step of the algorithm is to copy the original image and apply a Gaussian blur into it (Blur intensity is defined by a setting called Radius). If the blurred image is deducted from the original image, only the edges created by the blur are obtained, which is called the unsharped mask. The radius setting is related to the blur intensity because it defines the size of the edges. The amount, on the other hand, controls the intensity of the edges (how much dark or light it will be). The experimental results, observations and further developments are discussed in the subsequent sections. Finally, the enhanced image is collected after computed and visualized using Eq. (4.1).

$$\textit{sharpened_image} = \textit{original_image} + \textit{amount} * (\textit{unsharped_mask}) \quad (4.1)$$

In addition to this, bicubic interpolation was employed for upscaling the low-resolution (LR) image, that resulted to a high-resolution image where the dimension is similar to that of the reference image. Further, neural network SR methods like Very-Deep Super-Resolution (VDSR) (Kim *et al.*, 2016) and Single image super-resolution CNN (SRCNN) (Dong *et al.*, 2014) were utilized in developing a high-resolution (HR) image. VDSR network builds a HR image using a single LR image by learning and mapping the difference in its frequency. The network consists of an image input layer, then with a 2-D convolutional layer that consist of 64 filters. A total of 20 convolutional layers, each of which that follows a ReLU activation layer builds up the network, which introduces nonlinearity in the network. A image patch size of 41-by-41 is used and the network was trained for 100 epochs deploying stochastic gradient descent with momentum (SGDM) optimization. The learning rate was initially fixed to 0.1 and decreased with a factor of 10 for each 10 epochs. VDSR works with a surplus learning strategy, i.e., the network learns to assess with a surplus image. A surplus image informs regarding the high-frequency characteristics of an image. For super-resolution cases, a surplus image is a variation with a HR reference image and a LR image that has upscaled using bicubic interpolation by matching the dimension of the image to the reference image.

SRCNN learns pixel mapping between LR and HR images with pre-processing

optimization techniques. The first phase deals with patch extraction and representation, where, patches ($f_1 \times f_1$) from LR image are extracted and each patch is represented as a HD vector, which is a set of feature maps equivalent to its dimensions. The next operation is non-linear mapping, which performs a non-linear mapping of each HD vector (n_1) to another HD vector (n_2). Here, each mapped vector represents a HR patch. Finally, reconstruction operation combines all the HR patches ($f_2 \times f_2$) to generate the final HR image. We set the parameters as $f_1 = 9$, $f_2 = 5$, $n_1 = 64$ and $n_2 = 32$ while constructing HR image from a LR image. The learning rate is set to 10^{-4} in initial two layers, and 10^{-5} in the final layer. Empirically, a small learning rate was set to the last layer for the network to converge.

4.2.1 Experimental Results and Discussion

For experimental validation, a dataset of medical images from MedPix¹ consisting of 66 different body organ images was used. Standard evaluation metrics like PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) were used for measuring the quality of the reconstructed image. Given two images, I (ground truth image) and \hat{I} (reconstructed image), both of same size, the MSE and the PSNR (in dB) is given by Eq. (4.2) and (4.3), where, I is the maximum intensity of a grayscale image i.e. 256. X_{ij} and Y_{ij} are the intensity of original and reconstructed image. MSE is the Mean Square Error, M and N are the number of rows and columns in the image. Wang *et al.* (2004) SSIM is a visual metric that measures the quality of a reconstructed image with the effect of luminance, contrast and structure (as per Eq. 4.4, where, C_1 and C_2 are constants used to keep away the uncertainty when μ_x and μ_y are very close to zero. σ_x, σ_y are contrast comparison functions).

$$\text{PSNR} = 10 \log_{10} \left[\frac{I^2}{\text{MSE}} \right] \quad (4.2)$$

$$\text{MSE} = \frac{1}{[N \times M]^2} \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - Y_{ij})^2 \quad (4.3)$$

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4.4)$$

¹The National Library of Medicine MedPix[®], <https://medpix.nlm.nih.gov/home>

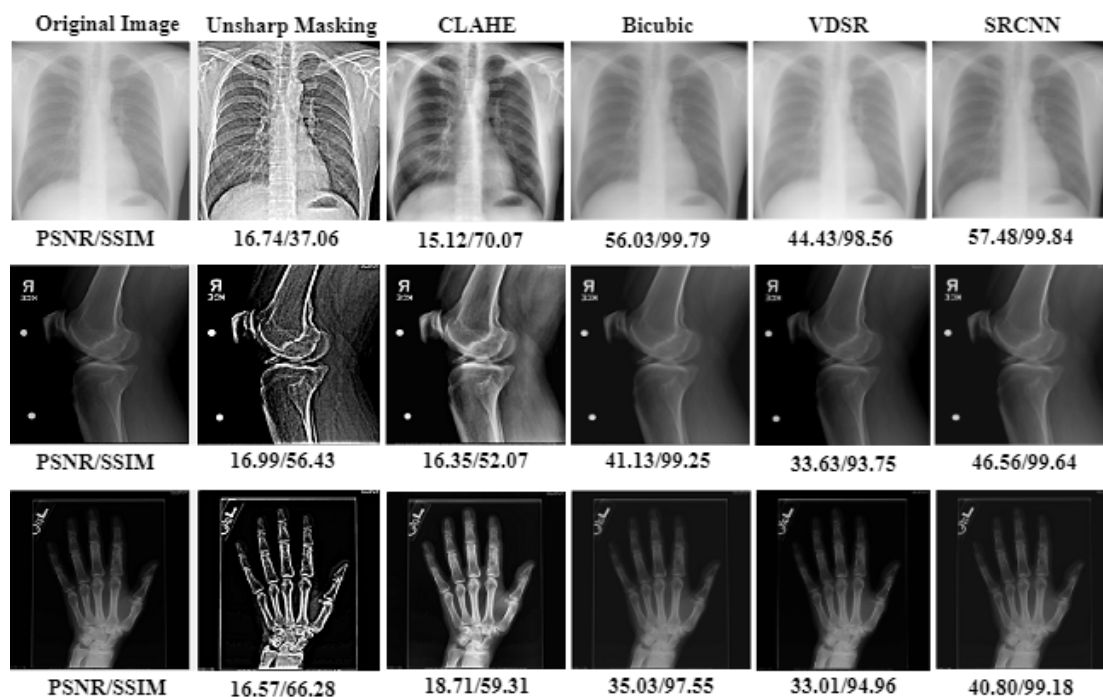


Figure 4.3: Comparative Evaluation of UM, CLAHE, Bicubic, VDSR and SRCNN for X-ray image enhancement.

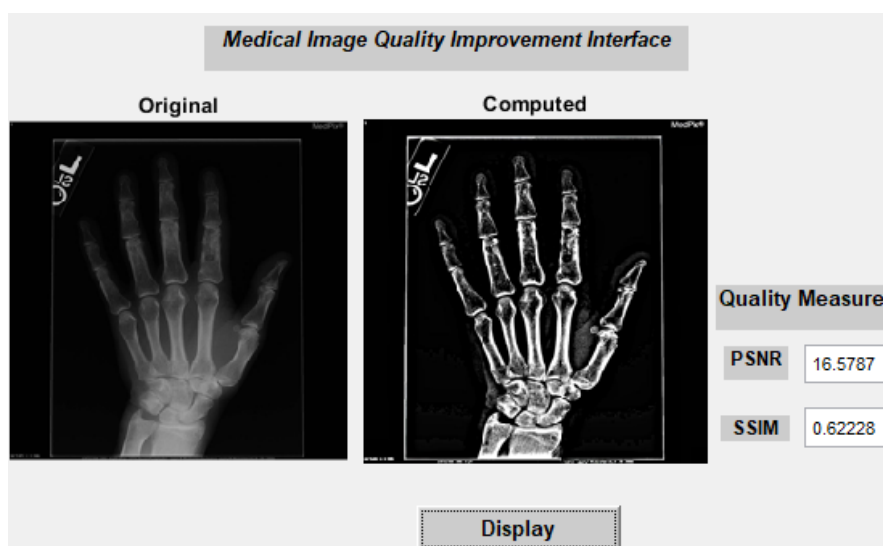


Figure 4.4: Illustration of X-ray Image Enhancement with quality metrics.

Fig. 4.3 shows the results of the comparative evaluation of the five X-ray image quality improvement methods on sample images. Based on the observations, it can be concluded that UM and CLAHE methods performed well in improving the quality of the image for a better visualization, whereas, in terms of HR image

display from a single LR image SRCNN outperformed among other SR methods by achieving good PSNR/SSIM index values. A simple GUI also was developed for illustrating the quality enhancement by visualizing the original and corrected image, which is shown in Fig. 4.4.

4.3 Automatic Quality Enhancement of Medical Diagnostic Scans with Deep Neural Super-Resolution Models

Based on the observations from the previous work, it can be seen that the latent features in the medical scans were captured well using edge and contrast enhancement, in turn amplifying the visibility of region of interest or artifacts. A patch size of 16 pixels (4×4) in Bicubic interpolation resulted in a smoother image, while VDSR showed better performance while transforming a LR image to HR. SRCNN outperformed all other methods due to its lightweight architecture and superior learning behavior. It was also understood from discussion with medical experts that, as end-users of these image quality enhancement system are human medical practitioners, modeling human perception accurately is an essential requirement. Hence, the next work is aimed at building a generalized pipeline for the reconstruction of medical images using different super-resolution techniques and to achieve a high correlation with human visual perception.

Fig. 4.5 illustrates the complete process designed for this evaluation. Single-Image Super-Resolution has played a vital role in Computer Vision related applications and led to the inception of many new algorithms. There are different variants of the models based on the approach they take, e.g., Prediction models, Edge-based models, Image Statistical models, Patch-based models. The benchmarking for all these model variants involves two sets of images used as the ground truth data, as seen in (Yang *et al.*, 2014). Each of these processes are discussed in further detail in this section.

Super-Resolution Reconstruction (SRR) methods consist of processing single or multiple images to increase their spatial resolution. Deployment of such techniques is particularly essential when high-resolution image acquisition is associated with high cost or risk, like medical or satellite imaging. Unfortunately, existing SRR techniques are not sufficiently robust to be deployed in real-world scenarios (Kostrzewa *et al.*, 2018). No real-life benchmark to validate multiple-image SRR has been published so far, to the best of our knowledge. As gathering a set of

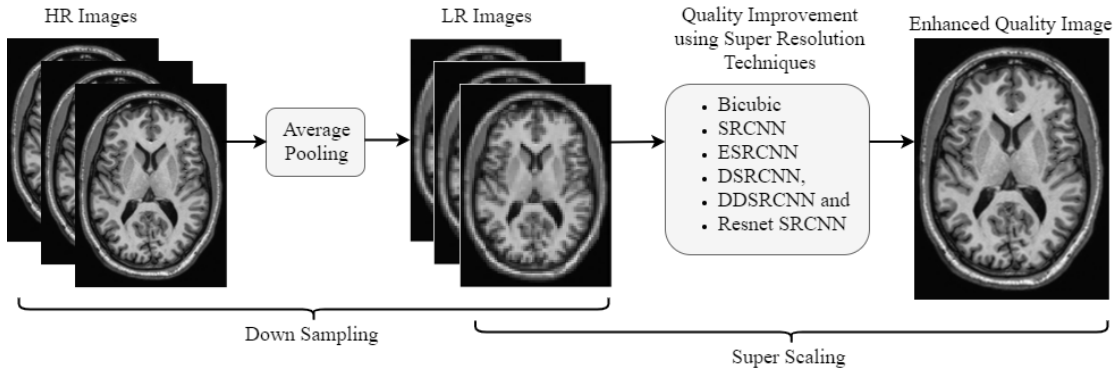


Figure 4.5: Deep CNN model for Automated Scan Quality Enhancement

images presenting the same scene at different spatial resolution is not a trivial task, the SRR methods were evaluated based on different assumptions, employing various metrics and datasets, often without using any ground-truth data.

Dong *et al.* (2015) proposed a novel method for the super-resolution of images using CNN, called SRCNN, which is one of the first models that applied CNN for image super-resolution. With full end-to-end utilization of the CNN, the SRCNN model takes an LR image as input while generating an HR image as output. Expanded Super Resolution CNN (ESRCNN) (Dong *et al.*, 2015) is an extension of the SRCNN model where the *Expansion* occurs in the intermediate hidden layer. Instead of using 1x1 kernels, kernels of order 3x3 and 5x5 kernels are used to maximize information learned from the layer. The outputs of this layer are then averaged in order to construct more robust upscaled images. Denoising (Auto Encoder) Super Resolution CNN (DSRCNN) (Dong *et al.*, 2015) is another extension of the SRCNN, which uses autoencoders as intermediate level layers. The models use bridge connections between the convolutional layers of the same level to speed up convergence and improve output results. The bridge connections are averaged to be more robust. Deep Denoising Super Resolution CNN (DDSRCNN) (Mao *et al.*, 2016) incorporates a framework where rectification layers are added after each convolution. Deconvolution and skip connections dividing the network into sequential blocks also give the model better element-wise correspondence ability. The ResNet Super Resolution network is derived from the SRResNet (Ledig *et al.*, 2017), which is intended to use the latest ResNet architecture for the Super-Resolution task by increasing the residual blocks and thus the upscaling capability of the network.

For the experimental evaluation, the cross-sectional images of the brain from

the IXI Dataset² were used, a sample of which is shown in Fig. 4.6, which were pre-processed and normalized. It is a publicly available dataset and it consists of nearly 600 Magnetic Resonance (MR) images. Each subject's MR image is further split into T1, T2, PD weighted and Magnetic Resonance Angiography (MRA) images. The individual images were 3-Dimensional and stored in the well-known nii format³. These 600 images were split into two sets of 550 and 50 images for training and evaluation purposes. The individual cross-sectional 2-D image obtained from these nii images has a dimension of 230X230. The SimpleITK⁴ library was used to process the nii format and obtain the cross-sectional images.

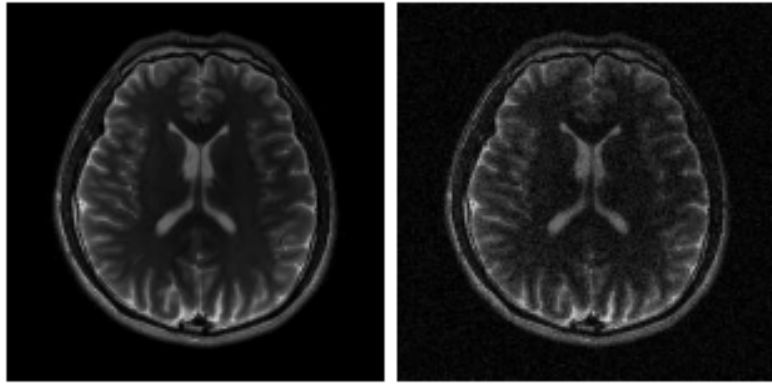


Figure 4.6: Sample images from IXI dataset²

The dataset consists of high-resolution images only, from which low-resolution images were recreated using average pooling. A tuple set for every image was created, which includes both the LR and HR image, as shown in Eq. (4.5). The images were normalized to zero mean and unit variance. It is crucial to normalize features (in our case, images) because if one of the features has a large range of values, then its contribution to the model parameters would be more. Hence, all features are squashed to a smaller range, which is done by centering all pixel values across the mean and dividing it by the standard deviation, as shown in Eq. (4.6), where μ and σ are the mean and standard deviation, respectively.

$$T_{im} = (LR_{sim} : 'x', HR_{GT} : 'y') \quad (4.5)$$

$$Im_{x,y} = (Im_{x,y} - \mu) / \sigma \quad (4.6)$$

The normalized images were used to build the dataset. The original input

²<http://brain-development.org/ixi-dataset>

³<https://nifti.nimh.nih.gov/nifti-1>

⁴<http://www.simpleitk.org>

images are treated as the ground truth and through the pooling operation, a low-resolution image is formed. Apart from passing this LR image to the models, bicubic interpolation (Keys, 1981) is also used to generate the HR image used to compare and test the other models' performance. For obtaining LR images from the ground truth image, a forward 3D Average pooling layer is used. A layer is a form of non-linear downsampling of an input tensor. The input tensor is partitioned and split further into 3D sub-tensors, where the average value of the sub-tensors is calculated and used to create the output tensor, as per Eq. (4.7). Here, x and p represent the size of the input image and the pooling sub-tensor, respectively. O is the output tensor and I is the input tensor. The obtained output tensor is the simulated LR image. The average pooling was done only to reduce the resolution by a factor of 2.

$$O_{x',y'} = \left(\sum_{i=1}^p I(x - p/2 + i, y - p/2 + i) \right) / p * p \quad (4.7)$$

After pre-processing, the 3-D images were represented as rank five tensors. A typical CNN architecture would accept a 4-D tensor. The extra dimension is due to the 3D nature of the image, which is represented as multiple cross-sections of the brain. These rank five tensors were input to the individual models. These images were trained for a scaling factor of 2. So the trained models were capable of doubling the resolution of any 3-D image. The models were created according to the individual models' architecture and modified accordingly to fit the 3-D images. These were then split into the training and evaluation dataset. The trained model was then fed LRsim images from the evaluation dataset and predicted high-resolution images were obtained for evaluation. A new tuple set consisting of the ground truth and the predicted image was created and fed into the models to be benchmarked for experimental evaluation, performed using standard metrics, the details of which are described in next section.

4.4 Experimental Evaluation and Results

For the validation of the proposed approach, four standard metrics were used to evaluate the models. These are – Peak Signal to Noise Ratio (PSNR) (Wang *et al.*, 2004), Structural Similarity Index Measure (SSIM) (Wang *et al.*, 2004), Multi-scale Structural Similarity Measure (MS-SSIM) (Wang *et al.*, 2003) and Visual Information Fidelity (VIF) (Sheikh and Bovik, 2006). Higher values for each of these metrics imply a greater similarity to the ground truth.

Peak Signal to Noise Ratio. PSNR (Wang *et al.*, 2004) is the ratio of the maximum value a pixel can take to the mean squared error. It is one of the easiest methods to calculate metrics and is expressed in the logarithmic decibel scale to accommodate the pixels' dynamic range. Although, in some cases, a higher value of PSNR might not indicate that the reconstruction is of a higher quality. PSNR loses out on other models when it comes to predicting human visual response to image quality. Two images can have the same PSNR but a pronounced difference in quality due to the objective nature of the way it is calculated. The PSNR (in dB) and MSE are computed as per Eq. (4.8) and (4.9), where I is the maximum intensity of a grayscale image i.e., 256. X_{ij} and Y_{ij} are the intensity of an original and reconstructed image. MSE is the Mean Square Error, M and N are the numbers of rows and columns in the image.

$$\text{PSNR} = 10 \log_{10} \left[\frac{I^2}{\text{MSE}} \right] \quad (4.8)$$

$$\text{MSE} = \frac{1}{[N \times M]^2} \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - Y_{ij})^2 \quad (4.9)$$

Structural Similarity Index Measure. SSIM (Wang *et al.*, 2004) is an improvement on PSNR and is similar to PSNR with regards to parameters used. Both use the same parameters but are combined differently. Also, both these measures depend on the absolute values of the input. SSIM is a composite measure of luminance, contrast coefficient and correlation coefficient, which is given by Eq. (4.10), (where C_1 and C_2 are constants used to keep away the uncertainty when μ_x and μ_y are very close to zero. σ_x, σ_y are contrast comparison functions).

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4.10)$$

Multi-scale Structural Similarity Measure. MS-SSIM (Wang *et al.*, 2003) is a variant of SSIM where the process of SSIM is applied over multiple scales through sub-sampling. It performs equally well or better than SSIM and also correlates better to human visual response to image quality, as per our observations. It is computed as per Eq. (4.11). Here, the constant $M=5$, and the exponents $\alpha_M, \beta_j, \gamma_j$ are selected such that $\alpha_M = \beta_j = \gamma_j=1$.

$$\text{MS-SSIM}(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=i}^M [c_j(x, y)]^{\beta_j} \cdot [s_j(x, y)]^{\gamma_j} \quad (4.11)$$

Visual Information Fidelity. VIF (Sheikh and Bovik, 2006) correlates best with the human visual perception with a Spearman’s rank coefficient⁵ of 0.96. Hence, this metric is comparatively much better than the other metrics considered. It uses natural scene statistics and tries to model a human visual system, under the assumption that the uncertainty in human perception of visual signals can be modeled as white Gaussian noise, and is given by Eq. (4.12).

$$VIF = \frac{I(C; F)}{I(C; E)} \quad (4.12)$$

where, C is the reference image, F is the distorted image and E is defined as the image that the Human Visual System (HVS) perceives. HVS tends to correlate between the real image and the distorted image, i.e., the visual quality. In this regard, VIF tests the algorithm using a human assigned assessment score. Thus, VIF proves to be a better visual quality assessment metric.

All the neural models were compared against the Simple Bicubic Interpolation model (Keys, 1981). Bicubic interpolation considers the nearest 4x4 neighborhood of known pixels for a total of 16 pixels than the bilinear model, which takes 2x2 neighborhoods. Bicubic produces noticeably sharper images and is perhaps the ideal method for processing time and output quality, which has been used as a base method for many image enhancement models so far. Compared to the other models, bicubic interpolation is much faster as there are no dense layers of neurons. Hence, there is a tradeoff between the time taken to resolve and the quality of resolution. In the medical domain, for most cases, there is a need for real-time super-resolution. In such scenarios, it might be more suitable to use a faster model than a slower model that provides high-quality images. The converse could also be true in some cases. It was found that the ResNet SRCNN (Ledig *et al.*, 2017) model outperformed the other models by far. The comparative evaluation of the six image quality enhancement methods, when applied to a sample image, as shown in Fig. 4.7.

Table 4.1 shows that ResNetSRCNN outperformed the other models by far, in terms of all considered metrics. The difference can be clearly observed in the PSNR and the VIF values. There was no much deviation in the values obtained for the other models compared to Bicubic Interpolation. Although the Deep Denoising SRCNN improved on DSRCNN, the margin of improvement was small and did not outperform much as per expectations. However, the ResNetSRCNN model significantly outperformed Bicubic by a large margin, as seen by the four metrics’

⁵<http://videoclarity.com/wp-content/uploads/2013/05/Statistic-of-Full-Reference-UT.pdf>

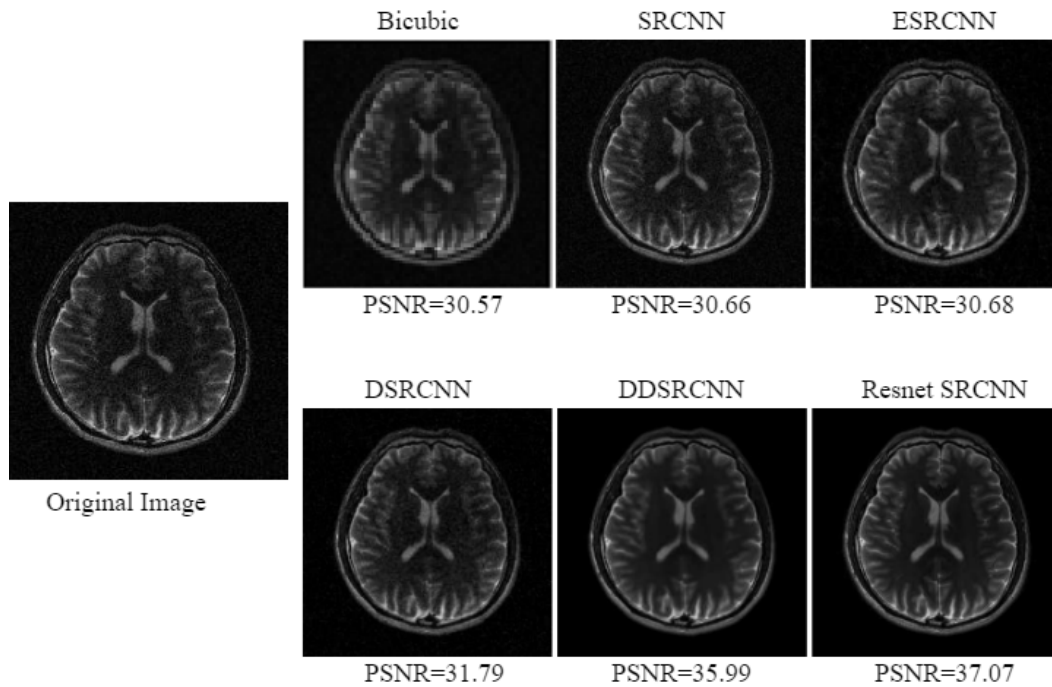


Figure 4.7: PSNR performance of Bicubic, SRCNN, ESRCNN, DSRCNN, DDSRCNN and ResNet SRCNN

values. An improvement of 6.5 is seen in PSNR values, which generally may not indicate superior image reconstruction performance, given the objective nature of PSNR.

Table 4.1: Quality Evaluation Metric Scores across different Models

Model	PSNR	SSIM	MSSIM	VIF
Bicubic	30.57	0.922	0.950	0.460
SRCNN	30.66	0.928	0.966	0.465
ESRCNN	30.68	0.929	0.969	0.467
DSRCNN	31.79	0.935	0.970	0.496
DDSRCNN	35.99	0.950	0.984	0.573
ResNet SRCNN	37.07	0.984	0.995	0.689

Furthermore, ResNetSRCNN achieved a VIF value of 0.689, which is a 22.9% improvement over Bicubic, thus showing significantly superior visual perception, as VIF correlates very well with human judgments of visual quality. VIF was the best metric to visualize the models' performance as the improvement could be clearly observed. The models' performance with respect to VIF scores is presented

Table 4.2: Benchmarking the proposed ResNet SRCNN model against State-of-the-art.

Approach	PSNR	SSIM
ResNet SRCNN (proposed)	37.07	0.984
Compressed Sensing MRI Recon (Abdullah <i>et al.</i> , 2019)	32.20	–
mustGAN (Yurt <i>et al.</i> , 2021)	29.45	0.947
GBWRT Recon (Abdullah <i>et al.</i> , 2019)	29.15	–
SIDWT Recon (Abdullah <i>et al.</i> , 2019)	29.04	–
pGAN _{many} (Yurt <i>et al.</i> , 2021)	28.80	0.940
pGAN _{one} -B (Yurt <i>et al.</i> , 2021)	28.73	0.936
pGAN _{one} -A (Yurt <i>et al.</i> , 2021)	28.39	0.934

*Note: SSIM values were not reported with approaches mentioned from (Abdullah *et al.*, 2019).*

in Table 4.1. The performance of the ResNet SRCNN is compared with other existing methods (Yurt *et al.*, 2021; Abdullah *et al.*, 2019), and tabulated in Table 4.2. The compared results also show a clear improvement over other methods with an improvement of 4.87 in the PSNR value. Based on observations during the experiments and comparing the time required by the different SRCNN models with the Bicubic model, it was concluded that the Bicubic model performs Super Resolution faster for medical images and its VIF values are pretty close to the SRCNN models implying that they would appear similar to doctors trained to employ the dominant features to diagnose the anomaly.

4.5 Summary

In this chapter, the works undertaken for addressing Objective 1, i.e., to improve the quality of the diagnostic scan images using super resolution techniques were presented. In the first work, image super-resolution algorithms, Unsharp mask using gaussian filter, CLAHE, Bicubic Interpolation, VDSR network and SRCNN models were implemented for image super-resolution, and evaluated using standard metrics like PSNR and SSIM for better visualization of X-ray images. Experiments were performed to comparatively evaluate the above five approaches. Based on the visualized enhanced images, VDSR showed better performance while transforming a LR image to HR. However, SRCNN outperformed all other methods due to its lightweight architecture and superior learning behavior. As part of the second work, five CNN based models were adapted for the medical scan qual-

ity enhancement task. An ensemble model ResNetSRCNN was designed, which showed good performance with reference to standard visual quality metrics. The proposed model outperformed other state-of-the-art models in terms of PSNR and SSIM by a large margin. The results of the studies emphasized the suitability of the proposed approaches for diagnostic scan image enhancement for better visualization, thus help in reducing the overall burden, time and cost in case of low quality diagnostic scan captures.

Publications

(based on work presented in this chapter)

1. Karthik K. and Sowmya Kamath S., “Improving Clinical Diagnosis Performance with Automated X-ray Scan Quality Enhancement Algorithms”, International Conference on Advances in Systems, Control and Computing (AISCC 2020), MNIT Jaipur (Springer proceedings). *(Status: Online)*
2. Karthik, K., Sowmya Kamath S. and Surendra U. Kamath, “Automatic Quality Enhancement of Medical Diagnostic Scans with Deep Neural Image Super-Resolution Models”, 15th International Conference on Industrial and Information Systems (ICIIS). IEEE, IIT Ropar, Punjab (CORE Ranked) *(Status: Online)*

PART III

Medical Image Modeling and Representation

Chapter 5

Medical Image Modeling and Representation

5.1 Introduction

With the proliferation of various imaging-based diagnostic procedures in the health-care field, patient-specific scan images constitute huge volumes of data that must be well-organized and managed to support clinical decision support applications. One such crucial application with a significant impact on point-of-care treatment quality is, content-based medical image retrieval (CBMIR) that can assist doctors in disease diagnosis based on similar image retrieval. The earliest approaches were keyword-based image querying systems, *i.e.*, a CBIR system that aims to capture the latent features from an image without requiring any external information (text metadata associated with images). Later, CBIR systems started to evolve based on the potential matching of images identified as per their actual visual content overlapped with a given query image. CBIR makes use of image-level features, where most of the CBIR systems are dependent on low-level features. As a result, a particular subset of features may be highly suitable for some image classes. In contrast, other features may be suitable for the remaining classes, thus making it difficult to pick out the most optimal feature attributes. In this chapter, approaches like hybrid feature modeling, bag of visual feature representation, and a convolutional neural network based modeling, to represent a rich set of visual attributes extracted from the medical images are presented in detail.

5.1.1 Problem Definition

Scan images like X-rays, CT scans etc can encompass several internal organs, it is essential to devise automatic classification approaches to deal with the diversity.

As medical images are multi-dimensional and often contain manifold information, due to which efficient techniques for optimal feature extraction from large-scale image collections are the need of the day. Despite significant improvements in medical image retrieval, conventional retrieval models that support querying based on text/keywords fail to capture the latent visual features in an image, paving the way to CBIR systems for medical images. Designing effective feature extraction mechanisms can help to improve overall retrieval accuracy. Thus, the problem to be addressed here is defined as follows:

“Given a large volume of medical image set consisting of several classes along with high variability in the images in each class, design and develop effective feature modeling and representation techniques for enabling intelligent clinical applications.”

In this chapter, the identified issues are addressed in two ways. As an initial work, an efficient CBMIR model built on multi-level feature sets extracted from medical images is presented. Four different feature extraction techniques are used to optimally represent images in a multi-dimensional feature space, for facilitating classification using supervised machine learning algorithms and top- k similar image retrieval. Next, a MedIR approach based on the Bag of Visual Words (BoVW) model for content-based medical image retrieval was designed. Further, a deep neural network-based approach for content-based image retrieval was also developed for demonstrating its suitability in efficient medical image retrieval.

5.2 Hybrid Feature Modeling for Content-Based Medical Image Retrieval

In this section, the proposed supervised learning framework that incorporates hybrid feature modeling techniques for large-scale medical image data, to enable content-based image retrieval is presented. For the experiments, the ImageCLEF 2009 dataset (Lehmann *et al.*, 2003c) was used, consisting of 12,560 images across 116 different categories. The processes involved in the proposed CBMIR system are illustrated in Fig. 5.1.

5.2.1 Noise Removal & Contrast Enhancement

Medical images often contain some visual noise. In x-Ray images specifically, the presence of noise is commonly due to random photon distributions. As this

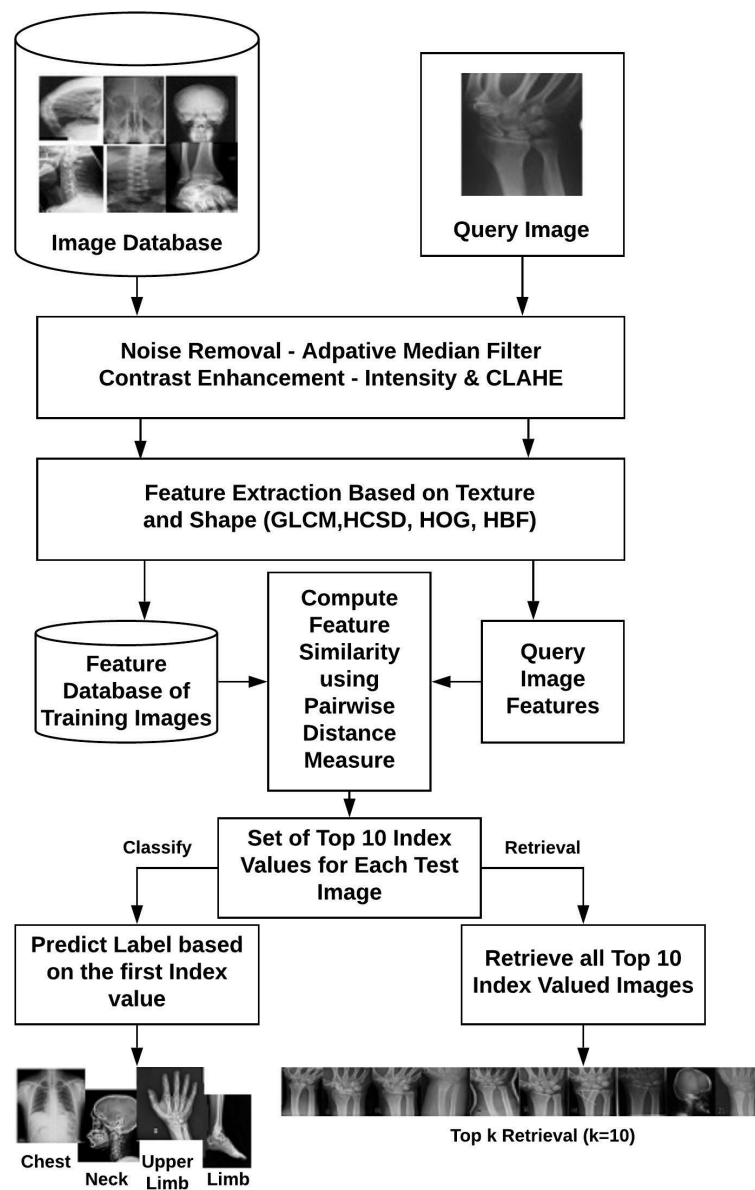


Figure 5.1: Workflow of the Proposed CBMIR Model

causes undesirable variations and also affects the visibility of objects of interest (like bones, organs etc.) in the image, it is vital to eliminate this noise before the process of feature extraction. We used an adaptive median filtering technique for removing inherent noise from the radiography images in the dataset. Adaptive median filtering (Hwang and Haddad, 1995) is an image enhancement technique that works well due to its low sensitivity to pixel value changes. It also does not affect the sharpness of the scan image, preserving even small objects-of-interest.

