TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO.: 075/MSICE/008

Attention-based Graph Convolutional Neural Network for Classification of Musculoskeletal Radiograph Images

by

Ganesh Singh Rawal

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

LALITPUR, NEPAL

AUGUST, 2021

Attention-based Graph Convolutional Neural Network for Classification of Musculoskeletal Radiograph Images

by

Ganesh Singh Rawal

075MSICE008

Thesis Supervisor

Prof. Dr. Shashidhar Ram Joshi

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Information and Communication Engineering

Department of Electronics and Computer Engineering
Institute of Engineering, Pulchowk Campus
Tribhuvan University
Lalitpur, Nepal

August, 2021

COPYRIGHT©

DECLARATION

I declare that the work hereby submitted for Master of Science in Information and Communication Engineering (MSICE) at IOE, Pulchowk Campus entitled "**Attention-based Graph Convolutional Neural Network for Classification of Musculoskeletal Radiograph Images**" is my own work and has not been previously submitted by me at any university for any academic award.

I authorize IOE, Pulchowk Campus to lend this thesis to other institution or individuals for the purpose of scholarly research.

Ganesh Singh Rawal
075MSICE008
Date: August, 2021

RECOMMENDATION

The undersigned certify that they have read, and recommended to the Institute of Engineering for acceptance, a thesis entitled **“Attention-based Graph Convolutional Neural Network for Classification of Musculoskeletal Radiograph Images”**, submitted by **Ganesh Singh Rawal** in partial fulfillment of the requirement for the award of the degree of **“Master of Science in Information and Communication Engineering”**.

…………………………………………………….

Supervisor: Prof. Dr. Shashidhar Ram Joshi,
Dean of Engineering,
Institute of Engineering, Tribhuvan University.

…………………………………………………….

External Examiner: Dr. Kamal Chapagain,
Assistant Professor,
Department of Electrical and Electronics Engineering,
School of Engineering, Kathmandu University.

…………………………………………………….

Committee Chairperson: Dr. Basanta Joshi,
Program Co-ordinator, MSc in Information and Communication Engineering,
Department of Electronics and Computer Engineering,
Institute of Engineering, Tribhuvan University.

Date: August, 2021

DEPARTMENTAL ACCEPTANCE

The thesis entitled "**Attention-based Graph Convolutional Neural Network for Classification of Musculoskeletal Radiograph Images**", submitted by **Ganesh Singh Rawal** in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Information and Communication Engineering**" has been accepted as a bonafide record of work independently carried out by him in the department.

……………………………………………

Prof. Dr. Ram Krishna Maharjan
Head of Department,
Department of Electronics and Computer Engineering,
Pulchowk Campus, Institute of Engineering,
Tribhuvan University,
Lalitpur, Nepal.

ACKNOWLEDGEMENT

I express my deepest sense of gratitude to my thesis supervisor, **Prof. Dr. Shashidhar Ram Joshi**, for providing me with all the feedbacks, suggestions and necessary guidance throughout the course of my thesis work.

It is my immense pleasure to mention **Assoc. Prof. Dr. Sanjeeb Prasad Panday**, for taking out some of his invaluable time listening to my queries and helping me deal with the problems encountered during this thesis work.

I am deeply indebted to **Assist. Prof. Dr. Basanta Joshi**, who is also our MSICE program co-ordinator, for hearing out the problems that I faced throughout the journey of my graduate study, and helping me overcome those problems.

I would like to present my sincere acknowledgements to Head of the Department **Prof. Dr. Ram Krishna Maharjan**, **Prof. Dr. Subarna Shakya**, **Assoc. Prof. Dr. Surendra Shrestha**, **Assoc. Prof. Dr. Dibakar Raj Pant**, **Assoc. Prof. Dr. Nanda Bikram Adhikari** and other faculties for their invaluable comments and suggestions which helped a lot in improving this thesis.

I am extremely thankful to the Department of Electronics and Computer Engineering for providing me the opportunity to conduct the research work through this thesis.

I would like to thank my family, friends and all who helped me directly or indirectly in conducting my research.

Sincerely,
Ganesh Singh Rawal
075MSICE008

ABSTRACT

Musculoskeletal Disorders (MSDs) are the abnormalities related to bones and muscles, affecting majority of the world population. Radiographic studies are the most common technique for the detection of these abnormalities as part of the medical diagnoses. An attention-based graph convolutional neural network (AGCNN) is implemented, in this thesis work, for the classification of such abnormalities in musculoskeletal radiograph images. The AGCNN network model is firstly implemented on the standard benchmark MURA dataset, consisting of 40,561 upper extremity radiograph images, for the binary classification of radiograph images into normal and abnormal. The performance of the network model is compared with that of the DenseNet169 baseline model. The network model showed improved performance results than the baseline model. The network model is then implemented on Xtremity dataset, consisting of 15,701 extremity radiograph images, for the multi-class classification of radiograph images into five different classes. The network model, that is implemented, is an ensembled network of soft attention-based Inception-ResNet-v2 network and graph convolutional network (GCN). Soft Attention map is used to localize the abnormality regions in the radiograph images for qualitative evaluation of the network. The network model achieved an accuracy of 0.884, average recall of 0.874, average F1 score of 0.876, and average AUC score of 0.976. The network model achieved above average results in the classification task. Furthermore, the performance results of the classification task by the ensembled AGCNN network are compared with that of different state-of-the-art pre-trained CNN architectures.

Keywords:

MSDs, AGCNN, MURA, Soft Attention, Inception-ResNet-v2, GCN, AUC

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AUC	Area Under the Curve
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CT	Computerized Tomography
FN	False Negative
Faster R-CNN	Faster Region-based Convolutional Neural Network
FP	False Positive
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GNN	Graph Neural Network
Grad-CAM	Gradient-weighted Class Activation Mapping
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
MRI	Medical Resonance Imaging
MSD	Musculoskeletal Disorder
MURA	Musculoskeletal Radiograph
NMC	Nepal Medical Council
PET	Positron Emission Tomography
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
ROC	Receiver Operating Characteristic
TN	True Negative
TP	True Positive
VGG	Visual Geometry Group

CHAPTER ONE: INTRODUCTION

1.1 Background

Musculoskeletal abnormalities involve problems related majorly to muscles, bones, and joints. These abnormalities include fractures, dislocations, degenerative joint diseases, lesions, etc. Some of these abnormalities such as fractures have a short-term effect only, but the conditions with prevalent pain or permanent disability have an effect for a lifetime. These abnormalities are broadly known as Musculoskeletal Disorders (MSDs). These disorders are very common, affecting the majority of world population. According to a recent study report on Global Burden of Disease in 2019 [1], over 1.7 billion people were affected worldwide due to musculoskeletal disorders. The conducted study found that musculoskeletal disorders were the second leading cause of global disability with almost 30% of the world's population suffering from such debilitating conditions. Age, family history, exercise level, and using bad gestures at work have all been linked to the development of musculoskeletal problems. Musculoskeletal problems can be cured with proper therapy based on the diagnosis. The diagnoses of such abnormalities often require physical examination by radiologists and their inspection of medical images such as X-ray, Ultrasonography, PET scan, CT scan, MRI, *etc*. Among all of the medical images used for examination, X-rays – also known as radiographs, are the most common and widely used. The cheaper cost and shorter examination time with availability of results within few hours are, most probably, the reasons for the popularity of radiographs for being used in examination by radiologists for the diagnosis of musculoskeletal abnormalities. Since these abnormalities affect a large population, a proportionally huge number of radiologists are required. However, this is not the case, as there are a limited number of radiologists available for examining a relatively large number of people with such disorders. Such huge workload can significantly affect the diagnostic performance of radiologist. According to a recent study report, over one billion radiologic tests were performed worldwide each year, the most of which were clarified by radiologists [2]. In most cases, the clinical affairs or conditions of patients, their relevant history, and earlier medical imaging had a substantial impact on radiological explanation. As a remedy, a system model that can perform automated detection of such abnormalities might be developed for radiologists with the goal of preventing issues from worsening as a result of failing to recognize warning indications. The automated system model can significantly reduce the radiologists' workloads and improve their diagnostic performance. Furthermore, the system takes relatively less time for detection as compared to the time-consuming manual detection.

The thesis work aims to develop a model with the application of deep learning that can classify musculoskeletal radiographs according to the abnormalities present. The application of deep learning in the medical images have been very effective, showing that the deep learning techniques can perform some medical tasks with comparable accuracy as that of medical experts. The classification of musculoskeletal abnormalities is done by the application of deep learning techniques in radiograph images. In this thesis work, an ensembled network, comprising of an inception residual neural network (Inception-ResNet-v2) [3] with attention mechanism and Graph Convolutional Network (GCN) [4], is trained on radiograph images to classify the images on the basis of musculoskeletal abnormalities present. The classification task takes X-ray images as input, and outputs the prediction score of the type of abnormality present in the image. Furthermore, qualitative and quantitative evaluations of the ensembled network are done by performing different visualizations and calculating the evaluation metrics related to the classification task.

1.2 Problem Statement

Musculoskeletal disorders involve pain and injuries related to bones, joints, muscles, ligaments, and tendons such as fractures, degenerative joint diseases, lesions, subluxations, etc. These disorders affect people of every age group from children to adults to old-aged people. There is a rapid increase in the number of people with musculoskeletal conditions because of the worldwide population increase and ageing. The detection of these kinds of disorders requires medical expertise examining the medical imaging, which are time-consuming procedures. There are an insignificant number of radiologists when compared to huge number of people with such disorders. As a result, with the majority of the world population affected by these disorders, the radiologists' workloads are massive and are increasing for the manual detection. Automated detection, with the usage of deep learning techniques for the probabilities of such disorders in radiographs, can significantly reduce the radiologists' workloads, and, therefore, reduce the rate of diagnostic error compared to tedious manual procedure. Moreover, the automated detection speeds up the process of diagnosis for the radiologists.

1.3 Thesis Objectives

The objectives of this thesis are:

1. To implement an ensembled network, comprising of an inception residual neural network with soft attention mechanism and graph convolutional network, for the multi-class classification of the radiograph images on the basis of musculoskeletal abnormalities present.
2. To validate the implemented network both qualitatively and quantitatively using different measures.
3. To compare the performance of the ensembled network with that of the state-of-the-art CNN architectures.

1.4 Contribution of the Thesis

The contribution of this thesis work involves the application of deep learning techniques for the multi-class classification of musculoskeletal abnormalities present in the radiograph images collected from multiple sources. The major contributions of this thesis are:

1. The radiographic images from various sources are collected forming the musculoskeletal radiograph image dataset which contains 15,701 labelled images and categorized into five different classes.
2. The application of graph convolutional network in image dataset is explored for extracting the latent correlational features among a group of images for the classification task.

1.5 Structure of the Thesis

The thesis is structured into six chapters. In chapter one, a brief overview of the problem is given along with motivations and importance of doing research in this particular field. Chapter two states the related works regarding this work and presents the research gap in those already done works. Chapter three describes the theories related to understand the concept behind the work. Chapter four includes the research methodology for performing the work to meet the objectives. Chapter five presents the experimental settings and implementation results along with some analytical discussions.

Finally, chapter six draws conclusion of the research done and mentions some future works that can be done in this subject area.

CHAPTER TWO: LITERATURE REVIEW

The application of Machine Learning (ML) has been common for the task of image analysis as they produce notable results in such situations. In image analysis, machine learning algorithms have achieved fair precision, however, they typically require very unique manually-engineered features to function, which dramatically reduces their capability to generalize to similar problems. In contrast, Deep Learning (DL) algorithms extract features on its own. The deep learning methods have relatively overshadowed the traditional machine learning methods in image processing and computer vision domains.

ImageNet project by Deng et al. [5] is a very large-scale image database designed for visual object recognition and consists of more than 14 million images that are categorized into more than 20,000 classes. The ImageNet project laid a foundation for the development of many state-of-the-art convolutional neural network architectures. The work of Krizhevsky et al. [6] can be seen as a landmark in computer-aided image analysis using deep neural networks. Their proposed deep convolutional neural network, named as AlexNet, won the ILSVRC-2012 challenge for classification of about 1.2 million images – a smaller ImageNet version of the larger ImageNet project database – into 1,000 distinct categories.

2.1 Deep Learning in Medical Imaging

Medical image analysis and their interpretation with significant accuracy are very crucial for better diagnoses. Deep Learning application in the analysis of medical images is gaining attention of many researchers worldwide. The widely used medical images for analysis are X-rays, Medical Resonance Imagings (MRIs), Computer Tomography (CT) scans, etc. These works of [7, 8, 9] laid significant foundations in the research world and paved a path for future work enhancements in the medical image analysis using deep learning techniques.

Gulshan et al. [7] implemented a deep learning algorithm in retinal fundus images for the detection of different grades of diabetic retinopathy and diabetic macular edema. They used a deep convolutional neural network and trained it on a large dataset of 128,175 retinal images. The implemented network was validated using 2 different datasets: EyePACS-1 dataset consisting of 9,963 retinal images and Messidor-2 consisting of 1,748 retinal images. Their network implementation achieved AUC score of 0.991, sensitivity of 0.903, and specificity of 0.981 on EyePACS-1 dataset. On Messidor-2 dataset, the network achieved AUC score of 0.990, sensitivity of 0.87, and specificity of 0.985.

Esteva et al. [8] implemented a deep convolutional neural network for the classification of skin cancer. They trained the network on a large dataset of 129,450 images of skin lesions consisting of more than 2,000 different diseases. They validated their results, from the tasks of binary classification of skin lesions on test set, by performing a comparative test with board-certified dermatologists. They claimed that their network achieved performance that is comparable to that of the dermatologists.

Wang et al. [9] released a huge medical dataset, named ChestX-ray8, and benchmarked on different CNN models pre-trained on ImageNet. The dataset consists of over 100,000 multi-labeled antero-posterior view of chest X-ray images. They later updated the dataset to include more images of different diseases and named the dataset as ChestX-ray14. The ChestX-ray dataset released by Wang et al. has been the most used medical dataset for research purposes. Rajpurkar et al. [10] used a 121-layered densely connected convolutional neural network for pneumonia detection with the network model trained on the ChestX-ray14 dataset. They compared the performance results of their implemented network model with that of the radiologist. They concluded that the performance of their network model for detecting pneumonia was beyond that of a radiologist.

2.2 Deep Learning in Musculoskeletal Abnormality Detection

Rajpurkar et al. [11] released a huge dataset, named as MURA, which consists of over 40,000 multi-view musculoskeletal radiographic images of seven study types of upper body extremities. They used a 169-layered densely connected convolutional network model – DenseNet169 – for the prediction of abnormality in radiograph images. The network was trained end-to-end on MURA dataset. They proposed an ensembled model by combining five models with the lowest validation losses. Their model attained an AUC score of 0.929, sensitivity of 0.815 and specificity of 0.887. They concluded that the performance of the model was comparable to the radiologist’s best performance in finger and wrist study parts, however, the model’s performance in detecting abnormalities in hand, humerus, forearm, elbow, and shoulder study parts was lower than the best performance of the radiologist.

For the identification of musculoskeletal anomalies in radiographs, Mondol et al. [12] proposed an ensemble model combining VGG-19 architecture [13] and ResNet-50 architecture [14]. The proposed ensembled model, named as Computer-Aided Diagnosis (CADx), was trained on four study types – Elbow, Finger, Humerus, and Wrist – of MURA dataset. The results showed that the ensembled model performed relatively better than the individual VGG-19 architecture and ResNet-50 architecture. They concluded their work by comparing the results of baseline model of Rajpurkar et al. [11] with their proposed ensembled model. They claimed their model's performance was better than the baseline model on most of the study types.