Next, intensity changes and contrast adjustment techniques are used to im-

prove image quality further. The intensity value of an image is adjusted by saturating all its pixel values by 1%. Similarly, image contrast is boosted by transforming its pixel values with a technique called Contrast-Limited Adaptive Histogram Equalization (CLAHE) (Zuiderveld, 1994). CLAHE operates on small regions in the image, called *tiles*, rather than the entire image, and each tile's contrast is enhanced individually. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image. The transformation of the image after applying these processes is shown in Fig.5.2. After these processes, feature modeling is performed for extracting texture and shape based features.

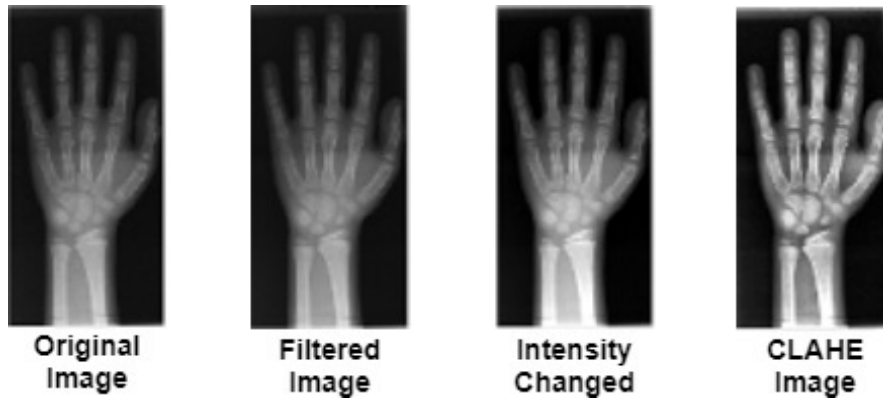


Figure 5.2: Radiographic image enhancement process

5.2.2 Feature Modeling

The performance of a CBIR system is highly dependent on the inherent visual properties of images, represented formally to facilitate retrieval. Most diagnostic scan images are generally monochromatic (e.g. X-ray). Hence, effective local and global-level analysis of the images are critical. Also, the dataset contains scans belonging to different classes, hence capturing this variation is also essential for dealing with the variety of images. In the proposed work, four different feature extraction techniques were used for obtaining a well-rounded image representation. As feature extraction plays an important role in image retrieval performance, we focused on texture and shape based feature extractions.

The Gray Level Co-occurrence Matrix (GLCM) (Haralick *et al.*, 1973) is a statistical method that examines the spatial relationship between the pixels in an image for generating features. The GLCM feature vector is composed of 14 features which are primarily texture based features. These are extracted for each

X-ray image in the dataset. These features are related to the specific characteristics of the image such as Angular Second Moment, Contrast, Correlation, Variance and Homogeneity. These features capture the individuality of an image in the form of probability of a pixel finding its gray level intensity i at distance d , which can be formulated as $P(i, j : d)$. Here, each pixel is associated with its 8 neighboring pixels except the edge pixels.

The Hierarchical Centroid Shape Descriptor (HCSD) (Ilunga-Mbuyamba *et al.*, 2016) generates a 124-dimensional feature space, and is augmented using a kd-tree technique proposed by (Sexton *et al.*, 2000). HCSD is a shape feature extraction method which starts by recursively decomposing an image into sub-images. Initially, it takes image I as the input and computes its transpose I^T , calculates the centroid $C(x_c, y_c)$ for each input with 1st order moment along x and y axis, where $x_c=m_{10}/m_{00}$ and $y_c=m_{01}/m_{00}$. The *order of moment* of a 2D function $f(x, y)$ is formulated as per Eq. (5.1). A digital image's raw moment m_{pq} with $I(i, j)$ pixel intensities is calculated as per Eq. (5.2). The *centroid* is calculated by dividing the image area recursively until the desired depth (here, we considered depth=7) is reached. Fig. 5.3 illustrates how the centroid values are calculated from the gray scale images. At each level, the axis of the coordinates is captured, after which the generated vector is normalized to $[-0.5, 0.5]$ range, point 0 being the root level. Finally, the features extracted are concatenated to the image feature.

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (5.1)$$

$$m_{pq} = \sum_{i=0}^M \sum_{j=0}^N i^p j^q I(i, j) \quad (5.2)$$

Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005) is implemented by dividing the image dimension into smaller regions and storing the histogram orientation of the pixel values of that region. Each pixel of the region has a weight as per the gradient L2-norm. Here, histogram channels are performed based on the rectangular region. As overlapping can occur, regions are contributed more than once to the final feature vector. Initially, this feature extraction was used for generic images, but it is found to be well-suited for medical images, as it can be used to extract the directional change of intensity levels in the image. A total of 9 rectangular regions are used here with 9 bin histogram per region. The 9x9 feature vector is now concatenated to the image's feature vector, thus resulting in 81 dimensional feature vector.

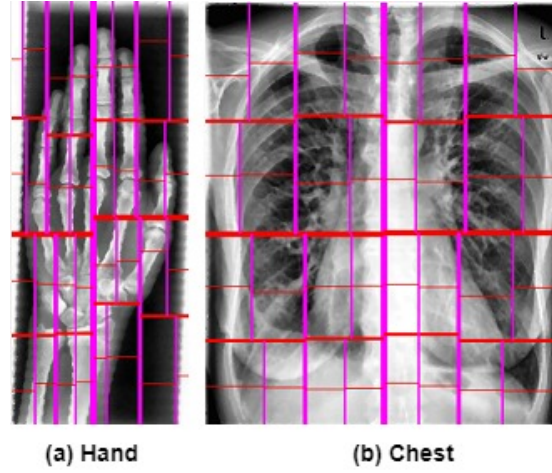


Figure 5.3: HCS D extraction from X-ray images

In case of Histogram Based Features (HBF), five different features, namely – mean, variance, standard deviation, median absolute and RMS values, were used. Let I be the variable indicating image gray levels, $p(z_i)$ be the histogram where $i = 0, 1, 2, \dots, L-1$, where L is the number of discrete gray levels. The average gray level of each region in the image (mean) is computed as per Eq. (5.3). Variance is the amount of the difference in the gray level (Eq. (5.4)).

$$\mu = \frac{1}{N} \sum_{i=1}^N A_i \quad (5.3)$$

$$S = \frac{1}{N-1} \sum_{i=1}^N |A_i - \mu|^2 \quad (5.4)$$

Standard deviation captures the amount of dispersion from the mean gray level of an image (Eq. (5.5)). A low value of standard deviation indicates that it is very close to the mean value, on the other hand a high value means that the data are spread over a range of values. The Median absolute is the measure of the variability in the image and it is computed as per Eq. (5.6). Root Mean Square (RMS) is used to compute the overall contrast of the image as per Eq. (5.7).

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N |A_i - \mu|^2} \quad (5.5)$$

$$M = \mu |X - \mu(X)| \quad (5.6)$$

$$I_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N |I_i|^2} \quad (5.7)$$

The feature values obtained after using the four feature extraction techniques described are fused to form a hybrid feature vector, which optimally represents each X-ray image. The hybrid feature vector is a 224-dimensional vector, composed of features as shown in Table 5.1. Thus, at this stage, all the images are processed and represented by their 224-dimensional feature vector. These are used for identifying how close the feature vectors of any two given images are, when a query image feature vector is submitted to the retrieval system.

Table 5.1: Summary of Feature Modeling processes

Method	Description	Total features
GLCM	Texture properties of the image	14
HCS D	Recursively decomposes the image into sub-images	124
HOG	9 rectangular regions with 9 bin histogram per region	81
HBF	Mean, variance, standard deviation, median absolute and RMS values of the input image	5
Hybrid	Combination of GLCM, HCS D, HOG and HBF features	224

5.2.3 Pair-wise Similarity & Class Label Prediction

Once the training and testing image feature vectors are generated, a pairwise distance calculation technique is used on them for determining the most relevant image index values for each of the testing image. As a distance measure, the Standard Euclidean pairwise distance method was adopted, which gave the best result, i.e., it gave the best nearest image index values for the testing image feature vectors to the training feature vectors. For top- k retrieval, where $k=10$, 10 image index values were generated from the training set for each of the testing image.

From the top-10 index values, the first index value is considered for the actual classification of the test image. The label of all images at the first index value is fetched from the training set which will be the predicted label for the classification task. For classification and label prediction, the k-nearest Neighbors (kNN) algorithm (Cover and Hart, 1967) was used. kNN classifies an unknown object into

its category among the training data and it uses the nearest neighbors between the two set of vectors i.e., between the test vector and the training vectors. kNN is a widely used machine learning algorithm due to its simplicity and the results that were obtained when this technique is applied was found to be excellent in many applications. Additionally, six variations of the kNN algorithms were experimented with, by varying the number of neighbors used. The kNN models – *fine*, *medium* and *coarse* are based on the nearest number distance calculations set to 1, 10 and 100 respectively. In the cosine, cubic and weighted kNNs, the respective distance metrics are used for comparing two dimensional vectors. The classification task was validated by using 10-fold cross validation.

5.2.4 Content-based Image Retrieval

After the pairwise distance calculation for all the testing images with the training image set, top 10 image index values were obtained, and their labels are also predicted. These image index values are now used for the content-based image retrieval. For each test image, its pairwise distance calculated with reference to the image index value is read from the training image set. Next, the test image and the read top 10 indexed images are displayed as the retrieval results, and the performance of the retrieval is observed.

5.3 Experimental Results and Discussion

For the experimental validation of the proposed approach, the ImageCLEF2009 dataset consisting of 12,560 labeled X-ray images of different body regions like face, nose area, shoulder, elbow, forearm, chest, arm, hand, wrist, finger, knee, leg, foot, ankle etc, was used. Each category of images was divided into training and testing with a ratio of 70:30. Some sample images from the dataset are shown in Fig. 5.4. The implementation was carried out with an Intel Xeon Workstation @3.31 GHz and 16 GB of memory running Matlab v.2017.

Standard Euclidean distance metric was used for pairwise distance measurement between the two image vectors, and the classification results are tabulated in Table 5.2. It can be seen that this achieved good accuracy of 85.91%, with significantly reduced false positive rate. With respect to this the performance of the proposed model with state of the art methods compared is reported in Table 5.3. The hybrid feature vector was fed to six variants of kNN classifiers, with 10-fold cross validation. From Table 5.4, it can be seen that the kNN classifier

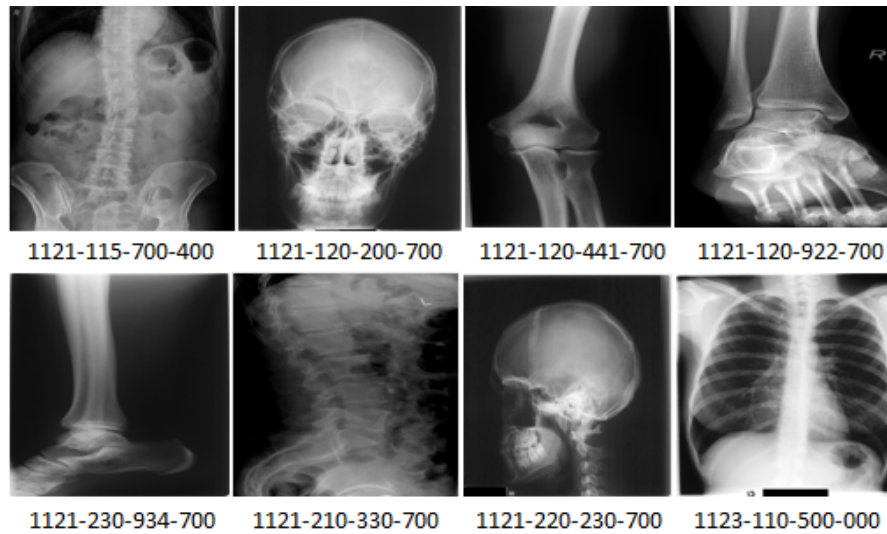


Figure 5.4: Sample IRMA dataset images and their corresponding IRMA codes (Lehmann *et al.*, 2003c)

model using Cosine Measure as the similarity metric achieved the best accuracy. To validate these results, the Receiver Operating Characteristic curves (ROC) were plotted for one versus other classes. The ROC curves for the class *hand* and *leg ankle joint* versus other classes plotted using cosine kNN classifier is shown in Fig. 5.5.

Table 5.2: Observed results for the Standard Euclidean Pairwise distance

Evaluation Metrics	Values (%)	Values (%)
	(Before Enhancement)	(After Enhancement)
Accuracy	82.97	85.91
Error	17.03	14.09
Precision	60.34	63.47
Recall	55.22	58.73
Specificity	99.64	99.70
False Positive Rate (FPR)	0.36	0.30
F1 Score	57.67	61.09

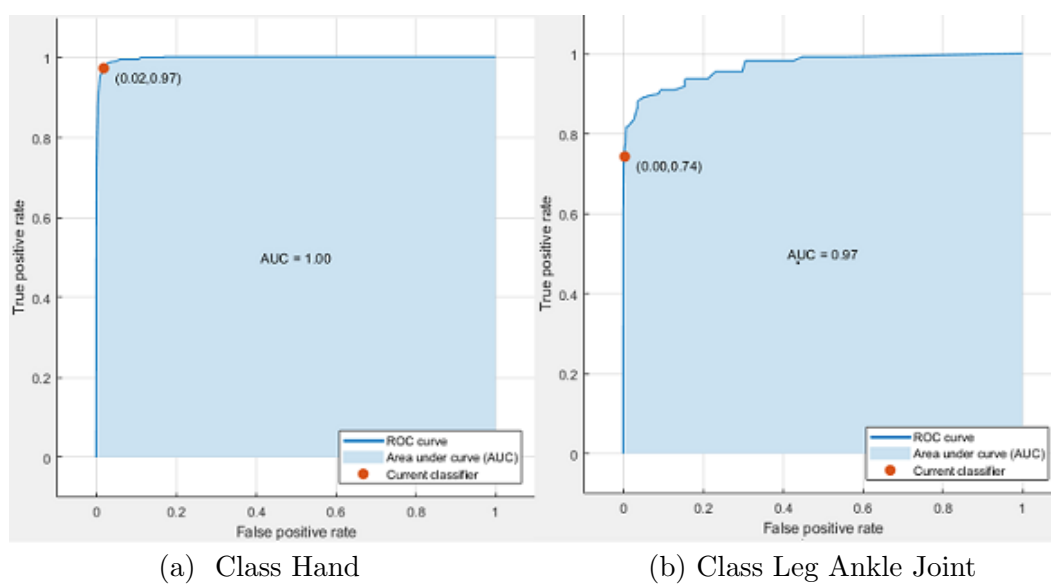
It was observed that most classes had a good retrieval performance while for some classes the retrieval was average. This is because, the ImageCLEF dataset displays high class imbalance, i.e., several classes have more than 200 images, while some classes have only 10-15 images. Thus, the classifier was trained on less data,

Table 5.3: Benchmarking the proposed approach with existing works

Approach	Total Classes	Accuracy (%)
Hybrid Approach (<i>Proposed</i>)	116	85.91
Shape Features and Bayesian Rule (Fesharaki and Pourghassem, 2012)	28	82.87
Merging Scheme (Pourghassem and Daneshvar, 2013b)	57	90.23
Combined Features (Zare et al., 2011)	116	$\geq 80\%$ for 70 classes

Table 5.4: Classification accuracy for different kNN variants

Classifier Model	Accuracy (%) (Before Enhancement)	Accuracy (%) (After Enhancement)
Fine kNN	79.1	81.6
Weighted kNN	79.3	81.4
Cosine kNN	78.5	81.3
Medium kNN	77.6	79.8
Cubic kNN	76.2	78.7
Coarse kNN	66.7	68.5

Figure 5.5: AUROC values for *one versus all* classes using Cosine kNN classifier

due to which label prediction accuracy and retrieval performance was low. Top- k retrieval results for some sample classes for which high accuracy was observed are shown in Fig. 5.6 and those classes where average accuracy was obtained are shown in Fig. 5.7. As per observations, out of 116 classes in ImageCLEF, label prediction accuracy was $> 90\%$ for 38 image classes, $> 60\%$ for 70 classes and the remaining classes showed $< 40\%$.

To evaluate retrieval performance, two popular IR metrics, *precision@k* ($p@k$) and *Mean Average Precision (MAP)* were used. Precision@ k is given by the ratio of images that are retrieved in top k set that are actually relevant. It can be computed as per Eq. (5.8). The MAP of a set of testing images is defined as the mean of the average precision scores for each query image (Eq. 5.9). MAP@ k can also be computed accordingly, by considering precision scores at k , where, Q is the number of query/test images, $q = 1, 2, 3, \dots, Q$.

$$p@k = \frac{\text{No. of retrieved images @k that are relevant}}{\text{Total images retrieved @k}} \quad (5.8)$$

$$MAP = \frac{\sum_{q=1}^Q AvgP(q)}{Q} \quad (5.9)$$

Table 5.5: Evaluation of retrieval with precision@ k for $k=3, 5, 10$

Test Image	k Value	Relevant images@ k	Precision@ k
Sample 1 (Class-Cranium)	k=3	3	100%
	k=5	5	100%
	k=10	10	100%
Sample 2 (Class-Arm)	k=3	3	100%
	k=5	5	100%
	k=10	9	90%
Sample 3 (Class-Right Leg)	k=3	3	100%
	k=5	5	100%
	k=10	9	90%
Sample 7 (Class-Nose Area)	k=3	3	100%
	k=5	5	100%
	k=10	8	80%
Sample 8 (Class-Umbar Spine)	k=3	3	100%
	k=5	5	100%
	k=10	8	80%

Table 5.6: Evaluation of retrieval performance for all 116 classes

k value	Observed MAP@k
k=3	87.06%
k=5	86.01%
k=10	83.91%

The image retrieval process was evaluated for various values of k , in order to conclusively evaluate the performance. In case of real-world medical diagnostics applications, high precision during top-3 and top-5 retrieval is really important, as the doctor can get clear information w.r.t to the submitted image, hence k values of 3, 5 and 10 were chosen for a comprehensive evaluation. The results of conducted experiments for some sample image classes are presented in Table 5.5 and the MAP@ k computed for all 116 classes is shown in Table 5.6 respectively. From the results, it can be clearly observed that the top-3 and top-5 retrieval performance of the proposed approach is almost equal. This can also be seen in the MAP@ k performance, as MAP@ k was 87.06% for $k=3$, even at $k=5$, MAP@5 was about 86.01%, which means the performance of the proposed approach was excellent and indicates a well-balance performance. As k increases, it can be seen that the class imbalance problem comes to play, due to which performance degrades.

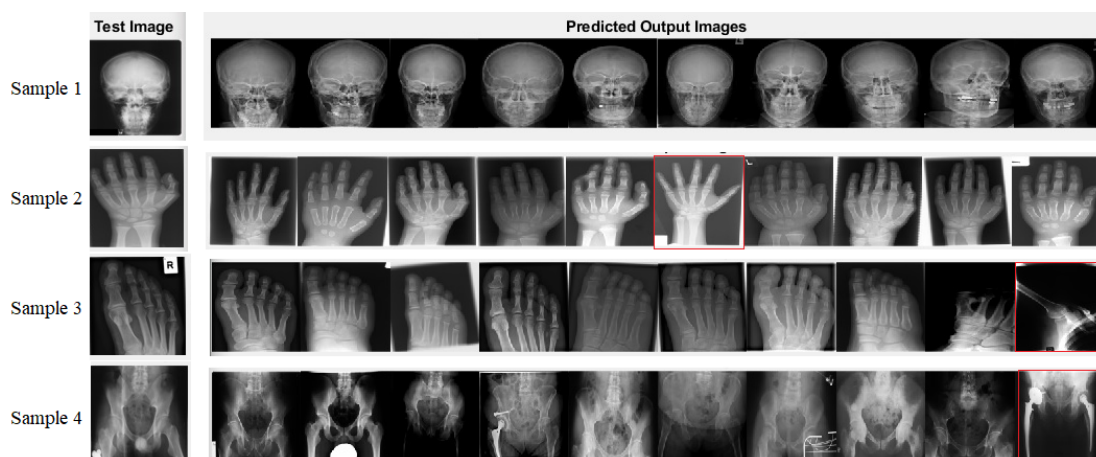


Figure 5.6: Observed retrieval results (classes with high accuracy)

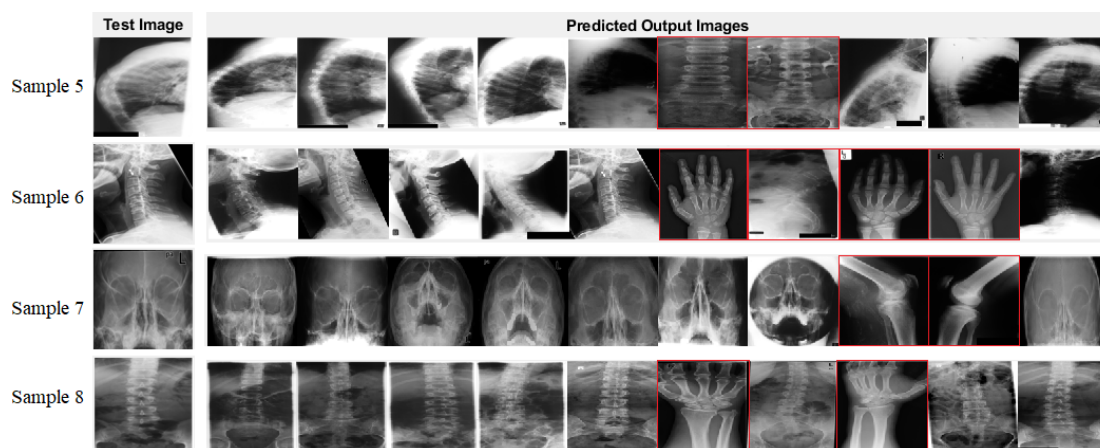


Figure 5.7: Observed retrieval results (classes with average accuracy)

5.4 Swarm Intelligence based Bag of Visual Words Model for CBIR

The proposed method depicted in Fig. 5.8, focuses on classification of X-ray images into different classes for enabling fast retrieval. The model is built on the Bag of features method for classification and retrieval of similar kinds of images for the given test image. The visual image category employed is a method to set a categorical label to a given test image.

5.4.1 Preprocessing and Feature Generation

During this phase, an ordered array of the extracted features from the images is constructed based on the image categories. The dataset is inherently balanced, hence thresholding on class probabilities is performed and set to 97 of images on each class in the dataset. These images are taken into account for all the classes, which makes an equal distribution of images per class. Classes that do not meet this threshold are not considered and which meets the threshold, a random of 97 images are taken into account. At this step, each class of image set contains an equal number of images. The image set is then separated, 70 for training and the remaining 27 for testing from each category. A total of 31 different categories of medical X-ray images have been used in the proposed framework and for the experiments conducted for evaluating classification and retrieval performance.

As X-ray image scans are taken from different angles on the test parts of the body, the SURF (Speeded Up Robust Features) feature extraction (Bay *et al.*, 2008) was performed. SURF performs an automatic selection of interest points

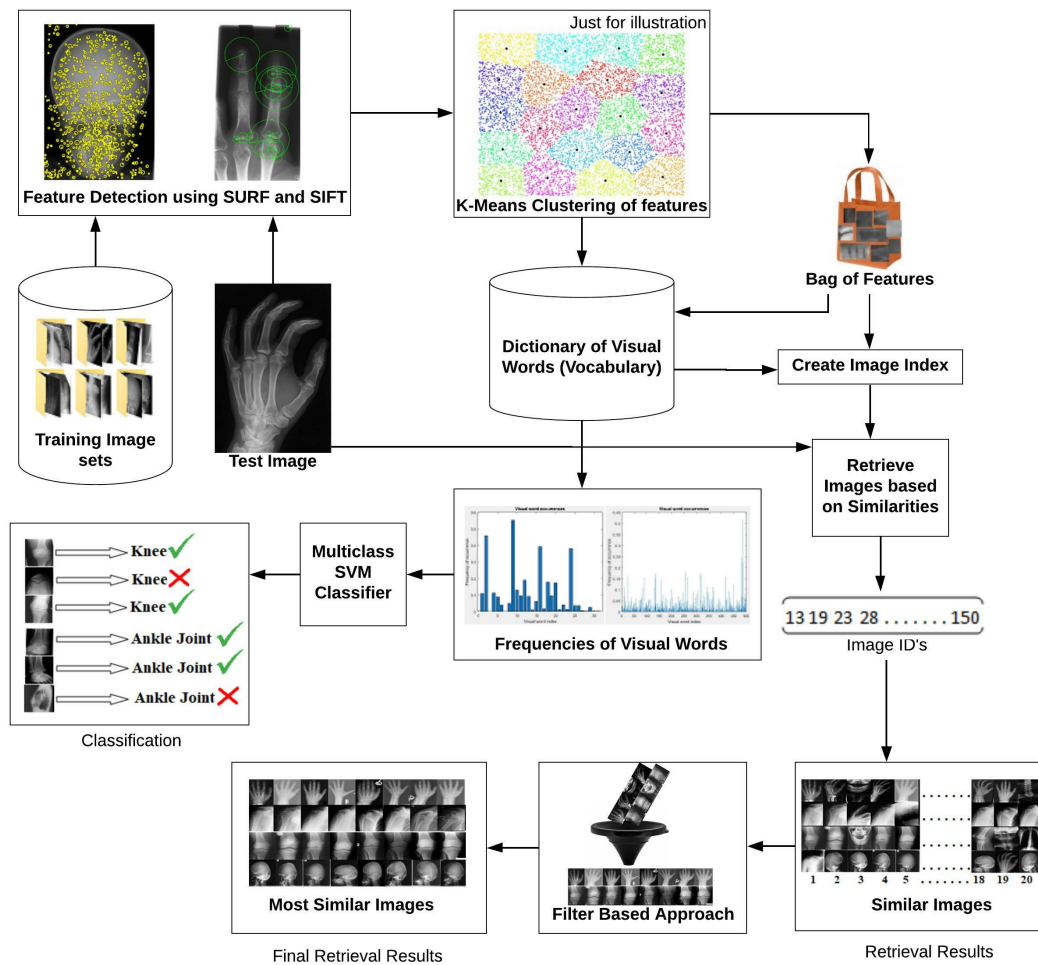


Figure 5.8: Proposed BoVW+PSO approach for CBMIR

that are found at different scales (Lindeberg, 1998). SURF descriptors provide a robust and unique representation of an image using the pixel intensity distribution within the interest points of their neighboring pixels. SURF is a popular local feature detector and descriptor, which is used for tasks such as object recognition, image registration, classification etc, and is based on the scale-invariant feature transform.

Next, rotational invariance of the image was found based on Haar wavelet in both $x - y$ directions in a circular radius of $6s$ ($s =$ scale at which interest point was detected). An illustration of the interest points located in a sample image of class *Finger* is shown in Fig. 5.9a. The SURF detector provides greater scale invariance and the algorithm runs on 'grid' method. The final SURF feature vector

length that is generated has 64 feature values, which are similar to the one when extracted from a different image of the same class (shown in Fig.5.9b). Thus, SURF provides a useful way for detecting features that have scale invariance, contrast invariance and rotation invariance in the image.

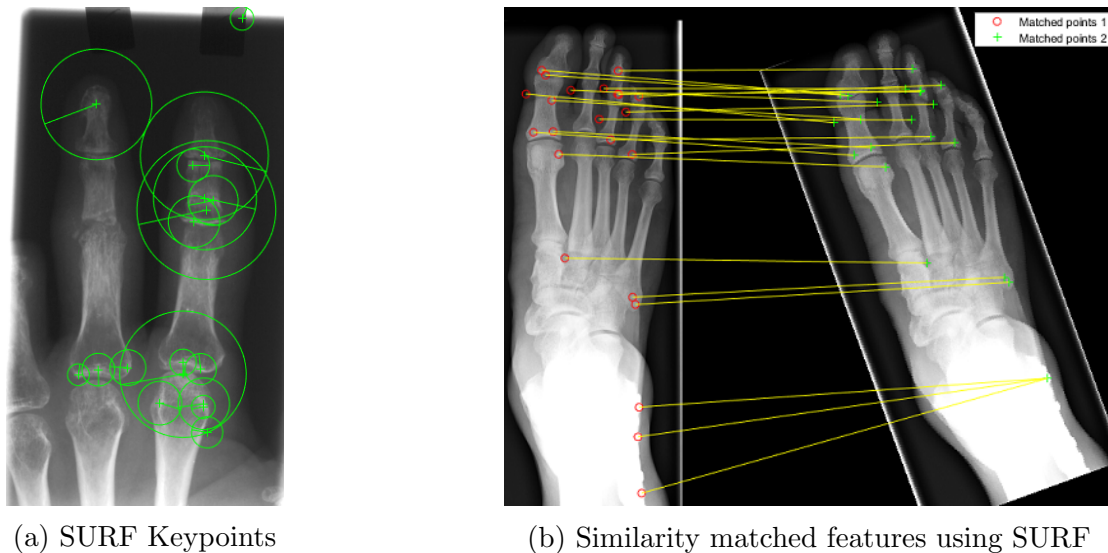


Figure 5.9: SURF feature generation process

5.4.2 Visual Vocabulary Construction and Training

The visual vocabulary helps in effective representation of the monochromatic X-ray images in a concept feature space. As images do not contain discrete words, a ‘vocabulary’ of visual words needs to be constructed by extracting features from a set of images of each category. Essential parts of an image form a feature vector with its dimensionality by the feature extraction process. The SURF detector is used to find the interesting points in the images and encode information about the area around the points as a feature vector. In our case, feature extraction is accomplished by applying the SURF descriptor to the X-ray images across all image categories, as per the defined experiment instances, which are then used for vocabulary construction.

The visual vocabulary is constructed by means of quantization using the K-means clustering technique (Lloyd, 1982) by a minimized set of features. To enhance clustering, the strongest features are kept from each image after feature extraction. Next, the visual vocabulary is built with K-Means clustering, and the number of clusters (K value) is determined. As developing a BOVW model is a user-dependent choice based on the dataset, the value of k is empirically

determined by varying k values from 30 to 600 with a step size of 25 at each experimental analysis to find the best results. In addition, the BoF object provides an encoding scheme to determine the visual word occurrences in an image. A histogram is produced, which becomes a reduced description of an image, as shown in Fig. 5.10. The histogram is then formed for training a classifier and also for the image classification. With reference to this, an image is encoded into its feature vector. Training images that are encoded from each category are then provided into the classifier for the purpose of training. The classifier was tested on the test set for observing the confusion matrix and for calculating the accuracy. The newly trained classifier was then used to categorize the test images and for predicting the label based on the index value.

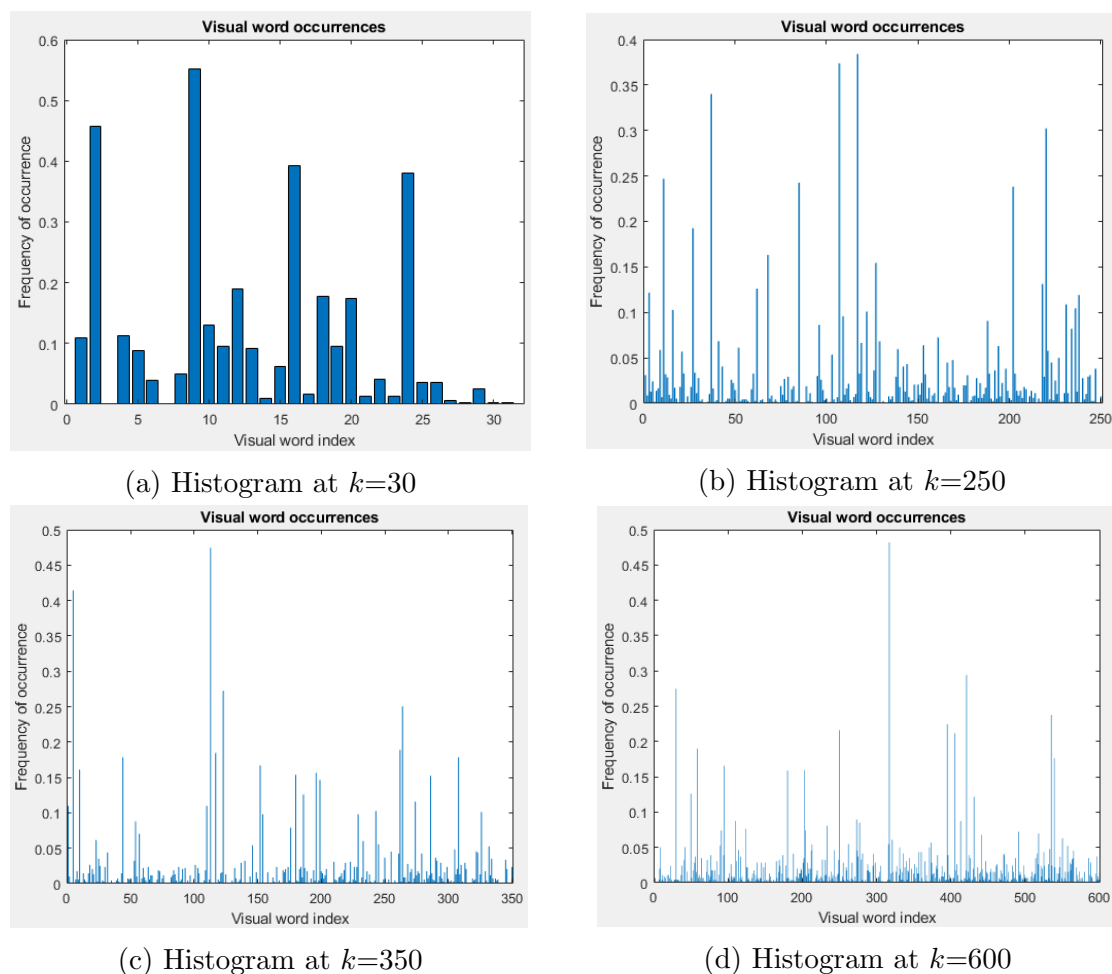


Figure 5.10: Histogram of visual word occurrences with different vocabulary sizes

5.4.3 Image Indexing and Retrieval

Once the vocabulary is constructed, the set of training images are indexed, i.e., a search index is created that maps the visual words to their occurrences in the image. Based on this index, a search of similar images is performed and matching images are retrieved based on the given query image. The number of identical images that need to be retrieved can be decided depending on how many similar images need to get displayed for the given test image. Now, when a test image is given, a similar image index is retrieved from the training set and these images are displayed as the initial retrieval results.