Thian et al. [15] proposed a model using object detection CNN for detecting radius and ulna fractures and localizing the areas of those fractures in radiographic images of the wrist. The proposed model was based on Inception-ResNet architecture and the final model was a Faster R-CNN [16] architecture. The model was evaluated on per-image and per-study basis achieving high sensitivity and specificity. Chung et al. [17] attempted to discover the capability of CNN to recognize and classify the humerus fractures using the dataset which contains 1,891 radiograph images of normal shoulder and 1,376 radiograph images with the proximal humerus fracture. These fractures were classified into four types and, then, evaluated to obtain the final results. After excluding the radiographs of normal shoulder, the fracture types were classified. Their implemented CNN model achieved high accuracy of 96%, AUC-ROC of 1.00, sensitivity of 0.99, and specificity of 0.97 to classify normal shoulder radiographs and radiographs with proximal humerus fractures.

Maya Varma et al. [18] used a 161-densely connected convolutional network for the detection of musculoskeletal abnormalities in lower extremity radiograph images. They used a large dataset of 93,455 radiograph images of multiple lower extremity body parts, labelled as abnormal or normal. Their model achieved an AUC score of 0.880, sensitivity of 0.714 and specificity of 0.961. Furthermore, they explored the effect that the size of the dataset, pretraining the model with relevant datasets, and model architecture have on the model's performance for performing deep learning analysis on extremity radiographs.

2.3 Graph Neural Networks for Medical Image Classification

Recent development in the medical image classification field explored the use of graph neural network (GNN) and its different variants such as graph convolutional network (GCN), graph attention network (GAT), *etc*. Xiang Yu et al. [19] proposed a graph neural network as classifier on the features extracted from ResNet101 network from the chest CT images for the COVID-19 detection. The model, named as ResGNet-C, performed binary classification of lung CT images into normal and COVID-19. The model achieved 96.6% accuracy, 97.3% sensitivity, and 95.9% specificity, using five-fold cross-validation on the dataset comprising of 296 lung CT images. They claimed their work as the first effort in combining knowledge of graph into the COVID-19 detection task. Graph structures were built on the basis of Euclidean distance metric calculated among the features extracted by ResNet101-C, and then, the graph structures are encoded with the extracted features to perform the prediction. They claimed their high-performance model surpassed all state-of-the-art methods.

Wang et al. [20] implemented Convolutional Neural Network and Graph Convolutional Network for the task of COVID-19 detection in chest CT images. They fed the individual image-level representation features extracted from self-created CNN to the GCN for the extraction of relation-aware representation features and then fused both the features. They compared their model, named as FGCNet, with 15 state-of-the-art methods, and concluded their model attained comparatively better performance in detecting COVID-19.

2.4 Research Gap

Many research works regarding the detection of the musculoskeletal abnormality in radiograph images has been done before. Those works dealt with only binary classification of radiograph images as normal and abnormal. After the consultation with few radiologists from local hospitals, it became evident that the common abnormalities that can be diagnosed from radiographic studies were fractures, dislocations, lesions, and degenerative joint diseases. Moreover, the radiographic studies were also used to observe the orientation of orthopedic hardware devices that were implanted as a treatment procedure of many musculoskeletal abnormalities. The research work with further classification, that is, multi-class classification of the specific classes of musculoskeletal abnormalities such as fracture, orthopedic hardware, lesion, and joint disorder had not been performed before. This issue was being addressed, in this thesis work, by classifying the radiograph images according to the major types of abnormalities present in the images.

Almost all of the works related to the classification of radiograph images involved the utilization of convolutional neural networks only. Convolutional neural networks are capable of capturing only the individual image representational features. However, they are not capable to capture the correlational representation features among a group of images. Graph Convolutional Network (GCN) have the capability of capturing the correlational features among images. This research gap was attempted to be filled by capturing relational features in addition to individual image features, by ensembling convolutional neural network and graph convolutional network together in hierarchical fashion for the classification of radiograph images.

CHAPTER THREE: THEORETICAL BACKGROUND

3.1 Inception-ResNet-v2 Network

The pre-trained Inception-ResNet-v2 network [3] was used in this thesis work as CNN sub-network since it achieved high performance results for the classification of images in the ImageNet dataset. It is a 164-layered deep convolutional neural network architecture pre-trained on ImageNet dataset. The network integrates the concept of residual connections into the Inception module structure. The Inception-ResNet network introduces residual connections that add the inception module's convolution output to the input. These connections, also called skip connections, help with vanishing gradient and exploding gradient problems. They also help in the reduction of training time. Figure 1 shows schematic diagram of compressed view of Inception-ResNet-v2 network.

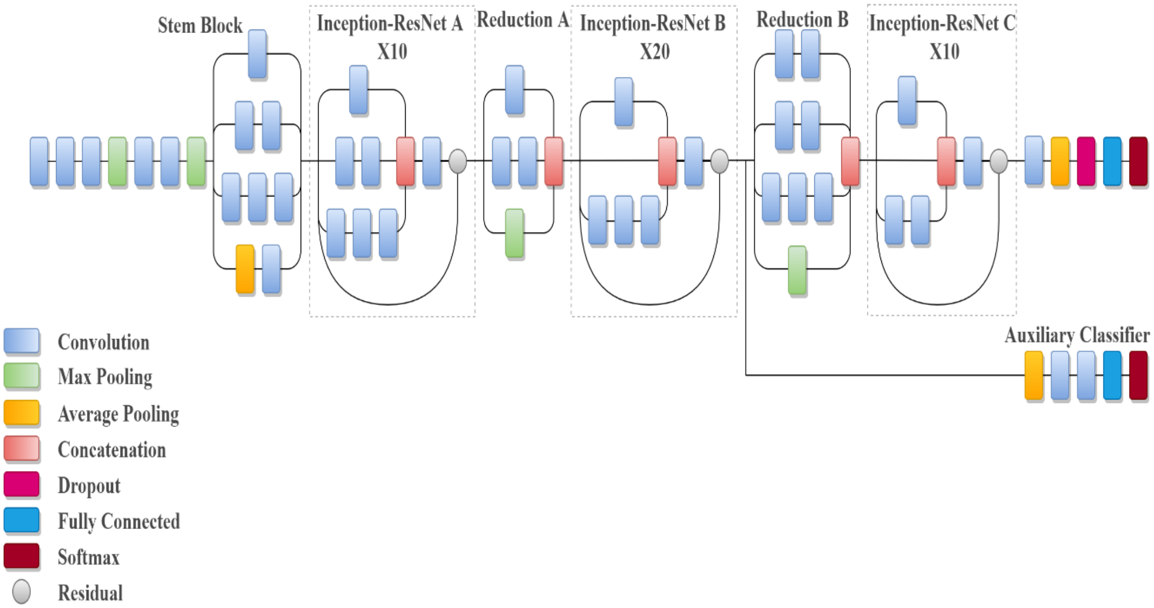


Figure 1: Schematic diagram of Inception-ResNet-v2 network.

The stem block comprises of initial set of operations that are needed to be performed on input before introducing the input to Inception modules. The concept of an inception module in the Inception-ResNet network incorporates convolutional kernels with multiple sizes operating on the same level so that a larger kernel and a smaller kernel can be effectively utilized for capturing information that are distributed both globally and locally, respectively. A filter expansion layer, which is actually a 1x1 convolution without activation function, follows each inception block in the network. The filter expansion layer is used for scaling up the dimensionality of the filter bank before performing addition so that the depth of the output matches the depth of the input. An auxiliary classifier is incorporated into the network which prevents the deep network from dying out. The auxiliary loss of the auxiliary classifier is only used for training purposes and is ignored during inference.

3.2 Soft Attention Mechanism

The concept of attention mechanism is employed in neural network architectures in order to focus on relevant features that contribute more to the results. One such technique is soft attention mechanism which was originally employed in image captioning task [22]. The concept is inspired from the implementation of skin lesion image classification [23] which showed improved performance results. Figure 2 shows the diagrammatic representation of soft attention block unit.

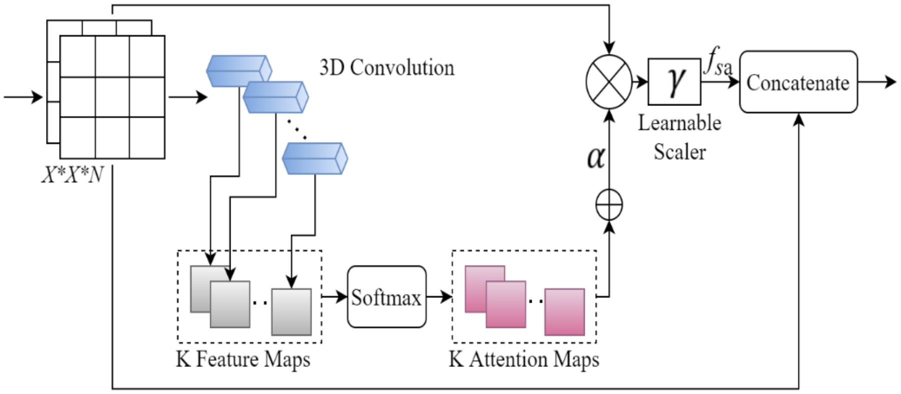


Figure 2: Diagrammatic representation of Soft Attention block unit.

The feature tensor (t) that streams down the convolutional neural network is fed as input to the soft-attention block unit. The soft attention map is calculated mathematically as:

$$f_{sa} = \gamma t(\sum_{k=1}^{K} softmax(W_k * t)) \qquad 3.1$$

Here,

$t \in \mathbb{R}^{hxwxd}$ represents an input feature tensor to 3D convolutional layer,

$W_k \in \mathbb{R}^{hxwxdxK}$ represents the k^{th} 3D weight, and

K represents the number of 3D weights.

The output from the 3D convolution is fed to the softmax activation function, which performs normalization operation, to produce K = 16 attention maps. As shown in Figure 2, the resulting attention maps are combined to yield an integrated attention map which performs as a weighting function (α). The resulting integrated attention map represented by (α) is then multiplied with the feature tensor (t) to scale the salient feature values attentively. The resulting feature values are further scaled by a learnable scalar parameter (γ). Finally, the resulting features (f_{sa}) that are attentively scaled are then concatenated with the input feature tensor (t) as a residual connection.

3.3 Graph Convolutional Network

Graph Convolutional Network (GCN) [4] is one of the many variants of Graph Neural Network family which operates on arbitrarily-structured graph data or simply graphs. Neural Networks could only be implemented on regular-structured data or, in other words, Euclidean data. However, most of the real-world data are non-regular in structure and can be represented by non-regular or graphical data. The non-regular data structures have led to recent improvements in Graph Neural Networks. Graph Convolutional Network is one of the most popular Graph Neural Network variants. The convolution operation in GCN is basically the same as in convolutional neural networks. However, the convolution in GCN is done on graph-structured data while the convolution in CNN is done on image which is grid-structured data. Figure 3 illustrates a schematic diagram of graph convolutional network.

The GCN learns the features by aggregating the features from the neighboring nodes. It takes the weighted average of neighbor’s feature vectors. The idea of weighted average is based on the assumption that low-degree nodes would have bigger influence on their neighbors whereas, high-degree nodes yield lower impact as they scatter their influence at a greater number of neighbors.

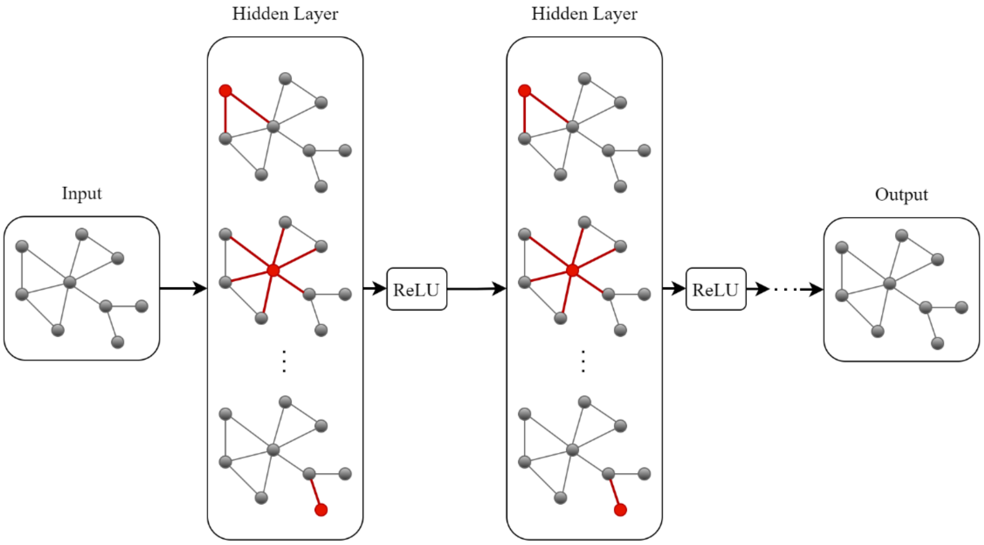


Figure 3: Schematic diagram of Graph Convolutional Network.

The propagation rule for each GCN layer is summarized as:

$$H^{(l+1)} = \sigma(\hat{A}H^{(l)}W^{(l)}) \qquad 3.2$$

Here,

H is the hidden state (or node features when layer, $l = 0$),

$\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ is the normalized version of adjacency matrix,

$\tilde{A}$ is the adjacency matrix taking individual self-nodes into account,

$\tilde{D}$ is the diagonal degree matrix of adjacency matrix $\tilde{A}$,

W is the trainable weight matrix,

σ is the activation function, and

l is the layer number.

The term $\hat{\mathrm{A}}$ represents the average of features of all the neighbors including the feature of itself. The adjacency matrix is normalized by both rows and columns to get the weighted average preferring features on low-degree nodes.