To obtain the best match retrieval results for a given test image, a novel technique called *Filter Based Image Retrieval Approach* was designed. In the filter based technique, the focus is on eliminating the least-useful retrievals and maximize the most-useful results. When a test image is provided as a query to retrieve similar images initially, it returns 20 images IDs that are similar to the given test image. All these 20 images are displayed as the first retrieval result (I_G in Algorithm 5.1). From this result, the images which do not belong to that particular class are further filtered based on the training image set index number. Each class of images holds a starting and an ending index number. Considering this as the reference range, the image IDs which do not belong to the range are excluded, while only those image IDs which are inside the range values are retained. These image IDs are now treated as the best match and are displayed as the final retrieval result.

5.4.4 PSO based Retrieval Optimization

Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995) is a population-based optimization algorithm, used for various optimization problems. The main strength of PSO is its fast convergence and suitability for diverse problem spaces. Each particle is considered from a population of particles, and each particle is an object which keeps moving in the search space, till it finally moves towards the best position. Using a fitness evaluation function, PSO decides the next best/better position based on the computed fitness value. Hence, the objective is to optimize the fitness function which is generally pre-defined based on the problem.

PSO based clustering generates a compact cluster from a low-dimensional dataset much more efficiently and in a better time when compared to the traditional K-means clustering method. However, while clustering a large dataset, the slow shift from the global searching phase to the local refining phase causes an

Algorithm 5.1 Filter based Retrieval process

```

1: Initialize  $N \leftarrow 0$ 
2:  $NSets \leftarrow \text{length}(\text{unique}(\text{Test\_Image\_Labels}))$ 
3:  $\text{TestImages} \leftarrow \text{length}(\text{TestImages})$ 
4:  $TIES = \text{TestImages}/NSets$ 
5: for  $i = 1$  to  $\text{length}(\text{TestImage})$  do
6:    $I \leftarrow \text{TestImage}$   $\triangleright$  Read the test image
7:    $I_R \leftarrow$  Resize the image.  $\triangleright$  Such that width=height
8:    $I_{ID} = \text{retrieve Images}(I, II)$   $\triangleright$  Get Image ID's
9:    $I_G \leftarrow []$   $\triangleright$  Initialize image gallery
10:  for  $j = 1$  to  $\text{length}(I_{ID})$  do
11:     $\text{MatchID} = I_{ID}[i]$ 
12:     $\text{Matched} = \text{ImageLocation}[\text{MatchID}]$ 
13:     $I_M \leftarrow \text{Matched}$   $\triangleright$  Read the Matched image
14:     $I_R \leftarrow$  Resize image.  $\triangleright$  Such that width=height
15:     $I_G = [I_G, I_R]$ 
16:  end for
17:  Display  $I_G$ 
18:   $\text{Rem} = \text{mod}(i, TIES)$ 
19:  if  $(\text{Rem} == 1)$  then
20:     $N = N + 1$ 
21:     $\text{mkdir}(\text{Output}(N))$   $\triangleright$  Create Directory 'Output(N)' to save each set of
 $i$  images
22:     $S \leftarrow$  Starting index number of  $N^{\text{th}}$  TrainImageSet
23:     $E \leftarrow$  Ending index number of  $N^{\text{th}}$  TrainImageSet
24:    end if
25:    Initialize  $m \leftarrow 1$ 
26:    for  $k = 1$  to  $\text{length}(I_{ID})$  do
27:      if  $(I_{ID}) > S$  & &  $(I_{ID}) < E$  then
28:         $\text{BestID}[j] = I_{ID}[k]$ 
29:         $m = m + 1$ 
30:      end if
31:    end for
32:     $BI_G \leftarrow []$   $\triangleright$  Initialize Best image gallery
33:    for  $l = 1$  to  $\text{length}(\text{BestID})$  do
34:       $\text{BestMatchID} = \text{BestID}[l]$ 
35:       $\text{BestMatch} = \text{ImageLocation}[\text{BestMatchID}]$ 
36:       $I_B \leftarrow \text{BestMatch}$   $\triangleright$  Read the BestMatch image
37:       $I_R \leftarrow$  Resize image.  $\triangleright$  Such that width=height
38:       $BI_G = [BI_G, I_R]$ 
39:    end for
40:    Display  $BI_G$ 
41:    Save  $I$ ,  $I_G$ , and  $BI_G$ 
42:  end for

```

increase in the number of iterations required to reach convergence at the optima in the refining phase compared to K-means algorithm. PSO is inherently parallel and can be implemented using parallel hardware, such as a computer cluster, the computation requirement for clustering large document dataset is still high. The aim of the PSO is to find the particle position that results in the best evaluation of a given fitness (objective) function. Each particle represents a position in N^{th} dimensional space, and moved through the search space, adjusting its position in the particle's best position found so far, and the best position in the neighborhood of that particle.

Algorithm 5.2 Retrieval Optimization with PSO

```

1: Initialize the parameters:
2:  $n \leftarrow$  Feature Vector.
3:  $dim \leftarrow$  Feature Dimension.
4: Load FeatureSet.
5: Current fitness position.
6: Initialize velocity and Position           $\triangleright$  Current position, Velocity, Local best position
7: for each particle do
8:   Calculate current fitness value  $p_i$ 
9:   if  $p_i > p_{best}$  then
10:    Set current value as the new fitness value.
11:   end if
12: end for
13: for each particle do
14:   Find the particle neighborhood           $\triangleright$  particle with best fitness value
15:   Calculate particle velocity:  $v_i$ , using Eq. (1)
16:   Apply the velocity constriction.
17:   Update particle position:  $x_i$ , using Eq. (2)
18:   Apply the position constriction.
19: end for
20: Return current position.                  $\triangleright$  Optimal visual vocabulary size

```

Algorithm 5.2 depicts the process of optimizing the retrieval results using PSO. Each particle i maintains the current position of the particle x_i , the current velocity of the particle v_i and the personal best position of the particle y_i . Using the above notation, a particle's position is adjusted as per Eq. (5.10).

$$v_1(t+1) = wv_1(t) + c_1r_1[p_1best - p_1(t)] + c_2r_2[gbest(t) - p_1(t)] \quad (5.10)$$

where, w is the inertia weight, c_1 and c_2 are the acceleration constants, $r_1(t)$; $r_2(t)$

$U(0, 1)$. Now, the particle position can be updated as per Eq. (5.11).

$$p(t + 1) = p_1(t) + v_1(t + 1) \quad (5.11)$$

During experiments, it was observed that Algorithm 5.2 requires more than 200 iterations to converge to the optimal result for the dataset used, which consists of 31 categories of X-ray scans, represented using SURF feature points. The various experiments conducted and the observations on performance of the proposed approach are presented in Section 5.5.

5.5 Experimental Results and Discussion

The implementation of the proposed approach was carried out on a high-end workstation with Intel processor 3.31 GHz speed and 16 GB of memory using Matlab v.2017. For this experiment, the IRMA ImageCLEFMed 2009 dataset (Tommasi *et al.*, 2009)¹ containing 31 unique categories with an equal distribution of 97 images per category was used. Labeled X-ray images of different organs like hand, chest, abdomen, knee, shoulder, ankle, foot and others were used for classification and retrieval of images. Each class of image has 13 digits specific IRMA code; a sample of which is shown in Fig. 5.11.

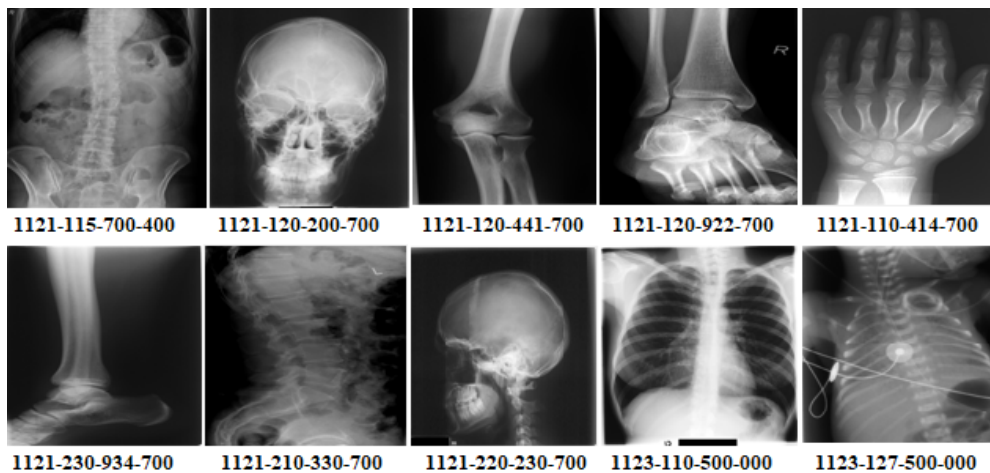


Figure 5.11: Sample images from IRMA dataset.

¹Available online at <https://www.imageclef.org/>

5.5.1 Classification and Retrieval Results

Experimental results revealed that the average classification accuracy on the training set was 96.45%, while the testing accuracy was 87.38% (shown in Table 5.7). The vocabulary size was varied from 31, 50, 75, 100 to 600, with a step size of 25 for each experimental analysis cycle. It can be seen that a vocabulary size of 350 was most optimal and resulted in the best overall accuracy rate. Also, it is noted that vocabulary sizes >200 always attained an accuracy of $>80\%$ (Fig. 5.12). During the iterations, the feature points after 200 iterations do not change much. The same can be noted from the performance curve graph shown in Fig. 5.12. The feature points before PSO and those feature points using PSO after 200 iterations are represented in Fig. 5.13a and 5.13b respectively. Accuracy rate of all 31 classes (See Fig. 5.14) shows that label prediction accuracy was $> 80\%$ for 26 classes, $> 60\%$ & $< 80\%$ for 4 classes and only 1 class showed $< 50\%$. This reveals that a vocabulary size of 200 and above is sufficient for the proposed model.

Table 5.7: Observed performance on Training and Test sets

Metrics	Training Set	Test Set
Accuracy	92.49	89.73
Error	7.51	10.27
Precision	92.66	90.01
Recall	92.49	89.73
Specificity	99.75	99.66
False Positive Rate	0.02	0.34
F1 Score	92.24	89.20

The top 20 retrieved images for some sample query images are shown in Fig. 5.15 without using the filter-based approach. Effective retrieval of the top 10 images for the same sample test images is shown in Fig. 5.16 when a filter-based approach is used. Standard metric Precision@ k was used to evaluate image retrieval performance at various valued of k ($k = 5, 10, 15$ and 20). In the case of real-world medical diagnosis, higher precision values for top-5 retrieval is important, so that the physician will be able to get accurate information w.r.t to the test image. Hence, these values were considered for extensive evaluation. The results of the conducted experiments for some sample image classes are presented in Table 5.8.

The performance of the proposed approach with other state-of-the-art methods is compared and tabulated in Table 5.9. The proposed approach outperformed all models. The performance improvement over Fesharaki and Pourghassem (2012)'s model was more than 7%, while, with reference to Mueen *et al.* (2007) and Tommasi *et al.* (2008)'s work, the proposed model built on BoVW and optimized by PSO, outperformed by a small margin of 1%. The state-of-the-art works used different features, like, shape features (Fesharaki and Pourghassem, 2012), visual features (Mueen *et al.*, 2007), and *local+global* features (Tommasi *et al.*, 2008) for classifying images into its categories. However, these methods are considered only classification as the primary task, and failed to report accuracy achieved by their models. The proposed method also incorporates rotational and scale-invariant features to compensate for the fact that medical scans are taken at different angles, which improves the feature matching strategy. Hence, improvements in terms of classification accuracy were achieved, and more significantly, the classified images are utilized for enhancing the performance during the retrieval phase. This is evident in the evaluation results presented in Table 2, with respect to standard metrics like Precision@ k for top-5, top-10, top-15 and top-20 image retrieval.

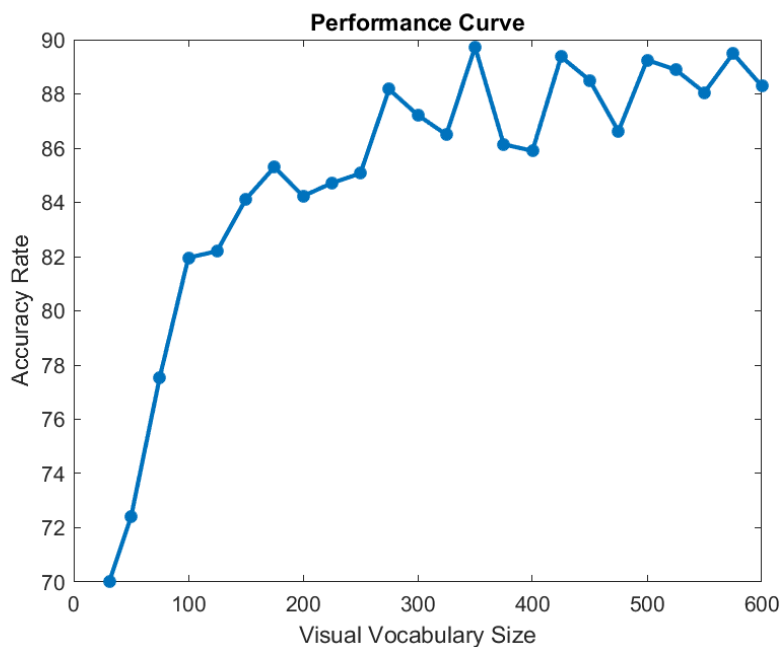
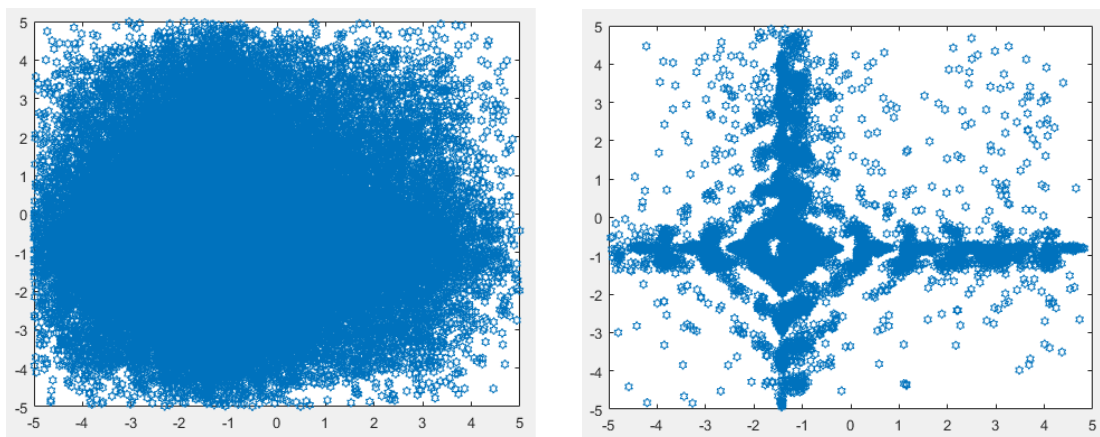


Figure 5.12: Observed performance for varying vocabulary sizes



(a) Feature points before PSO

(b) Feature points after 200 iterations

Figure 5.13: Feature Point Representation using PSO

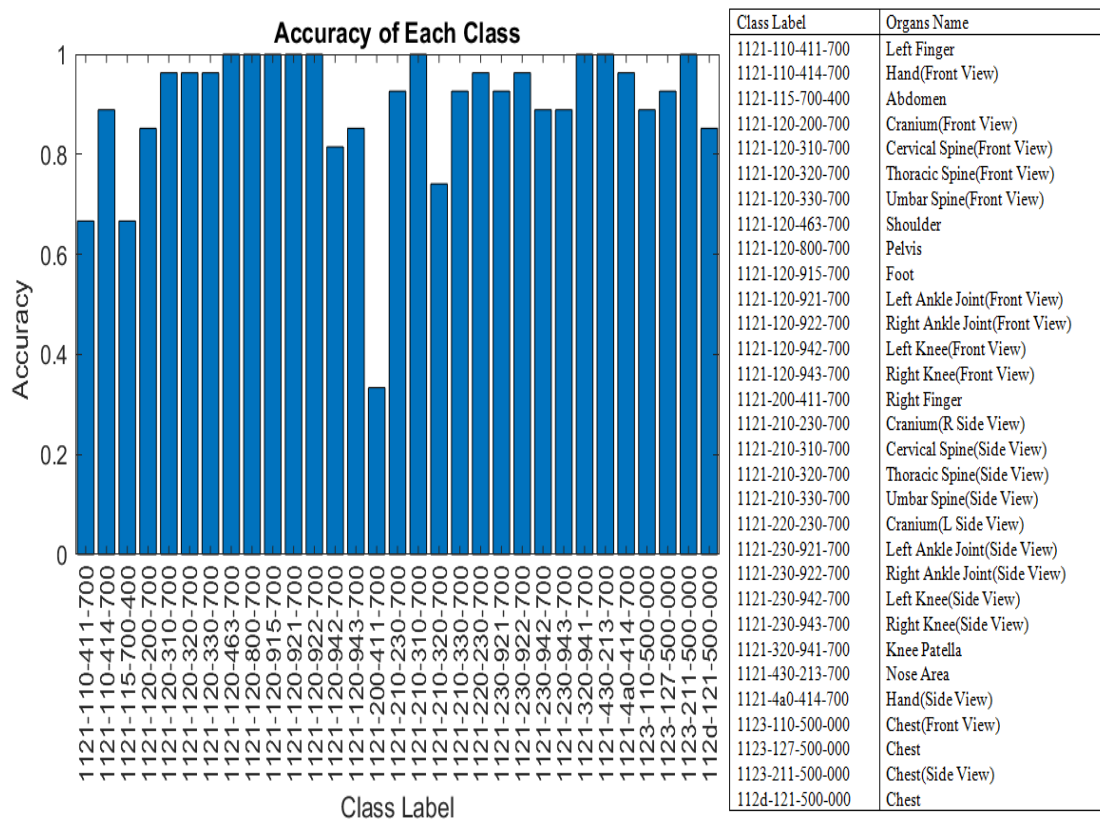


Figure 5.14: Classification Results for different classes

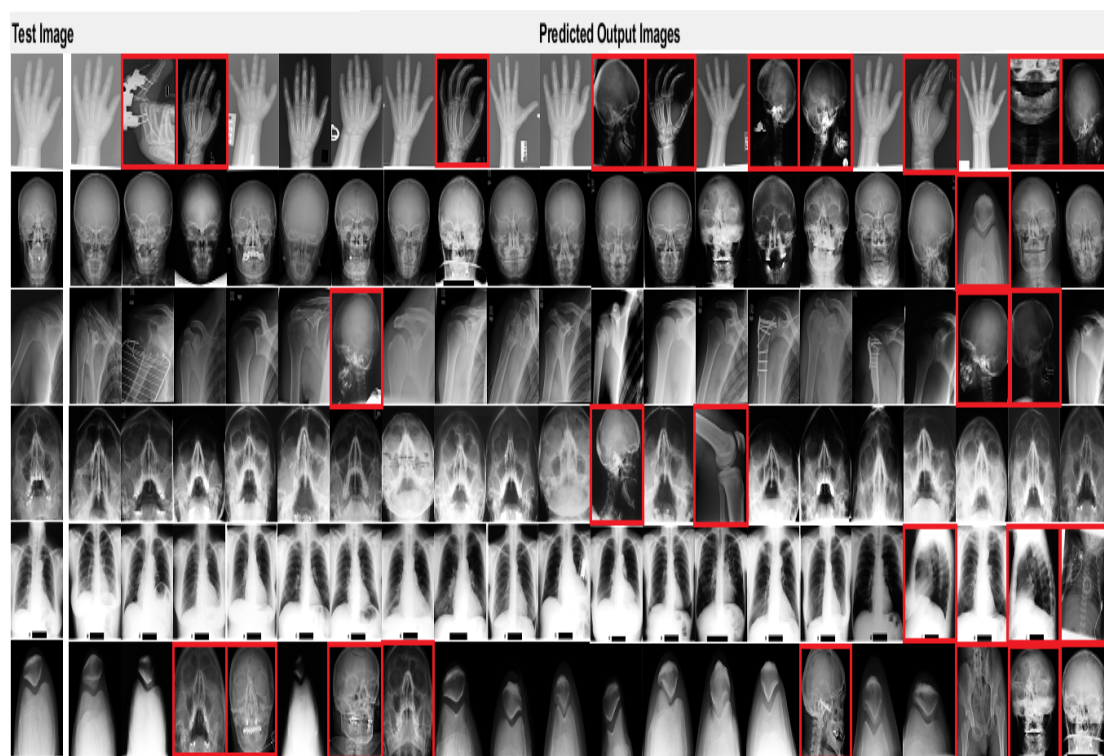


Figure 5.15: Image Retrieval results without using Filter approach and PSO

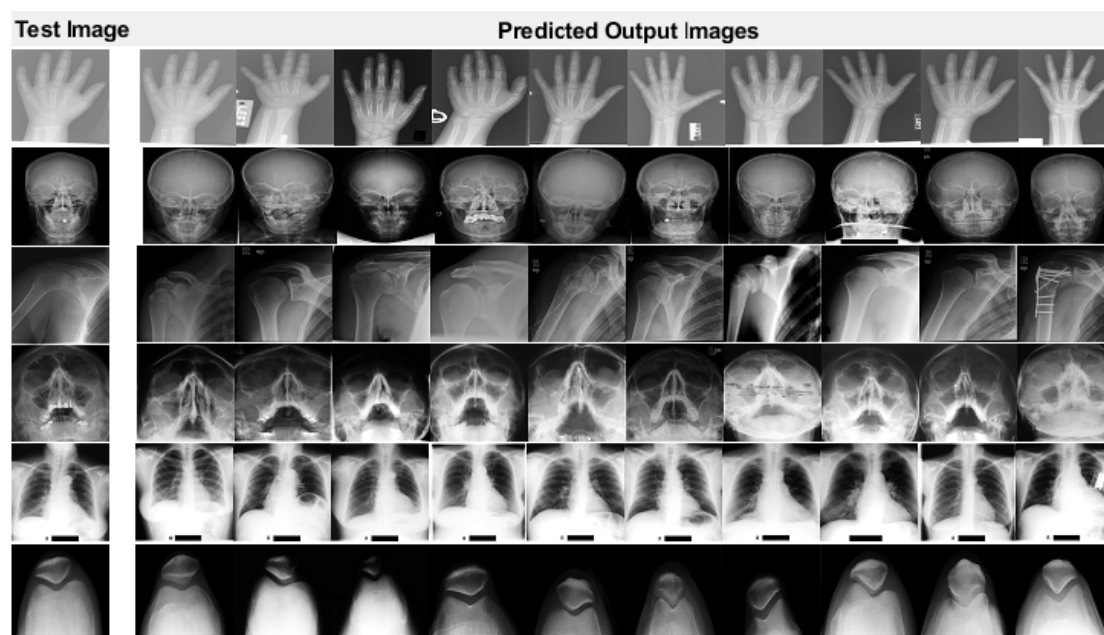


Figure 5.16: Image Retrieval results using Filter approach and PSO

Table 5.8: Evaluation of retrieval with precision@ k for $k = 5, 10, 15, 20$

Test Image	k Value	Relevant images@ k	Precision@ k
Sample 1 (Class-Hand) 1121-110-414-700	k=5	3	60%
	k=10	7	70%
	k=15	8	53.33%
	k=20	11	55%
Sample 2 (Class-Cranium) 1121-120-200-700	k=5	5	100%
	k=10	10	100%
	k=15	15	100%
	k=20	19	95%
Sample 3 (Class-Shoulder) 1121-120-463-700	k=5	5	100%
	k=10	9	90%
	k=15	14	93.33%
	k=20	17	85%
Sample 4 (Class-Nose Area) 1121-430-213-700	k=5	5	100%
	k=10	10	100%
	k=15	13	86.66%
	k=20	18	90%
Sample 5 (Class-Chest) 1123-110-500-000	k=5	5	100%
	k=10	10	100%
	k=15	15	100%
	k=20	17	85%
Sample 6 (Class-Knee Patella) 1121-320-941-700	k=5	3	60%
	k=10	6	60%
	k=15	10	66.66%
	k=20	12	60%

Table 5.9: Benchmarking the proposed approach against State-of-the-art models

Approach	Total Classes	Accuracy
BoVW+PSO (<i>proposed</i>)	31	89.73%
Shape Features, Bayesian Rule (Fesharaki and Pourghassem, 2012)	28	82.87%
Multiple visual features (Mueen <i>et al.</i> , 2007)	57	89%
Local & Global Features (Tommasi <i>et al.</i> , 2008)	-	89.7%

5.6 Deep Neural Models for Effective CBMIR

Designing deep neural models for a particular application is a non-trivial task, mainly due to the number of parameter choices that have to be made when developing such an architecture (Anthimopoulos *et al.*, 2016). Extensive research has been undertaken on assessing the adaptability of deep CNNs for color image classification on large-scale natural-image datasets like ImageNet. However, there is limited research work on texture recognition towards medical image analysis. The proposed approach is based on first classifying medical images using a Convolutional Neural Network (CNN), the results of which are utilized for supporting content-based medical image retrieval. Finally, a fully connected layer and an output layer provides the predictions. In contrast, the layers in CNNs are organized in 3 dimensions: *width*, *height* and *depth*. Additionally, the neurons of one layer do not link or tie up to all the neurons to its preceding layer, with the exception of a small region which is connected, as per the dropout strategy defined. Ultimately, the final output can be minimized to a single vector of probability scores, structured in a systematic way along the depth dimension, which can then be used for other problem-specific tasks. Fig. 5.17 illustrates the CNN architecture designed for the classification task.

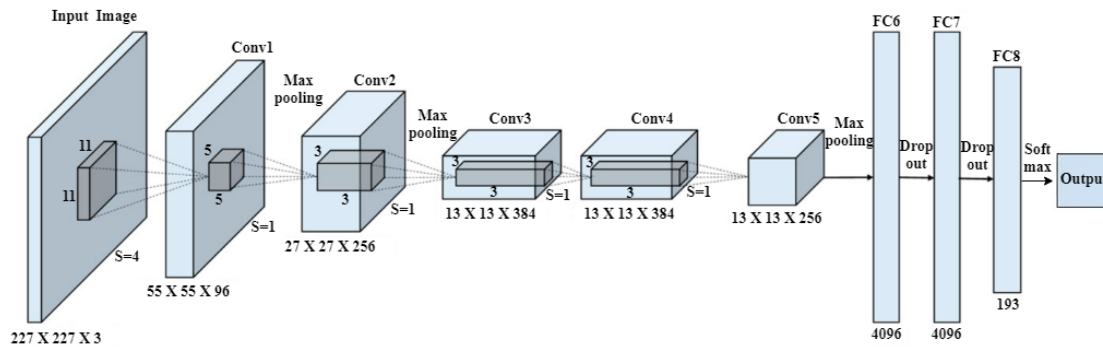


Figure 5.17: Proposed CNN model for Radiograph classification

For experimental evaluation, the ImageCLEF 2009 dataset (Tommasi *et al.*, 2009) was used, which consists of a collection of 14,410 radiographs over 193 classes, collected from the Department of Radiology, Aachen University of Technology, Germany. Due to variation in the X-ray images provided in the dataset, each image was first preprocessed before feeding into the neural network. It was found that the images were of different dimensions, hence it was necessary to normalize the size of these images, to a standard resolution of 227x227. Then, after concatenating each to the required depth, they are stored separately with

the associated class label. These images are fed as input to the input layer of the CNN model. A CNN based approach was designed for efficient retrieval, as it can learn the visual features more efficiently for improved classification. From this classification, an attempt is made to label the orientation view of the specific image under test, for which a novel algorithm was developed (described in Section 6.2, Algorithm 5).

The CNN model designed used for classification and retrieval tasks consists of five convolutional layers followed by 3 fully connected layers. Each convolutional layer is followed by a ReLu activation layer and a max pooling layer, except at Conv3 and Conv4 layers. In between the first and second ReLu and max pooling layers, a normalization layer is added. The last three layers must be fine-tuned as per the requirements of the classification task. Hence, the last three layers of the model were replaced with a fully connected layer, a softmax layer, and a pixel classification output layer. The newly added fully connected layer is configured with the number of classes in the ImageCLEF dataset. Pixel classification layers helps in classifying the images in a better way by ignoring the undefined pixel labels during training. For training the model, a learning rate of 0.001 with a batch size of 50 was chosen, while the number of epochs was set to 10.

The proposed work deals with multi-class classification task, so a modified transfer learning AlexNet approach was incorporated, where Categorical Cross Entropy (CCE) loss function was used. The requirement of CCE is to have each output nodes for each class. Hence, the number of output nodes to 193, same as the total number of classes in the dataset. At the end of the network architecture, the final output layer uses a softmax activation function, so that each node outputs a probability value between (0–1). The classification model performs two functions - feature extraction and classification. In the hidden layers, the process of feature extraction from the X-ray images is performed during a sequence of convolutions and pooling operations. The convolution is performed on the input images with the use of a filter or kernel to produce a feature map that represents each medical image in a secondary feature space. A convolution is executed by sliding the filter over the input images. At each location, matrix multiplication is carried out to sum up the results onto the feature map. While one filter is slid over the entire image, the process of identifying image-specific features is done in this phase, regardless of their discriminating size and position. As the first convolution layer accepts the image itself as the input and later layers take the output activation of the previous layer as their input, the feature extraction process is inherently hierarchical in nature.

After each convolution layer, a max pooling layer is added for continuously reducing the dimensionality of the previously processed X-ray image, *i.e.*, to reduce the number of parameters and also the computation load on the network. This reduces the feature map size, simultaneously keeping the significant information that is captured from the image at the earlier layer. Also, both training time and issues due to over-fitting can be reduced. The max pooling layer accepts two parameters as input: the filter size and the stride, which decide the extent of down-sampling to be performed. For example, a [2x2] pooling size with stride 2 will result in a spatial resolution reduction of 50% of the input size. In this work, a 3x3 max pooling with stride 2 and 0 padding in all max pooling layers were used.

The fully connected layers in the model shown in Fig. 5.17, serve as a classifier on top of the features extracted at previous layers. Each fully connected layer maps the features extracted from convolutions to the corresponding class score outputs. Thus, a value that indicates the classification probability with respect to the class label given as input during the training is assigned to each image at each layer. We used ReLU (defined as shown in Eq. 5.12) as the activation function for the intermediate layers, as our classification task is a multi-class problem.

$$f(z) = \begin{cases} 0 & \text{when } z < 0 \\ z & \text{when } z \geq 0 \end{cases} \quad (5.12)$$

Dropout layers were also incorporated in the model, for reducing the number of units considered during a particular forward or backward pass during training. At each dropout layer, certain units (*i.e.* neurons) are ignored, choosing such neurons at random. Therefore, at each training pass, individual nodes are either dropped out of the network with probability $1 - p$ (or retained with a probability p), so that a reduced network is obtained. This also means that the incoming and outgoing edges of a dropped-out node are also removed, thus resulting in a reduced network architecture. During experimental validation, it was observed that introducing dropout helped the model learn more robust features that helped improve the classification accuracy. In the proposed work, the dropout layer was set with a probability of 0.5 dropouts.

The final layer in the model is a soft-max layer that outputs a probability distribution, *i.e.*, the values of the output sum to 1. The soft-max activation function is basically equivalent to a Logistic Regression over the image features extracted from the layer before the fully connected layer *FC8*. The soft-max layer

was used on an assumption that the classes are mutually exclusive for the multi-class classification problem. During training, L2 regularization was applied by adding a regularization term to the weights of the loss function $E(\theta)$, to reduce problems due to over-fitting. The regularization term performs a weight decay function. For every weight w (here 0.004) in the network, a value equal to $\frac{1}{2}\lambda w^2$ is added, where, λ is the regularization strength.