CHAPTER FOUR: RESEARCH METHODOLOGY

4.1 Block Diagram

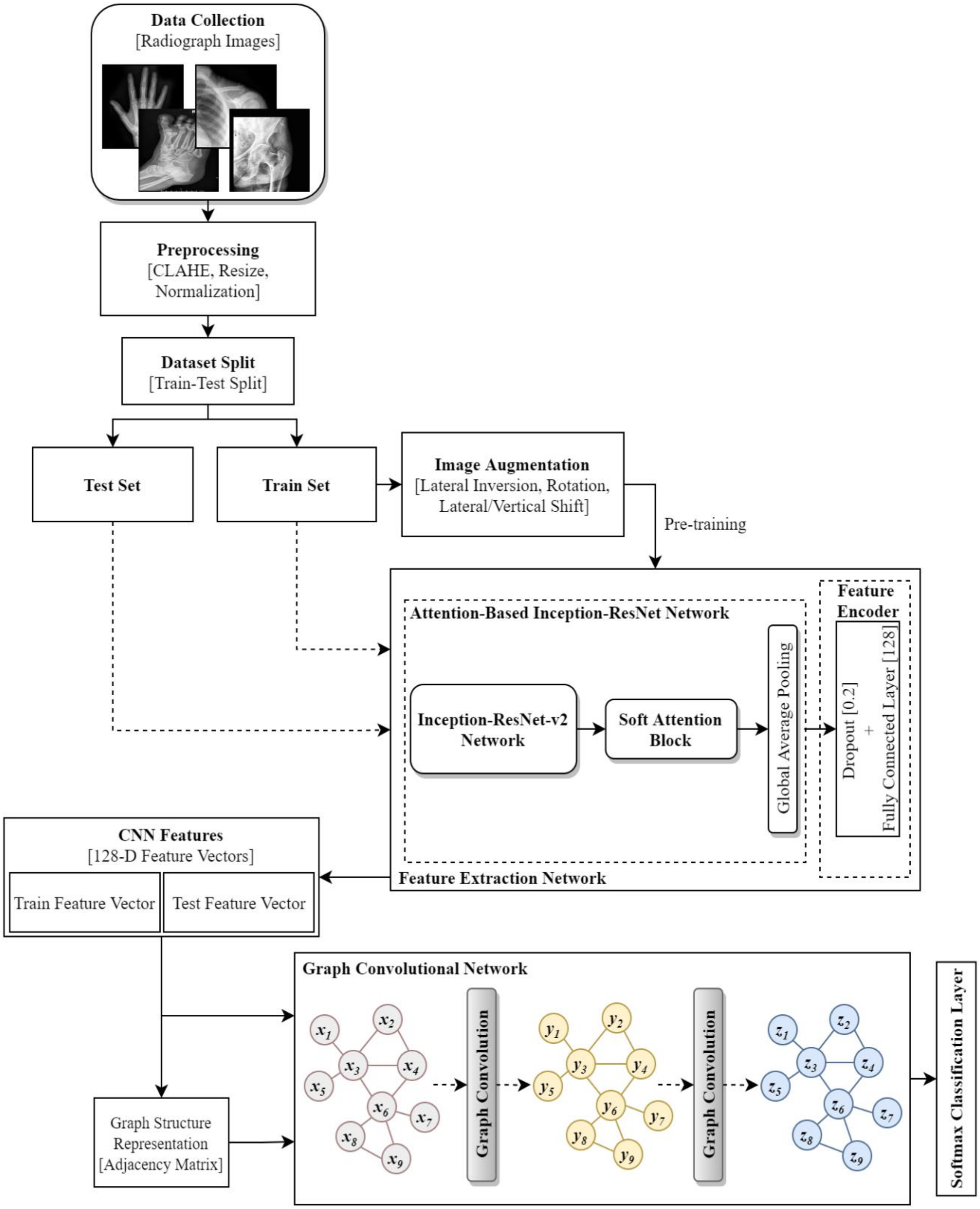

Data Collection
[Radiograph Images]
Preprocessing
[CLAHE, Resize, Normalization]
Dataset Split
[Train-Test Split]
Test Set
Train Set
Image Augmentation
[Lateral Inversion, Rotation, Lateral/Vertical Shift]
Pre-training
Attention-Based Inception-ResNet Network
Inception-ResNet-v2 Network
Soft Attention Block
Global Average Pooling
Feature Encoder
Dropout [0.2]
+
Fully Connected Layer [128]
Feature Extraction Network
CNN Features
[128-D Feature Vectors]
Train Feature Vector
Test Feature Vector
Graph Convolutional Network
x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9
Graph Convolution
y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9
Graph Convolution
z_1 z_2 z_3 z_4 z_5 z_6 z_7 z_8 z_9
Graph Structure Representation
[Adjacency Matrix]
Softmax Classification Layer

Figure 4: Diagrammatic representation of methodology for the classification task.

4.2 Data Collection

The radiographic image data were curatively collected from various local hospitals of Nepal and public repositories of Artificial Intelligence in Medicine & Imaging Center of Stanford University[1], Radiopaedia[2], and Medpix[3]. The collected dataset is comprised of high-quality extremity radiograph images of patients who went under radiographic examination for the diagnosis of musculoskeletal disorder. The dataset, henceforth, named as Xtremity dataset, contains radiograph images of upper and lower extremities – ankle, elbow, finger, foot, hand, hip, knee, and shoulder. The collected radiograph images were labelled manually with the help of NMC-certified radiologist having professional experience of more than five years. The radiograph images were categorized, according to the presence or absence of major musculoskeletal abnormalities in the images, into five classes: Normal, Fracture, Lesion, Arthritis, and Hardware. Figure 5 shows samples of radiograph images of each class from the Xtremity dataset.

The standard benchmark dataset – MURA dataset [11], was collected from the official repository of Stanford Machine Learning Group[4]. The dataset consists of 40,561 multi-view radiograph images, collected from 14,863 studies of 12,173 patients. Each study, comprising of one or more views (images), were labeled manually as either abnormal or normal by expert radiologists from the Stanford Hospital. The dataset included radiograph images of the upper extremity – wrist, humerus, hand, shoulder, finger, elbow, and forearm. The dataset was partitioned into training set, validation set and test set where the training set consisted of 36,808 images from 13,457 studies of 11,184 patients, the validation set consisted of 3,197 images from 1,199 studies of 783 patients, and the test consisted of 556 images from 207 studies of 206 patients. The usage of MURA dataset, in this thesis work, was two folds. Firstly, the MURA dataset being the standard benchmark dataset was used to evaluate the classification results with the implemented network. The evaluated classification results were compared to the results of baseline model. Secondly, the large-scale MURA dataset was used to explore the effect of pre-training the CNN architecture.

[1] https://aimi.stanford.edu/lera-lower-extremity-radiographs-2
[2] https://radiopaedia.org
[3] https://medpix.nlm.nih.gov
[4] https://stanfordmlgroup.github.io/competitions/mura

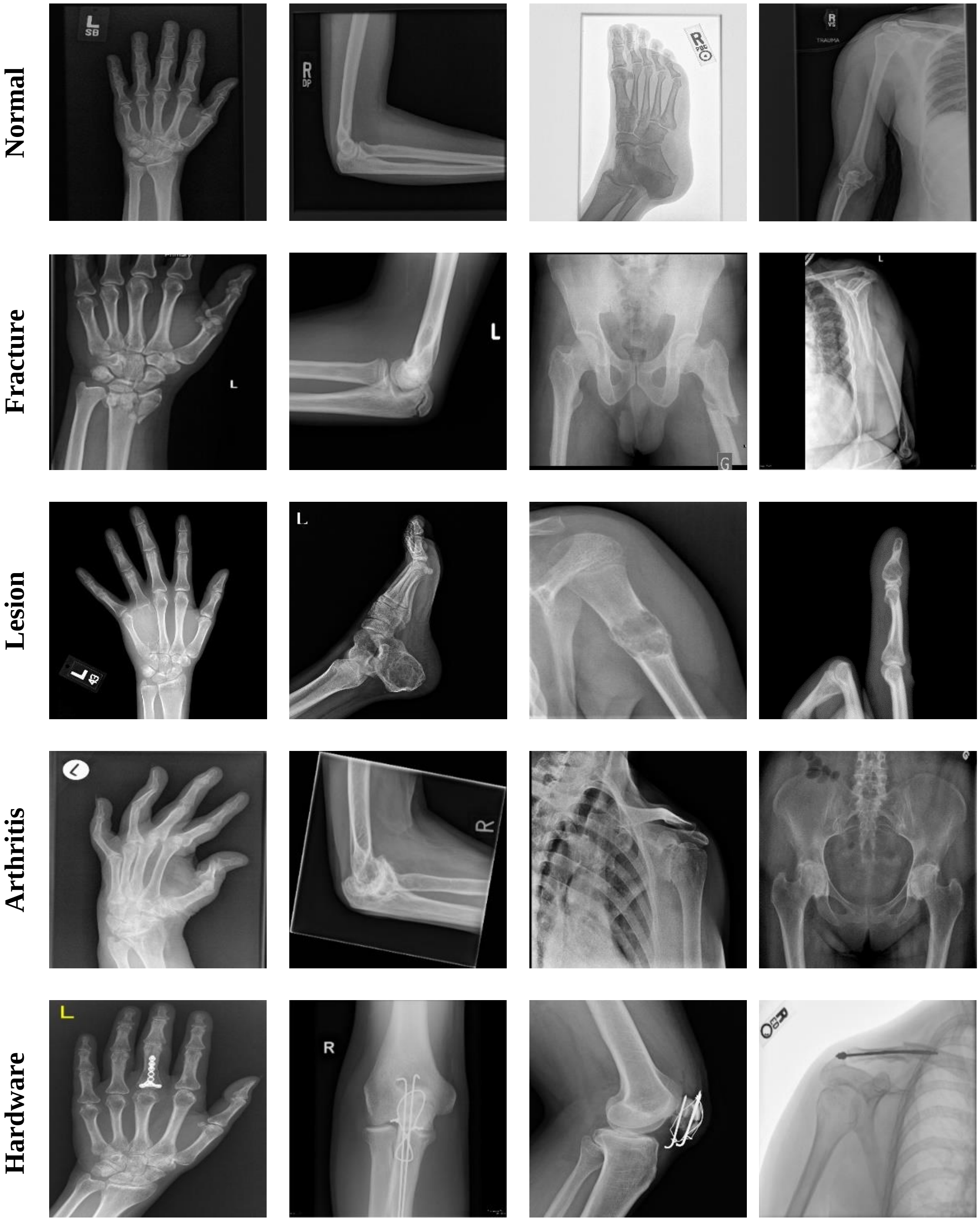


Figure 5: Sample radiograph images from each class of the Xtremity dataset.

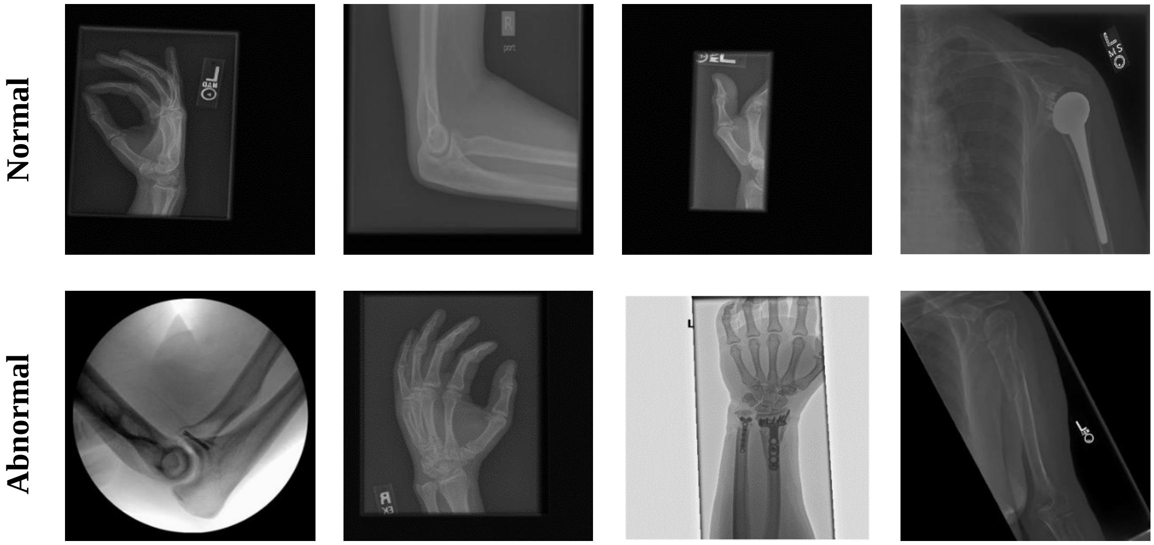


Figure 6: Sample radiograph images from each class of the MURA dataset.

4.3 Pre-processing

As the radiographic images were collected from multiple sources, they had varying sizes, resolutions, and colors. Therefore, comprehensive pre-processing techniques were applied to standardize all images.

- CLAHE (Contrast Limited Adaptive Histogram Equalization) [21] transformation technique was applied to enhance the contrast of radiograph images. It is a modification of the adaptive histogram equalization technique to prevent the tendency to overamplify noise in relatively homogeneous regions of an image by restricting the amplification. First of all, the neighborhood histogram for each pixel was computed in the image. Each histogram was clipped at a predefined value and the clipped histogram was redistributed equally among all the histogram bins. The Cumulative Distribution Function (CDF) and transformation function were computed for each pixel using the clipped histogram. Finally, the transformation function was applied to each pixel to get the equalized image.
- The variable-sized radiographic images were re-scaled to $299x299$ image size. The rescaling was done since the CNN sub-network only accepts the square-shaped images and the particular $299x299$ image format was selected because the

Inception-ResNet-v2 network was trained on ImageNet images of that very image size. The mathematical interpretation of scaling operation on an image is given as:

$$x' = S_x * x \qquad 4.1$$

$$y' = S_y * y \qquad 4.2$$

Here,

(x, y) are the spatial co-ordinates of a pixel in the image,

(x', y') are the spatial co-ordinates after scaling, and

S_x and S_y are scaling factors.

$$S_x = \frac{New\ width\ of\ the\ rescaled\ image}{Actual\ width\ of\ the\ original\ image} \qquad 4.3$$

$$S_y = \frac{New\ height\ of\ the\ rescaled\ image}{Actual\ height\ of\ the\ original\ image} \qquad 4.4$$

- After re-scaling, all the images were normalized pixel-wise so the pixel values in the image ranges between 0 and 1. This process is called min-max normalization. The process of normalization helps to reduce the computational complexity during training the model.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad 4.5$$

Here,

x represents the value of pixels in the image,

$x_{min} = 0$, represents the minimum pixel value, and

$x_{max} = 255$, represents the maximum pixel value in the image.

4.4 Dataset Split

After preprocessing, the curatively collected Xtremity dataset, consisting of 15,701 radiograph images, was divided into approximately 90% train set and 10% test set. The test set was carefully created with best efforts so that the test set comprised of radiograph images of all the extremity body parts in equal proportion. Table 1 shows the distribution of radiograph images in the Xtremity dataset.

Table 1: Distribution of images in Xtremity Dataset.

Class	**Train Set**	**Test Set**	**Total**
Normal	4,138	348	**4,486**
Fracture	2,643	294	**2,937**
Lesion	2,210	246	**2,456**
Arthritis	2,312	257	**2,569**
Hardware	2,927	326	**3,253**
Total	**14,230**	**1,471**	**15,701**

The standard benchmark MURA dataset was partitioned into train set, validation set and test set. Table 2 shows the distribution of radiograph images in MURA dataset.

Table 2: Distribution of images in MURA dataset

Class	**Train Set**	**Validation Set**	**Test Set**	**Total**
Normal	21,935	1,667	290	**23,892**
Abnormal	14,873	1,530	266	**16,669**
Total	**36,808**	**3,197**	**556**	**40,561**

4.5 Augmentation

During the training phase, in order to prevent the model from the problem of overfitting, augmentation technique was applied to introduce diversity to the images in the dataset. The techniques that were applied for augmentation were:

- The images in the train set were laterally inverted, that is, horizontally flipped with random probability of 0.5. The mathematical interpretation is given as:

$$x' = -1 * x \tag{4.6}$$

$$y' = y \tag{4.7}$$

- The radiograph images were randomly rotated up to ±30 degrees. The mathematical representation for rotation is:

$$x' = \cos\theta \ * x + \ \sin\theta \ * y \qquad 4.8$$

$$y' = -\sin\theta \ * x + \ \cos\theta * y \qquad 4.9$$

  Here, θ is the angle of rotation (θ = 30^{o}).

- The radiograph images were shifted laterally and vertically with shift range in the interval [-0.2, +0.2] of the total width and height, respectively.

$$x' = x + \ t_x \qquad 4.10$$

$$y' = y + \ t_y \qquad 4.11$$

  Here, t_x and t_y are the translational factors in horizontal and vertical directions, respectively.

4.6 Ensembled Network Model

An ensembled network of convolutional neural network with attention mechanism and graph convolutional network, hereafter, named as AGCNN network model, was used for the radiograph image classification task. Convolutional neural network was employed for pre-training on the radiograph image dataset and extracting the visual features of individual images. The extracted features from CNN were then fed to the graph convolutional network for exploring the latent correlation among visual features.

The pre-processed radiograph images were fed to an Inception-ResNet-v2 network integrated with soft attention block for the extraction of features, after pre-training on the images. The final classification layer of the Inception-ResNet-v2 network was removed. A soft attention block unit was added to the truncated network. The soft attention block unit was used to focus on the more salient features that are related to the classification task. This was achieved by providing higher weights to feature maps that are more relevant and lower weights to the feature maps that are less relevant to the prediction. After the soft attention block, a dropout layer [24] with drop rate of 0.2 was added. The dropout layer prevents the model from overfitting during training phase by making the neurons less dependent on each other. The dropout layer was then followed by a fully connected layer consisting of 128 neurons. The fully connected layer was used as a feature encoder which converts the higher dimensional feature vectors of the network to 128-dimensional feature vectors. This process of encoding for dimensionality reduction was employed for decreasing the computational complexity.

The final fully connected layer of the modified Inception-ResNet-v2 network, pre-trained by radiograph image classification task, was used to extract the individual image representation features. However, the CNN leaves out the relational representation among a group of images. In contrast, the latent relational representation between radiograph images can be captured by implementing graph convolutional network. Therefore, a GCN network was used to establish the connectivity analysis and augment the relational representation features to the CNN extracted individual image-level features. A two-layered GCN was used to learn over the graph structure and node features, targeting to generate the relational representations of nodes.

The feature vectors with 128-dimensions were extracted from the final fully connected layer after feeding the Inception-ResNet network with radiograph images. Each feature vector which represents an image was considered as a node in graph G for building the graph structure representation for GCN input. Graph (G) is represented by $G = (V, E)$, where V represents the set of nodes (or vertices) in the graph, and E represents the set of edges. Edges in the graph were represented by the adjacency matrix (A). The corresponding element in the adjacency matrix, $A(i, j)$, was set to one when there falls an edge between nodes i and j, otherwise it was set zero. It was assumed that there exists a connection or edge when the node falls into the top *k* nearest neighbors of another node. The nearest neighbors were calculated according to the cosine similarity metric. It characterizes the latent correlations of nodes and discovers the possible relationships among images. The cosine similarity between node *i* and *j* is calculated as:

$$cosine(X_i . X_j) = \frac{X_i . X_j}{|X_i| * |X_j|} \qquad 4.12$$

Here, $X_i \in \mathbb{R}^{1xM}$ and $X_j \in \mathbb{R}^{1xM}$ represent feature vectors of node *i* and *j* of extracted features $X \in \mathbb{R}^{NxM}$.

The adjacency matrix representing the graph structure was constructed as:

$$A_{ij} = \begin{cases} 1, & if\ X_j \in knn(X_i)\ or\ X_i \in knn(X_j) \\ 0, & otherwise \end{cases} \qquad 4.13$$

Here, $knn(X_i)$ represents the *k* nearest neighbors of node X_i based on cosine similarity.

Correspondingly, a degree matrix D, having dimensions NxN which is same as that of adjacency matrix (A), can be calculated as:

$$D_{ii} = \sum_{j=1}^{N} A_{ij} \tag{4.14}$$

Here, D_{ii} is an element of the diagonal degree matrix D.

With the graph structure representation by normalized adjacency matrix and feature vectors, the convolution operation was performed in GCN as defined in equation 3.2. The node representation was improved by the GCN layer by taking the average of all neighbors' features including itself. GCN with two stacked layers were used to capture the latent relational representations out of the CNN extracted features. After each convolution layer, ReLU activation function was applied. Mathematically, ReLU activation function is defined as:

$$ReLU(z) = \max{(0, z)} \tag{4.15}$$

Here, z represents the input vector.

After performing convolution on graph, the nodes were classified into different classes by using a dense layer with softmax function. The mathematical interpretation of softmax activation function is defined as:

$$softmax(z) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{4.16}$$

Here, z represents the input vector, and C represents the number of classes.

The algorithm of the ensembled AGCNN network model is summarized as:

Step-1: Load the 164-layered ImageNet pre-trained Inception Residual neural network – Inception-ResNetV2;

Step-2: Modify the network by replacing the top layer with new layers that are soft attention layer, dropout layer, fully connected layer and softmax classification layer for the radiograph image classification task;

Step-3: Train the modified Inception-ResNet network with the train set of the dataset;

Step-4: Generate the feature vectors through the final fully connected layer in the trained network;

Step-5: Find the top k nearest neighbors for each feature vector based on cosine similarity and build the graph structure which is represented by adjacency matrix;

Step-6: Combine the feature vectors with graph structure representation by multiplicative fusion of feature vectors with the normalized version of adjacency matrix;

Step-7: Train the two-layered GCN with combined feature representation and update the parameters;

Step-8: Finally, classify the nodes representing the images by a dense layer with softmax activation function into different classes.

4.7 Evaluation Metrics

4.7.1 Qualitative Evaluation

The qualitative evaluation of the ensembled AGCNN model was done in two stages. First, attention map was extracted from the soft attention block of the network for visualizing the abnormality region in radiograph image that the network was focusing on for making the prediction for the classification task. Second, the node feature representations that were learned by each node in the GCN network are visualized using t-SNE visualization technique [25].

4.7.1.1 Soft Attention Map

The radiograph images consisted of specific regions that were prominent for the classification task. The soft attention mechanism that was implemented focused on salient regions that contribute more to the prediction score related to the classification task. The attention map from the soft attention module showed where the implemented network was looking before making the prediction. The visualization of attention map represents the visual interpretation of the Inception-Resnet network. Furthermore, rectangular bounding box for localizing the abnormality region was constructed from the contour of the generated attention map. The visualization results of soft attention map were compared to the results that were achieved by implementing Grad-CAM [26] technique.

4.7.1.2 t-SNE Visualization of Node Embeddings

Node Embedding represents the embedding of the nodes into a latent lower-dimensional vector space that captures the information that the network has learnt about the nodes and their neighborhoods. The visualization was done to illustrate the feature representations that were learned from the graph convolutional network. The underlying motivation for node embeddings was to capture the features learned by the graph nodes after training the graph convolutional network. The t-SNE visualization technique was used to represent the node embeddings. The t-SNE visualization is a tool to visualize high-dimensional data. It converts similarities between nodes to joint probabilities and tries to minimize the Kullback-Leibler (KL) divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. The KL divergence is a measure of how similar or different two probability distributions are. The higher the value of the divergence, the more dissimilar are the distributions.

4.7.2 Quantitative Evaluation

The quantitative evaluation of the network signifies the ability of generalization of the network. The network model that is evaluated using one metric may give satisfactory results, however, when evaluated using another metric, it may give unsatisfactory results. The network model was, therefore, assessed in terms of several evaluation metrics to test the model with respect to diversity.

Table 3: Confusion Matrix

		Predicted Class	
		Negative	Positive
True Class	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

True Positive (TP): It represents the number of positive samples that the classifier identified correctly.

True Negative (TN): It represents the number of negative samples that the classifier identified correctly.

False Positive (FP): It represents the number of negative samples that the classifier identified incorrectly.

False Negative (FN): It represents the number of positive samples that the classifier identified incorrectly.

Accuracy: It simply measures the ratio of number of samples that are identified correctly to the overall number of samples.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad 4.17$$

Precision: It is the ratio of number of samples that the classifier model predicted as true, which were actually true.

$$Precision = \frac{TP}{TP + FP} \qquad 4.18$$

Recall: It is the ratio of number of positive samples that the classifier model correctly predicted to the total number of actual positive samples. It is also referred to as sensitivity or hit rate or true positive rate.

$$Recall = \frac{TP}{TP + FN} \qquad 4.19$$

F₁ Score: It is the harmonic mean between recall and precision, which gives a measure of balance between them. The F_1 score ranges from the worst value 0 to the best value 1.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \qquad 4.20$$

AUC Score: The AUC score is equal to the probability that a classifier would rank an arbitrarily chosen positive sample higher than an arbitrarily chosen negative one. The score can be obtained by plotting True Positive Rate (TPR) against False Positive Rate (FPR) at varying classification thresholds. The AUC score ranges from 0 to 1, where 1 means a perfect classifier, 0.5 a random classifier, and 0 a completely wrong classifier.

Cohen's Kappa Score: The Cohen's Kappa score is a more robust metric which aims to measure the degree of agreement between the input and the predictions, excluding the agreement by chance [27].

$$k = \frac{p_o - p_e}{1 - p_e} \qquad 4.21$$

The term $\mathrm{p_o}$ is the observed proportion of agreement, which is same as the accuracy, and the term $\mathrm{p_e}$ is the expected proportion of agreements by chance. For c classes, N samples to classify and $\mathrm{n_{ci}}$the number of times rater *i* predicted class c:

$$p_e = \frac{1}{N}\sum_c n_{c1}n_{c2} \qquad 4.22$$

The Kappa score value ranges between -1 and 1, where -1 represents complete disagreement, 0 represents no agreement or disagreement, and 1 represent perfect agreement.

4.8 Tools and Resources

The tools and resources that were used in this thesis work were:

- Google Colaboratory resource provides free browser-based Jupyter notebook environment. Colab is used because it provides free NVIDIA Tesla K80 GPU with 12 GB RAM.
- Python is used as the programming language for implementation of the thesis work. The libraries used in the work were:
    - **Keras**: The Keras library provides high-level APIs for neural networks. It executes on top of Tensorflow which is an open-source machine learning platform.
    - **NetworkX**: This python library is used for creating, manipulating, and studying the structure, dynamics, and functions of networks.
    - **StellarGraph**: This python library is used for machine learning implementation on graphs.
    - **Scikit-Learn**: This library features various Machine Learning algorithms such as classification, regression, and clustering.
    - **Matplotlib**: It is used for creating static, animated, interactive visualizations. The pyplot module of this library consists of functions that make matplotlib plotting like that of MATLAB.
    - **OpenCV**: This library is used to solve image processing and computer vision related problems.

- **Scipy**: It is used for scientific computing purposes. It contains modules for linear algebra, optimization, integration, interpolation, signal and image processing.
- **Pandas**: It is a python library that provides high-performance data structures and analysis tools that are easy-to-use.
- **Numpy**: The Numerical Python library provides multi-dimensional array objects and functions for their processing.

CHAPTER FIVE: RESULTS AND DISCUSSION

5.1 Preprocessing Results

The radiograph images were pre-processed for making the images appropriate as inputs to the network model. The preprocessing techniques that were applied, in this thesis work, were CLAHE, Rescaling and Normalization.

5.1.1 Contrast Limited Adaptive Histogram Equalization

The radiograph images were, at first, pre-processed by applying Contrast Limited Adaptive Histogram Equalization (CLAHE) transformation technique for enhancing the contrast of the images. Figure 7 illustrates the CLAHE transformation results on the radiograph images. The images in the left indicate the input images before transformation and the images in the right indicate the images after transformation.

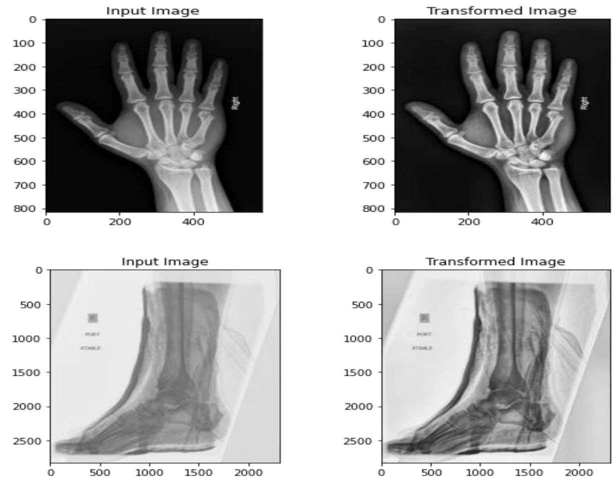


Figure 7: CLAHE transformation of the radiograph images.

5.1.2 Rescaling

The CLAHE transformed radiograph images were, then, rescaled to *299x299* pixel format. Figure 8 shows the rescaled radiograph images of the variable-sized images. The images in the left indicate the input images before performing rescaling operation and the images in the right indicate the output images after rescaling.

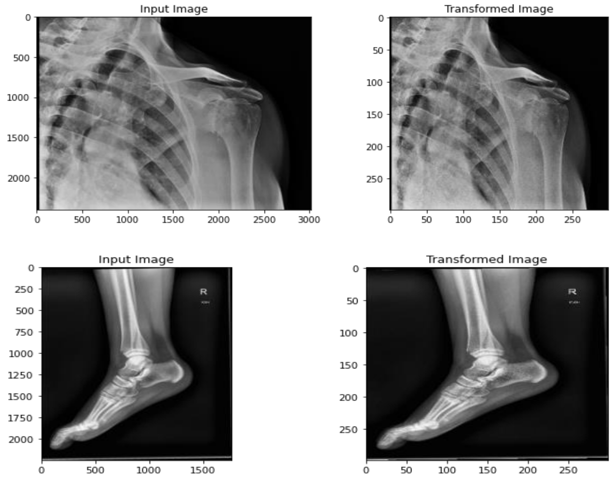


Figure 8: Rescaling of the radiograph images.

5.1.3 Normalization

The pixels of the rescaled radiographic images were, then, normalized so that the values range between 0 to 1. There were no visual changes in the normalized output images as compared to the input images.

The preprocessed radiograph images were then partitioned into train set and test. The samples of the resulting preprocessed radiograph images after the partition are illustrated in Appendix A.

5.2 Augmentation Results

The radiograph images, after pre-processing, were augmented on the fly during training to introduce diversity in the dataset. The techniques adopted for the augmentation were Lateral Inversion, Rotation, and Shifting of the radiograph images.

5.2.1 Lateral Inversion

The preprocessed radiograph images were augmented, during training stage, by randomly flipping the images horizontally. Figure 9 illustrates the horizontally flipped radiograph images. The images in the left indicate the input images before performing horizontal flip and the images in the right indicate the output images after the flip.

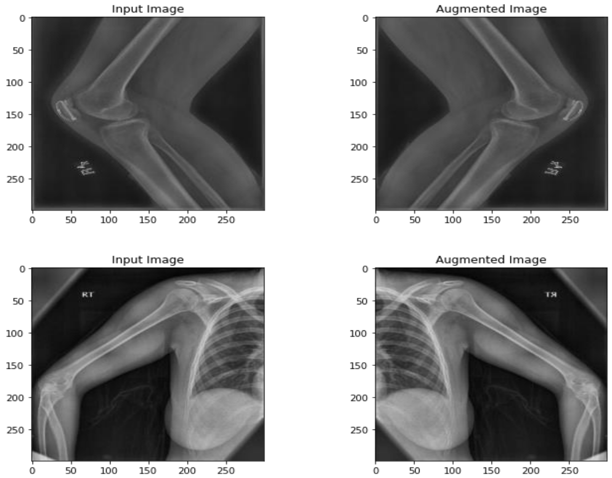


Figure 9: Horizontal flip of the radiograph images.

5.2.2 Rotation

The preprocessed images were augmented by applying random rotation up to 30 degrees. Figure 10 illustrates the radiograph images after performing rotation at different angles. The radiograph images in the left most indicate the input images before performing rotation, the images in the middle indicate the output images after performing anti-clockwise rotation of 15 and 30 degrees, respectively, and the images in the right indicate the output images after performing clockwise rotation of 15 and 30 degrees, respectively.

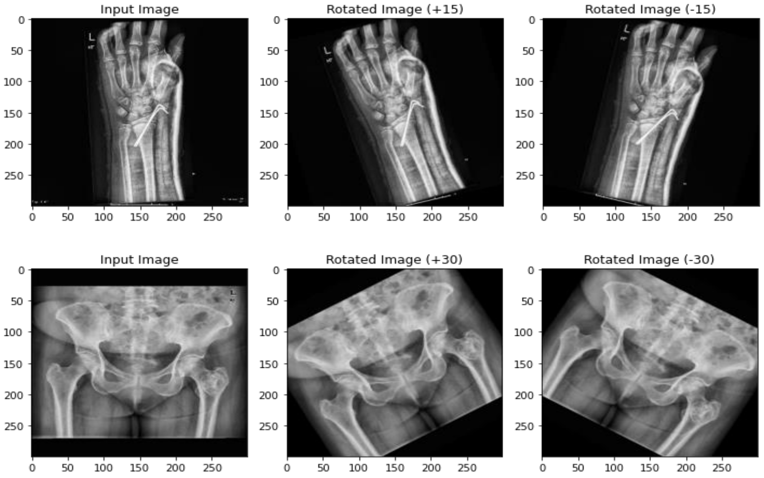


Figure 10: Rotation of the radiograph images.