5.6.1 Content-Based Image Retrieval Task

The quality of extracted features which represent an image dataset in a semantically rich feature space plays a vital role in any image retrieval task. The next objective is to use the features obtained from the CNN along with their class labels for extending CBMIR capabilities. To obtain the required feature sets from the training and test images, an activation layer was added to the fully connected layer *FC7*. The network presents a hierarchical feature representation for each of the input images. The deeper layers of the network provide higher-level features, which are obtained from the previous layers using low-level features. These are now used for creating a feature vector for each image, in training and test sets. When a test image feature set is given as QBE (Query By Example), the pairwise distance measure is used to compute distances that can be used to obtain matching feature sets with the smallest distance from the images in the training set. Using this, the training set feature index numbers are obtained for the given test image. The number of indexes obtained depends on the number of images k that need to be retrieved by the system (here we used $k=10$).

The images that match with these index numbers from the training sets are retrieved and considered for experimental evaluation of the retrieval results using standard metrics. Eight different distance measures – Correlation, Spearman, Cosine, Euclidean, Minkowski, Cityblock, Standard Euclidean and Chebychev, were used for the experiments. These measures are usually used to find the similarity or dissimilarity between two data objects. For each observation in Y (*Test image features*), the pairwise distance method finds the smallest distances by computing and comparing the distance values to all the observations in X (*Training image features*).

Distance correlation or distance co-variance is a measure of dependence between two paired random vectors of arbitrary, not necessarily equal dimension and is measured as per Eq.(5.13). Spearman distance is a square of Euclidean distance between two data points X and Y (Eq. 5.14). Cosine similarity (Eq.

5.15) computes the similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. The Euclidean distance between two data points X and Y are given by computing the sum of squares of the differences between the corresponding values as per Eq.(5.16), where, μ_i and μ_j are the means of i and j respectively.

$$d(i, j) = \frac{1 - (i - \mu_i) \cdot (j - \mu_j)}{\|i - \mu_i\| \|j - \mu_j\|} \quad (5.13)$$

$$d(i, j) = \sum_{k=1}^n [X_{ik} - Y_{jk}]^2 \quad (5.14)$$

$$d(i, j) = \frac{1 - i \cdot j}{\|i\| \|j\|} \quad (5.15)$$

$$d(i, j) = \sqrt{\sum_{k=1}^n [X_{ik} - Y_{jk}]^2} \quad (5.16)$$

Minkowski distance (Eq. 5.17) is a generalization of Euclidean distance, hence from the classification results, it can be seen that both Euclidean and Minkowski give the same results. City block distance between two points X and Y, with k dimensions, is calculated as per Eq. (5.18). In most cases, this distance measure gives a similar result to that of the Euclidean distance. Chebychev distance is a metric defined on a vector space, such that the distance between two vectors is the maximum difference along any coordinate dimension and is given by Eq. (5.19), where, q is a positive integer.

$$d(i, j) = \sum_{k=1}^n (|X_{ik} - Y_{jk}|)^{\frac{1}{q}} \quad (5.17)$$

$$d(i, j) = \sum_{k=1}^n [X_{ik} - Y_{jk}] \quad (5.18)$$

$$d_{ij} = \max_k |X_{ik} - Y_{jk}| \quad (5.19)$$

Often, standardization is necessary to balance the contributions of individual feature sets, one way to do this is to transform variable values such that they all have the same variance of 1. At the same time, the variables are centered at their means. This centering is not necessary for calculating distance, but this sets all variables variances to mean zero and thus easier to compare. The transformation

is commonly called Standard Euclidean and is calculated as follows:

$$r(X, Y) = \frac{X_i Y_i - \mu_X \mu_Y}{\sigma_X \sigma_Y} \quad (5.20)$$

where, μ_X and μ_Y are the means of X and Y respectively, and σ_X and σ_Y are the standard deviations of X and Y.

To measure the similarity between images in the training dataset distance measures were used to measure the similarity between images in the training dataset and those in the test dataset, so that class-labels can be predicted for unseen images, and similar images can be retrieved during the retrieval phase. Various experiments were conducted to validate the performance of the proposed approach with respect to both the classification task and retrieval task, using standard metrics. In addition to this, the proposed approach was benchmarked against state-of-the-art works, the details of which are presented in Section 5.7.

5.7 Experimental Results and Discussion

The proposed methodology was implemented on a high-end workstation and server equipped with NVIDIA P40 GPUs using Matlab v.2017. For the experimental validation, the ImageCLEFMed 2009² (Tommasi *et al.*, 2009) dataset was considered, containing 14,410 X-ray images belonging to 193 classes. Among these images, 12,677 images are to be used for training and 1,733 images are for testing, as provided in the dataset. As per the IRMA code, the images in the ImageCLEF dataset were classified and grouped into their class labels. Fig. 5.18a shows 10 sample images from the IRMA dataset with their related IRMA code (shown in Fig. 5.18b).

5.7.1 Classification Task

Once the network is trained, the testing images were classified using the learned features from the model. A classifier based on the pairwise distance calculations between two sets - testing and training observations is carried out. The predicted labels are compared with the actual label of the test image. Experiments with 8 different distance measures (discussed in Section 5.6.1) were performed, to compare the classification with its retrieval performance of the proposed CNN based model. Classification performance was evaluated using metrics like accuracy,

²<https://www.imageclef.org/>

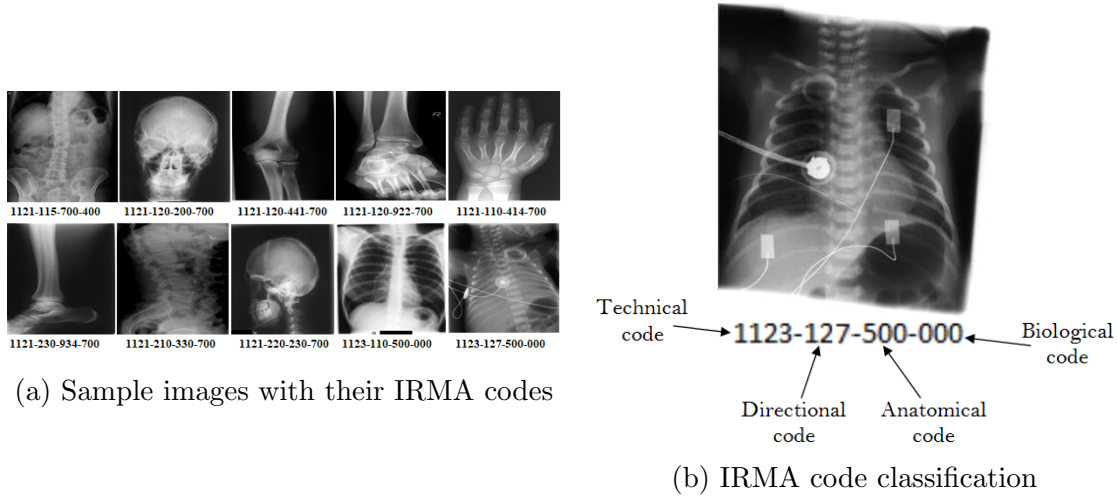


Figure 5.18: IRMA dataset - Image sample and specifics (Lehmann *et al.*, 2003c)

precision, recall and f-score. Classification accuracy is computed as the ratio of a number of correctly predicted samples to the total number of input samples, as given by Eq. (5.21).

$$Accuracy = \frac{\text{Number of correctly predicted class samples}}{\text{Number of input samples}} \quad (5.21)$$

Precision is the ratio of the number of correctly classified results to the total correctly and incorrectly classified predictions by the classifier (Eq. 5.22). The recall is defined as the number of correct positive results divided by the number of all relevant samples, given by Eq. (5.23). False Positive Rate (FP Rate) corresponds to the proportion of negative image samples that are mistakenly considered as positive, with respect to all negative image sample (as per Eq. 5.24). F-score is a balanced measure of how precise a classifier is (how many instances it classifies correctly), as well as how robust it is (not missing a significant number of instances), as per Eq. (5.25).

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (5.22)$$

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (5.23)$$

$$FPRate = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}} \quad (5.24)$$

$$F - score = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.25)$$

Experiments were conducted to observe the classification performance when different distance measures are used, and the results are tabulated in Table 5.10. It can be seen that the proposed CNN features when classified with neural network classifier, achieved better classification accuracy when compared to different standard distance measures. Experiments revealed that the proposed CNN model outperformed the traditional distance measures by a margin of 1 to 5%. This may be due to the introduction of the pixel classification layer and the additional standardization measures applied to the data during the computation phase. As the dataset has inherent data imbalance, the classification accuracy also was found to vary accordingly.

Table 5.10: Classification performance of proposed CNN based model with reference to various distance measures.

Technique	Accuracy	Precision	Recall	FPR	F-score
Deep CNN (<i>Proposed</i>)	0.6456	0.5184	0.5437	0.0021	0.4904
Correlation	0.6408	0.5144	0.5434	0.0022	0.4900
Spearman	0.6375	0.5140	0.5442	0.0022	0.4901
Cosine	0.6318	0.5183	0.5326	0.0022	0.4833
Euclidean	0.6191	0.4838	0.5175	0.0023	0.4704
Minkowski	0.6191	0.4838	0.5175	0.0023	0.4704
Cityblock	0.6151	0.4905	0.5188	0.0023	0.4717
Standard Euclidean	0.6138	0.4737	0.5131	0.0023	0.4643
Chebyshev	0.6006	0.4952	0.5068	0.0024	0.4661

In addition to this, experiments were also performed for benchmarking the base AlexNet architecture against the modified architecture adapted in this work. Here, the base AlexNet model with the default values for the various hyperparameters was applied to the dataset consisting of all 193 categories. In comparison to the performance achieved by the proposed modified CNN model, the base model failed to perform well, as is evident from the much lower classification accuracy of 57.37%. The observed performance of the base AlexNet model with neural network classifier, in comparison to the traditional distance measures is shown in Table 5.11. Since feature values are different in both cases, a significant difference was observed in classification accuracy of both models. The objective of this comparison was to highlight the feature modeling performance of the proposed

CNN model when compared to the base AlexNet model.

Table 5.11: Classification results of Base AlexNet Model in comparison to various distance measures.

Technique	Accuracy	Precision	Recall	FPR	F-score
Base AlexNet	0.5737	0.3943	0.4114	0.0022	0.3759
Correlation	0.5707	0.3913	0.4101	0.0022	0.3729
Spearman	0.5684	0.3868	0.4055	0.0023	0.3686
Cosine	0.5667	0.3859	0.4014	0.0023	0.3680
Euclidean	0.5642	0.3842	0.4004	0.0023	0.3664
Minkowski	0.5532	0.3831	0.3984	0.0024	0.3644
Cityblock	0.5419	0.3814	0.3913	0.0024	0.3585
Std. Eucl.	0.5331	0.3808	0.3868	0.0024	0.3522
Chebychev	0.5290	0.3803	0.3757	0.0025	0.3475

Experimental evaluation showed that the proposed model achieved an accuracy rate of $>80\%$ for about 70 classes, $< 80\%$ and $> 40\%$ for 46 classes and the remaining 52 classes showed $< 40\%$ accuracy. The performance was also evaluated using the F-score metric, which is considered a more balanced measure when there is an uneven class distribution, as a harmonic mean of Precision and Recall. The proposed distance measure achieved a f-score value of 0.4904, which shows that there is a sufficient weight distribution among the classes. Further investigations revealed that, 52 classes in the dataset had less than 10 images available for training. The classification performance for a sample of classes having an accuracy rate of $>80\%$ is shown in Fig. 5.19.

5.7.2 Retrieval Task

To evaluate the retrieval performance, two popular IR metrics, *precision@k* (p at k) and *Mean Average Precision (MAP)* were employed. Precision@ k is given by the ratio of images that are retrieved in the top k set, which are actually relevant (computed as per Eq. 5.26). MAP for a set of testing images is defined as the mean of the average precision scores for each query image. MAP@ k can be also computed accordingly, by considering precision scores at k , as per Eq. (5.27), where, Q is the number of query/test images and $q = 1, 2, 3, \dots, Q$.

$$p \text{ at } k = \frac{\text{No. of retrieved images at } k \text{ that are relevant}}{\text{Total images retrieved at } k} \quad (5.26)$$

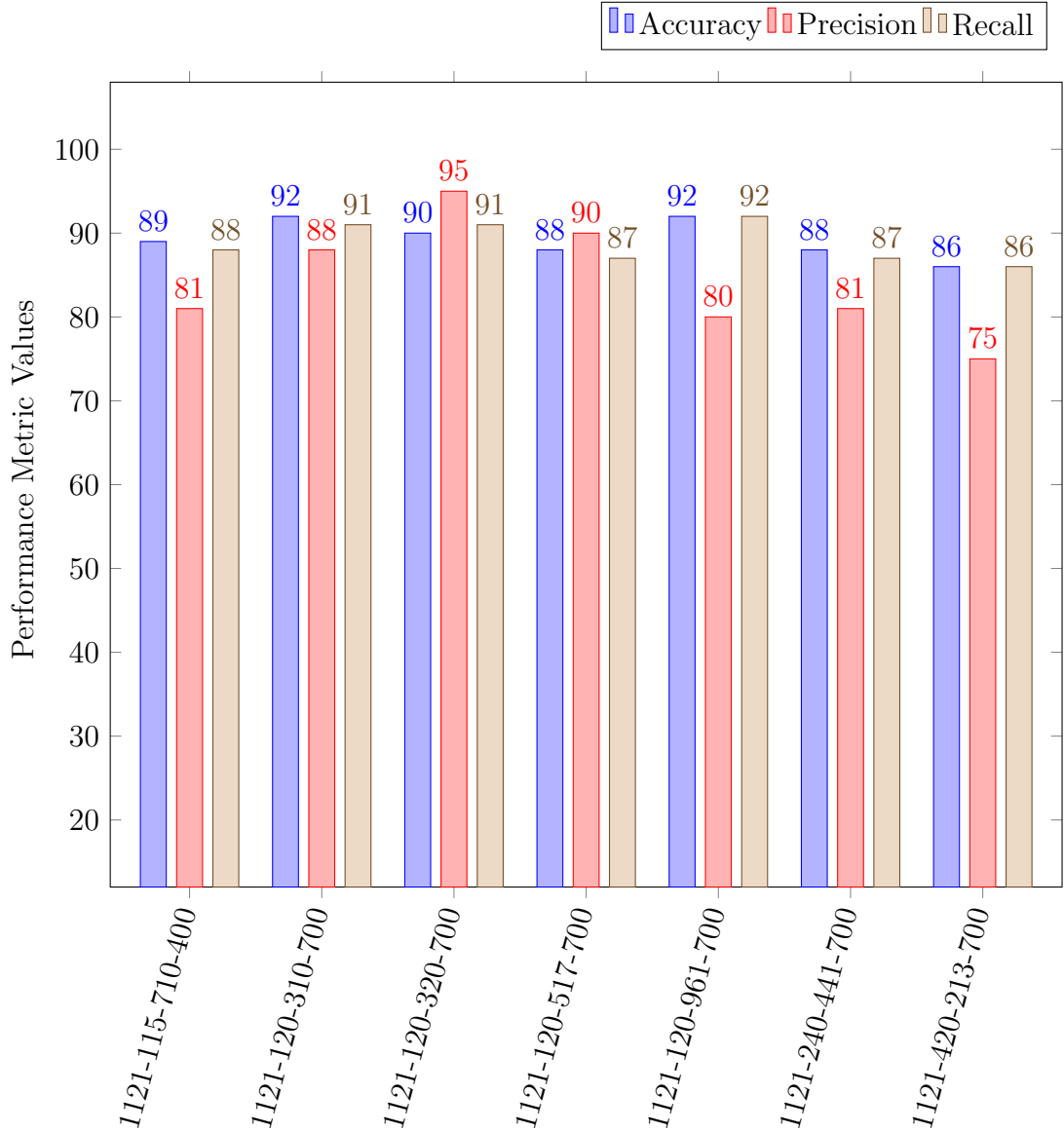


Figure 5.19: Classification performance of proposed CNN model for various classes

$$MAP = \frac{\sum_{q=1}^Q AvgP(q)}{Q} \quad (5.27)$$

For evaluating the retrieval performance, all 1,733 test images in the dataset were considered. It was observed that, for some classes the retrieval results were very good, while for some others, the test images did not give the best match. The reason for this is the significant class imbalance in the dataset, which makes feature learning difficult. The results are tabulated in Table 5.12, from which, it can be observed that the top- k retrieval results *i.e.*, Precision@ k for $k = 3, 5, 10$ for the

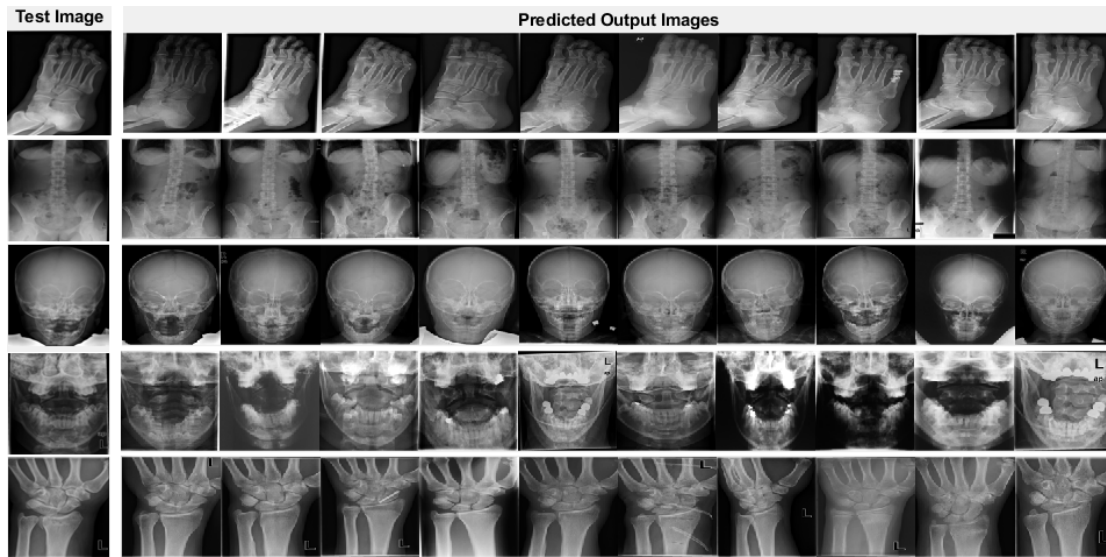


Figure 5.20: Observed retrieval results for Best-match classes

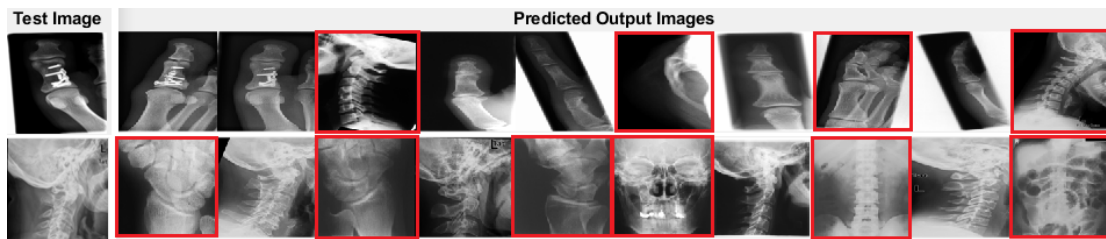


Figure 5.21: Observed retrieval results for Average-match classes (*red marked images are wrongly retrieved images*)

best match (shown in Fig. 5.20) was found to be 100% for most of the test images in best-match classes like *foot*, *lumbar spine*, *skull*, *jaw* and *wrist*. A mean average precision (MAP@10) of 63.34 is achieved for the retrieval task. However, in the case of average-match classes (shown in Fig. 5.21), it can be seen the retrieval performance suffers significantly (Table 5.12) due to an insufficient number of images in those classes, thus, precision@ k and MAP values also deteriorate. The results of the retrieval experiments are shown in Fig. 5.20 and Fig. 5.21 for the test images belonging to best-match classes and average-match classes respectively.

5.7.3 Benchmarking against State-of-the-art Works

To benchmark the proposed work, experiments were performed for assessing its performance against other state-of-the-art works (Liu *et al.*, 2016; Tommasi *et al.*, 2009). A metric called Error score (Tizhoosh, 2015) was used to benchmark the proposed model's performance against that of other contemporary works. The error score is used to calculate the total error (as per Eq. 5.28) in image retrieval

Table 5.12: Evaluation of retrieval with precision@ k for $k=3, 5, 10$. (Average-match classes)

Test Class	k value	Relevant images	Precision@ k
Sample 1 (Class - Finger) 1121-110-411-700	k=3	2	66.66%
	k=5	4	80%
	k=10	6	60%
Sample 2 (Class - Cervical Spine) 1121-210-313-700	k=3	1	33.33%
	k=5	2	40%
	k=10	4	40%

over 1,733 test images, each characterized by their IRMA codes with n_d digits, where, $n_d \in \{3,4\}$, and B_j^{ik} is the number of possible labels at position i , and δ is a decision measure which is 1 for an incorrect label and 0 for a correct one, when image l_i is compared with the image \hat{l}_i . The retrieval accuracy can also be obtained from the calculated IRMA error (Khatami *et al.*, 2018b), which is computed as per Eq. (5.29). In this case, the total no. of test images in the dataset considered is 1,733.

$$E_{total}(l^{query}) = \sum_{i=1}^{1733} \sum_{k=1}^4 \sum_{j=1}^{n_d} \frac{1}{B_j^{ik}} \frac{1}{j} \delta(l_i, \hat{l}_i) \quad (5.28)$$

$$Accuracy = 1 - \frac{Error}{Total\ number\ of\ test\ images} \quad (5.29)$$

The performance of the proposed model in comparison to several other works in terms of computed Error score is reported in Table 5.13. It can be seen that the proposed model outperformed the state-of-the-art, (Camlica *et al.*, 2015)'s model, by achieving a significantly lower IRMA error score, with a reduction of over 14 points, at 132.45. More significantly, the proposed approach achieved the lowest error score value in comparison to 14 other state-of-the-art models (as listed in Table 5.13), thus, proving the effectiveness of the modified CNN based model over others, in terms of lowest error score on the IRMA ImageCLEF Med. dataset.

Table 5.13: Benchmarking the proposed model against state-of-the-art approaches, using Error Score and Retrieval accuracy metrics.

Approach	Error	Accuracy
Deep CNN (proposed)	132.45	92.35%
Camlica <i>et al.</i> (2015)	146.55	91.54%
Khatami <i>et al.</i> (2018a)	165.5	90.45%
Shrinking search space with LBP (Khatami <i>et al.</i> , 2018b)	168.05	90.30%
TAUbiomed (Avni <i>et al.</i> , 2009)	169.50	90.22%
Idiap*	178.93	89.68%
CNNC (128x128, binary)+RBC**	224.13	87.06%
CNNC (96x96, binary)+RBC**	237.93	86.27%
FEITLJS*	242.46	86.00%
VPA SabanciUniv*	261.16	84.93%
CNNC (96x96, no binarization) + RBC**	270.12	84.41%
Randon Barcodes via SVM #	294.83	82.98%
SP-R*	311.8	82.01%
MedGIFT*	317.53	81.68%
IRMA*	359.29	79.27%

Note: Approaches marked with * were reported in Tommasi *et al.* (2009). Approaches indicated by ** were reported in Liu *et al.* (2016), while approach marked with # is reported in Zhu and Tizhoosh (2016).

5.8 Summary

In this chapter, three different works undertaken as part of the defined objective were presented. The first is a hybrid-feature modeling approach for content-based medical image retrieval. The experimental results showed that the proposed approach was very suitable for real-world medical image retrieval applications used for disease diagnosis and decision support, due to its excellent top-3 and top-5 retrieval performance. Second approach is a PSO enhanced Content-Based Image Retrieval approach built on the Bag of Visual Words Model. PSO was incorporated to optimize retrieval performance for a given query image, and was used to gain insights into the optimal clustering value. Further, a filtering approach was designed to obtain the best matches in the retrieval task. The third work is a deep CNN model designed for classification of medical images, the results of which are

used for supporting similar image retrieval. By using CNN's feature extraction and with similarity distance calculation between the feature vectors, it was observed that the model achieved good retrieval results. When benchmarked against several state-of-the-art CBMIR approaches, the proposed model outperformed the others with the lowest error score and highest retrieval accuracy.

Publications

(based on work presented in this chapter)

1. Karthik, K. and Sowmya Kamath, S, “*A Hybrid Feature Modeling Approach for Content-Based Medical Image Retrieval*”, In 13th International Conference on Industrial and Information Systems (ICIIS). IEEE, IIT Ropar, Punjab (CORE Ranked) *(Status: Online)*
2. Karthik K., Sowmya Kamath S., “*A Deep Neural Network Model for Content-Based Medical Image Retrieval with Multi-View Classification*”, The Visual Computer Journal (TVCJ), Springer Nature, DOI: 10.1007/s00371-020-01941-2 [SCIE & Scopus, IF: 2.601] *(Status: Online)*
3. Karthik K., Sowmya Kamath S., “*Swarm Optimization Based Bag of Visual Words Model for Content-Based X-Ray Scan Retrieval*”, Intl. Journal of Biomedical Engineering and Technology (IJBET), Inderscience. [ESCI & Scopus] *(Status: Abstract online, Article in press)*

PART IV

Dealing with Variance in Body Orientation Views

Chapter 6

Medical Image View Classification

6.1 Introduction

Radiological procedures like X-rays have evolved as a crucial diagnostic imaging tool for identifying abnormalities in different body parts. Typically, medical personnel require insights derived from various views/body orientations of the patient, in order to assess the disease physiology completely. In these situations, *frontal*, *lateral* and *sagittal views* are commonly used for overall assessment. For computer-aided diagnosis (CAD), internal and external shapes are very important in identifying the abnormality. Currently, the projection view/ image orientation of radiographs are labeled manually by radiologists and technicians. Manual corrections for wrongly labelled views makes it impractical in PACS and digital imaging systems, as it involves cost and time of human resources. Instead of manually labeling such multi-oriented images, it can be accomplished automatically by intelligent algorithms that are trained to understand the patterns with large-scale images. Methods that can assess this automatically and provide the necessary information regarding the view of the organ at which the scan is taken can be beneficial. Therefore, a classifier model developed for categorizing the disease according to the image view is of great importance. Further, this helps in providing a proper description of the image in a overall clinical workflow management system.

The objective of this work is to design effective models for view classification with reference to the different orientations in which the patients' diagnostic scans are performed, during the initial process of radiographic profiling. The impact of such automated techniques in large-scale medical image management systems. Towards this objective, the suitability of different deep neural architectures are explored and designed models are benchmarked against existing works using standard evaluation metrics. The proposed approach is intended for use in real-time

applications to enhance preprocessing and facilitate its use in building intelligent applications like Clinical Decision Support Systems (CDSS) and predictive analytics systems.

6.1.1 Problem Definition

In diagnostic medical images, the patient body orientation or view of the scanning posture is often not recorded explicitly during storage in digital archival systems like PACS. Different orientations like anterior or frontal view, posterior or back view and the lateral or side views (also known as left lateral or right lateral) can be used during scanning. However, computer-aided diagnosis systems do not record this additional header information of the image. Automated orientation identification for images is required for quality and quantitative analysis of the image in many diagnostic applications. If such patient body orientations are not recorded or are documented using an incorrect label, automated system indexing may be inconsistent, and may also result in improper interpretation by computers and radiologists. Thus, the problem to be addressed here is defined as follows:

“Given a set of medical images consisting of multiple views based on patient posture, design effective models for enabling automated view classification based on the image orientation, for intelligent clinical tasks.”

In this chapter, the research undertaken for addressing this problem is presented. The focus is on modeling the image variances by giving attention to the body view positioning, using the designed multi-view classification algorithms. The algorithm contributes positively towards effective classification labeling. Next, an automatic system that recognizes the orientation view label of different parts of the body soon after the scan is taken is presented, for aiding efficient indexing, categorization and storage.

6.2 Body Orientation: Multi-View Classification

A detailed examination of the ImageCLEF IRMA dataset revealed that several classes contained images with variance in body orientation. It was observed that classes like *Neuro Cranium* (IRMA code: 1121-220-230-700, 1121-210-230-700, 1121-4b0-233-700) and *Cervical Spine* (IRMA code: 1121-210-310-700, 1121-120-310-700, 1122-220-310-700) had three different views, where as, classes like *Chest*,

Hand and *Leg* had two different views. This orientation view is valuable during Computer-Aided Diagnosis (CAD) based medical systems and could be potentially useful for categorization. To address this problem, a table body orientation labeling and identification algorithm is designed to improve the performance of view-specific medical image classification requirements. The models designed for identifying and classifying all three types of views – Left lateral view, Right lateral view and Frontal view are discussed in detail. Algorithm 6.1 illustrates the body orientation view classification process.

Algorithm 6.1 Body Orientation Classification Process

```

1: Initialize image gallery  $I_G \leftarrow []$ 
2: for  $i = 1$  to  $\text{length}(\text{TestImage})$  do
3:    $\text{Img} \leftarrow \text{Read the test image}$ 
4:    $\text{LowerRow} = \text{round}(\text{size}(\text{Img}, 1) * 0.75)$      $\triangleright$  To get the lower part of the image
5:    $\text{subImage} = \text{Image}(\text{LowerRow}:\text{end}, :)$ 
6:    $\text{mask} = \text{true}(\text{size}(\text{subImage}))$      $\triangleright$  Mask the lower part of the image, to get its centroid
7:    $\text{props} = \text{regionprops}(\text{mask}, \text{subImage}, \text{'WeightedCentroid'})$ 
8:    $\text{xCentroid} = \text{props}.\text{WeightedCentroid}(1)$      $\triangleright$  Get x centroid
9:    $\text{columns} = \text{size}(\text{subImage}, 2)$ 
10:  if  $(\text{xCentroid} < 0.42 * \text{columns})$  then
11:     $\text{label} = \text{'Facing Left'}$ 
12:  else
13:    if  $(\text{xCentroid} > 0.46 * \text{columns})$  then
14:       $\text{label} = \text{'Facing Right'}$ 
15:    else
16:       $\text{label} = \text{'Facing forward'}$ 
17:    end if
18:  end if
19:   $\text{img} = \text{imresize}(\text{Img}, [128\ 128])$      $\triangleright$  Resize image, such that width = height
20:   $\text{position} = [1\ 100]$      $\triangleright$  Set the position where label needs to get displayed
21:   $\text{value} = \text{label}$ 
22:   $\text{IMG} = \text{insertText}(\text{img}, \text{position}, \text{value})$      $\triangleright$  Insert the label on the image.
23:   $I_G = [I_G, \text{IMG}]$      $\triangleright$  Update images
24: end for
25: Return processed images

```

The process starts with the extraction of the lower portion of the image, after which the location of the weighted centroid of each image is calculated. If the weighted centroid of an image is way off to the left, its orientation is left facing. If it is to the right of the midline, then the image is right facing. If it is regionally close to the middle, then we can determine that the image is a frontal view. The

position of the centroid was empirically determined to be close to 0.42 and 0.46 values, after several trials with a set of sample images to see where the centroid lies. We tested this procedure on the classes of images that have all the three different views, which gave good results. Once the model classifies the images to its corresponding class label, further classifying the image with reference to the body orientation view can help in improving hospital CAD systems.

6.2.1 Experimental Results and Discussion

To evaluate the classification performance when medical images have associated body orientation views, various experiments were conducted. The observed results for a given orientation view with reference to its directional view of some sample images from the *Cervical-Spine* and *Neuro-Cranium* classes are shown in Fig. 6.1 and 6.2. It can be seen that the proposed body orientation classification algorithm achieved good orientation label prediction for three different type of views - facing forward, facing left and facing right. The orientation label highlighted in red color indicates a wrong orientation view label of the corresponding image.