5.2.3 Shifting

The images were randomly shifted both in horizontal and vertical directions. Figure 11 illustrates shifting operation on the radiograph images in horizontal and vertical directions.

The images in the left indicate the input images before performing shifting operation, the images in the middle indicate the output images after performing horizontal and vertical shifting of 20 pixels, and the images in the right indicate the output images after performing horizontal and vertical shifting by 20 pixels.

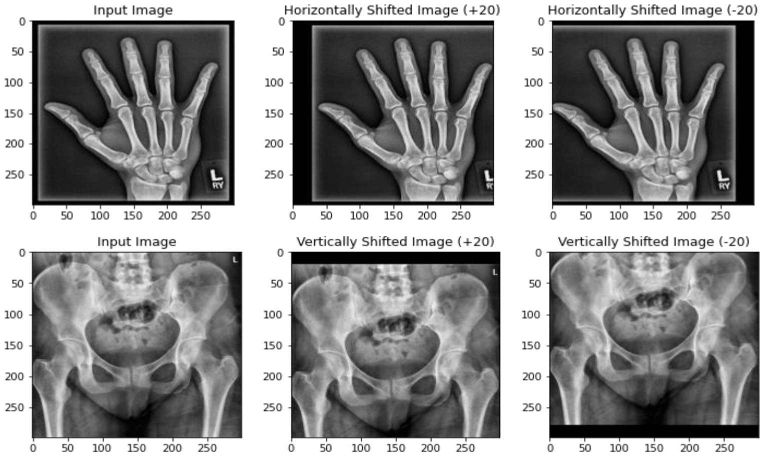

Input Image
Horizontally Shifted Image (+20)
Horizontally Shifted Image (-20)
Input Image
Vertically Shifted Image (+20)
Vertically Shifted Image (-20)
0
50
100
150
200
250

Figure 11: Shifting of the radiograph images.

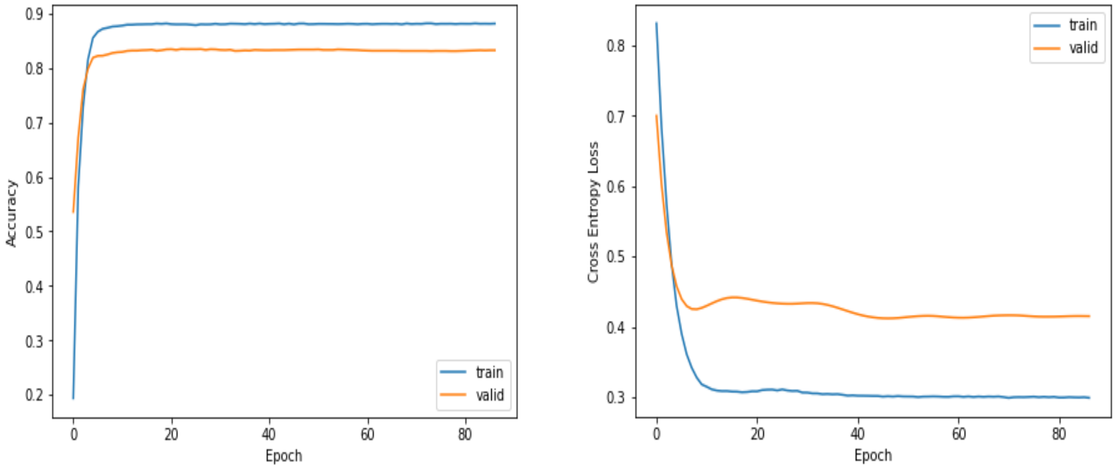
5.3 Experimental Settings

The pre-processed radiograph images were fed to the attention-based inception residual neural network model for pre-training. The network model was trained with batch size of 32. Adam optimizer [28] with cross-entropy loss function was used with an early learning rate of 10^{-4}. After every epoch, the value of learning rate was set to decrease by a factor of 10 whenever there seem no improvement in the validation loss. The early stopping technique was used to prevent the model from overfitting.

After training the modified Inception-ResNet network model with the radiograph images, 128-dimensional feature vectors were extracted from the final fully connected layer – named as feature extraction layer. The adjacency matrix representing the graph structure was constructed by performing *k*-nearest neighbors (*k*-nn) search on every node based on cosine similarity metric. The value of *k* that achieved best results was explored by trying out different values. The extracted features characterizing the individual image-level representation and adjacency matrix representing the graph structure were fed as inputs to the graph convolutional network with two stacked layers of size 128 each for capturing the relational representation. Finally, a dense layer with softmax activation function was used to classify the nodes which represent the radiograph images. The GCN was trained with Adam optimizer with learning rate of 10^{-3}.

5.4 Implementation on MURA dataset

The ensembled graph convolutional neural network was firstly implemented on the standard benchmark MURA dataset. Figure 12 shows the graphical representation of accuracy and loss curves obtained by the GCN sub-network on training set and validation set of MURA dataset.



(a) (b)

Figure 12: (a) Accuracy and (b) Loss curve plots against Epoch.

The graphical representations illustrate that the GCN model converged well achieving stable accuracy and loss. The best model, which is the model with minimum value of validation loss, was taken into consideration for evaluation.

5.4.1 Qualitative Results

5.4.1.1 Localization of Abnormality Regions

The key areas of the radiograph images highlighting the regions of musculoskeletal abnormality were localized by extracting Soft Attention Map from the output of soft attention block of the network. The key area localization was done, by highlighting the class discriminative region that the network focuses, with heatmap and bounding box. Bounding boxes were constructed from the contour of normalized heatmaps to represent real clinical work environment and to make the localization results more evident. The jet color map was used in the heatmap in which the high intensity red color indicates the most salient region where the network actually focused for making the prediction. The localization results of soft attention map were compared with that of Grad-CAM implementation as shown in Figure 13. The images in the leftmost column indicate the input images to the network model. The images in the second column represent the respective soft attention heatmap and Grad-CAM heatmap of the input images. The images in the third column indicate the superimposed images resulting from the combination of input images and their respective heatmaps. The rightmost column comprises the images with bounding boxes enclosing the abnormality regions in the radiograph images.

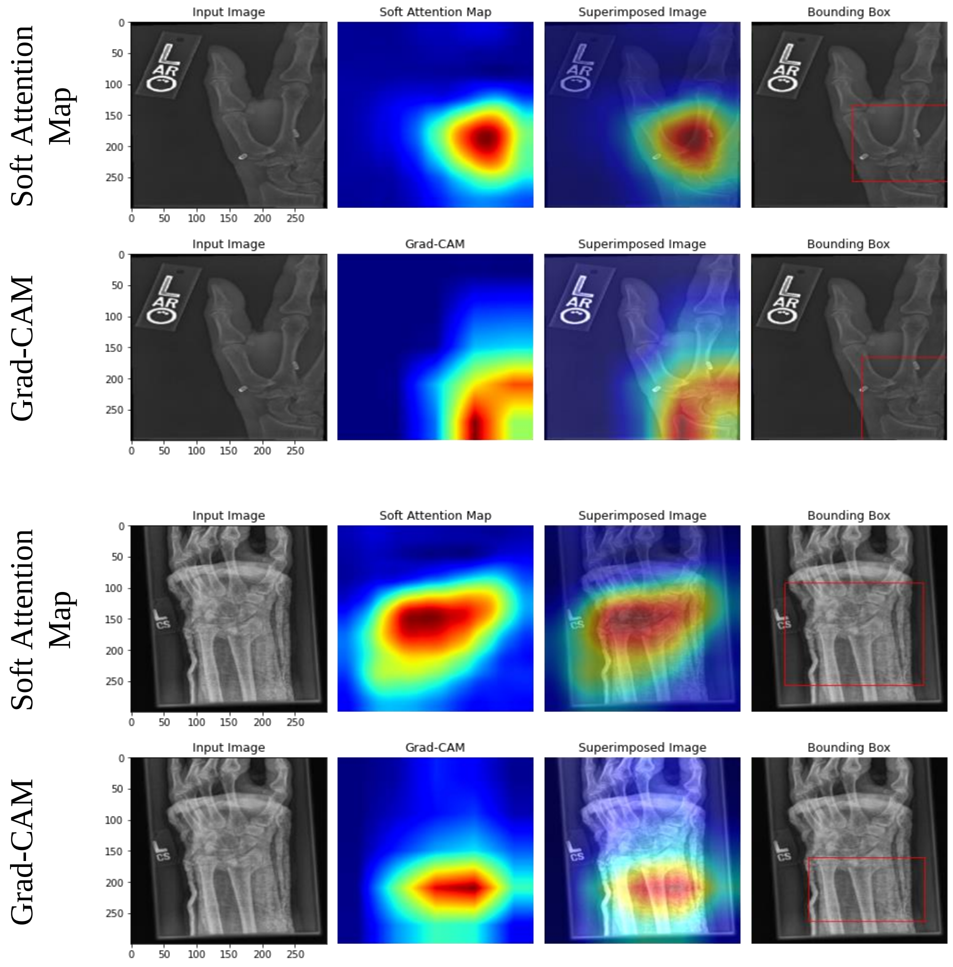
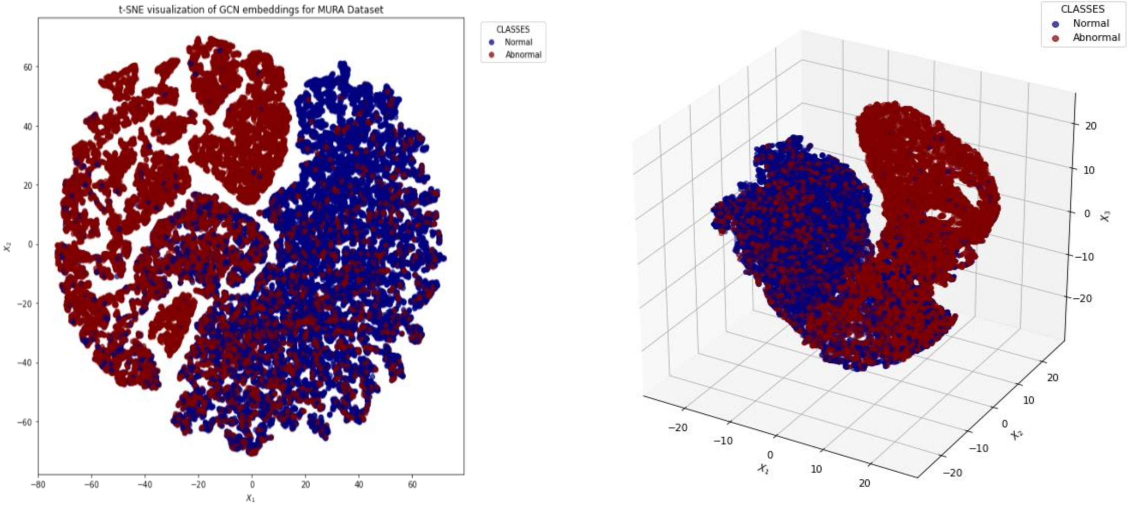


Figure 13: Localization of abnormality regions in the radiograph images.

The key area localization results on the radiograph images of the MURA dataset depicts that the soft attention mechanism improved the network's focusing ability on relevant features of the images. The localization using soft attention maps showed better results than the localization with standard Grad-CAM technique. The key area localization results of some additional radiograph images are illustrated in Appendix B.

5.4.1.2 t-SNE visualization of Node Embeddings

The t-SNE visualization of the node embeddings was done which illustrates the feature representations of nodes that were learned by the Graph Convolutional Network. The visualization was done to get a detailed picture of information that the network has learnt about the nodes and their neighborhoods. Figure 14 demonstrates t-SNE visualization of the GCN node embeddings. The features of all nodes were extracted from the final graph convolution layer of the GCN sub-network. Each node in the visualization represents an individual radiograph image.



(a) 2D embedded space (b) 3D embedded space

Figure 14: t-SNE visualization of GCN embeddings for MURA dataset.

The two well-separated clusters in the t-SNE visualization of the node embeddings represents that the model classified the radiograph images of the MURA dataset efficiently as normal and abnormal.

5.4.2 Quantitative Results

The performance of the graph convolutional network was evaluated on the validation set of the MURA dataset by varying the hyperparameter values. One such hyperparameter considered was the number of nearest neighbors (k) for constructing the graph structure, which was one of the two inputs to the GCN sub-network. Table 4 represents the performance results achieved with varying the value of *k*. The values of *k* considered were first three multiples of 10.

Table 4: Performance results of the network when varying the value of *k*.

	Accuracy	**Sensitivity**	**Specificity**	**AUC score**
k=10	0.835±0.013	**0.789±0.014**	0.876±0.011	**0.897±0.011**
k=20	**0.839±0.013**	0.763±0.015	0.889±0.011	0.893±0.011
k=30	0.838±0.013	0.762±0.015	**0.908±0.01**	0.893±0.011

The values that are highlighted in bold represent the best results of the specific metrics. The values were reported with the 95% confidence interval. The maximum value of sensitivity and AUC score was achieved when the value of *k* was set to 10. Therefore, the graph structure, that was constructed by setting the value of *k* equal to 10, was used as input to the GCN sub-network for further evaluation.

The value of the evaluation metrics in Table 4 were calculated on validation set of 3,197 radiograph images. However, the DenseNet169 baseline model [11] was implemented on holdout test set of 556 images. The baseline model was formed by ensembling the five best models which achieved the lowest validation loss. The test set representations each consisting of 556 images were created by performing random stratified sampling for ten times on the validation set. The performance of the ensembled AGCNN model was calculated by averaging the results obtained on those samples. Table 5 represents the comparison of different values of evaluation metrics achieved by the ensembled network with that of DenseNet169 baseline model.

Table 5: Comparison of the network performance with the baseline implementation.

	Image Size	**Accuracy**	**Sensitivity**	**Specificity**	**AUC score**	**Kappa score**
Baseline [11]	*320x320*	**-**	0.815±0.013	0.887±0.011	**0.929±0.009**	0.705±0.016
AGCNN	*299x299*	**0.856±0.012**	**0.82±0.013**	**0.89±0.011**	0.902±0.01	**0.711±0.016**

The symbol '-' in the table represents the value of accuracy of the baseline model was not mentioned in the baseline implementation [11]. The implemented ensembled network when compared with the DenseNet169 baseline model showed better performance results in most of the metrics even with the smaller input image size. This showed that the ensembled AGCNN model performed better on standard benchmark MURA dataset than the baseline model.

5.5 Implementation on Xtremity dataset

The ensembled AGCNN model was then applied on the Xtremity dataset. Figure 15 shows the graphical representation of accuracy curves and loss curves in both train and test sets of the Xtremity dataset.

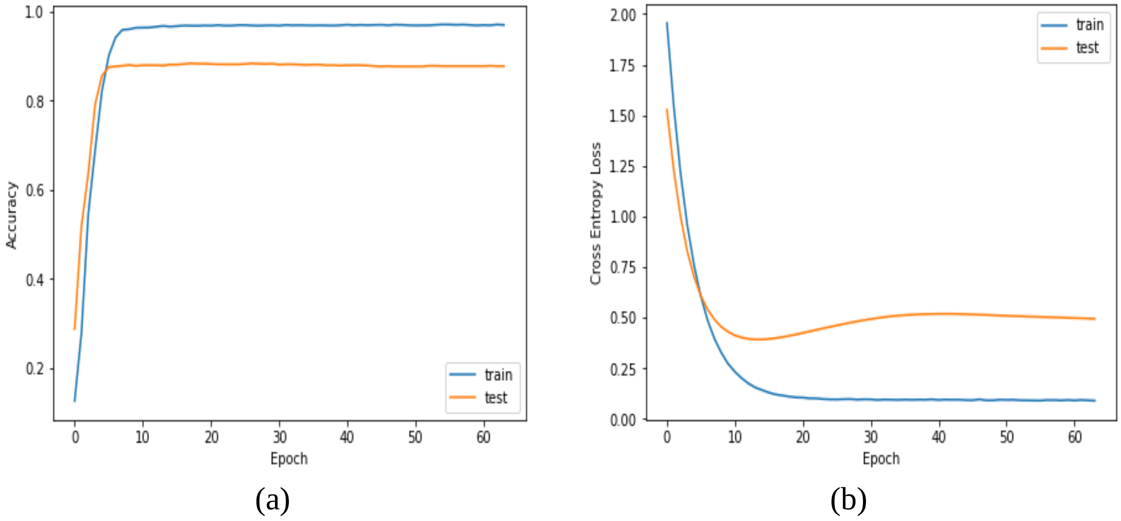


Figure 15: (a) Accuracy and (b) Loss curve plots against Epoch.

The graph curves demonstrate that the GCN model achieved steady results which represent the convergency of the network model. The best model with minimum value of loss was taken into consideration for evaluation.

5.5.1 Qualitative Results

5.5.1.1 Localization of Abnormality Regions

The prominent parts of the radiograph images of Xtremity dataset that illustrate the regions of abnormalities were localized with the Soft Attention Map. Figure 16 demonstrates the localization results, highlighting the discriminative regions that the network concentrates, with heatmaps and bounding boxes. The localization results with soft attention maps were again compared with that of Grad-CAM implementation.

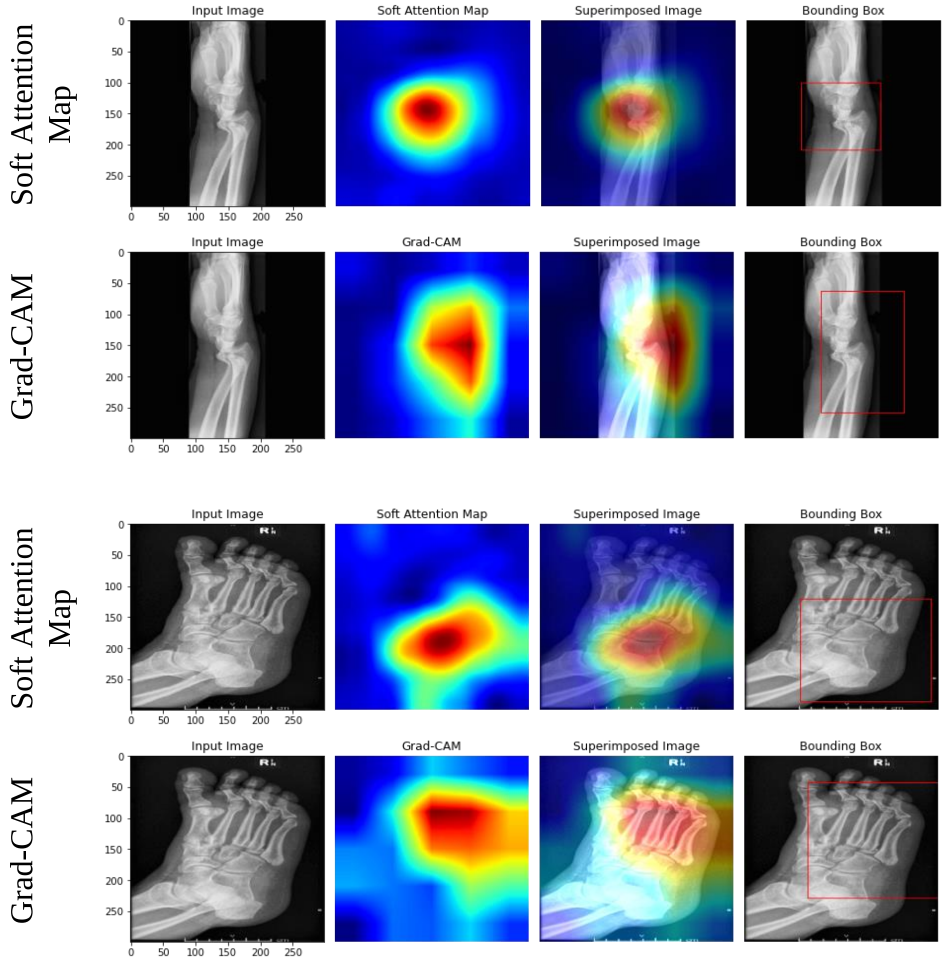
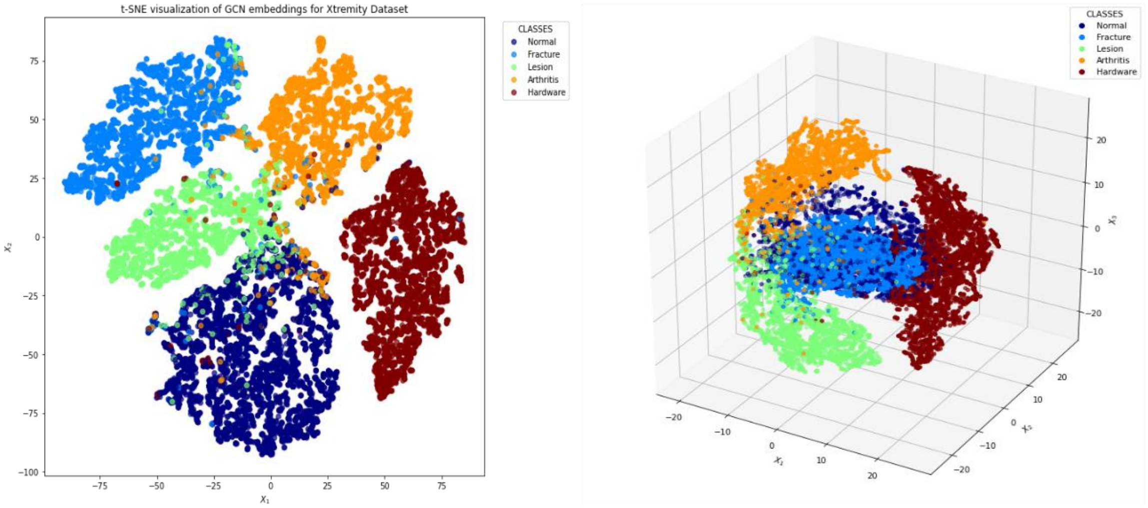


Figure 16: Localization of abnormality regions in the radiograph images.

The localization results highlighting the abnormality regions showed that the soft attention implementation improved the network’s focusing ability of relevant features in the radiograph images of Xtremity dataset as well. The key area localization with soft attention maps showed more compact and accurate results signifying improved results than the localization with standard Grad-CAM technique. The visualization results of some additional radiograph images are illustrated in Appendix B.

5.5.1.2 t-SNE Visualization of Node Embeddings

The t-SNE visualization of the node embeddings was done on Xtremity dataset. Figure 17 illustrates t-SNE visualization of the GCN node embeddings in 2D embedded space (left) as well as 3D embedded space (right).



(a) 2D embedded space (b) 3D embedded space

Figure 17: t-SNE visualization of GCN node embeddings for Xtremity dataset.

The t-SNE visualization illustrates the GCN features formed five distinguishable clusters representing Normal, Fracture, Lesion, Arthritis, and Hardware classes. The fine-partitioned clusters represented that the network classified the nodes, which represent the radiograph images, into five different classes very efficiently.

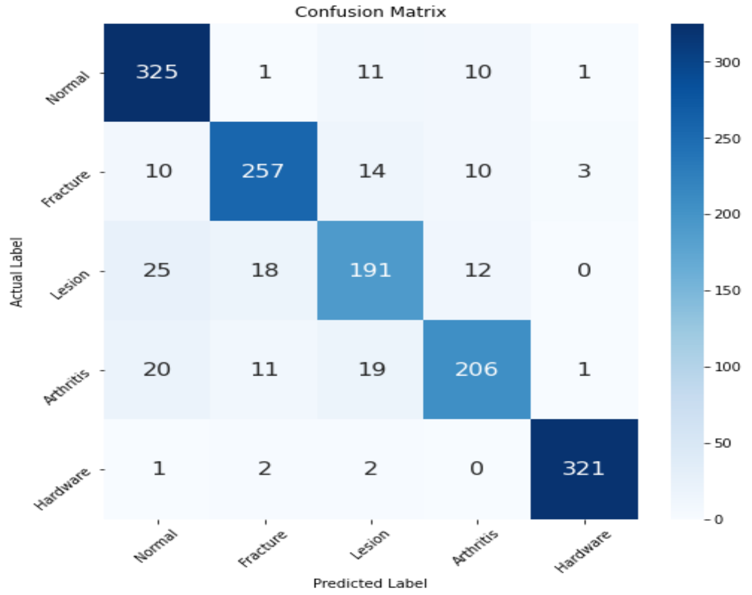
5.5.2 Quantitative Results

The performance of the GCN sub-network was evaluated by varying the hyperparameter value *k*, representing the number of nearest neighbors for constructing the graph structure. Table 6 represents the results achieved with varying the value of *k*. The values of *k* that were considered in the study were first four multiples of 5.

Table 6: Performance results of the network with varying *k*.

k	**Accuracy**	**Precision**	**Recall**	**F$_1$ score**	**AUC score**
5	0.8763±0.0168	0.8733±0.017	0.8652±0.0175	0.8680±0.0173	0.9768±0.0077
10	**0.8838±0.0164**	**0.8797±0.0166**	**0.8741±0.017**	**0.8762±0.0168**	0.9764±0.0078
15	0.8797±0.0166	0.8764±0.0168	0.8681±0.0173	0.8709±0.0171	**0.9769±0.0077**
20	0.8783±0.0167	0.8747±0.0169	0.8677±0.0173	0.8702±0.0172	0.9763±0.0078

The best results were achieved when the number of nearest neighbors for graph structure construction was set to 10. The ensembled graph convolutional neural network model achieved an accuracy of 88.4% and Cohen's Kappa score of 85.38% when evaluated on the test set of Xtremity dataset containing 1,471 radiograph images. The confusion matrix heatmap is shown in Figure 18.



	Normal	Fracture	Lesion	Arthritis	Hardware
Normal	325	1	11	10	1
Fracture	10	257	14	10	3
Lesion	25	18	191	12	0
Arthritis	20	11	19	206	1
Hardware	1	2	2	0	321

Figure 18: Confusion Matrix Heatmap.

The ROC curve obtained after the evaluation is shown in Figure 19.

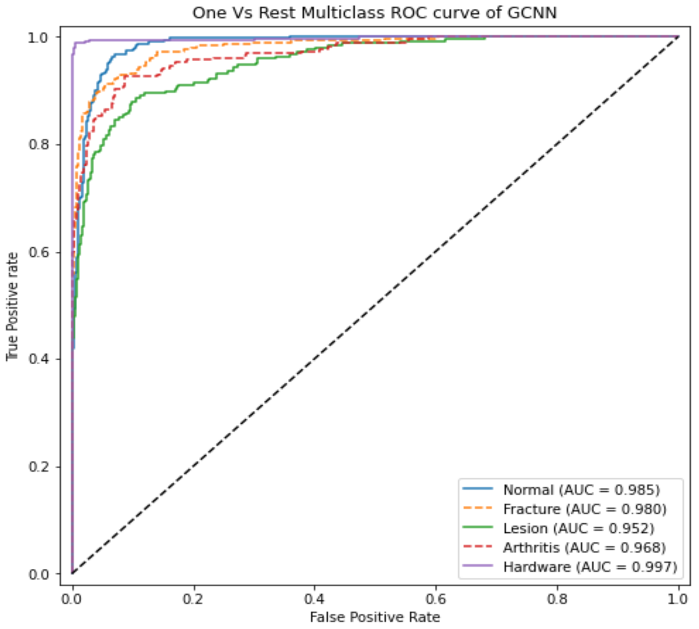


Figure 19: One *vs.* Rest ROC curve for multi-class classification.

The ROC curves of every single class with respect to rest of the classes were incorporated in a single representation with their respective AUC scores. The AUC score ranges between 0.952 of class Lesion and 0.997 of class Hardware.

The results depicted in the confusion matrix heatmap and ROC curves show the ensembled network model classified the radiograph images of class Hardware excellently. The model performed above average in classification of radiograph images of classes Normal and Fracture. However, the model showed average results in classifying radiograph images of classes Lesion and Arthritis. The average classification performance of radiograph images with Lesion and Arthritis might be because of two reasons. Firstly, the number of images of classes Lesion and Arthritis are less compared to images of other classes. It is evident that the performance of deep learning models is directly proportionate to the number of images in the dataset. Secondly, the radiograph images of class Lesion incorporates images of multiple lesion types such as bone cyst, giant cell tumor, osteochondroma, *etc*. Likewise, the radiograph images of class Arthritis incorporates images of three arthritis types – osteoarthritis, rheumatoid arthritis and psoriatic arthritis. The subtle difference in radiographs with these abnormalities might have adversely affected the performance of the ensembled network model.

The mean value of precision, recall and F_1 score evaluation metrics for quantitative evaluation are depicted in the bar graph as shown in Figure 20.

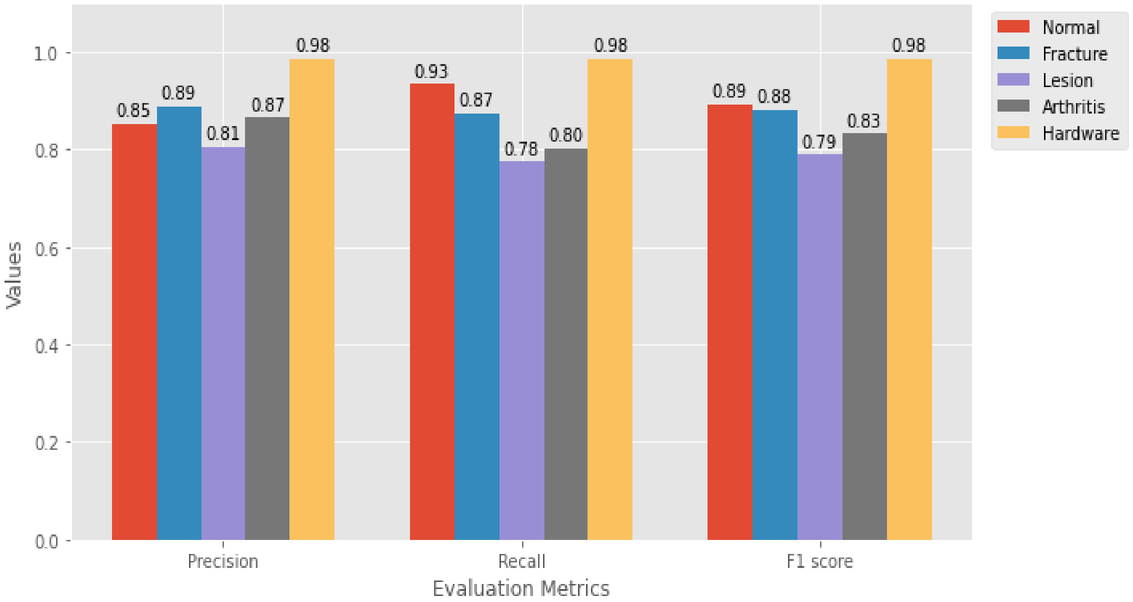


Figure 20: Bar graph showing the performance results of individual classes.