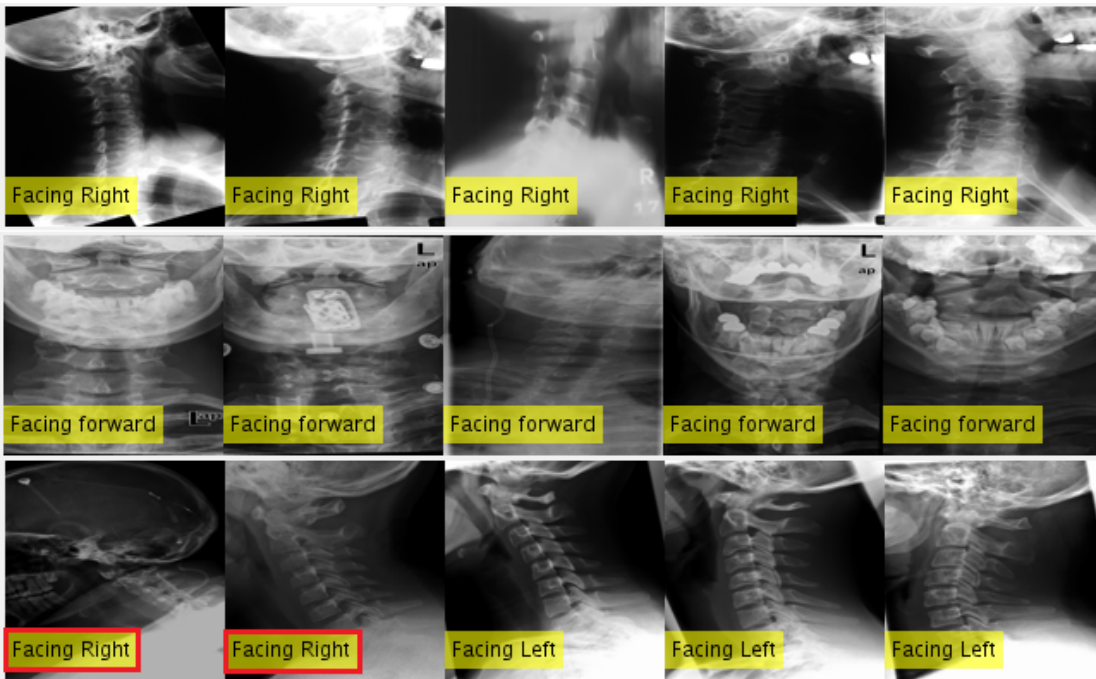


Figure 6.1: Predicted orientation label on images of *Cervical Spine* class

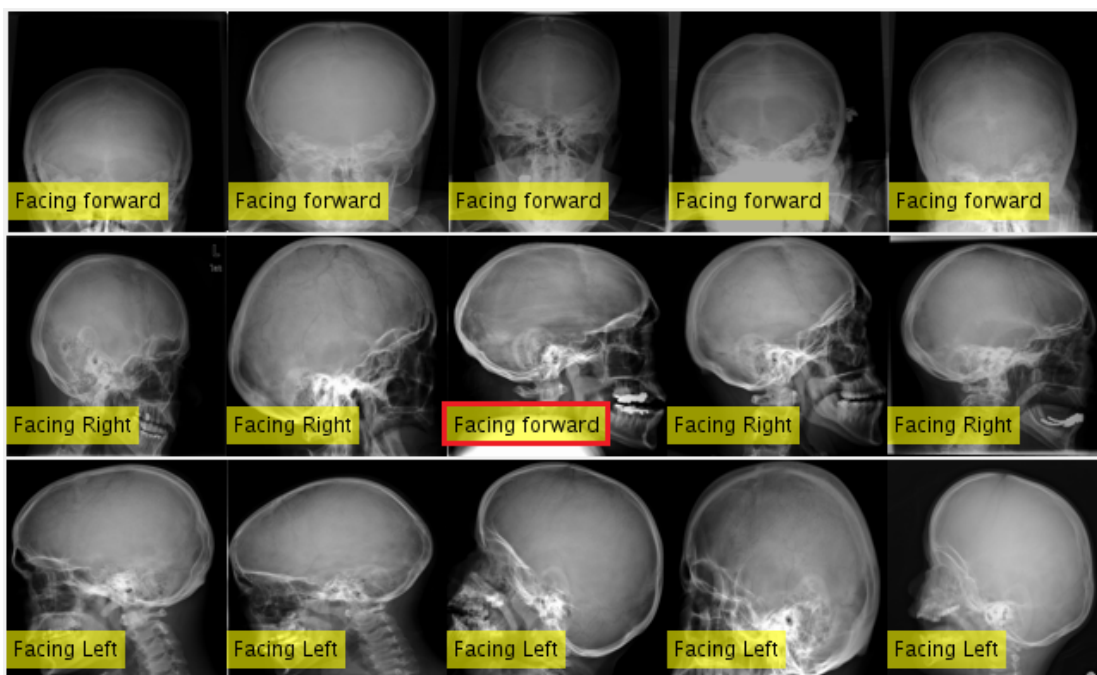


Figure 6.2: Predicted orientation label on images of *Neuro Cranium* class

6.3 Deep Neural Models for X-ray View Orientation Classification

The sequence of tasks involved in the overall progress of the proposed view orientation classification model using pre-trained network is shown in Fig. 6.3. Neural-network architectures were adopted as an alternative to traditional supervised learning based systems, due their adaptive learning behavior and semi-supervised nature. Four different convolutional neural networks were used as Transfer Learning approaches for optimizing the training process. Apart from this, a novel neural architecture called ViewNet was also designed for the task of medical image view classification. The objective is to use neural models that are capable of learning rich latent feature representations for a large number of image views, by incorporating transfer learning approaches for identification of the orientation label for different body part images. Transfer learning techniques are generally used in most deep learning applications. The advantage of using a pre-trained network is to learn a modern task, making it easier and quicker than learning the features from scratch. Another benefit here is that, the model learns better with even lesser number of training images, when trained for newer tasks with the transfer learning approaches.

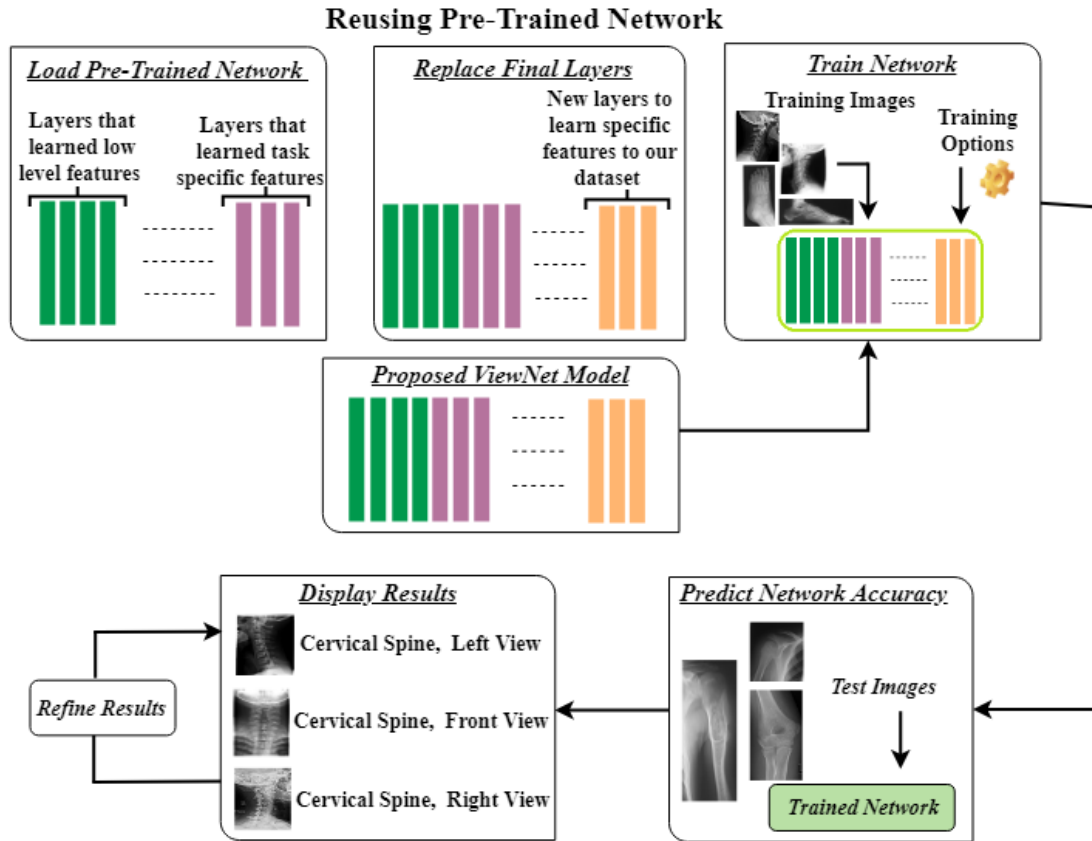


Figure 6.3: Proposed Approach for View/Body Orientation Classification

For the orientation identification task, a subset of orientation identified classes from ImageCLEF dataset are used (Lehmann *et al.*, 2003c). A total of 41 different classes consisting of different organs of the body like spine (cervical spine, Thoracic spine, Umbar spine), leg, hand, carpal bones, nose and eye area were used for the experiments. Each class had its own IRMA code, which are considered for orientation classification task (described in detail in Section 6.4). Images of size $m \times n$ are fed into the neural network for training. Each dataset image had different image dimensions, therefore resizing was performed as soon as the images were read from the datastore i.e, before feeding it into the network. Augmentation was also performed on the training dataset through operations like random flip and translation along the vertical and horizontal axis, which prevents the network from overfitting. While training different deep neural network architectures using a transfer learning approach, some of the final layers should be changed to a fully connected, softmax, and a classification output layers by observing the dataset used. The newly replaced fully connected layer parameters need to be specified according to the new dataset for a new classification model. Increasing

the *WeightLearnRateFactor* and *BiasLearnRateFactor* values helps the network learn features faster, with the addition of the new layers. Also, while training the model, several hyperparameter values like mini-batch, epochs, batch normalization, learning rate, regularizations, optimizers and activators were varied and finally the best suitable values were chosen in building the final model for the orientation identification task.

As part of early experiments, four different neural models were applied to the dataset for understanding the suitability of state-of-the-art deep neural models for the view orientation classification task. These four models are AlexNet, ResNet-18, GoogleNet and SqueezeNet. The specifics of these models are discussed here.

1. AlexNet (Krizhevsky *et al.*, 2012) comprises of 25 layers, with the first layer as the image input layer having 227×227 dimension, followed by the first convolution layer with window shape 11×11 . Since most medical scan images have larger dimensions than natural photographic images, more pixels are required to copy the data present in the image. Consequently, a larger convolution window is used in the next layer for handling the input data. Later, the convolution window size is reduced to 5×5 and 3×3 respectively. Next to the convolutional layer, ReLU activation and a max-pooling layer are placed excluding at Conv3 and Conv4. A max-pooling and a normalization layer reside in between the Conv1 and Conv2. Two fully-connected layers are present after the last convolutional layer that produces feature maps with size 4096. Finally, a softmax followed by a classification layer is used for the prediction. Compared to other CNNs, the main difference is that AlexNet comprises of more convolution channels.
2. ResNet-18 (He *et al.*, 2016) is 71 layers deep with 224×224 dimension for image input, followed by the first convolution layer with window shape 7×7 , and a max-pooling layer. A total of 20 such convolution layers are present in this network with different batches. The convolution window shape in the first batch layer is of 1×1 , whereas in the second batch layer is of 3×3 . Between each convolution layer, batch normalization and ReLU activation function are placed, except at the first and second convolution layer which has an additional max-pooling layer. Only at the last convolution layer pooling layer changes as an average pooling layer. The network architecture concludes by a fully-connected, softmax and a classification output layer.
3. GoogleNet (Szegedy *et al.*, 2015) is 144 layers deep with 224×224 dimension for image input. Convolution layers of conv1 and conv3 have ReLU

activation function, max-pooling and cross channel normalization layers in batches. A total of 57 such convolution layers are present in this network with different batches. Every convolution layer has a ReLU activation function, the max-pooling layer is added for a batch of each sixth convolutions starting from the eighth convolution layer. The next preceding convolution layer has a depth concatenation layer, with the same height and width that of conv layer and concatenates them along the channel dimension. The convolution layer window size varies from 5×5 , 1×1 , 1×1 , 1×1 , 3×3 , 1×1 in the batch of each six conv layers. The last convolution block ends with an average pooling layer and a dropout by 40%. The network architecture ends with fully-connected, softmax and a classification output layer.

4. SqueezeNet (Iandola *et al.*, 2016) is 68 layers deep with 227×227 dimension for image input. The first convolution layer has a window shape of 3×3 , with a ReLU activation function along with a max-pooling layer. A total of 26 such convolution layers are present in this network with different batches. Starting from the Conv4 layer for every third convolution layer a depth concatenation layer is added. The convolution layer's window size varies in its number from 3×3 , 1×1 , 1×1 . The last convolution block ends with a ReLU activation function and an average pooling layer. The architecture of the network completes by adding a fully-connected, a softmax and a classification output layer.

Based on the results of the experiments and observations with respect to the efficacy of each deep neural model considered for early experimentation, a novel deep CNN architecture adapted from AlexNet and ResNet-18, called *ViewNet* was designed, and used for classification of orientation views. A complete architectural model of the ViewNet is represented in Fig. 6.4. The network is comprised of 35 layers with an image input layer with a dimension of 227×227 . The first convolution layer has a window shape of 7×7 , followed by a ReLU activation function, cross channel normalization and a max-pooling layer. Further, the CNN is further built up with a grouped convolution layer, batch normalization, ReLU activation function, Cross Channel Normalization and a Max Pooling layer. Next, convolution, batch normalization and ReLU activation function are used five times and finally a max-pooling layer is added prior to the first fully connected layer. In between the fully connected layers, dropout, the ReLU activation function, global average pooling are also added. The network architecture ends with fully-connected, softmax and a classification output layer.

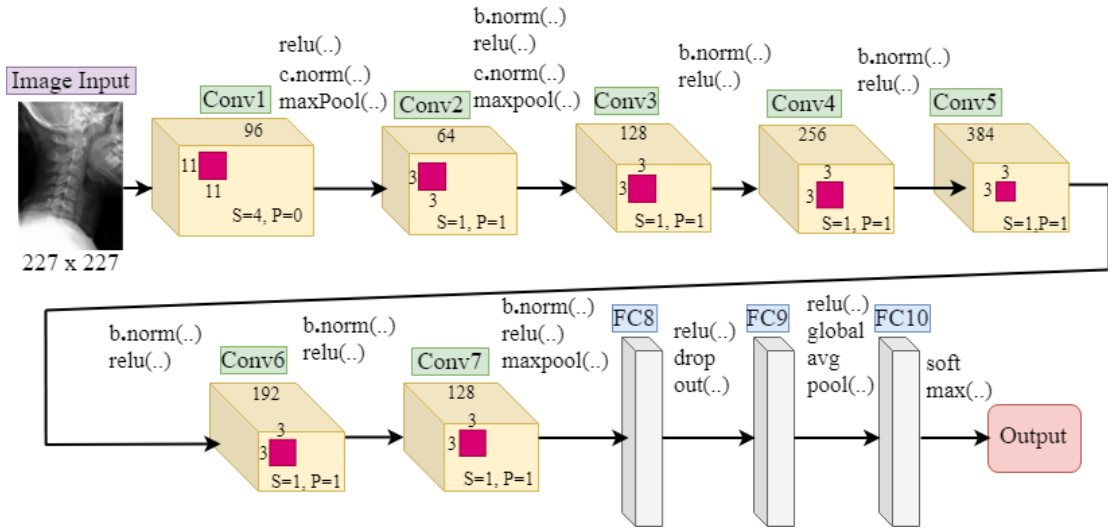


Figure 6.4: Architecture of the proposed ViewNet model

Table 6.1 lists the hyperparameter values determined during training for each of deep neural network models as discussed in Section 6.3. The optimization function used in all four networks used for the benchmarking experiments and the proposed ViewNet is SGDM (Stochastic Gradient Descent Momentum) algorithm, which works faster and better than Stochastic Gradient Descent. The main advantage of SGDM is that, it helps to accelerate gradient’s vectors in the right direction, leading to faster convergence. The experiments conducted to validate the proposed approach and observations regarding the performance are presented in detail in the next section.

Table 6.1: Classification Model parameters

Model	Learning Rate	Weight/ Bias	Batch Size	Epochs
ViewNet (<i>proposed</i>)	0.0001	20	8	10
AlexNet (Krizhevsky <i>et al.</i> , 2012)	0.0001	20	8	10
ResNet18 (He <i>et al.</i> , 2016)	0.0001	10	8	8
GoogleNet (Szegedy <i>et al.</i> , 2015)	0.0003	10	10	6
SqueezeNet (Iandola <i>et al.</i> , 2016)	0.0001	20	10	8

6.4 Experimental Results and Discussion

For the experimental validation, the ImageCLEF 2009 dataset (Lehmann *et al.*, 2003c) was used again, which consists of 41 unique classes of different orientations of body organs like spine (i.e., cervical spine, Thoracic spine, Umbar spine), leg, hand, carpal bones, nose and eye area. Some classes had all three different orientations (e.g. spine) while some classes had only two orientations (e.g. nose and eye area). A set of sample images that are taken for this work from the dataset are shown in Fig. 6.5.

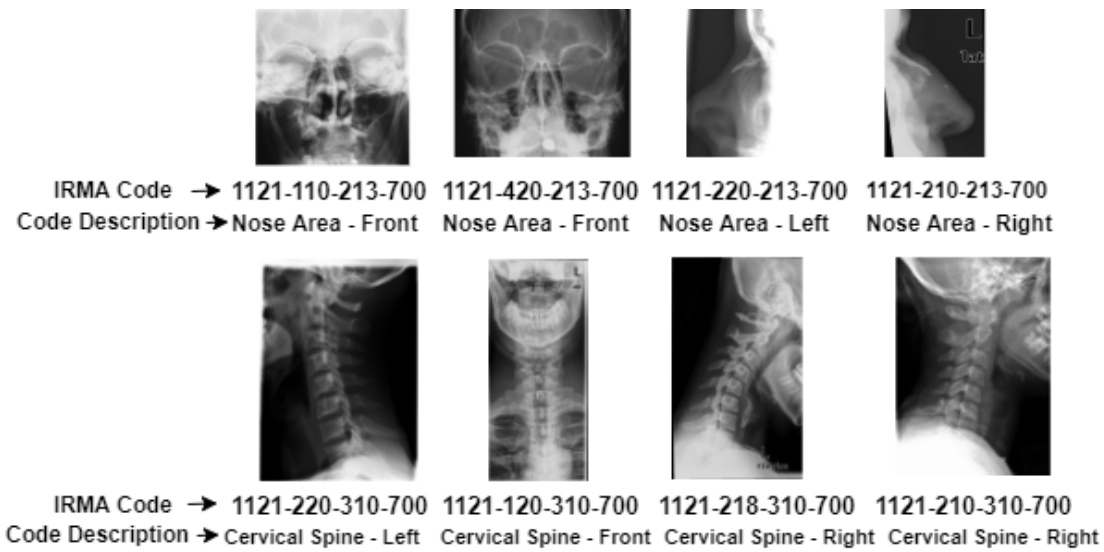


Figure 6.5: Sample dataset images showing the IRMA class code and code description.

The IRMA code is a 13 character unique code, which is subdivided into four parts along its axis. The notation appears as (T-D-A-B), where *T* refers to *technical* (4 digits), *D* - *directional* (3 digits), *A* - *anatomical* (3 digits) and *B* - *biological* (3 digits). Considering *D* - *directional* (i.e., body orientation) and *A* - *anatomical* (i.e., examined part of the body) sub-parts of the IRMA code, the directional/view orientation identification label is predicted for various body parts by using different neural network models. The three-digit directional code (DDD) gives a detailed description of the image orientation view. The first digit gives details about the orientation (e.g.: 1-coronal, 2-sagittal, 3-axial). The second digit gives more information about the position (e.g.: 11-posteroanterior (PA), 12-anteroposterior (AP)) and the third digit details the direction on the orientation of the type of the organ examined (e.g.: 218-inclination).

Similarly, the three-digit anatomical code (AAA) gives a detailed description of

which part of the body was examined. Major regions are coded as (e.g.: *1-whole body, 2-cranium, 3-spine, 4-upper extremity/arm, 5-chest, 6-breast, 7-abdomen, 8-pelvis, 9-lower extremity/leg*) following a two hierarchical sub-codes (e.g.: *7-abdomen, 71-upper abdomen, 711-upper right quadrant, 712-upper middle quadrant, 713-upper left quadrant*). Using these codes of observation helps in building a proper model for the orientation identification system.

The proposed CNN model's performance was measured using standard metrics like accuracy, sensitivity, specificity and F1-score as per Eq. (6.1) to (6.4). Accuracy is computed as the total number of accurate predictions on view class labels divided by the total number of images under test classification. Precision is the number of correctly predicted orientation view labels from a set of selected images. In contrast, recall is the total number of correctly predicted orientation view labels from the entire test set. F1-Score helps to have a measurement that represents both i.e., TPR and TNR, it is the weighted average of the true positive rate and true negative rate.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (6.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (6.3)$$

$$F_1score = \frac{2 * (TP)}{(2 * TP + FP + FN)} \quad (6.4)$$

The classification results before refining *i.e.*, the classes that have three different orientations/views (front view, lateral views - right and left) are depicted in Table 6.2. However, it was observed that some of the class IRMA codes are different but, the orientations/views do not vary. In such cases, both the IRMA codes were combined into a single label (*i.e.*, label refinement was performed). After all such class labels are combined, the test images are classified again, after which a significant improvement in accuracy is observed. The classification results observed after the process of refining the classes was performed are shown in Table 6.3. Experimental evaluation revealed that AlexNet and ResNet18 achieved good classification accuracy of 85.49% & 85.21% (before merging the same orientation classes) and 91.45% & 91.82% (after merging).

The ViewNet model, adapted from AlexNet and ResNet CNN models, out-

Table 6.2: Observed classification performance w.r.t different CNN Models (*before class label refinement*)

NN Model	Accuracy	Precision	Recall	F1-Score
ViewNet (<i>proposed</i>)	0.8571	0.5387	0.5841	0.5298
AlexNet (Krizhevsky <i>et al.</i> , 2012)	0.8549	0.5636	0.5882	0.5634
ResNet18 (He <i>et al.</i> , 2016)	0.8521	0.5033	0.5161	0.4987
GoogLeNet (Szegedy <i>et al.</i> , 2015)	0.6543	0.3788	0.3994	0.3609
SqueezeNet (Iandola <i>et al.</i> , 2016)	0.6502	0.3332	0.3420	0.3082

Table 6.3: Observed classification performance w.r.t different CNN Models (*After class label refinement*)

NN Model	Accuracy	Precision	Recall	F1-Score
ViewNet (<i>proposed</i>)	0.9151	0.5743	0.5996	0.5414
AlexNet (Krizhevsky <i>et al.</i> , 2012)	0.9145	0.6361	0.7011	0.7411
ResNet18 (He <i>et al.</i> , 2016)	0.9182	0.5731	0.5971	0.5741
GoogLeNet (Szegedy <i>et al.</i> , 2015)	0.9053	0.5491	0.6191	0.5531
SqueezeNet (Iandola <i>et al.</i> , 2016)	0.8868	0.5141	0.5921	0.5161

performed all the other CNN models used in the benchmarking experiments. It was found that the proposed ViewNet reached an accuracy of 85.71% with a small improvement over AlexNet. This improvement was attained because the layers that are incorporated in the proposed ViewNet model are taken from both of the top resulted models of AlexNet and ResNet. More importantly, the complexity of the network is significantly lower than AlexNet and ResNet, while still achieving comparable results. The compact architecture actually helps achieve significant optimization in training time, which can be highly significant when the size of medical image repositories that are to be is large. The ViewNet model can be fin-tuned for further improvements in performance, so that it can accurately predict diagnostic image's body orientation for large-scale HIMS as well.

6.5 Summary

In this chapter, a centroid based algorithm designed to detect the orientation of radiography images based on feature extracted were presented. Experimental

results showed that the algorithm achieved good orientation label prediction for three different type of views - facing forward, facing left and facing right. Another work presented the details of an efficient and accurate method for identifying the orientation label of the body organ positioned at the time of the scan. Four transfer learning based neural models were used in early experiments for benchmarking the orientation classification task on the standard open dataset, ImageCLEF 2009. A novel architecture, ViewNet was also proposed for the task of view classification. These neural models were validated on the images available from dataset, with its IRMA code specifics, and achieved promising results when measured in terms of accuracy, sensitivity, specificity and F1-score. The possible applications of this work are in the context of HIMS, for effective and automated view labelling of the scan images after the scan process, to optimize the indexing process, thus streamlining the workflow of HIMS.

Publications

(based on work presented in this chapter)

1. Karthik K., Sowmya Kamath S., “*A Deep Neural Network Model for Content-Based Medical Image Retrieval with Multi-View Classification*”, The Visual Computer Journal (TVCJ), Springer Nature, DOI: 10.1007/s00371-020-01941-2 [SCIE & Scopus, IF: 2.601] *(Status: Online)*
2. Karthik K. and Sowmya Kamath S., “*Automated View Orientation Classification for X-ray images using Deep Neural Networks*”, Smart Computational Intelligence in Biomedical and Health Informatics, CRC Press, Taylor & Francis UK, 2021, DOI: 10.1201/9781003109327. ISBN: 9781000434378 *(Status: Online)*

PART V

Automatic Generation of Medical Image Descriptions

Chapter 7

Generating Medical Image Description

7.1 Introduction

Recent advancement in applications in AI based models for healthcare, research in Machine learning (ML) and deep learning (DL) have shown great promise in accurate diagnosis in tasks like disease prediction, image classification, caption generation (Kumar *et al.*, 2016; García-Floriano *et al.*, 2019; Faes *et al.*, 2019) among others. The performance of these systems in clinical settings can revolutionise the way healthcare services are delivered, especially in a labor-intensive field like radiography, where, the radiologist is expected to manually check each scan and write a list of observations, for enabling diagnosis by the referring doctors. Image processing and Computer Vision (CV) based techniques have been applied to design surgical and imaging intervention applications. Such systems extend clinical decision-making capabilities to healthcare professionals by automating certain tasks related to diagnosis or forecasting the severity of several abnormalities.

Incorporating AI in these systems to support learning behaviour so that systems can detect abnormalities at the earliest disease onset in a wide variety of diagnostic media are of critical importance. The radiologist can utilize these insights for enabling and optimizing the quality of diagnosis. The objective of this work is to design neural ensemble models that effectively combine the latent image features and semantic information from the clinical text reports for enabling improved context inference for automatic generation of diagnostic reports for new X-ray images.

7.1.1 Problem Definition

Automatically generating a medical image report is a challenging task, and requires both computer vision and natural language processing insights. However, such a capability has a huge impact on medical data management, and could significantly benefit clinicians for valuable insights and reduce the overall burden on patient care's workflow. Recent advancement in machine learning and deep learning has resulted in design of applications for disease prediction, image classification, caption generation, etc. In some existing approaches, attention models are used in addition to encoder-decoder models and are pre-trained with convolutional neural networks like VGG16, VGG19 and Resnet50. The focus of this work is to design automated methods for radiographic image examination for identifying abnormalities and generating reliable radiology reports. Thus, the problem to be addressed here is defined as follows:

“Given a set of diagnostic images containing latent visual information and a set of corresponding diagnostic text reports, designing multimodal models for automatically identifying anomalies in the diagnostic images for generating their natural language descriptions w.r.t findings.”

In view of these observations, the work presented in this chapter encompasses development of ensemble deep neural models for automated abnormality detection and classification. The abnormalities present in the images are identified using the developed abnormal region detection algorithm. Further, the features generated by the ensemble neural model are used for the automated generation of radiological text reports, thus reducing the radiologists' workload and also streamlining the diagnosis process.

7.2 Abnormality Detection and Classification of Plain Radiographs

The proposed approach for abnormality classification, abnormality localization and automated diagnostic report retrieval for X-ray images is illustrated in Fig. 7.1. Two publicly available, standard datasets were used for the experimental validation of the proposed approach. The first one, the MURA dataset (Rajpurkar *et al.*, 2017a) provides musculoskeletal radiograph images of seven upper extremity classes like Hand, Forearm, Wrist, Finger, Shoulder, Humerus and Elbow. The second dataset, the Indiana University dataset (Demner-Fushman *et al.*, 2016)

consists of chest X-ray images, along with indications, findings, and impressions in a textual form for each image. These two datasets were together used for addressing the three different clinical tasks, each involving data belonging to two data modalities, X-ray images and clinical text reports, thus resulting in multimodal datasets. Due to this, the proposed ensemble models are more robust, capable of dealing with varied types of abnormalities in underlying radiographic images. The three clinical tasks undertaken form a significant part of a typical clinical workflow that is managed on a daily basis by a radiologist in hospital scenarios. They are –

1. Classifying a given diagnostic image as either normal or abnormal.
2. Abnormal region detection for localize and visualizing identified abnormal areas.
3. Automated diagnostic text report generation.

The methodology adopted for addressing each of these tasks are discussed in the subsequent sections. The proposed MSDNet ensemble model is the core of the pipeline that is used for the three clinical tasks, which is trained to extract features and image index values that will facilitate the classification, identification and localization of anomalies, and also automated diagnostic text report generation for a given input image.

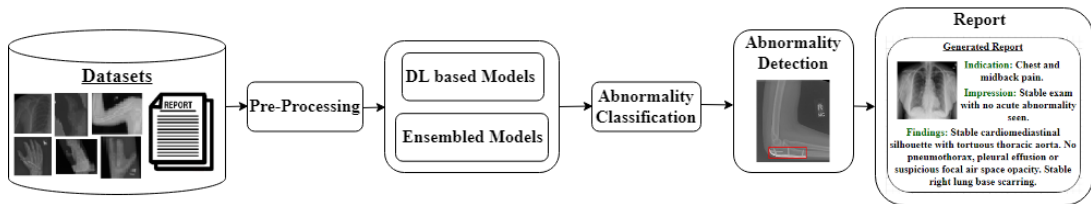


Figure 7.1: Abnormality Classification and Report Generation process.

7.2.1 Abnormality Classification

The architecture of the proposed MSDNet model is depicted in Fig. 7.2. The proposed network is built as an ensemble of the AlexNet (Krizhevsky *et al.*, 2012) and ResNet18 (He *et al.*, 2016) architectures for initial classification of the category of the image, i.e. abnormal or normal. The global features are obtained from AlexNet, while ResNet18 is used to generate the local features, which are combined to form a fused feature set. Concatenated features are then fed into the fully-connected layer for final abnormality classification. If the probability value is

equal to or higher than 0.5, then the image is classified as an abnormal study. Here, the task is to classify the given image into the category to which it belongs. Hence, it is a single-label categorical classification (i.e., softmax activation), and the standard weighted categorical cross-entropy loss is given by:

$$J_{wcce} = -\frac{1}{M} \sum_{k=1}^K \sum_{m=1}^M w_k \times y_m^k \times \log(h_\theta(x_m, k)) \quad (7.1)$$

where, M gives the number of training examples; K gives the number of classes; w_k is the weight for class k ; y_m^k is the target label for training example m for class k ; x_m is the input for training example m and h_θ represents a model with neural network weights θ .

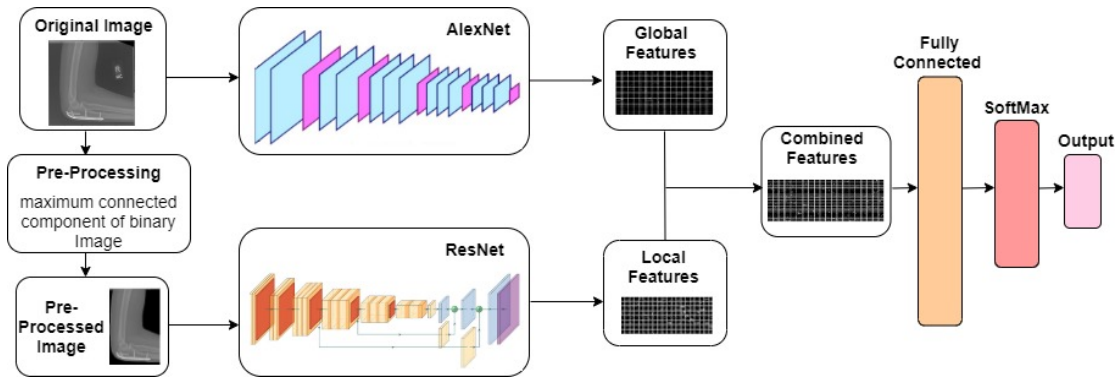


Figure 7.2: Architecture of the proposed MSDNet model

For the training process, initially different batch sizes like 8, 16, 32, 64 and finally the learning rate was set to 0.0001, with a batch size of 16, while the number of epochs for training the network were set to 10. Stochastic Gradient Descent with Momentum (SGDM) was used as the solver optimizer because of its ability to switch back and forth to reach the optimum path. Hence, momentum parameters were added to reduce the switching problem (McHugh, 2012). The values of SGDM were calculated and updated as per Eq. (7.2), where, ℓ is the iteration number, $\alpha > 0$ is the learning rate, θ is the parameter vector, $E(\theta)$ is the loss function and γ determines the contribution of the previous gradient step to the current iteration.

$$\theta_{\ell+1} = \theta_\ell - \alpha \nabla E(\theta_\ell) + \gamma(\theta_\ell - \theta_{\ell-1}) \quad (7.2)$$

To avoid considering the same data segments at every epoch, a shuffling parameter value was set before each training epoch. i.e., if the mini-batch size cannot

uniformly distribute the training samples, then *trainNetwork* discards the training data that does not fit into the final complete mini-batch of each epoch. To learn faster in the new layers, the *WeightLearnRateFactor* and *BiasLearnRateFactor* values of the fully connected layer were increased. This was heuristically set to 20, the learning rate is determined by multiplying this factor by the global learning rate, for deriving the biases in the fully connected layer. The cross-entropy loss function is used as an encoded output. For a single image this loss is computed as per Eq. (7.3),

$$\text{Cross-entropy loss} = \sum_{c=1}^M (y_c \cdot \log \hat{y}_c) \quad (7.3)$$

where, M is the number of classes and \hat{y}_c is the model's prediction for that class (i.e. the output of the softmax for class c). y is a (2×1) vector of one's and zero's, y_c is either 1 or 0. Finally, the predicted class label of an input X-ray image is obtained at the output layer.

7.2.2 Abnormal Region Detection

After an image is classified as normal or abnormal, an automated analysis of the type of abnormality present in the image is of critical importance. The objective is to identify potential abnormality findings like hardware artifacts and the existence of fractures in the scan image. Boundary detection in abnormal images is one of the crucial steps while generating x-ray scan reports. Currently the abnormal regions are manually marked by expert clinicians/physicians. The developed algorithm can make a change over in the Computer Aided Diagnosis medical system, where the boundary will be marked by the system itself if it finds any abnormalities present in the image.

In this work, a boundary detection algorithm is incorporated to detect the abnormalities present in the image. For each image, the histogram of the abnormal image is plotted and the image is binarized with less than the bin location, after which the largest blobs/regions in the image are determined. Next, the rightmost connected components are obtained, after which, the centroid and bounding box of the masked image are captured. Using these, a bounding box is marked with the final values. Algorithm 7.1 illustrates the process of identifying the abnormal regions in the radiograph images.