The network achieved lowest value of precision, recall and F_1 score of Lesion classification and highest value of Hardware classification. The evaluation scores achieved by the network for other classes were in between the results of these two classes.

The implemented ensembled graph convolutional neural network achieved par results in overall for the radiograph image classification task.

5.5.3 Ablation Study

The ablation study of the AGCNN network model was also carried out. Different evaluation metrices were taken into consideration for individual network performance analysis. Table 7 shows the performance results of the network with the integration of soft attention mechanism and graph convolutional network into the pre-trained Inception-ResNet-v2 network.

Table 7: Ablation Study

Network	**Accuracy**	**Precision**	**Recall**	**F_1 score**	**AUC score**
Inception-ResNet-v2 (IRv2)	0.853±0.018	0.848±0.018	0.843±0.019	0.844±0.019	0.971±0.009
Soft Attention-based Inception-ResNet	0.872±0.017	0.868±0.017	0.862±0.018	0.864±0.018	0.975±0.008
AGCNN	**0.884±0.016**	**0.879±0.017**	**0.874±0.017**	**0.876±0.017**	**0.976±0.008**

The analytical study showed the soft attention mechanism integration into the pre-trained Inception-ResNet-v2 network improved the classification accuracy by 1.9%. Furthermore, the addition of Graph Convolutional Network resulted in an improvement of overall accuracy by 1.2%. This individual network analysis showed that the ensemble of soft attention mechanism and graph convolutional network achieved improved performance results for the classification.

5.5.4 Effect of pre-training with MURA dataset

The performance of ImageNet pre-trained Inception-ResNet-v2 network was evaluated by pre-training the network with MURA dataset. The evaluation was based on the assumption that the MURA dataset comprised of radiograph images of upper extremity which were more relevant than the natural images of ImageNet dataset. However, the performance of the Inception-ResNet-v2 network degraded when pre-trained with MURA dataset. There might be two main reasons for the degraded performance. First, the MURA dataset consisted of about 40,000 images, however, ImageNet dataset consisted of about 1.4 million images. The deep neural network perform well on large datasets and ImageNet is the largest of them all. Second, MURA dataset included radiograph images of upper extremity only, however, the Xtremity dataset included radiograph images of both upper and lower extremities. The images in MURA dataset were of low resolution whereas the images in Xtremity dataset were of high resolution.

5.5.5 Comparative Study

The comparative study was performed on seven different state-of-the-art pre-trained CNN architectures. The pre-trained architectures were evaluated on different evaluation metrics. All the architectures were trained up to 10 epochs with batch size of 32. Table 8 shows the performance results of different network architectures that were considered in the study.

Table 8: Comparison of the ensembled network with SOTA CNN architectures

Network	**Accuracy**	**Precision**	**Recall**	**F$_1$ score**	**AUC score**
AlexNet *[6]*	0.719±0.023	0.723±0.023	0.709±0.023	0.711±0.023	0.922±0.014
VGG16 *[13]*	0.759±0.022	0.748±0.022	0.739±0.022	0.739±0.022	0.93±0.013
GoogLeNet *[29]*	0.792±0.021	0.787±0.021	0.781±0.021	0.781±0.021	0.954±0.011
ResNet50v2 *[14]*	0.806±0.02	0.803±0.02	0.795±0.021	0.80±0.02	0.957±0.01
Xception *[30]*	0.823±0.02	0.820±0.02	0.810±0.02	0.812±0.02	0.964±0.01
DenseNet121 *[31]*	0.820±0.02	0.824±0.019	0.811±0.02	0.811±0.02	0.964±0.01
IRv2$_{224x224}$	0.831±0.019	0.831±0.019	0.821±0.02	0.822±0.02	0.967±0.009
IRv2$_{299x299}$	0.853±0.018	0.848±0.018	0.843±0.019	0.844±0.019	0.971±0.009
AGCNN	**0.884±0.016**	**0.879±0.017**	**0.874±0.017**	**0.876±0.017**	**0.976±0.008**

After observing the results from the comparison table of different pre-trained architectures, three findings can be inferred. Firstly, the network models performed the classification task better on increasing the number of layers, that is, the depth of the network. In addition to the depth of the network, width of the network also played crucial role in the improvement of the network performance which was illustrated by the better results of wider Xception model than the DenseNet121 model, even though DenseNet121 model is deeper network. Secondly, the input image size to the network model when increased from $224x224$ to $299x299$ format showed better performance results. Lastly, the ensembled graph convolutional neural network showed better classification results which proved that the ensembled network can outperform any single end-to-end pre-trained architectures.

CHAPTER SIX: CONCLUSIONS AND FUTURE WORKS

6.1 Conclusions

An ensembled attention-based graph convolutional neural network (AGCNN) model is effectively implemented for the multi-class classification of musculoskeletal abnormalities in extremity radiographs. The ensembled network, when first implemented on the standard benchmark MURA dataset for the binary classification of radiograph images, achieved better performance results than the baseline implementation. The ensembled network achieved above average performance results with the implementation on Xtremity dataset for the multi-class classification of radiograph images. The ensembled network model achieved an accuracy of 88.38%, average precision of 87.9%, and average recall of 87.4% on musculoskeletal radiograph image classification task. The AGCNN network model achieved high performance results in the classification task despite the large variation of radiograph images of upper and lower extremities. The abnormality region localized on the radiographic images, using soft attention map extracted from the network, were precise and accurate. The localization results of abnormality regions using soft attention mechanism on the radiograph images, when compared to the standard Grad-CAM technique, showed better results. This indicates that automated classification of musculoskeletal abnormalities and their localization has strong potential application in real clinical environments. The automated abnormality classification helps medical professionals to prioritize their worklist giving quicker diagnosis and treatment to patients with critical conditions. The localization of abnormality in the radiographs helps radiologists combat fatigue, which in turn helps them increase their performance.

6.2 Challenges

The most challenging part that was faced, during this thesis work, was the radiograph image dataset collection from various local hospitals and public repositories, and labelling them with the help of radiologists. It took approximately seven weeks to collect the radiograph images from multiple sources and label them with the help radiologists. The current COVID-19 pandemic situation laid many obstacles in this process of data collection. Another challenging part was the implementation of graph convolutional network for image classification task since this is a new approach to the image classification task.

6.2 Future Works

The collection of image dataset is a time consuming and tedious task. Moreover, the collection of medical image dataset is much more challenging. The labelling of the collected medical images is yet another difficult task. Since it is difficult to collect such images in large scale, it might be better opting for the classification technique that can be applied on limited images. One such technique in deep learning is the Few-shot Learning technique. The ensembled graph convolutional neural network can be used for Few-shot classification of images as future work scope.

The X-ray medical imaging is not enough for the accurate detection of wider range of abnormalities. In real clinical environment, orthopedic doctors prefer CT scan and MRI images for the accurate confirmation of more complicated abnormalities. Therefore, such CT scan and MRI images can also be considered as possible future study.

REFERENCES

[1] A. Cieza, K. Causey, K. Kamenov, S. W. Hanson, S. Chatterji, and T. Vos, "Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019", *The Lancet*, 396(10267), 2006-2017, 2020.

[2] M. A. Bruno, E. A. Walker and H. H. Abujudeh, "Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction", *Radiographics,* 35(6), 1668-1676, 2015.

[3] C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, "Inception-v4 Inception-ResNet and the Impact of Residual Connections on Learning," *in AAAI* 2017, AAAI Press, pp. 4278-4284.

[4] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks", *Proc. Int. Conf. Learn. Represent. (ICLR)*, pp. 1-13, 2017.

[5] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[6] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems 25*, pp. 97-1105, 2012.

[7] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs", *Jama*, 316(22), 2402-2410, 2016.

[8] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks", *Nature*, vol. 542, no. 7639, pp. 115-118, 2017.

[9] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3462-3471.

[10] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren and A.Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning", Dec. 2017, [online] Available: https://arxiv.org/abs/1711.05225.

[11] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, et al., "MURA: Large dataset for abnormality detection in musculoskeletal radiographs", *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, pp. 1-10, 2017, [online] Available: https://arxiv.org/abs/1712.06957.

[12] T. C. Mondol, H. Iqbal and M. Hashem, "Deep CNN-Based Ensemble CADx Model for Musculoskeletal Abnormality Detection from Radiographs," *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, Dhaka, Bangladesh, 2019, pp. 392-397. doi: 10.1109/ICAEE48663.2019.8975455

[13] K. Simonyan and A Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *Computing Research Repository*, 2014.

[14] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770-778, Jun. 2016.

[15] Y. L. Thian, Y. Li, P. Jagmohan, D. Sia, V. E. Y. Chan and R. T. Tan, "Convolutional neural networks for automated fracture detection and localization on wrist radiographs", *Radiology: Artificial Intelligence* 1, no. 1 (2019): e180001.

[16] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[17] S. W. Chung, S. S. Han, J. W. Lee, K. Oh, N. R. Kim, J. P. Yoon, J. Y. Kim, S. H. Moon, J. Kwon, H. Lee, Y. Noh and Y. Kim, "Automated detection and classification of the proximal humerus fracture by using deep learning algorithm", *Acta Orthopaedica*, 89:4, 468-473, 2018, doi: 10.1080/17453674.2018.1453714

[18] M. Varma, M. Lu, R. Gardner et al. "Automated abnormality detection in lower extremity radiographs using deep learning", *Nat Mach Intell* **1,** 578–583, 2019. https://doi.org/10.1038/s42256-019-0126-0

[19] X. Yu, S. Lu, L. Guo, S. H. Wang and Y. D. Zhang, "ResGNet-C: A graph convolutional neural network for detection of COVID-19", *Neurocomputing*, 2020.

[20] S. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang and Y. Zhang, "Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network", *Information Fusion* 67, 2021.

[21] Karel Zuiderveld, "Contrast limited adaptive histogram equalization", *Graphics gems*, pages 474–485, 1994.

[22] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention" *In International conference on machine* learning (pp. 2048-2057). PMLR, 2015.

[23] S. K. Datta, M. A. Shaikh, H. Srihari and M. Gao, "Soft-Attention Improves Skin Cancer Classification Performance", 2021.

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting", *The journal of machine learning research*, 15(1), 1929-1958, 2014.

[25] L.v.d. Maaten and G. Hinton, "Visualizing Data using t-SNE", *Journal of Machine Learning Research*, 2008.

[26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[27] McHugh, M. L, "Interrater reliability: the kappa statistic", *Biochemia medica*, 22(3), 276-282, 2012.

[28] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization", *Proc. Int. Conf. Learn. Represent.*, pp. 1-15, 2015.

[29] S. Christian, 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*, pp. 1-9, 2015.

[30] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.

[31] G. Huang, Z. Liu, L. van der Maaten and K. Q. Weinberger, "Densely connected convolutional networks", *arXiv:1608.06993*, Sep. 2016, [online] Available: https://arxiv.org/abs/1608.06993.

APPENDIX A: Pre-processed Images

A.1 Preprocessed radiograph images from MURA dataset

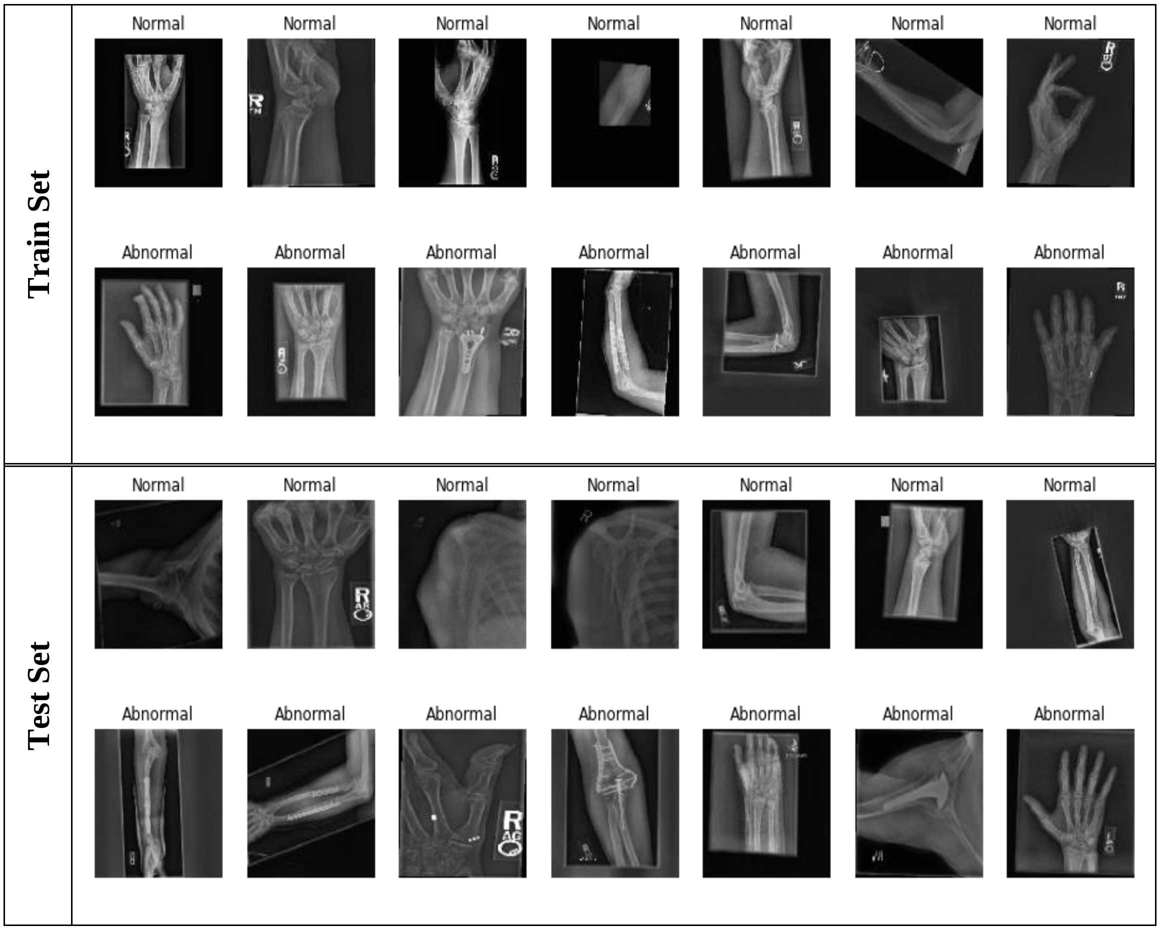


Figure 21: Samples of preprocessed radiograph images from MURA dataset.

A.2: Preprocessed radiograph images from Xtremity dataset

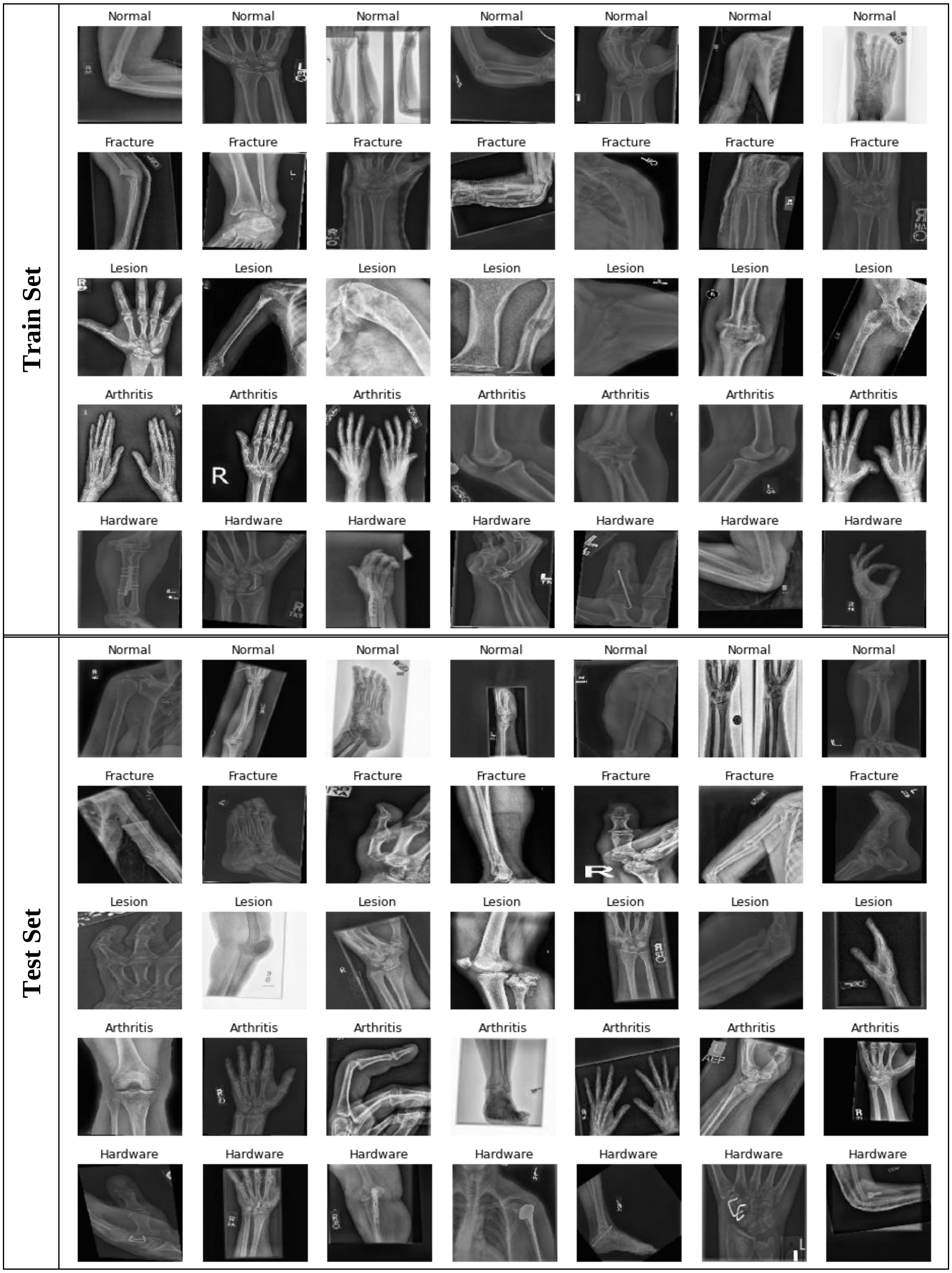


Figure 22: Samples of preprocessed radiograph images from Xtremity dataset.

APPENDIX B: Localization Results

B.1 Localization results on MURA dataset

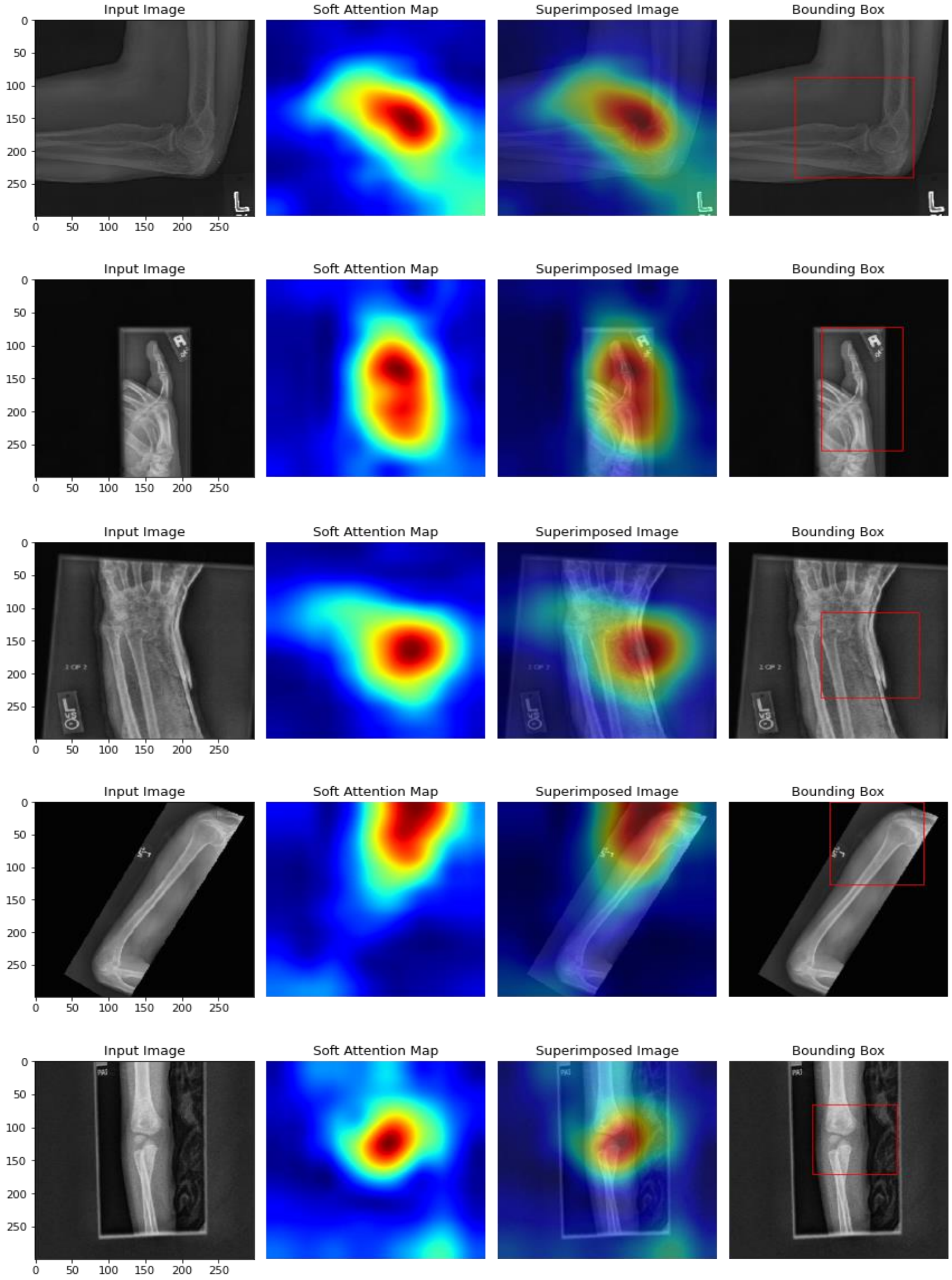


Figure 23: Localization of abnormality regions in radiographs of MURA dataset.

B.2: Localization results on Xtremity dataset

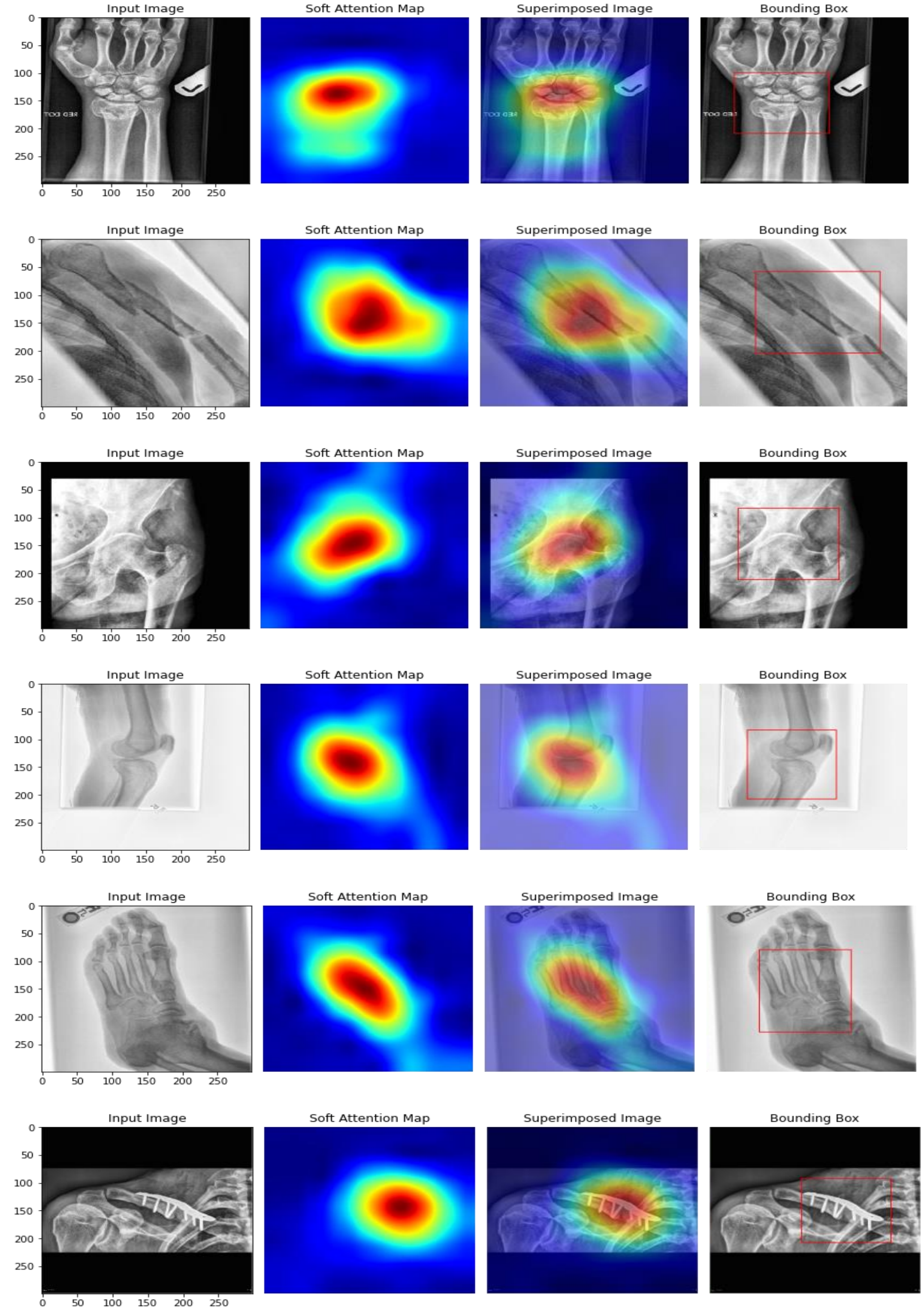
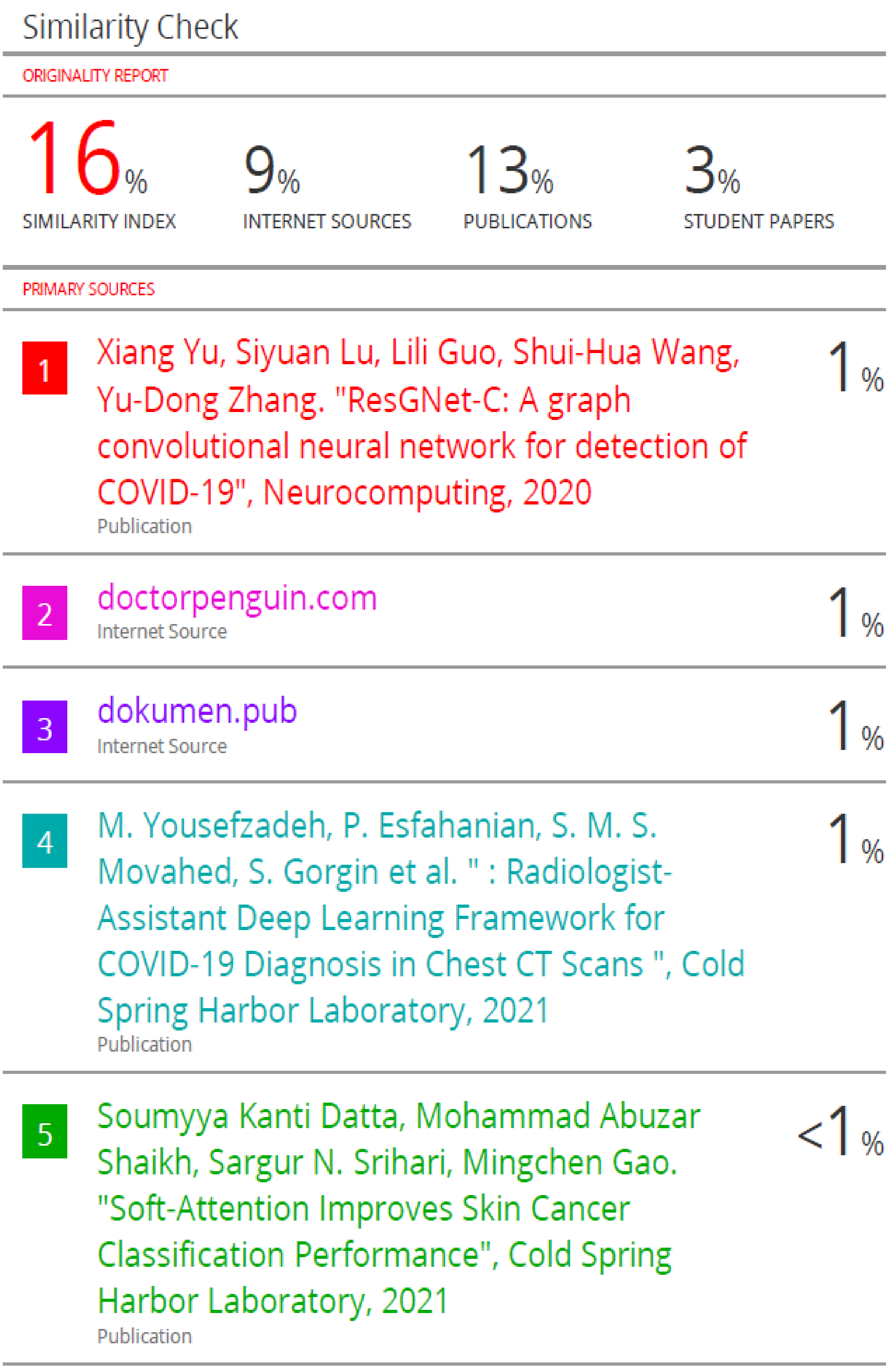


Figure 24: Localization of abnormality regions in radiographs of Xtremity dataset.



Similarity Check

ORIGINALITY REPORT

16%	9%	13%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Xiang Yu, Siyuan Lu, Lili Guo, Shui-Hua Wang, Yu-Dong Zhang. "ResGNet-C: A graph convolutional neural network for detection of COVID-19", Neurocomputing, 2020 Publication	1%
2	doctorpenguin.com Internet Source	1%
3	dokumen.pub Internet Source	1%
4	M. Yousefzadeh, P. Esfahanian, S. M. S. Movahed, S. Gorgin et al. " : Radiologist-Assistant Deep Learning Framework for COVID-19 Diagnosis in Chest CT Scans ", Cold Spring Harbor Laboratory, 2021 Publication	1%
5	Soumyya Kanti Datta, Mohammad Abuzar Shaikh, Sargur N. Srihari, Mingchen Gao. "Soft-Attention Improves Skin Cancer Classification Performance", Cold Spring Harbor Laboratory, 2021 Publication	<1%