Algorithm 7.1 Abnormal Region Detection Algorithm

Input: An Abnormal Image
Output: Boundary region on Abnormal Image

- 1: **for** $i = 1$ to $\text{length}(\text{TestImage})$ **do**
- 2: $I_T \leftarrow \text{TestImage}$ \triangleright Read the test image
- 3: Compute the histogram of the image.
- 4: Find total histogram values > 1000 .
- 5: $\text{binaryImage} \leftarrow \text{grayImage} < \text{x-axis value with last peak from histogram}$
 \triangleright Binarize the image
- 6: $\text{binaryImage} \leftarrow \text{bwareafilt}(\text{binaryImage}, 2)$ \triangleright Extract only the two largest blobs
- 7: $\text{labeledImage} \leftarrow \text{bwlabel}(\text{binaryImage})$ \triangleright Label Connected Components
- 8: $\text{binaryImage2} = \text{labeledImage} \leftarrow 0$
- 9: $\text{binaryImage2} \leftarrow \text{imfill}(\text{binaryImage2}, \text{'holes'})$ \triangleright Fill holes
- 10: $\text{TestImage}(\sim \text{binaryImage2}) \leftarrow 0$ \triangleright zero out the other parts of the image
- 11: $\text{Mask} \leftarrow \text{grayImage} > \text{call}$ \triangleright Get a new binary image
- 12: $\text{Mask} \leftarrow \text{imfill}(\text{Mask}, \text{'holes'})$ \triangleright Fill holes
- 13: $\text{Mask} \leftarrow \text{bwareafilt}(\text{Mask}, 1)$ \triangleright largest blob selected
- 14: $\text{Mask} \leftarrow \text{bwconvhull}(\text{Mask})$ \triangleright Take convex hull
- 15: Get the Centroid points.
- 16: Mark the abnormal region with the bounding box and centroid points.
- 17: Output abnormalities.
- 18: **end for**

7.2.3 Automatic Diagnostic Text Report Generation

For this task, the modeled features extracted from the X-ray images and the expert-written diagnosis reports are utilized for capturing the findings and impressions as a text report for the identified abnormal chest X-rays taken from the Indiana University dataset. The convolutional layers that build up each of the two adapted models based on the architectures of AlexNet and ResNet-18 were trained to extract disease-specific features from the X-ray images. Using these extracted image features and the image IDs, the findings and impressions of the image readings are incorporated for generation of a natural language text report for a test chest X-ray image. Fig. 7.3 illustrates the process of automatic diagnostic report retrieval.

Algorithm 7.2 illustrates the report retrieval procedure after the image classification process. During the feature extraction process, each image in the training and test sets is processed for generating a feature vector. When a test image feature set is given as a query, the pairwise distance measure is used to compute distances that can be used to obtain the matching feature set with the smallest

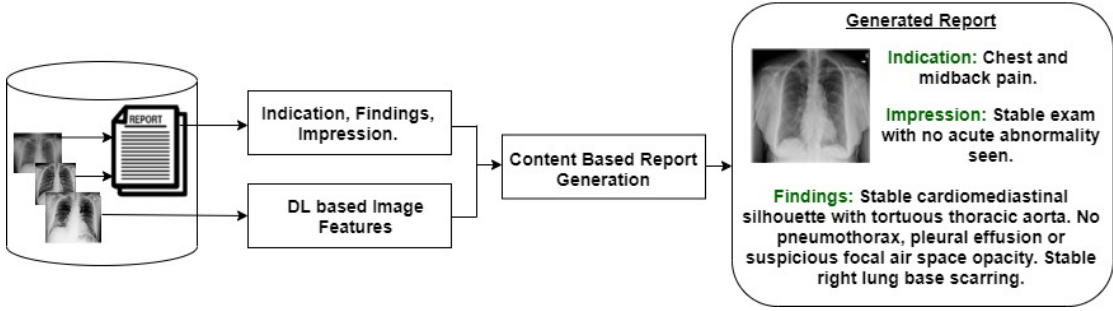


Figure 7.3: Automatic Report Generation Process for Chest X-ray Images.

Algorithm 7.2 Automatic Report Generation Algorithm

Input: A Sequence of Images and its corresponding Report.

Output: Image and its corresponding Report.

For each test feature set

$ID's \leftarrow$ Find the nearest distance in train feature set.

Read text report file. \triangleright *Image's Report.*

$Text \leftarrow I^{ID}$, Indication, Findings & Impression. \triangleright *Read I^{ID} , Indication, Findings & Impression from Report.*

for $i = 1$ to $length(TestImage)$ **do**

$I \leftarrow TestImage$ \triangleright *Read the test image*

$I_l \leftarrow TestImageLabel$ \triangleright *Get the abnormality label of the test image.*

for $j = 1$ to $length(ID)$ **do**

$M_I \leftarrow TrainImage(ID[i,j])$ \triangleright *Read the nearest matched train image.*

$Lbl \leftarrow TrainImageLabel(ID[i,j])$ \triangleright *Get the abnormality label of the*

matched image.

$Index \leftarrow ID[i,j]$ \triangleright *Get the Index number of the matched image.*

$Ind = Text(Index,a)$

$Find = Text(Index,b)$

$Imp = Text(Index,c)$ \triangleright *Get*

Indication, Findings & Impression of the relevant Index row., where a, b & c are the Indication, Findings & Impression column number of the Report file.

end for

Display $I, I_l, M_I, Lbl, Ind, Find$ and Imp .

distance from the images in the training set. These measures are usually used to find the similarity between two data objects. For each observation in Y (*Test image features*), the pairwise distance method finds the smallest distances by computing and comparing the distance values to all the observations in X (*Training image features*).

During experiments, it was observed that Cosine and Standard Euclidean achieved the smallest distance measure with an overall accuracy of 78.03% and 77.26%. Hence, for this report generation experiment, we used Cosine as the dis-

tance measure. When a test image is fed into the model, the Cosine distance is computed to find the training set's nearest match. Its equivalent index number is the closest reference, using which the image and its description (Indication, Findings & Impression) that match are retrieved and displayed.

7.3 Experimental Evaluation and Results

For the experimental evaluation of the proposed model, the MURA and Indiana University datasets were used. Both are open access, publicly available datasets, the specifics of which are listed below.




1. The MURA dataset ([Rajpurkar *et al.*, 2017a](#)) is made available by the Stanford Group, and consists of musculoskeletal radiography images of seven upper extremity classes. The scan images are represented as per the PACS specifications. MURA consists of 14,863 studies taken from 12,173 patients, consisting of 40,561 radiography images in total. The dataset is split into training (36,808 images, 13,457 studies from 11,184 patients), testing (3,197 images, 1,199 studies from 783 patients) sets. There is no overlap in patients between any of the sets and each study has been labeled manually as either normal or abnormal, by Stanford Hospital's board-certified radiologists. Abnormal images in the dataset contain anomalies like fractures, hardware artifacts, degenerative joint diseases and other miscellaneous abnormalities, including lesions and subluxations. A sample of normal and abnormal images from the dataset is shown in Fig. 7.4.
2. The Indiana Dataset ([Demner Fushman *et al.*, 2016](#)) includes 7,470 chest X-Ray images having both frontal and lateral images with annotations, which consist of indications, findings, & impressions in a textual form. We use this dataset for the clinical task of abnormality classification and for retrieving the reports for a given test image. Hence, only the frontal chest X-ray images (comprising of 4,000 images) to extract the image's relevant features at the training phase. A sample of this is shown in Table 7.1.

First, the predicted class labels are obtained from the deep ensemble model. It is observed that, for the seven classes of the MURA dataset, a total of 3197 images - 1,667 images were normal and 1,530 images contain abnormal findings (fractures, hardware artifacts and joint diseases). In the Indiana dataset, abnormal/ disease annotations like cardiomegaly, opacity, pleural effusion, pneumothorax, pulmonary



Figure 7.4: Sample images of MURA dataset in Hand, Forearm, Wrist, Finger, Shoulder, Humerus and Elbow classes (*Upper row - abnormal images; Lower row - normal images*)

Table 7.1: Sample images from the Indiana Chest X-ray dataset, along with the associated indications, findings and impressions

Image	Indication	Findings	Impression
	Preoperative renal transplant.	The lungs and pleural spaces show no acute abnormality. Stable left upper lobe calcified granuloma. Heart size is mildly enlarged, pulmonary vascularity within normal limits. Mild tortuosity of the descending thoracic aorta.	No acute pulmonary findings. Mild cardiomegaly.
	Chest and midback pain.	Stable cardiomeastinal silhouette with tortuous thoracic aorta. No pneumothorax, pleural effusion or suspicious focal air space opacity. Stable right lung base scarring.	Stable exam with no acute abnormality seen.
	Shortness of breath.	The cardiac contours are normal. The lungs are hyperinflated with flattening of the diaphragms and tapering of the distal pulmonary vasculature. There is no focal consolidation. Thoracic spondylosis. Mild dextroscoliosis of the spine. Prior anterior cervical fusion.	Emphysema without superimposed pneumonia.

edema, and shortness of breath were identified in the study. Of these sets, 81.79% of images were correctly classified as normal and 76.62% of images were predicted to be abnormal. The proposed approach’s performance is measured using standard evaluation metrics like accuracy, sensitivity, specificity, and kappa statistics.

Cohen’s kappa statistics (McHugh, 2012) measures the inter-observer agreement or precision and is especially useful when the same score is assigned to the same data items. Hence, the outcome is to predict whether the data sample under test is normal or abnormal. The importance of this metric lies in the correct representation of the data measured. Its range is -1 to +1; a value of 1 indicates a “perfect agreement” and a value less than 1 shows “less than a perfect agreement”. In some rare situations, the Kappa value can also be negative, signifying that the agreement score is much lower than expected. It is computed as per Eq. (7.4), where $Pr(a)$ is the actual observed agreement, and $Pr(c)$ is the chance agreement.

$$Kappa = \frac{Pr(a) - Pr(c)}{1 - Pr(c)} \quad (7.4)$$

The results of the experimental evaluations conducted with each neural model selected for the comparison are tabulated in Table 7.2 and 7.3. From these results, it was observed that ResNet outperforms AlexNet, which can be attributed to its 71 layer deep architecture, in contrast to AlexNet’s 25 layer architecture. The feature representation learnt by the convolution layers of ResNet18 with different batch sizes also contributed to greater accuracy. Another significant reason is the inclusion of local features, by locating the maximum connected component on the binary map. The local area is cropped from the input image (preprocessed image), and it fed through the subsequent layers, to finally produce the local features. In view of this, the global features extracted by AlexNet and the local features generated by ResNet were concatenated and fed into the fully-connected layer for final ensemble classification model that forms MSDNet. The results of this were evident, as the proposed MSDNet models outperformed both ResNet and AlexNet, emphasizing the effectiveness of the global+local feature representations towards anomaly classification.

The models performed best for the *Elbow*, *Forearm*, *Humerus*, *Wrist* and *Chest* classes achieving >80% accuracy, while, they showed satisfactory results for the *Finger*, *Hand* and *Shoulder* classes. The other two metrics, sensitivity and specificity, capture some additional aspects of the classification performance. Sensitivity is a measure of the true positive rate or probability of detection, i.e., it

Table 7.2: Classification Performance w.r.t different classes for the proposed MSDNet model

Classes	Accuracy	Sensitivity	Specificity	Kappa
Elbow	0.8317	0.8383	0.7004	0.736
Finger	0.7994	0.7897	0.7263	0.671
Forearm	0.8394	0.8533	0.6894	0.792
Hand	0.7832	0.8413	0.6732	0.754
Humerus	0.8586	0.7905	0.8086	0.676
Shoulder	0.7735	0.6877	0.7635	0.731
Wrist	0.8447	0.8874	0.7458	0.855
Chest	0.8827	0.8911	0.7231	0.758
Overall	0.8269	0.8179	0.7662	0.746

Table 7.3: Classification Accuracy of various models

Classes	AlexNet	ResNet18	MSDNet
Elbow	0.7867	0.8218	0.8317
Finger	0.7113	0.8059	0.7994
Forearm	0.7450	0.8112	0.8394
Hand	0.7082	0.7603	0.7832
Humerus	0.7985	0.8549	0.8586
Shoulder	0.6956	0.7937	0.7935
Wrist	0.8067	0.8469	0.8447
Chest	0.8221	0.8571	0.8827
Overall	0.7625	0.8218	0.8269

indicates the percentage of medical scans correctly identified as abnormal. Specificity or the true negative rate gives the percentage of normal medical scans that were correctly classified as normal.

The observed sensitivity scores for the *Elbow*, *Forearm*, *Hand*, *Wrist* and *Chest* classes were in the range of 83% to 89%, indicating that the abnormal samples were correctly classified for these classes to a larger extent. However, the specificity scores of classes like *Forearm* and *Hand* indicate that the percentage of normal

scans correctly classified as normal was lower than that of the other classes. The lowest sensitivity score was observed for the *shoulder* class, indicating that the model could not very well distinguish between normal and abnormal images, thus requiring more detailed scrutiny and analysis. On average, the proposed approach achieved an accuracy rate of 82.69%, with sensitivity and specificity scores of 81.79% and 76.62%, respectively, which indicates good classification performance.

A graphical plot that illustrates the diagnostic ability of a classifier system in terms of the Area under the ROC¹ curve (AUC) is illustrated in Fig. 7.5. An AUC value of 0.9038 is achieved, indicating good performance in distinguishing anomalous and non-anomalous radiographical scans. As discussed earlier, the average Kappa statistic value was 0.746, which indicates a substantial agreement on the test samples with the expected values. However, radiograph readings and their findings are often judged subjectively, hence we also used others metrics like accuracy, specificity and sensitivity to gain more refined insights into the proposed model's performance.

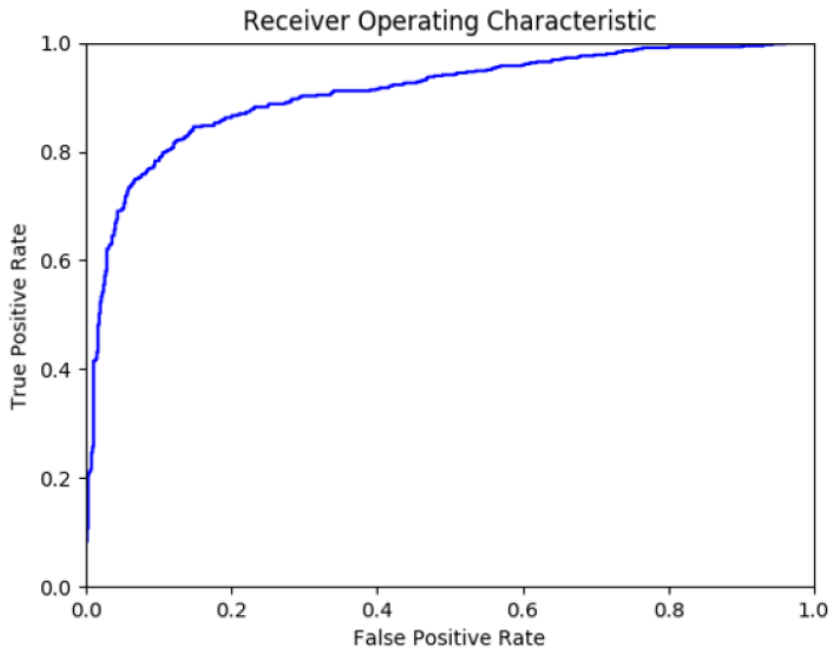


Figure 7.5: AUROC performance of the proposed Ensemble Model.

The process of identification of abnormalities in the scanned image using the proposed algorithm is one of the potential outcomes noted here. Hardware artifacts used for setting bones like metal inserts and screws are automatically detected and correctly classified as a type of abnormality. Similarly, even fractures and

¹Receiver Operating Characteristic

cracks were also captured accurately as abnormalities using the proposed abnormal detection algorithm. It was found that the abnormal region detection algorithm performed well during experimental validation, evident which is shown in Fig. 7.6. A sample report that is retrieved from the model for a given test image is shown in Fig. 7.7.

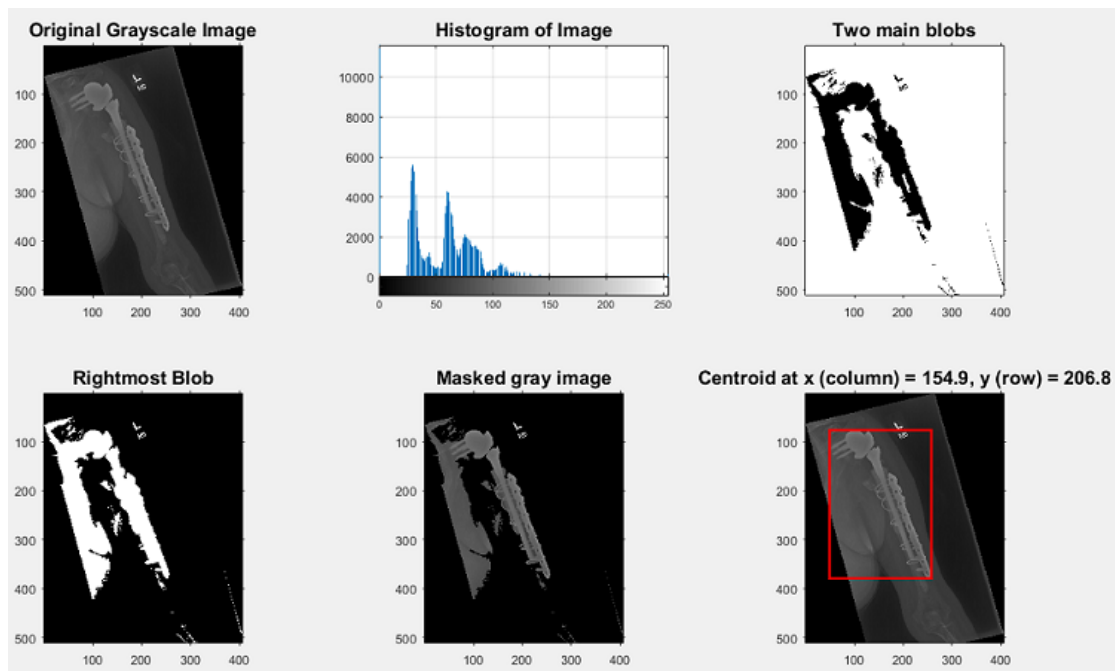


Figure 7.6: Illustration of the abnormal area detection process for sample images from the *Shoulder* class.

UI
- □ ×

Diagnostic Text Report

<p>Test Image</p>	<p>Actual Report</p> <p>Indication: XXXX-year-old female with chest pain.</p> <p>Findings: The heart size is enlarged. Tortuous aorta. Otherwise the mediastinal contour is within normal limits. The lungs are free of any focal infiltrates. There are no nodules or masses. No visible pneumothorax. No visible pleural fluid. The XXXX are grossly normal. There is no visible free intraperitoneal air under the diaphragm.</p> <p>Impression: Cardiomegaly without lung infiltrates.</p>
<p>Model's Report</p> <p>Indication: Chest pain</p> <p>Findings: The heart size is enlarged. The aorta is tortuous. The pulmonary vasculature appears normal. Lungs are otherwise clear bilaterally. No pleural effusions or pneumothorax. No bony abnormalities.</p> <p>Impression: Cardiomegaly</p>	

Figure 7.7: Sample model generated report, with the ground-truth data.

Table 7.4: Proposed model’s performance w.r.t Kappa Score against State-of-the-art models

Classes	Rajpurkar <i>et al.</i> (2017a)	Saif <i>et al.</i> (2019)*	Banga and Waiganjo (2019)	Solovyova (2020)	Proposed
Elbow	0.710	0.733	0.617	0.715	0.736
Finger	0.389	0.735	0.653	0.395	0.671
Forearm	0.737	0.785	0.695	0.730	0.792
Hand	0.851	0.835	0.584	0.862	0.754
Humerus	0.600	0.754	0.599	0.602	0.676
Shoulder	0.729	0.855	0.659	0.735	0.731
Wrist	0.931	0.907	0.740	0.942	0.855
Chest	—	—	—	—	0.758

Note: Approach marked with * used 50% of the data.

Table 7.5: Benchmarking proposed model against state-of-the-art models using standard metrics.

Models	Accuracy	Sensitivity	Specificity	Kappa
Proposed Model	0.82	0.81	0.76	0.74
DenseNet-169 (Chada, 2019) *	0.79	0.72	0.88	0.60
DenseNet-201 (Chada, 2019) *	0.82	0.81	0.84	0.64
InceptionResNetV2 (Chada, 2019) *	0.82	0.81	0.83	0.64
EnsembleD [Dense, MobileN] (Banga and Waiganjo, 2019)	0.83	0.92	0.73	0.66
EnsembleE [Xcep, Dense] (Banga and Waiganjo, 2019)	0.71	0.77	0.63	0.41
MobileNet (Single) (Banga and Waiganjo, 2019)	0.67	0.73	0.61	0.34
EnsembleD [Xcep, MobileN] (Banga and Waiganjo, 2019)	0.65	0.73	0.56	0.29

Note: Approach marked with * used Finger and Humerus class only. Average of two class results is presented here.

The proposed model was benchmarked against state-of-the-art models using Kappa score, the results of which are tabulated in Table 7.4. It can be observed from the table that the proposed MSDNet model showed substantial improvement over other models. However, it underperformed slightly for the *Hand* and *Wrist* classes when compared to the DenseNet-169 model (Rajpurkar *et al.*, 2017a). Since, Rajpurkar *et al.* (2017a) used only one evaluation metric, i.e., *Kappa*, it is difficult to analyze the result because other standard evaluation metrics like Accuracy, Sensitivity and Specificity were not observed for their work. However, the deeper layers might have contributed more to those specific classes as is evidenced in the better kappa scores. Several other state-of-the-art models were considered for the benchmarking experiments, and were compared with the proposed model, the results of which are presented in Table 7.5. The proposed model outperformed all these state-of-the-art models, showing its dominance in the abnormality prediction clinical task.

7.4 Deep Neural Models for Automated Multi-task Diagnostic Scan Management

This section details an approach comprising a multiple clinical tasks like scan quality enhancement, image orientation view identification and generation of diagnostic radiology reports for chest X-ray images. The overall framework of the proposed approach is illustrated in Fig. 7.8. In the first phase, a X-ray image of size $M \times N$ is fed into the adversarial network to produce a super-resolution image of size $P \times Q$ for enabling high-quality diagnosis (if size of $M = N = 256$, then $P = Q = 1024$, i.e., 4 times the original size after enhancement). In the next step, this enhanced image is passed into the proposed neural network model (ViewNet) to predict the orientation label of the image that was acquired during the scanning process. Additionally, a natural language report generation model is incorporated in this overall clinical diagnostic application, trained on existing text reports. The methodologies proposed here for addressing each of these tasks are discussed in the subsequent sections.

7.4.1 Scan Quality Enhancement

Image quality affects the ease of extracting information from an image. Good image quality will ensure for the maximum amount of diagnostic details is gained from the image or not. Medical images having image quality problems like con-

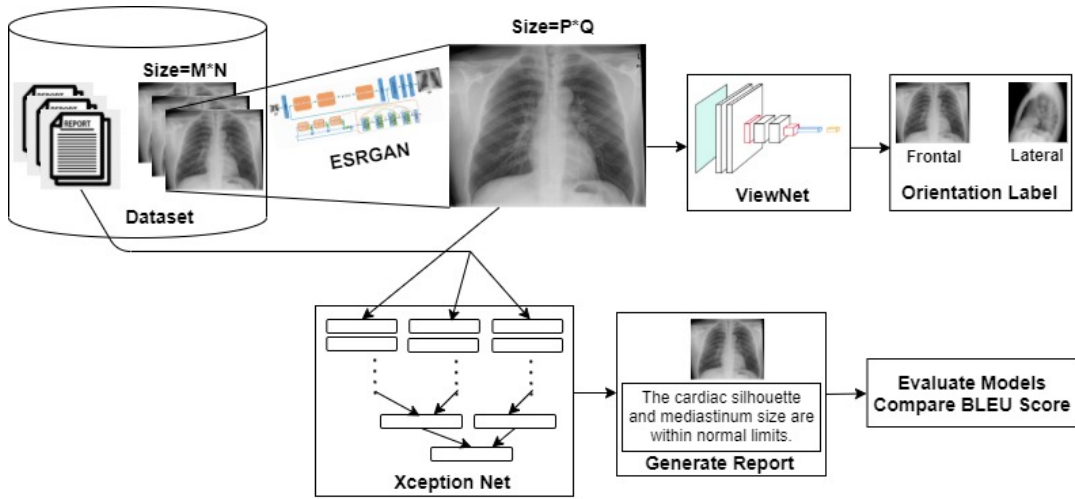


Figure 7.8: Proposed Automated Multi-task Diagnostic Scan Management.

trast, blur, sharpness, and exposure levels make it difficult to properly examine the body's biological parts. Hence, improving the image's quality with enhancement helps physicians to study the body's internal factors for proper diagnosis. During enhancement of the images, the size and dimensionality of the data enhance, so accurate and novel computer-aided methods need to be properly modeled since there is a dependency between the medical data and the model design. Based on these criteria, the aim is to improve the perceptual quality of the image for improved diagnosis by incorporating ESRGAN technique (Wang *et al.*, 2018b).

The architecture used for enhancement is the basic design of SRGAN (Fig. 7.9) with some modifications in its layers. It is observed that the BN layers introduce artifacts when the model architecture tends to be deeper and violates a stable performance during training (Wang *et al.*, 2018b) reducing computational complexity. Using residual-in-residual dense structure improves the performance, supporting the network capacity for dense connections.

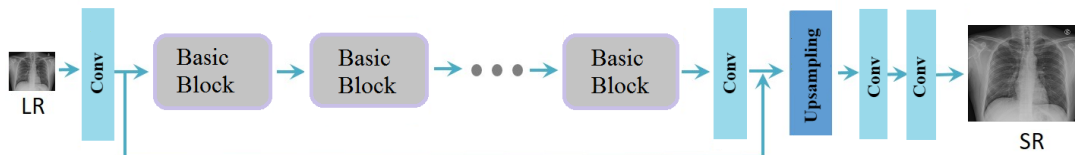


Figure 7.9: Architecture of proposed SRResNet.

To improve the visual quality of the chest X-ray images, the SRGAN architecture was restructured. Firstly, all the batch normalization layers were removed, as it has proven that this increases the performance and reduces the complexity in different image resolution techniques like super-resolution and deblurring. The

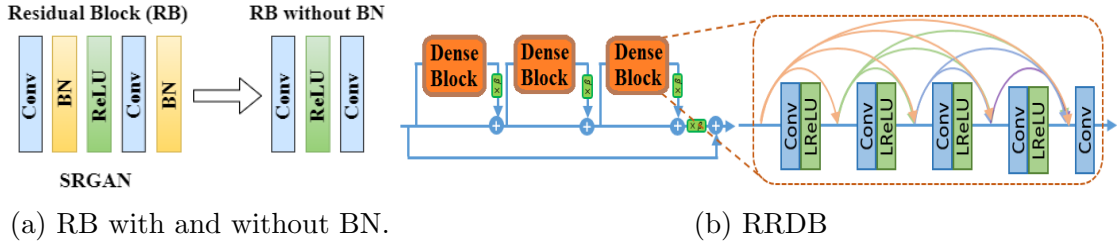


Figure 7.10: RB and RRDB in SRGAN.

basic blocks were replaced with residual dense blocks that combine residual network with dense connections, as shown in Fig 7.10. In addition to the generator, the discriminator part was also enhanced by adapting the relativistic discriminator instead of the standard discriminator. The key difference between the two is that the prior one estimates the real image’s probability “is the image comparatively more real than the fake”. The later, i.e., the standard discriminator calculates the probability between the real and the fake images. In ESRGAN, the features were driven before the activation layer compared to SRGAN, which was positioned after the activation. This design’s change is mainly to overcome the sparse representation of activated features and the unstable brightness when tried to match with the ground-truth image. Therefore, the total loss for the generator is computed as per Eq. (7.5), where L_1 is the content loss that evaluates the 1-norm distance between the recovered image and the ground-truth image. λ and η are the coefficients to balance different loss terms.

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \quad (7.5)$$

Sometimes, GAN architectural models produce undesirable noise, which degrades the quality of network performance. The proposed ESRGAN is built with an easily adjustable and efficient approach called network interpolation, which is defined as per Eq. (7.6), where, θ_G^{INTERP} , θ_G^{PSNR} and θ_G^{GAN} are the parameters of G_{INTERP} , G_{PSNR} and G_{GAN} , respectively, and $\alpha \in [0, 1]$ is the interpolation parameter. An advantage of using this network interpolation is that the model produces a meaningful result without any artifacts. Also, balanced perceptual quality of the image can be maintained consistently. A learning rate 0.0002 with a batch size of 16 was set for training the network.

$$\theta_G^{INTERP} = (1 - \alpha)\theta_G^{PSNR} + \alpha\theta_G^{GAN} \quad (7.6)$$

7.4.1.1 Image Quality Assessment

For this task, different evaluation metrics were explored to evaluate the quality of the enhanced image produced. Most of the evaluation metrics used for measuring super-resolution performance were based on PSNR and SSIM. Since they can be used when the ground-truth and the super-resolution images have same pixel resolution. But in this case, it is an enhancement of the original image, thus there can be difference in the resolutions of the actual and computed image. From the image quality metric analysis study, it is found that, when evaluating the quality of the computed image, there are two types of quality metric evaluations, which are listed below -

1. *Full-reference Quality Metrics* - PSNR, SSIM, MSSIM.
2. *No-reference Quality Metrics* - Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), Natural Image Quality Evaluator (NIQE), Perception-based Image Quality Evaluator (PIQE).

The BRISQUE and the PIQE algorithms calculate an image's quality score with good computational efficiency, after the model is trained. PIQE is less computationally efficient, but it is a measure of local quality in addition to being a global quality score. All no-reference quality metrics usually outperform full-reference metrics in terms of agreement with a subjective human quality score. In view of this, the metrics BRISQUE and PIQE were chosen for evaluating the quality of the enhanced images.

7.4.2 Orientation Classification

The image orientation label is a crucial requirement for effective medical image management. Currently, the orientation view is identified by a single character during the analysis of the scanned part of the body, which is typically noted by the radiologist during scanning. However, this is often overlooked due to the vast, continuous workload that scanning equipment is subjected to in large hospitals and by the busy schedules of scanning technicians and doctors. In the proposed work, the body orientation label identification is attempted by training deep neural models to predict the orientation label and automating the learnings from image observations and findings. Two different neural models were considered during the initial benchmarking phases and the *ViewNet* model is used for the body orientation view classification task. The architecture of ViewNet (as depicted in

Fig. 7.11) is composed of a combination of the layers from AlexNet and ResNet-18 to classify the input scan images based on the scan orientations.

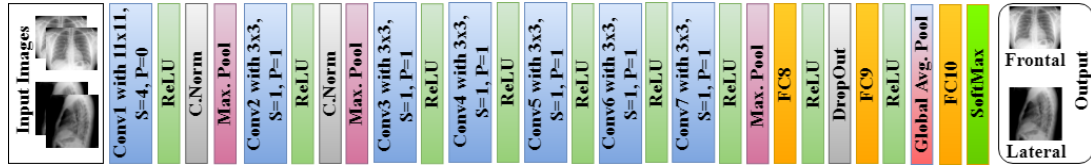


Figure 7.11: ViewNet Model for Body Orientation Classification.

The ViewNet model is a CNN based network which comprises 29 layers, with an image input layer with a dimension of 227×227 . The first convolution layer consists of a window shape of 7×7 , followed by a ReLU activation function, cross channel normalization and a max-pooling layer. Further, the CNN is built up with a grouped convolution layer, ReLU activation function, Cross Channel Normalization and a Max Pooling layer. Next, convolution and ReLU activation function are used five times and finally, a max-pooling layer is added before the first fully connected layer. In between the fully connected layers, dropout, the ReLU activation function, global average pooling are also added. The network architecture ends with fully-connected, softmax and a classification output layer. All batch normalization layers are removed because it has proven that it increases the performance and reduces the complexity in different image resolution techniques (Nah *et al.*, 2017). As per the findings of Nah *et al.* (2017), the batch normalization enables higher learning rates during initialization. As an example, they demonstrated this phenomenon through a training network that stopped producing deterministic values. This effect is found to be advantageous to the generalization of the network either with or without Batch Normalization.

7.4.3 Diagnostic Report Generation

During medical check-ups or surgery of any chest-related issues, doctors recommend diagnostic scanning to diagnose the problem using modalities like chest x-rays and CT scans. After inspecting the chest x-ray images, doctors generate radiology reports containing summarized information essential for further diagnosis and follow-ups. Although deep learning techniques have been successfully applied to image classification and image captioning tasks, radiology report generation remains challenging in understanding and linking complicated medical visual contents with accurate natural language descriptions. Considering the demands of accurately interpreting medical images in large amounts, a medical imaging report

generation model can be helpful.

In this approach, the medical report generation is based on findings detected using a deep learning model. Here, for a given chest X-ray image, its equivalent report is generated. Data Augmentation technique has been used to generate an extensive dataset to better enable deep learning models to learn data features well. The Encoder-decoder model is ensembled with the Xception model as a pre-trained model for extracting diagnostic image features, which are then utilized for generating the diagnostic report. The performance of the model for this task is evaluated using the BLEU (Bilingual Evaluation Understudy) metric. The proposed model is shown in Fig. 7.12, where the first input is of image feature and the second input is findings and the third one is for partial impressions that extracts a length vector of 2048, 166 and 114 respectively from each input layer.

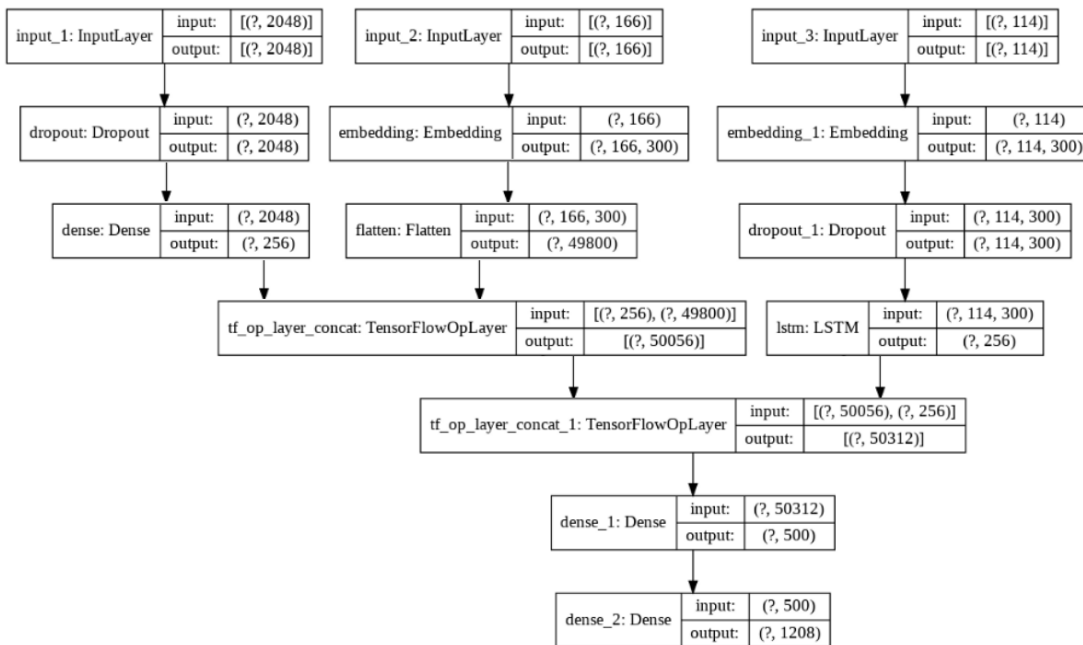


Figure 7.12: Architecture of the Automated Diagnostic Report Generation Model

Every image is converted into a fixed-sized vector which is then fed as input to the neural network. For this purpose, transfer learning is utilized by using the Xception model (Convolutional Neural Network) trained on the ImageNet dataset to perform image classification on 1,000 different classes of images. However, the purpose here is not to classify the image, but to get a fixed-length feature vector for each image. Hence, the final softmax layer is removed from the model and a 2048 length vector is extracted for every image. A total of 22,410 images are thus

available for training and testing the neural model.

Next, the text data from XML report files is subjected to text pre-processing. The dataset report contains everyday radiographic observations, which are listed as: Cardiomegaly, Pneumonia, Pulmonary edema, Lung opacity, Pleural effusion, Pneumothorax. For text data, some basic cleaning processes like lower-casing all the words, removing special tokens (like ‘%’, ‘\$’, ‘#’, etc.) are applied. In addition to this, erroneous data (“XXXX”, “X-XXX”) from *Impressions* and *Findings* features are also dealt with.

The proposed model consists of an encoder-decoder model, where image features are given to the encoder as input and as partial input to the decoder part that predicts the succeeding words in sequence. Two different approaches are experimented with – in the first one, only the image impressions are used, while in the second, the findings are also used. The text data is prepared by converting each sentence to integer sequences, for which a tokenization module is used. Integer sequences are padded to a fixed length so that all inputs will be of the same size. A weight matrix of the embedding layer in which every word (or index) is mapped (embedded) to a higher dimensional space (300-long vector) using a pre-trained GLOVE word embedding model. The neural network was trained with a batch size of 512 for 30 epochs using ReLu activation function and categorical cross entropy loss with adam optimizer.

7.5 Experiments Results and Discussion

For the experimental validation of the proposed methodology, the open-source dataset released by the Indiana University containing chest X-ray images for research purpose was used. The data is provided in 2 folders: one containing image files, a total of 7,470 chest x-ray images, both frontal and lateral. Another containing a total of 3,955 report files in XML format. Each report is related to one set of images (frontal and lateral), and contains fields like *Indication*, *Findings* and *Impression* (Refer Table 7.1).

Evaluation of the X-ray image enhancement methods using the image quality metrics like BRISQUE, PIQE and Perceptual Index are tabulated in Table 7.6 and a visual result on a sample set of three images is shown in Fig. 7.13. Many models use metrics like PSNR, SSIM when evaluating the quality of the super-resolution images. But here, the aim is to enhance the given image to a high-resolution space for better visualization, resolution change exists between the original and the computed image. So PSNR, SSIM cannot be used as evaluation metrics; hence

BRISQUE, PIQE and Perceptual Index metrics were used for quality evaluation. The BRISQUE score is usually in the range $[0 - 100]$ and lower values reflect better perceptual quality of the generated image with respect to the input image. For PIQE, values in the range of $[0 - 20]$ are considered *excellent* performance and $[21 - 35]$ is considered to be *good*. Similarly, lower values of PI indicate better perceptual quality in the generated image. From Table 7.6, it can be observed that the proposed ESRGAN achieved good performance in terms of all three metrics for both the *frontal* and *lateral* classes. To the best of our knowledge, no other works have attempted quality enhancement task on the Indiana dataset, hence benchmarking against state-of-the-art works could not be performed.

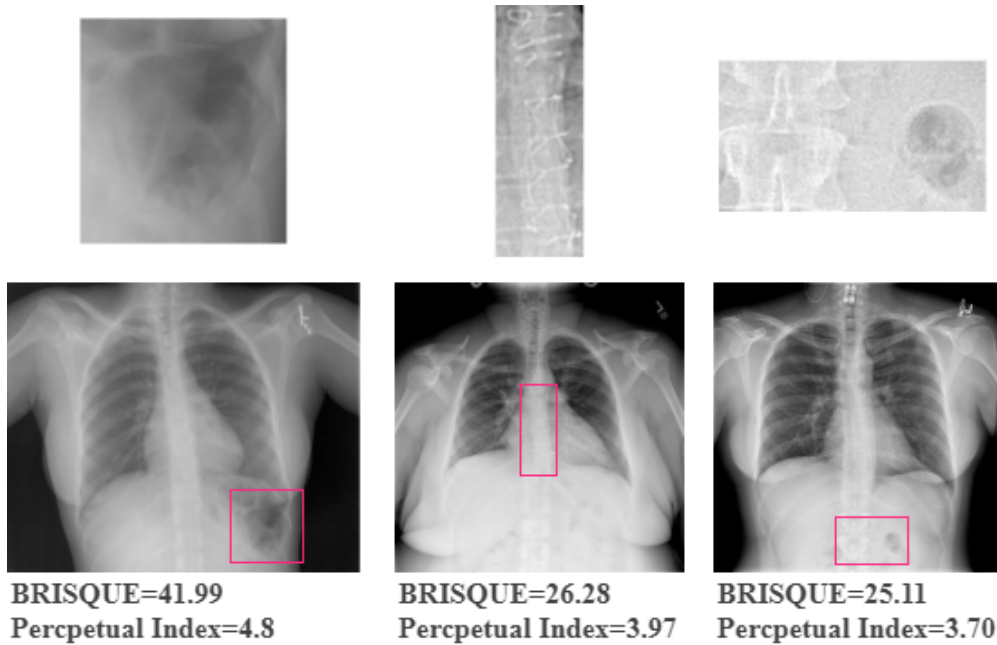


Figure 7.13: ESRGAN performance evaluation. (RoIs indicated using a box and the enhanced region generated by ESRGAN is shown in the first row)

Table 7.6: Performance of ESRGAN for the *Frontal* and *Lateral* classes

Class	BRISQUE	PIQE	Perceptual Index
Frontal	29.16	27.09	4.72
Lateral	36.90	32.37	5.98

For the body orientation view classification, the proposed *ViewNet* model's performance was measured using standard metrics like accuracy, sensitivity, specificity and F1-score. The classification results of different orientations/views (front view, lateral view) is depicted in Table 7.7. It is observed that the proposed

method shows best accuracy when compared to the state-of-the-art (Xue *et al.*, 2015). However, they have used only accuracy for their model, but we have also reported precision, recall and F1-score.

Table 7.7: Orientation classification performance with different CNN Models

NN Model	Accuracy	Precision	Recall	F1-Score
ViewNet without BN (proposed)	0.9866	0.9819	0.9922	0.9870
ViewNet with BN	0.9506	0.9631	0.9515	0.9665
AlexNet (Krizhevsky <i>et al.</i> , 2012)	0.9763	0.9779	0.9848	0.9863
ResNet18 (He <i>et al.</i> , 2016)	0.9643	0.9764	0.9686	0.9835
Image profile based (Xue <i>et al.</i> , 2015)	0.9840	-	-	-

For assessing the quality of the domain-specific text generated automatically by the proposed models, the BLEU score (BiLingual Evaluation Understudy) proposed by (Papineni *et al.*, 2002). It evaluates the similarity between a candidate document and a collection of reference documents. In short, BLEU is a metric used to evaluate a generated sentence in comparison to a reference sentence. As per an ordering from 1 to n , cumulative scores of individual n -grams can be calculated. n -gram is an evaluation of matching terms i.e., single word (1-gram), two word (2-gram or bigram) and so on. Weighing them together gives the geometric mean. In other words, for each i -gram where $i = 1, 2, 3 \dots N$, the percentage of i -gram tuples in the candidate document that also occur in the reference document represented as BLEU- i is given by Eq. (7.7), where, $C(i)$ is the number of i -gram tuples in the candidate document. In this work, suppose C = “the lungs are clear” then $C(1)=4$, $C(2)=3 \dots C(4)=1$. Here, (t_i) is an i -gram tuple in candidate C ; $H_c(t_i)$ is the number of times (t_i) occurs in the candidate; $H_{c_j}(t_i)$ is the number of times (t_i) occurs in reference j of this candidate.

$$BLEU - (i) = \frac{Matched(i)}{C(i)} \quad (7.7)$$

$$Matched(i) = \sum_{i=1} \min \{H_c(t_i), \max_j H_{c_j}(t_i)\} \quad (7.8)$$

The proposed model is a pipeline of models combining all three tasks: the

image enhancement evaluation shows that the BRISQUE and PIQE give the best performance with good scores subjective to human visual perceptions. During analysis, it is noticed that the image quality is enhanced by more than 4 times the original size (from Fig. 7.8, if $M = N=256$, then $P = Q=1024$), thus maintaining an excellent spatial resolution subjective to the human visual system. In view of the orientation classification model, the proposed MSDNet CNN architecture outperformed the state-of-the-art methods based on feature-based techniques (Xue *et al.*, 2015) and other CNN models. A combination of AlexNet and Resnet-18 layers have contributed more to achieve this higher accuracy score. Also, the removal of batch normalization layers helped train the model in lesser time, thus reducing the computational time requirement of the proposed approach.

The proposed model performed well for the report generation task also, and achieved good BLEU scores, which emphasizes that the text report generated is most accurate. Observing the performance of the proposed model in terms of BLEU values obtained for 1-gram to 4-gram approaches, the score obtained is almost at the same level in the case of all 4-grams, indicating that the evaluation of matching grams between the candidate and reference is nearly a perfect match. Most available models underperform when 3-gram and 4-gram matches are used to compute BLEU scores. However, the proposed model showed very good performance, while matching the candidate sentences with reference text, even in case of 3-grams and 4-grams matching (an example of which is shown Fig. 7.14), indicating a near-perfect match. Further, these experiments also revealed the effect of the Batch Normalization (BN) layers while training the network.

REFERENCE:- The lungs are clear bilaterally. Specifically, no evidence of focal consolidation, pneumothorax, or pleural effusion. Minimal right basilar subsegmental atelectasis noted. Cardio mediastinal silhouette is unremarkable. Tortuosity of the thoracic aorta noted. Scattered calcified granulomas are seen without evidence of active granulomatous/tuberculous process. Visualized osseous structures of the thorax are without acute abnormality.

MODEL GENERATED/CANDIDATE:-

1-GRAM MATCH 3-GRAM MATCH

2-GRAM MATCH 4-GRAM MATCH

Specifically, no evidence of focal consolidation, pneumothorax, or pleural effusion. Stable small right basilar calcified granuloma. Cardio mediastinal silhouette is unremarkable. Visualized osseous structures of the thorax are without acute abnormality.

Figure 7.14: Example of BLEU match with n-gram approach.

7.5.1 Ablation Study

To understand the effects of certain components employed in the proposed multi-task deep neural model pipeline, we performed additional experiments in the form of an ablation study. Each component was considered during experimentation to observe its particular effects on the model's performance. A visual interpretation of our observations are demonstrated in Fig. 7.15. Each column represents the model with a change in its configuration. A column with green indicates that an improvement is observed to its previous model. There are two key components here, which we used while evaluating model performance – the first part is removing the batch normalization layer; another is using the features before activation. Removing the BN layers improved the model's performance significantly, with the added advantages of lower computational resource consumption and reduction in memory usage. Another observation noted is that, in the resulting image, features when used after the activation lead to an indistinct image, but using features before activation brings out a clear, sharp-edged and brightened image (See columns no. 2 and 4).

Further, it was observed that, without the use of BN layer, the ESRGAN model's performance increased, which also conforms to the observations reported by (Nah *et al.*, 2017) in their work. During training, the batch normalization process normalizes features based on the mean and variance in each batch. Due to this, the features do not have significant contribution when BN layer is employed, thus the best performance is observed when the BN layer is removed. Thus, to further study the completeness of the overall system architecture, i.e., the proposed MSDNet model for the body orientation classification task and in the report generation model, we further experimented with and without the use of the BN layer, to test the model's performance. As can be seen from the results of these experiments reported in Table 7.7 and 7.8, there is a clear improvement in the model's performance.

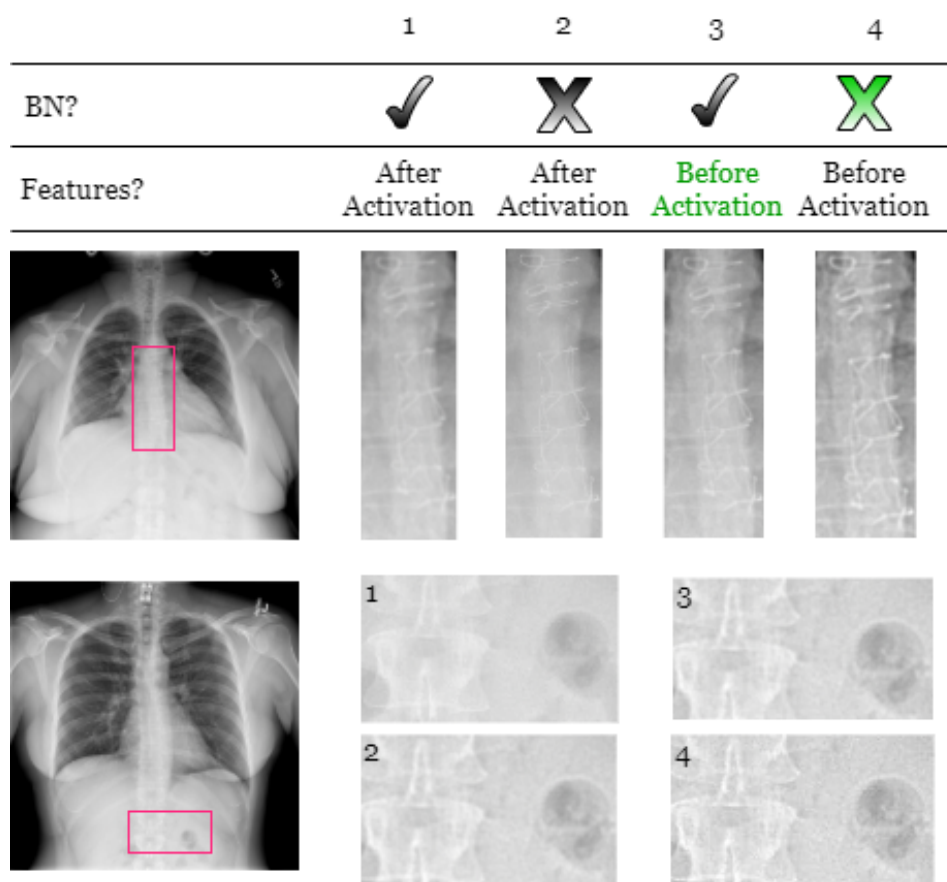


Figure 7.15: Visual Comparison representing the outcome of each component in ESRGAN. (Each column represents a model's output with its configurations mentioned at the top (first and second row). The green sign indicates an improvement compared to its previous model.)

Table 7.8: Comparison of Report generation performance with and without BN layer.

Method	BLEU1	BLEU2	BLEU3	BLEU4
Proposed + BN Layer	0.9502	0.9460	0.9414	0.9265
Proposed (without BN layer)	0.9735	0.9679	0.9650	0.9500
CNN-RNN (Vinyals <i>et al.</i> , 2015)	0.333	0.205	0.136	0.094
CoAtt (Jing <i>et al.</i> , 2017)	0.455	0.288	0.205	0.154
HLSTM+att+Dual (Harzig <i>et al.</i> , 2019)	0.469	0.335	0.249	0.183
KERP (Li <i>et al.</i> , 2019)	0.482	0.325	0.226	0.162

7.6 Learning COVID-19 Disease Representations from Multimodal Data

This task aims to alleviate the cognitive burden of radiologists and clinicians for segregating actual COVID-19 cases from other types of Shortness of Breath (SoB) cases. Multiple deep neural models are experimented with for chest X-ray classification and automatic diagnosis report generation tasks. The classification task aims to distinguish between actual COVID-19 and other types of SoB cases. It also uses X-ray images and diagnostic reports (English natural language texts) to model the data. For classifying the X-ray reports, various machine learning algorithms were adopted for observing the performance. Finally, the task of automatic generation of reports for the input X-ray images is addressed.

7.6.1 Chest X-ray based Screening of COVID-19

This phase aims to analyze available chest X-ray images for automated identification of those with COVID-19 infections indications and those belonging to other types of lung-related diseases that may cause shortness of breath, like pneumonia, lung inflammation, and enlarged lymph nodes. Four different deep neural models (AlexNet, DenseNet-201, Inception v3 and Resnet-18) are experimented for classifying the chest X-ray images into COVID-19 and Shortness of Breath (SoB) cases. Certain changes are made in the base neural models to adapt to the given task. For AlexNet, the final three layers are replaced with a fully connected layer, a softmax layer, and a classification output layer, whereas, in ResNet-18, DenseNet-201 and Inception v3, the last learning layer and the final classification layer are replaced with global average pooling, fully connected, softmax and classification output layers for accurate classification of COVID-19 instances. The learning rate factor was also heuristically set to enable CNN to learn the disease-specific features more effectively during the transfer learning phase. The network hyper-parameters are chosen based on experiments and are shown in Table 7.9.

7.6.2 Automatic Diagnostic Report Generation

For this task, the modeled features extracted from x-ray images and expert-written diagnosis reports were used for automatically generating the reports of identified COVID-19 patients. The features are extracted using the models that are previously trained for classification task (AlexNet, Inception v3 and ResNet-18).

Table 7.9: Classification model parameters.

NN Model	Learning Rate	Weight/ Bias	Batch Size	Epochs
AlexNet	0.0001	20	8	10
DenseNet-201	0.0001	10	8	10
Inception v3	0.001	8	10	10
ResNet-18	0.001	10	8	8

The features are extracted from the last average pooling layer (before the first Dense layer) for all the networks other than DenseNet. DenseNet does not sum up the output feature maps of the layer with the incoming features; instead, these are concatenated. Thus, its equation is $x_l = H_l([x_0, x_1, \dots, x_{l-1}])$, where, l is the index of each layer, H is the non-linear operation, x_0, x_1, \dots, x_{l-1} are the feature values from each layer and x_l is the output of the l^{th} layer. The dimensions of the feature maps remain constant within a DenseBlock, but the number of filters varies between them. As all the feature maps are concatenated, combining them with different sizes would thus be impractical. Hence, a transition layer is present between two DenseBlock, which carries out down-sampling by applying a batch normalization, a 1x1 convolution and a 2x2 pooling layer. The growth rate k which normalizes how much information is added to the network at each layer as per $k_l = k_0 + k \times (l - 1)$. Every layer adds its information, making it a collective knowledge. Deeper layers produce the higher-level features, constructed using the lower-level features of earlier layers. To get the feature representations of the training and test images, activation on the global average pooling is applied, 'avg pool' is used at the last layer of the network, giving 1024 features in total.

During the feature extraction process, each image in the training and test sets is used for creating a feature vector. When a test image feature set is given as a query, the pairwise distance measure is used to compute distances that can be used to obtain matching feature sets with the smallest distance from the images in the training set. Initially, eight different distance measures, Cosine, Correlation, Cityblock, Euclidean, Spearman, Minkowski, Standard Euclidean, and Chebychev, are used to check the closest distance measure among the test and training feature sets. For each observation in Y (*Test image features*), the pairwise distance method finds the smallest distances by computing and comparing the distance values to all the observations in X (*Training image features*). Based on experimentation performance, Cosine and Standard Euclidean similarity measures performed best.

7.7 Experimental Results & Discussion

The dataset consists of 200 cases collected from publicly available dataset created by [Cohen *et al.* \(2020\)](#). For curating the dataset, a total of 100 confirmed COVID-19 patient cases were collected from publicly available open access sources. Each patient report also consisted of additional background details like age, gender, clinical history and image findings. For cases where patients suffered from shortness of breath (SoB), a set of scan images along with their expert-generated diagnosis descriptions available from the IU dataset provided by Indiana University ([Demner-Fushman *et al.*, 2016](#)) were considered. Here too, a total of 100 cases are used to put together a balanced dataset containing an equal number of COVID-19 and SoB cases. Combining these two datasets, a total of 200 scan images along with their descriptions is considered for this research work.

The dataset is split as per the 70:30 ratio, i.e., 70% of the input records are used for training, and the remaining 30% of the records are utilized as a testset. For assessing the performance of the proposed models, standard metrics like accuracy, sensitivity, specificity and F1-score are utilized. These metrics are calculated based on the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) cases predicted by a particular neural model. Here, TP is the number of cases that are correctly identified by the prediction model to be COVID-19 positives, which matches with experts' opinion, while FN are incorrectly rejected cases. TN is the number of correctly identified non-COVID-19 cases, and FP is the number of incorrectly identified COVID-19 cases. Sensitivity or Recall (also called True Positive Rate) is a measure of the percentage of correctly identified COVID-19 cases. Higher the value, better is the system's prediction performance. Specificity provides the percentage of correctly identified non COVID-19 cases, and higher values indicate a good prediction performance. F1-score is the harmonic mean of precision and recall, and high values indicate a balanced performance by the model. Finally, accuracy is the ratio of correctly predicted COVID-19 cases to the total number of cases.

7.7.1 Chest X-ray Classification for COVID-19 Diagnosis.

The results of the experiments and the performance achieved by the various models when applied to the test X-ray images are shown in [Table 7.10](#). DenseNet-201 achieved the best overall accuracy on correctly classifying both COVID-19 and SoB cases. In the case of COVID-19, Inception v3 performed well by properly predicting all the test cases as COVID-19. Thus, the sensitivity is 100%. Also, it is

noted that the features extracted using AlexNet and ResNet-18 models contributed greatly towards the report generation task.

Table 7.10: Performance evaluation of the chest X-ray image classification task

NN Model	Accuracy	Sensitivity	Specificity	F1-Score
AlexNet	0.8095	0.9333	0.7368	0.8235
ResNet-18	0.8730	0.9091	0.8333	0.8621
DenseNet-201	0.9048	0.9000	0.8788	0.8850
Inception v3	0.8889	1.0	0.7667	0.8679

7.7.2 Automated chest X-ray text report generation task.

For assessing the quality of the domain-specific text generated, we used the BLEU score (BiLingual Evaluation Understudy) proposed by Papineni *et al.* (2002). It evaluates the similarity between a candidate document and a collection of reference documents. As per an ordering from 1 to n , cumulative scores of individual n-grams can be calculated. N-gram is an evaluation of matching terms i.e., single word (1-gram), two word (2-gram or bigram) and so on. Weighing them together gives the geometric mean. In other words, for each i -gram where $i = 1, 2, 3 \dots N$, the percentage of i -gram tuples in the candidate document that also occur in the reference document represented as BLEU- i is given by Eq. (7.7), section 7.5. For calculating the BLEU score, each report in the training dataset is considered as a potential candidate, and the report generated by each ML/DL model for each test image is taken as a reference. The number of distinct sentences generated for the whole set (training and testing cases) is calculated separately. References are then evaluated with each of the different candidate sets. Finally, the model with the highest BLEU-4 score is depicted in Table 7.11.

Table 7.11: Performance evaluation of chest X-ray report generation

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	Mean
AlexNet	0.9148	0.9008	0.6848	0.6312	0.5484
ResNet-18	0.9125	0.8872	0.6251	0.6211	0.5171
DenseNet-201	0.8879	0.7790	0.6184	0.5867	0.5006
Inception v3	0.8984	0.7008	0.6304	0.5637	0.4896

After several experiments, the best values for the hyper-parameters of the proposed models are heuristically determined. As can be seen from the tabulated values (Refer Table 7.10), the best overall accuracy is achieved with DenseNet-201 with its 90.48% success rate in correctly classifying both COVID-19 and SoB images. This may be due to the feature values concatenated from each layer up to the final layer, in contrast to the other DL models, which sum up the feature values. This concatenation enables DenseNet to model the disease-specific feature maps for the most accurate classification with reference to both classes. In the case of COVID-19, the Inception v3 model attained the highest peak sensitivity value in finely predicting the COVID-19 cases. Due to a rapid increase in the number of the infected patient and suspected cases, scalability of any new diagnosis procedures is of significant interest to medical professionals. To address this challenge, content based techniques were explored to automatically generate diagnosis reports for an input chest X-ray image, and the proposed models achieved good results (as discussed in Section 7.7). From Table 7.11, it is evident that AlexNet features contributed more in matching the retrieved text with that of the reference text reports in the proposed content based report generation task.

7.8 Summary

In this chapter, an ensemble deep neural model called *MSDNet*, that combines global and local features for the clinical task of radiological abnormality detection and classification is detailed. The model is built on an abnormal region detection algorithm which is used for identifying the anomalous regions in the image. The proposed model achieved an accuracy of 82.69%, along with promising sensitivity, specificity and Kappa statistic value of 0.746, indicating good performance. The overall computation time was also reduced during training, as the proposed model uses a comparatively shallow architecture compared to the other state-of-the-art models. These models used a much deeper architecture; however, the proposed model outperformed them being less computationally expensive due to shorter training time. An automated diagnostic text report generation algorithm was also designed, which further extended the proposed model pipeline for the identified abnormal images, to alleviate radiologists' cognitive burden and improve the overall efficiency of the diagnosis.

In the next work, multiple tasks like image enhancement, orientation classification, and report generation were attempted for enabling a practical framework for chest X-ray image management. An efficient image enhancement technique

called ESRGAN was presented for enhancing medical X-ray image quality and RoI visualization. The quality enhancement helps to increase the image resolution by a factor of 4, and this dramatically improves the diagnosis effectiveness due to enhanced visualization of scanned body part's overall structure. To the best of our knowledge, we are the first to attempt image enhancement using ESRGAN techniques for the Indiana dataset. Next, the proposed MSDNet model for automatically detecting the scan view and predicting the orientation label of the image is discussed. The input images are classified as per their orientation and this orientation label gives more information about the image in which orientation the image was captured. The enhanced image is then utilized to automate the generation of findings/observation report, thus alleviating the cognitive load of radiologists. The proposed model was based on the Xception model.

The proposed models were experimentally validated with standard metrics suitable for the individual sub-task. The quality enhancement model gained a better quality score in terms of visual perceptions; while a lower score indicates a better perception of the human visual system. View orientation classification showed promising results with over 98.40% accuracy, outperforming state-of-the-art methods. Further, the text report generation model attained a 0.9735 BLEU score, which signifies that the exceptional performance in generating accurate diagnosis reports automatically. In another work, the automated report generation task was an attempt at reducing the cognitive burden of medical professionals, given the rapid increase in COVID-19 cases. For this task, the features extracted by DL models were employed to generate diagnosis reports for test X-ray images.

Publications

(based on work presented in this chapter)

1. Karthik K., Sowmya Kamath S., “*MSDNet: A Deep Neural Ensemble Model for Abnormality Detection and Classification from Plain Radiographs*”, Journal of Ambient Intelligence and Humanized Computing (JAIHC), Springer Nature DOI: 10.1007/s12652-022-03835-8 [SCIE & Scopus, IF: 7.104] (*Status: Online*)
2. Karthik K., Sowmya Kamath S., “*Deep Neural Models for Automated Multi-task Diagnostic Scan Management - Quality Enhancement, View Classification and Report Generation*”, International Journal of Biomedical Physics

and Engineering Express, IOP, DOI: 10.1088/2057-1976/ac3add [ESCI & Scopus] (*Status: Online*)

3. Mayya V*, Karthik K*, Karadka K., and Kamath S., “*Multi-task Deep CNN Models for Learning COVID-19 Disease Representations from Multimodal Data*”, Intl. Journal of Medical Engineering and Informatics (IJMEI), Inderscience [Scopus] (*Status: In Press*)

*Equal contribution

Chapter 8

Conclusion & Future Work

Medical imaging technologies are an integral part of the healthcare ecosystem, facilitating non-invasive diagnostic capabilities over a wide variety of modalities. Several challenges are prevalent despite major advancements in diagnostic imaging systems, specifically in the areas of automated quality management, preprocessing, modeling, representation, categorization, retrieval and others. Other contributing factors include the sheer volume and the streaming nature of medical scan image generation, which make diagnostic image management a very challenging task. With the availability of high computational power and advanced AI algorithms, the advent of computer-aided medical diagnostics and decision support systems in clinical environments is inevitable. The efficiency of the healthcare processes (e.g., diagnosis, prognosis, and screening) can be enhanced by using computational intelligence and predictive analytics applications, that leverage large-scale diagnostic data for generating actionable insights. The work presented in this thesis focused on addressing several highlighted challenges, for improving diagnostic accuracy through effective management of diagnostic image data with intelligent AI models, for enabling various clinical tasks.

The initial work focuses on identifying the medical image scan quality issues and improving the quality of the image. The first task involved five image super-resolution algorithms - Unsharp mask using Gaussian filter, CLAHE, Bicubic Interpolation, VDSR and SRCNN were implemented for image quality enhancement and evaluated for better visualization of X-ray images. Experiments were performed to comparatively evaluate these 5 approaches. Based on the visualized enhanced images, it was observed that the processed image captured hidden features well through edge and contrast enhancement, in turn amplifying the visibility of regions of interest. A patch size of 16 pixels (4×4) in Bicubic interpolation resulted in a smoother image, while VDSR showed better performance while transforming

a LR image to HR. SRCNN outperformed all other methods due to its lightweight architecture and superior learning behavior. In another work, five CNN based models were experimented with for efficient medical image enhancement. An ensemble model, ResNetSRCNN, was designed which showed good performance with reference to standard visual quality metrics and outperformed state-of-the-art models by a large margin. Additionally, an efficient image enhancement model called ESRGAN was developed, for enhancing medical X-ray image quality and RoI visualization. The model was able to achieve a better quality score in terms of visual perception, as a lower score indicates a better perception of the human visual system.

The second phase of the work addressed effective feature representation and modeling of medical images. A hybrid feature modeling approach developed for content-based medical image retrieval showed a promising result. Most CBMIR models are restricted to a particular class or modality, however, the experiments were conducted on the large-scale ImageCLEF 2009 dataset consisting of X-ray images spanning 116 classes. The experimental results showed that the proposed approach was very suitable for real-world medical image retrieval applications used for disease diagnosis and decision support, due to its excellent top-3 and top-5 retrieval performance. In the next approach, a PSO enhanced CBMIR approach built on the Bag of Visual Words Model was presented. The SURF algorithm was used for generation of features from the medical images, which were represented as BOF to classify and retrieve the medical X-ray images. PSO was incorporated to optimize retrieval performance for a given query image. PSO was used to gain insights into the optimal clustering value. Further, a filtering approach was designed to obtain best matches. Results showed that the filter approach achieved 100% precision when used for top-10 retrieval for given test images. Additionally, a CNN based model was designed for classification of medical images, the results of which are used for supporting similar image retrieval. By using CNN's feature extraction and with similarity distance calculation between the feature vectors, it was observed that the model achieved good retrieval results. Another significant observation was that, the model was able to retrieve similar images even for classes where X-ray images had different body orientations, underscoring its ability to learn features well despite data variance.

Four types of transfer learning-based neural network models have been experimented with, for body orientations consisting of different orientation view classification tasks on the standard open dataset, ImageCLEF 2009. A novel architecture, *ViewNet* was also proposed for the task of view classification and achieved

promising results when measured in terms of accuracy, sensitivity, specificity and F1-score. Further, an ensemble deep neural model called as *MSDNet*, that combines global and local features for the clinical task of radiological abnormality detection and classification was progressed. The model is built on an abnormal region detection algorithm which is used for identifying the anomalous regions in the image. The overall computation time was also reduced during training, as the proposed model has a comparatively shallow architecture compared to the other state-of-the-art models. An automatic text report generation model was developed which signifies its exceptional performance in generating accurate diagnosis reports. The text report generation model attained a 0.9735 BLEU score, which signifies its exceptional performance in generating accurate diagnosis report.

Another work, dealing with the problem of diagnosing COVID-19 using multi-modal patient data was also presented. The experiments was performed on a collated dataset consisting of 210 images and the associated expert-written diagnosis reports of 100 unique patient cases for COVID-19 and SoB. Chest X-ray image classification using neural network models was experimented with, and it was observed that DenseNet-201 achieved the best overall accuracy on correctly classifying both COVID-19 and SoB cases. In the case of COVID-19, Inception v3 performed well by correctly predicting all relevant test cases as COVID-19, achieving a sensitivity of 100%. Also, it is noted that the features extracted using AlexNet and ResNet-18 models contributed greatly towards the report generation task. The automated report generation task was an attempt at reducing the cognitive burden of medical professionals, given the rapid increase in COVID-19 cases. For this task, the features extracted by DL models were employed to generate diagnosis reports for test X-ray images.

8.1 Future Work

In medical image analysis, useful information is not just contained within the images itself, it may be required to observe other contents too. To get a clear decision on the patient's health, physicians often hold a wealth of data on patient history, age, and other health issues¹. Physicians often also need to use anatomical information to come to an accurate diagnosis. Some researchers have already explored combining this information into deep learning networks in a straightforward manner (Su and Liu, 2018; Harzig *et al.*, 2019; Estiri *et al.*, 2021). However,

¹<https://www.nia.nih.gov/health/obtaining-older-patients-medical-history>

it is observed that the improvements that were obtained were not as significant as expected. One of the challenges is balancing the number of imaging features in the deep learning network with the number of clinical features to prevent the clinical features from obscuring. However, many deep learning systems in medical imaging are still based on classification, where the anatomical location is often unknown to the network.

The works that are developed can be extended for other diagnostic imaging modalities like MRI and CT, and super-resolution models can be adapted for enhancing the quality of such multi-dimensional data also. Identifying effective shape-based algorithms and incorporating it with a deep neural network to solve body orientation changes will also be explored as an extension of the current work. Further, segmenting the image or considering only the region of interest and feeding that to a neural network can also further enhance the accuracy, which we intend to experiment and benchmark. It is also crucial to design real-time applications based on the novel architectures designed as part of this work to extend support to the medical personnel, which still balances the need for achieving good computation time and good performance.

Publications based on Research Work

Journal Publications

1. Karthik K., Sowmya Kamath S., “*MSDNet: A Deep Neural Ensemble Model for Abnormality Detection and Classification from Plain Radiographs*”, Journal of Ambient Intelligence and Humanized Computing (JAIHC), Springer Nature DOI: 10.1007/s12652-022-03835-8 [SCIE, IF: 7.104] (*Status: Online*)
2. Karthik K., Sowmya Kamath S., “*A Deep Neural Network Model for Content-Based Medical Image Retrieval with Multi-View Classification*”, The Visual Computer Journal (TVCJ), Springer Nature, DOI: 10.1007/s00371-020-01941-2 [SCIE, IF: 2.601] (*Status: Online*)
3. Karthik K., Sowmya Kamath S., “*Deep Neural Models for Automated Multi-task Diagnostic Scan Management - Quality Enhancement, View Classification and Report Generation*”, International Journal of Biomedical Physics and Engineering Express, IOP, DOI: 10.1088/2057-1976/ac3add [ESCI & Scopus] (*Status: Online*)
4. Karthik K., Sowmya Kamath S., “*Swarm Optimization Based Bag of Visual Words Model for Content-Based X-Ray Scan Retrieval*”, Intl. Journal of Biomedical Engineering and Technology (IJBET), Inderscience. [ESCI & Scopus] (*Status: Abstract online, Article in press*)
5. Mayya V*, Karthik K*, Karadka K., and Kamath S., “*Multi-task Deep CNN Models for Learning COVID-19 Disease Representations from Multimodal Data*”, Intl. Journal of Medical Engineering and Informatics (IJMEI), Inderscience [Scopus] (*Status: In Press*)

*Equal contribution

Conference Publications

1. Karthik, K. and Sowmya Kamath S., “*A Hybrid Feature Modeling Approach for Content-Based Medical Image Retrieval*”, 13th International Conference on Industrial and Information Systems (ICIIS 2018). IEEE, IIT Ropar, Punjab [CORE Ranked] (*Status: Online*)
2. Karthik K., Sowmya Kamath S. and Surendra U. Kamath, “*Automatic Quality Enhancement of Medical Diagnostic Scans with Deep Neural Image Super-Resolution Models*”, 15th International Conference on Industrial and Information Systems (ICIIS 2020). IEEE, IIT Ropar, Punjab [CORE Ranked] (*Status: Online*)
3. Mayya V, Karthik K., Kamath S S., Karadka K, and Jeganathan, J. “*COVID-DX: AI-based clinical decision support system for learning COVID-19 disease representations from multimodal patient data*”, 14th International Conference on Healthcare Informatics (HEALTHINF 2021), Feb 11-13, 2021, pages 659–666. [CORE Ranked] (*Status: Online*)
4. Karthik K. and Sowmya Kamath S., “*Improving Clinical Diagnosis Performance with Automated X-ray Scan Quality Enhancement Algorithms*”, International Conference on Advances in Systems, Control and Computing (AISCC 2020), MNIT Jaipur (Springer). (*Status: In press*)

Invited Book Chapters

1. Karthik K. and Sowmya Kamath S., “*Automated View Orientation Classification for X-ray images using Deep Neural Networks*”, Smart Computational Intelligence in Biomedical and Health Informatics, CRC Press, Taylor & Francis UK, 2021, DOI: 10.1201/9781003109327, ISBN: 9781000434378 (*Status: Online*)

References

- Abdullah, S., O. Arif, M. B. Arif, and T. Mahmood (2019). Mri reconstruction from sparse k-space data using low dimensional manifold model. *IEEE Access*, 7, 88072–88081.
- Aggarwal, P., H. Sardana, and R. Vig, Content-based medical image retrieval using patient’s semantics with proven pathology for lung cancer diagnosis. *In Fifth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2013)*. IET, 2013.
- Ahmad, J., K. Muhammad, and S. W. Baik (2018). Medical image retrieval with compact binary codes generated in frequency domain using highly reactive convolutional features. *Journal of medical systems*, 42(2), 24.
- Ahmed, N., W. Ahmed, and S. M. Arshad (2011). Digital radiographic image enhancement for improved visualization. *proceedings COMSATS Institute of Information Technology*.
- Antani, S. K., L. R. Long, and G. R. Thoma, A biomedical information system for combined content-based retrieval of spine x-ray images, associated text information. *In ICVGIP*. 2002.
- Anthimopoulos, M., S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE transactions on medical imaging*, 35(5), 1207–1216.
- Arimura, H., S. Katsuragawa, Q. Li, T. Ishida, and K. Doi (2002). Development of a computerized method for identifying the posteroanterior and lateral views of chest radiographs by use of a template matching technique. *Medical Physics*, 29(7), 1556–1561.

- Avni, U., J. Goldberger, and H. Greenspan, Addressing the imageclef 2009 challenge using a patch-based visual words representation. *In CLEF (Working Notes)*. 2009.
- Avni, U., H. Greenspan, E. Konen, M. Sharon, and J. Goldberger (2010). X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE transactions on medical imaging*, 30(3), 733–746.
- AyushmanBharat (2018). Press Information Bureau, Government of India. Ayushman Bharat for a New India-2022; 2018. <http://www.pib.nic.in/newsite/PrintRelease.aspx?relid=176049>. [Online; accessed March 2021].
- Banga, D. and P. Waiganjo (2019). Abnormality detection in musculoskeletal radiographs with convolutional neural networks (ensembles) and performance optimization. *arXiv preprint arXiv:1908.02170*.
- Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3), 346–359.
- Bay, H., T. Tuytelaars, and L. Van Gool, Surf: Speeded up robust features. *In European conference on computer vision*. Springer, 2006.
- Berman, P., R. Ahuja, and L. Bhandari (2010). The impoverishing effect of health-care payments in india: new methodology and findings. *Economic and Political Weekly*, 65–71.
- Bertrand, H., M. Hashir, and J. P. Cohen (2019). Do lateral views help automated chest x-ray predictions? *arXiv preprint arXiv:1904.08534*.
- Bhairannawar, S. S., Efficient medical image enhancement technique using transform hsv space and adaptive histogram equalization. *In Soft Computing Based Medical Image Analysis*. Elsevier, 2018, 51–60.
- Binh, N. and H. Tuyet, Improving quality of medical images in shearlet domain. *In Intl. Conf. on Advanced Technologies for Communications*. IEEE, 2015.
- Boita, J., R. E. van Engen, A. Mackenzie, A. Tingberg, H. Bosmans, A. Bolejko, S. Zackrisson, M. G. Wallis, D. M. Ikeda, C. Van Ongeval, *et al.* (2021). How does image quality affect radiologists’ perceived ability for image interpretation and lesion detection in digital mammography? *European radiology*, 31(7), 5335–5343.

- Boone, J. M., G. S. Hurlock, J. A. Seibert, and R. L. Kennedy (2003). Automated recognition of lateral from pa chest radiographs: saving seconds in a pacs environment. *Journal of Digital Imaging*, 16(4), 345–349.
- Britnell, M., *In search of the perfect health system*. Macmillan International Higher Education, 2015.
- Bustos, A., A. Pertusa, J.-M. Salinas, and M. de la Iglesia-Vayá (2020). Padchest: A large chest x-ray image dataset with multi-label annotated reports. *Medical image analysis*, 66, 101797.
- Camlica, Z., H. R. Tizhoosh, and F. Khalvati, Autoencoding the retrieval relevance of medical images. *In 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2015.
- Chada, G. (2019). Machine learning models for abnormality detection in musculoskeletal radiographs. *Reports—Medical Cases, Images, and Videos*, 2(4), 26.
- Choplin, R. H. (1992). Picture archiving and communication systems: an overview. *Radiographics*, 12(1), 127–129.
- Cohen, J. P., P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi (2020). COVID-19 image data collection: Prospective predictions are the future. *arXiv 2006.11988*.
- Cover, T. and P. Hart (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21–27.
- Dalal, N. and B. Triggs, Histograms of oriented gradients for human detection. *In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on volume1*. IEEE, 2005.
- Datta, R., D. Joshi, J. Li, and J. Z. Wang (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)*, 40(2), 5.
- de Herrera, A. G. S., C. Eickhoff, V. Andrearczyk, and H. Müller (2018). Overview of the imageclef 2018 caption prediction tasks. *CLEF working notes, CEUR*.
- Demner-Fushman, D., M. Kohli, M. Rosenman, S. E. Shooshan, L. M. Rodriguez, S. Antani, G. Thoma, and C. McDonald (2016). Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association : JAMIA*, 23(2), 304–10.

- Demner Fushman, D., M. D. Kohli, M. B. Rosenman, S. E. Shooshan, L. Rodriguez, S. Antani, G. R. Thoma, and C. J. McDonald (2016). Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2), 304–310.
- Dimitrovski, I., D. Kocev, S. Loskovska, and S. Džeroski (2011). Hierarchical annotation of medical images. *Pattern Recognition*, 44(10-11), 2436–2449.
- Dong, C., C. C. Loy, K. He, and X. Tang, Learning a deep convolutional network for image super-resolution. *In European conference on computer vision*. Springer, 2014.
- Dong, C., C. C. Loy, K. He, and X. Tang (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2), 295–307.
- Elliott, D. and F. Keller, Image description using visual dependency representations. *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 2013.
- Estiri, H., Z. H. Strasser, J. G. Klann, P. Naseri, K. B. Waghlikar, and S. N. Murphy (2021). Predicting covid-19 mortality with electronic medical records. *NPJ digital medicine*, 4(1), 1–10.
- Faes, L., S. K. Wagner, D. J. Fu, X. Liu, E. Korot, J. R. Ledsam, T. Back, R. Chopra, N. Pontikos, C. Kern, G. Moraes, M. K. Schmid, D. Sim, K. Balaskas, L. M. Bachmann, A. K. Denniston, and P. A. Keane (2019). Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *The Lancet Digital Health*, 1(5), e232–e242.
- Feng, J., N. Xiong, and B. Shuoben, X-ray image enhancement based on wavelet transform. *In Asia-Pacific Services Computing Conference, 2008. APSCC'08. IEEE*. IEEE, 2008.
- Fesharaki, N. J. and H. Pourghassem, Medical x-ray images classification based on shape features and bayesian rule. *In Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on*. IEEE, 2012.

- Gao, Y., H. Li, J. Dong, and G. Feng, A deep convolutional network for medical image super-resolution. *In Chinese Automation Congress (CAC), 2017*. IEEE, 2017.
- García-Floriano, A., Á. Ferreira-Santiago, O. Camacho-Nieto, and C. Yáñez-Márquez (2019). A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images. *Computers & Electrical Engineering*, 75, 218–229.
- Georgieva, V., R. Kountchev, and I. Draganov, *An Adaptive Enhancement of X-Ray Images*. Springer International Publishing, Heidelberg, 2013, 79–88.
- Getto, R., A. Kuijper, and T. von Landesberger (2015). Extended surface distance for local evaluation of 3d medical image segmentations. *The Visual Computer*, 31(6-8), 989–999.
- Guillaumin, M., T. Mensink, J. Verbeek, and C. Schmid, Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. *In Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009.
- Ha, V. K., J. Ren, X. Xu, S. Zhao, G. Xie, and V. M. Vargas, Deep learning based single image super-resolution: A survey. *In International Conference on Brain Inspired Cognitive Systems*. Springer, 2018.
- Haralick, R. *et al.* (1973). Textural features for image classification. *IEEE Transactions on systems, man & cybernetics*, (6), 610–621.
- Harzig, P., Y. Y. Chen, F. Chen, and R. Lienhart (2019). Addressing data bias problems for chest x-ray image report generation. *arXiv preprint arXiv:1908.02123*.
- He, K., X. Zhang, S. Ren, and J. Sun, Identity mappings in deep residual networks. *In European conference on computer vision*. Springer, 2016.
- Hillman, B. J., E. S. Amis Jr, and H. L. Neiman (2004). The future quality and safety of medical imaging: proceedings of the third annual acr forum. *Journal of the American College of Radiology*, 1(1), 33–39.
- Huang, G., 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*. 2017.

- Huang, R. Y., L. R. Dung, C. F. Chu, and Y. Y. Wu (2016). Noise removal and contrast enhancement for x-ray images. *Journal of Biomedical Engineering and Medical Imaging*, 3(1), 56.
- Hwang, H. and R. A. Haddad (1995). Adaptive median filters: new algorithms and results. *IEEE Transactions on image processing*, 4(4), 499–502.
- Iakovidis, D. K., N. Pelekis, E. E. Kotsifakos, I. Kopanakis, H. Karanikas, and Y. Theodoridis (2009). A pattern similarity scheme for medical image retrieval. *IEEE Transactions on Information Technology in Biomedicine*, 13(4), 442–450.
- Iandola, F. N., S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer (2016). Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. *arXiv preprint arXiv:1602.07360*.
- Ilunga-Mbuyamba, E., J. G. Avina-Cervantes, D. Lindner, J. Guerrero-Turrubiates, and C. Chalopin, Automatic brain tumor tissue detection based on hierarchical centroid shape descriptor in t1-weighted mr images. In *2016 International Conference on electronics, communications and computers (CONI-ELECOMP)*. IEEE, 2016.
- Isaac, J. S. and R. Kulkarni, Super resolution techniques for medical image processing. In *Technologies for Sustainable Development (ICTSD), 2015 International Conference on*. IEEE, 2015.
- Ittyachen, A. M., A. Vijayan, and M. Isac (2017). The forgotten view: Chest x-ray-lateral view. *Respiratory medicine case reports*, 22, 257–259.
- Jing, B., P. Xie, and E. Xing (2017). On the automatic generation of medical imaging reports. *arXiv preprint arXiv:1711.08195*.
- Johnson, C. D., K. N. Krecke, R. Miranda, C. C. Roberts, and C. Denham (2009). Developing a radiology quality and safety program: a primer. *Radiographics*, 29(4), 951–959.
- Kao, E. F., C. Lee, T. S. Jaw, J. S. Hsu, and G. C. Liu (2006). Projection profile analysis for identifying different views of chest radiographs. *Academic radiology*, 13(4), 518–525.
- Kao, E. F., W. C. Lin, J. S. Hsu, M. C. Chou, T. S. Jaw, and G. C. Liu (2011). A computerized method for automated identification of erect posteroanterior

- and supine anteroposterior chest radiographs. *Physics in Medicine & Biology*, 56(24), 7737.
- Kawamura, T., S. NAITO, K. OKANO, and M. YAMADA (2015). Improvement in image quality and workflow of x-ray examinations using a new image processing method, “virtual grid technology”. *Fujifilm Research & Development*, 60, 21–27.
- Kennedy, J. and R. Eberhart (1995). Particle swarm optimization [a] proceedings of the IEEE international conference on neural networks [c] Piscataway, NJ, USA: *IEEE*.
- Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6), 1153–1160.
- Khatami, A., M. Babaie, A. Khosravi, H. R. Tizhoosh, and S. Nahavandi (2018a). Parallel deep solutions for image retrieval from imbalanced medical imaging archives. *Applied Soft Computing*, 63, 197–205.
- Khatami, A., M. Babaie, H. R. Tizhoosh, A. Khosravi, T. Nguyen, and S. Nahavandi (2018b). A sequential search-space shrinking using CNN transfer learning and a radon projection pool for medical image retrieval. *Expert Systems with Applications*, 100, 224–233.
- Kim, J., J. Kwon Lee, and K. Mu Lee, Accurate image super-resolution using very deep convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Kim, Y. T. (1997). Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE transactions on Consumer Electronics*, 43(1), 1–8.
- Kitamura, G., C. Y. Chung, and B. E. Moore (2019). Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation. *Journal of digital imaging*, 32(4), 672–677.
- Kostrzewa, D., Ł. Skonieczny, P. Benecki, and M. Kawulok, B4multisr: a benchmark for multiple-image super-resolution reconstruction. *In International Conference: Beyond Databases, Architectures and Structures*. Springer, 2018.

- Krizhevsky, A., I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems*. 2012.
- Kumar, A., J. Kim, D. Lyndon, M. Fulham, and D. Feng (2016). An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE journal of biomedical and health informatics*, 21(1), 31–40.
- Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, Photo-realistic single image super-resolution using a generative adversarial network. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- Lehmann, T. M., M. O. Guld, C. Thies, B. Fischer, D. Keysers, M. Kohlen, H. Schubert, and B. B. Wein, Content-based image retrieval in medical applications for picture archiving and communication systems. *In Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation* volume5033. International Society for Optics and Photonics, 2003a.
- Lehmann, T. M., O. Guld, D. Keysers, H. Schubert, M. Kohlen, and B. B. Wein (2003b). Determining the view of chest radiographs. *Journal of Digital Imaging*, 16(3), 280–291.
- Lehmann, T. M., H. Schubert, D. Keysers, M. Kohlen, and B. B. Wein, The irma code for unique classification of medical images. *In Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation* volume5033. International Society for Optics and Photonics, 2003c.
- Li, C. Y., X. Liang, Z. Hu, and E. P. Xing (2018). Hybrid retrieval-generation reinforced agent for medical image report generation. *arXiv preprint arXiv:1805.08298*.
- Li, C. Y., X. Liang, Z. Hu, and E. P. Xing, Knowledge-driven encode, retrieve, paraphrase for medical image report generation. *In Proceedings of the AAAI Conference on Artificial Intelligence* volume33. 2019.
- Lindeberg, T. (1998). Feature detection with automatic scale selection. *International journal of computer vision*, 30(2), 79–116.
- Liu, H., Q. Guo, G. Wang, B. Gupta, and C. Zhang (2017). Medical image resolution enhancement for healthcare using nonlocal self-similarity and low-rank prior. *Multimedia Tools and Applications*, 1–18.

- Liu, X., H. R. Tizhoosh, and J. Kofman, Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. *In Neural Networks (IJCNN), 2016 International Joint Conference on.* IEEE, 2016.
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2), 129–137.
- Lowe, D. G., Object recognition from local scale-invariant features. *In Computer vision, 1999. The proceedings of the seventh IEEE international conference on volume 2.* Ieee, 1999.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110.
- Luo, H., W. Hao, D. H. Foos, and C. W. Cornelius (2006). Automatic image hanging protocol for chest radiographs in pacs. *IEEE Transactions on Information Technology in Biomedicine*, 10(2), 302–311.
- Madani, A., M. Moradi, A. Karargyris, and T. Syeda-Mahmood, Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation. *In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* IEEE, 2018.
- Mao, X. J., C. Shen, and Y. B. Yang (2016). Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*.
- Matas, J., O. Chum, M. Urban, and T. Pajdla (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10), 761–767.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3), 276–282.
- Mifflin, J. (2007). Visual archives in perspective: enlarging on historical medical photographs. *The American Archivist*, 70(1), 32–69.
- Mikolajczyk, K., T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool (2005). A comparison of affine region detectors. *International journal of computer vision*, 65(1-2), 43–72.

- Mildenberger, P., M. Eichelberg, and E. Martin (2002). Introduction to the dicom standard. *European radiology*, 12(4), 920–927.
- Mueen, A., S. Baba, and R. Zainuddin (2007). Multilevel feature extraction and x-ray image classification. *Journal of Applied Sciences*, 7(8), 1224–1229.
- Müller, H., J. Kalpathy-Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke, C. E. Kahn, and W. Hersh, Overview of the clef 2009 medical image retrieval track. *In Workshop of the Cross-Language Evaluation Forum for European Languages*. Springer, 2009.
- Nah, S., T. Hyun Kim, and K. Mu Lee, Deep multi-scale convolutional neural network for dynamic scene deblurring. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- O’Hara, S. and B. A. Draper (2011). Introduction to the bag of features paradigm for image classification and retrieval. *arXiv preprint arXiv:1101.3354*.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62–66.
- Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu, Bleu: a method for automatic evaluation of machine translation. *In Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 2002.
- Pourghassem, H. and S. Daneshvar, A framework for medical image retrieval using merging-based classification with dependency probability-based relevance feedback. 2013a.
- Pourghassem, H. and S. Daneshvar (2013b). A framework for medical image retrieval using merging-based classification with dependency probability-based relevance feedback. *Turkish Journal of Electrical Engineering & Computer Sciences*, 21(3), 882–896.
- Qayyum, A., S. M. Anwar, M. Awais, and M. Majid (2017). Medical image retrieval using deep convolutional neural network. *Neurocomputing*, 266, 8–20.
- Rajpurkar, P., J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng (2017a). Mura dataset: Towards radiologist-level abnormality detection in musculoskeletal radiographs. *arXiv preprint arXiv:1712.06957*.

- Rajpurkar, P., J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, L. B. Robyn, C. Langlotz, K. Shpanskaya, P. L. Matthew, and Y. N. Andrew (2017*b*). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
- Ramakrishna, B., W. Liu, G. Saiprasad, N. Safdar, C. I. Chang, K. Siddiqui, W. Kim, E. Siegel, J. W. Chai, C. C. C. Chen, and S. K. Lee (2009). An automatic computer-aided detection system for meniscal tears on magnetic resonance images. *IEEE Transactions on medical imaging*, 28(8), 1308–1316.
- Ren, Y., S. Wu, M. Wang, and Z. Cen (2014). Study on construction of a medical x-ray direct digital radiography system and hybrid preprocessing methods. *Computational and mathematical methods in medicine*, 2014.
- Report, A. (2017). Medical Imaging Market Outlook To 2024: Key Product Categories, Applications, Regional Segmentation, Competitive Dynamics, Pricing Analysis and Segment Forecast. <https://www.ameriresearch.com/medical-imaging-market/>. [Online; accessed April 13, 2017].
- Report, G. (2021). U.S. Imaging Services Market Size, Share Trends Analysis Report. <https://www.grandviewresearch.com/industry-analysis/us-imaging-services-market/methodology>. [Online; accessed Sep 2021].
- Report, M. (2019). Diagnostic Imaging Services Market. <https://www.marketsandmarkets.com/Market-Reports/diagnostic-imaging-service-market-17157849.html>. [Online; accessed Aug 2021].
- Rubin, J., D. Sanghavi, C. Zhao, K. Lee, A. Qadir, and M. Xu-Wilson (2018). Large scale automated reading of frontal and lateral chest x-rays using dual convolutional neural networks. *arXiv preprint arXiv:1804.07839*.
- Rui, W. and W. Guoyu, Medical x-ray image enhancement method based on dark channel prior. *In Proceedings of the 5th International Conference on Bioinformatics and Computational Biology*. ACM, 2017.
- Saif, A., C. Shahnaz, W. P. Zhu, and M. O. Ahmad (2019). Abnormality detection in musculoskeletal radiographs using capsule network. *IEEE Access*, 7, 81494–81503.
- Saleem, A., A. Beghdadi, and B. Boashash (2012). Image fusion-based contrast enhancement. *EURASIP Journal on Image and Video Processing*, 2012(1), 10.

- Santosh, K., S. Candemir, S. Jaeger, A. Karargyris, S. Antani, G. R. Thoma, and L. Folio (2015). Automatically detecting rotation in chest radiographs using principal rib-orientation measure for quality control. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(02), 1557001.
- Santosh, K., S. Vajda, S. Antani, and G. R. Thoma (2016). Edge map analysis in chest x-rays for automatic pulmonary abnormality screening. *International journal of computer assisted radiology and surgery*, 11(9), 1637–1646.
- Santosh, K. and L. Wendling (2018). Angular relational signature-based chest radiograph image view classification. *Medical & biological engineering & computing*, 56(8), 1447–1458.
- Saunders Jr, R. S., J. A. Baker, D. M. DeLong, J. P. Johnson, and E. Samei (2007). Does image quality matter? impact of resolution and noise on mammographic task performance. *Medical physics*, 34(10), 3971–3981.
- Seco, G., A. de Herrera, R. Schaer, S. Bromuri, and H. Müller (2016). Overview of the imageclef 2016 medical task. *Working Notes of CLEF*.
- Sekher, T., Catastrophic health expenditure and poor in india: health insurance is the answer. In *Proceedings of the 27th IUSSP International Population Conference* volume2013. 2013.
- Sexton, A., A. Todman, and K. Woodward, Font recognition using shape-based quad-tree and kd-tree decomposition. In *3rd International Conference on Computer Vision, Pattern Recognition and Image Processing*. 2000.
- Shapiro, M., D. Johnston, J. Wald, and D. Mon (2012). Patient-generated health data. *RTI International*, April.
- Sharma, S., I. Umar, L. Ospina, D. Wong, and H. R. Tizhoosh, Stacked autoencoders for medical image search. In *International Symposium on Visual Computing*. Springer, 2016.
- Sheikh, H. R. and A. C. Bovik (2006). Image information and visual quality. *IEEE Transactions on image processing*, 15(2), 430–444.
- Shin, H. C., L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, Interleaved text/image deep mining on a very large-scale radiology database. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

- Shiraishi, J., F. Li, and K. Doi (2007). Computer-aided diagnosis for improved detection of lung nodules by use of posterior-anterior and lateral chest radiographs. *Academic radiology*, 14(1), 28–37.
- Shyu, C. R., C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick (1999). Assert: A physician-in-the-loop content-based retrieval system for hrct image databases. *Computer vision and image understanding*, 75(1-2), 111–132.
- Smeulders, A. W., M. Worring, S. Santini, A. Gupta, and R. Jain (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12), 1349–1380.
- Solovyova, A. (2020). X-ray bone abnormalities detection using mura dataset. *arXiv preprint arXiv:2008.03356*.
- Srinivas, M., R. R. Naidu, C. S. Sastry, and C. K. Mohan (2015). Content based medical image retrieval using dictionary learning. *Neurocomputing*, 168, 880–895.
- Stefan, L.-D., B. Ionescu, and H. Müller (2017). Generating captions for medical images with a deep learning multi-hypothesis approach: Medgift–upb participation in the imageclef 2017 caption task.
- Su, Y. and F. Liu, Umass at imageclef caption prediction 2018 task. In *CLEF2018 Working Notes. CEUR Workshop Proceedings, Avignon, France*. 2018.
- Summers, R. M. (2012). Evaluation of computer-aided detection devices: consensus is developing. *Academic radiology*, 19(4), 377–379.
- Sze To, A., H. R. Tizhoosh, and A. K. Wong, Binary codes for tagging x-ray images via deep de-noising autoencoders. In *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- Takeuchi, D., R. Thai, and K. Tran (2019). Exploring model architectures and view-specific models for chest radiograph diagnoses. <http://cs229.stanford.edu/proj2019spr/report/17.pdf>, 1–6.

- Tang, L. H., R. Hanka, H. H. Ip, and R. Lam, Extraction of semantic features of histological images for content-based retrieval of images. *In Medical Imaging 1999: PACS Design and Evaluation: Engineering and Clinical Issues* volume 3662. SPIE, 1999.
- Tizhoosh, H. R., Barcode annotations for medical image retrieval: A preliminary investigation. *In Image Processing (ICIP), 2015 IEEE International Conference on.* IEEE, 2015.
- Tommasi, T., B. Caputo, P. Welter, M. O. Güld, and T. M. Deserno, Overview of the clef 2009 medical image annotation track. *In Workshop of the Cross-Language Evaluation Forum for European Languages.* Springer, 2009.
- Tommasi, T. and F. Orabona, Idiap on medical image classification. *In ImageCLEF.* Springer, 2010, 453–465.
- Tommasi, T., F. Orabona, and B. Caputo (2008). Discriminative cue integration for medical image annotation. *Pattern Recognition Letters*, 29(15), 1996–2002.
- Trapp, M., F. Schulze, K. Bühler, T. Liu, and B. J. Dickson (2013). 3d object retrieval in an atlas of neuronal structures. *The Visual Computer*, 29(12), 1363–1373.
- Unay, D., O. Soldea, A. Ekin, M. Cetin, and A. Ercil, Automatic annotation of x-ray images: a study on attribute selection. *In MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support.* Springer, 2009.
- Villegas, M., H. Müller, A. Gilbert, L. Piras, J. Wang, K. Mikolajczyk, A. G. Herrera, S. Bromuri, M. A. Amin, and M. K. Mohammed, General overview of imageclef at the clef 2015 labs. *In International conference of the cross-language evaluation forum for European languages.* Springer, 2015.
- Vinyals, O., A. Toshev, S. Bengio, and D. Erhan, Show and tell: A neural image caption generator. *In Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015.
- Wang, J., Y. Li, Y. Zhang, C. Wang, H. Xie, G. Chen, and X. Gao (2007). Bag-of-features based medical image retrieval via multiple assignment and visual words weighting. *IEEE Transactions on Medical Imaging*, 6(1), 1.

- Wang, J. Z., Pathfinder: multiresolution region-based searching of pathology images using irm. *In Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2000.
- Wang, X., Y. Peng, L. Lu, Z. Lu, and R. M. Summers, Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018a.
- Wang, X., K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, Esrgan: Enhanced super-resolution generative adversarial networks. *In Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018b.
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600–612.
- Wang, Z., E. P. Simoncelli, and A. C. Bovik, Multiscale structural similarity for image quality assessment. *In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003* volume2. Ieee, 2003.
- Xi, P., H. Guan, C. Shu, L. Borgeat, and R. Goubran (2019). An integrated approach for medical abnormality detection using deep patch convolutional neural networks. *The Visual Computer*, 1–14.
- Xue, Y., T. Xu, L. R. Long, Z. Xue, S. Antani, G. R. Thoma, and X. Huang, Multimodal recurrent model with attention for automated radiology report generation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018.
- Xue, Z., D. You, S. Candemir, S. Jaeger, S. Antani, L. R. Long, and G. R. Thoma, Chest x-ray image view classification. *In 2015 IEEE 28th International Symposium on Computer-Based Medical Systems*. IEEE, 2015.
- Yang, C. Y., C. Ma, and M. H. Yang, Single-image super-resolution: A benchmark. *In European conference on computer vision*. Springer, 2014.
- Yurt, M., S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur (2021). Mustgan: Multi-stream generative adversarial networks for mr image synthesis. *Medical Image Analysis*, 70, 101944.

- Zare, M. R., A. Mueen, and W. C. Seng (2013). Automatic classification of medical x-ray images using a bag of visual words. *IET Computer Vision*, 7(2), 105–114.
- Zare, M. R., A. Mueen, W. C. Seng, and M. H. Awedh, Combined feature extraction on medical x-ray images. In *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE, 2011.
- Zhang, Y. and M. An (2017). Deep learning-and transfer learning-based super resolution reconstruction from single medical image. *Journal of healthcare engineering*, 2017.
- Zhao, C., Z. Wang, H. Li, X. Wu, S. Qiao, and J. Sun (2019). A new approach for medical image enhancement based on luminance-level modulation and gradient modulation. *Biomedical Signal Processing and Control*, 48, 189–196.
- Zhou, W., X. Li, and D. S. Reynolds, Nonlinear image interpolation via deep neural network. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017.
- Zhu, B., J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen (2018). Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697), 487–492.
- Zhu, S. and H. R. Tizhoosh (2016). Radon features and barcodes for medical image retrieval via svm. *arXiv preprint arXiv:1604.04675*.
- Zodpey, S. and H. H. Farooqui (2018). Universal health coverage in india: Progress achieved & the way forward. *The Indian journal of medical research*, 147(4), 327.
- Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. *Graphics gems*, 474–485.

Bio-data

Name: Karthik K.

Current Address: Research Scholar,
Department of Information Technology,
NITK Surathkal
Mangaluru, Karnataka
India - 575025.

Permanent Address: K.P. II-215/A, Shree Karthikeya Nilaya,
Near Mundol Temple, Neerolipara,
Mulleria PO, Mulleria
Kasaragod, Kerala
India - 671543.

Email: 2karthik.bhat@gmail.com

Mobile No: +91 84960 89879

Qualification: Ph.D. in Information Technology
Department of Information Technology
National Institute of Technology Karnataka, Surathkal
Mangaluru, Karnataka
India - 575025.

M.Tech in Computer Science & Engineering
NMAM Institute of Technology, Nitte, Karkala,
Udupi, Karnataka
India - 574110.

B.Tech in Information Science & Engineering
Acharya Institute of Technology, Soldevanahalli,
Bengaluru, Karnataka
India - 560107.

Research Area: Medical Image Retrieval, Machine Learning, Deep Learning, Healthcare Analytics