Proceedings of 10th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2021 Month: October Volume: 10

Attention-based Graph Convolutional Neural Network for Classification of Musculoskeletal Radiographs

Ganesh Singh Rawal [a], Shashidhar Ram Joshi [b]

[a, b] *Department of Electronics and Communication Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
Corresponding Email: [a] gjanakrwl@gmail.com, [b] srjoshi@ioe.edu.np

Abstract

Musculoskeletal Disorders (MSDs), affecting majority of the world population, are the abnormalities related to bones, muscles and joints. 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 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. The performance of the network model, when compared with the DenseNet169 baseline model, showed improved performance results. The network is then implemented on Xtremity dataset, consisting of 15,701 extremity radiograph images, for the multi-class classification of radiograph images. The implemented network 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 representing 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. Furthermore, the performance results of the ensembled network is compared with that of various state-of-the-art CNN architectures.

Keywords

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

1. Introduction

Musculoskeletal abnormalities involve pain or injuries related to muscles, bones, and joints. These abnormalities, which are broadly known as Musculoskeletal Disorders (MSDs), include fractures, dislocations, degenerative joint diseases, lesions, etc. 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 diagnoses of such abnormalities often require physical examination by radiologists and the inspection of medical images such as X-ray, Ultrasonography, PET scan, CT scan and MRI. Among all of the medical images used for examination, Radiographs (or X-rays), 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 in such examinations.

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 huge number of people with such disorders. Such huge workload can significantly affect the diagnostic performance of radiologist. 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.

2. Related Works

Medical image classification took its pace and attracted many researchers with the advancement of deep learning techniques. The works done in [2, 3, 4] laid significant foundations in the research world and paved a path for future work enhancements on the medical image analysis using deep learning techniques. Gulshan et al. [2] implemented a deep Convolutional Neual Network (CNN) in 128,175 retinal fundus images for the detection of different grades of diabetic retinopathy and diabetic macular edema. The implemented network was validated using 2 different datasets: EyePACS-1 dataset consisting of 9,963 and Messidor-2 consisting of 1,748 retinal images. Their network implementation achieved high performance results on both datasets. Esteva et al. [3] used a deep CNN network for the classification of skin cancer 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. [4] 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. Rajpurkar et al. [5] used a 121-layered densely connected CNN 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.

Rajpurkar et al. [6] released huge MURA dataset, which consists of over 40,000 multi-view radiographic images of 7 study types of upper body extremity. They used a DenseNet169 network for the prediction of abnormality in radiograph images. 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. Varma et al. [7] used a DenseNet161 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.

Almost all of the works related to the classification of radiograph images involved the utilization of CNNs only. CNNs are capable of capturing only the individual image-level representation features. However, they are unable to capture the correlational representation features among a group of images. Graph Convolutional Networks (GCNs) have the capability of capturing the correlational features. The main objectives of this research work are to explore the application of GCN for the improvement in the classification task and to implement the Soft Attention mechanism for the localization of musculoskeletal abnormalities in the radiograph images.

3. Research methodology

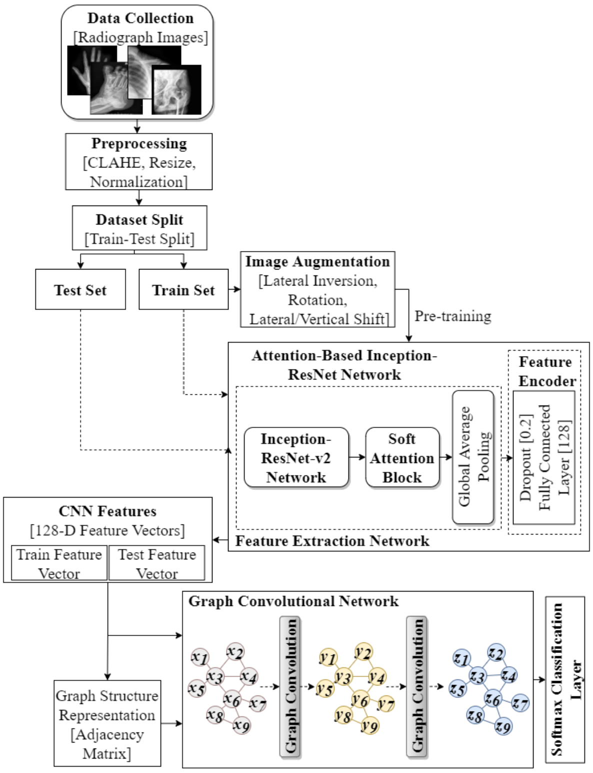
Data Collection
[Radiograph Images]
Preprocessing
[CLAHE, Resize, Normalization]
Dataset Split
[Train-Test Split]
Test Set
Train Set
Image Augmentation
[Lateral Inversion, Rotation, Lateral/Vertical Shift]
Pre-training
Attention-Based Inception-ResNet Network
Inception-ResNet-v2 Network
Soft Attention 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
x1 x2 x3 x4 x5 x6 x7 x8 x9
Graph Convolution
y1 y2 y3 y4 y5 y6 y7 y8 y9
Graph Convolution
z1 z2 z3 z4 z5 z6 z7 z8 z9
Graph Structure Representation [Adjacency Matrix]
Softmax Classification Layer

Figure 1: Methodology for classification of radiograph images

3.1 Dataset Description

The radiograph images were curatively collected from various local hospitals of Nepal [Dr. Iwamura Memorial Hospital[1], BPKIHS[2] and BnB Hospital[3]]

[1] https://iwamurahospital.com

[2] bpkihs.edu

[3] https://bbhospital.com.np

and online public repositories [AIMI[4], Radiopaedia[5], and Medpix[6]]. The collected dataset, henceforth, named as Xtremity dataset, is comprised of high-quality extremity radiograph images of patients who went under radiographic examination for the diagnosis of musculoskeletal disorders. The radiograph images in Xtremity dataset, with the help of radiologists, were categorized into five different classes on the basis of musculoskeletal abnormalities.

Table 1: Distribution of radiograph 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 [6], collected from the official repository of Stanford ML Group[7], was used for comparing the performance of the network model with the baseline implementation. The dataset consists of 40,561 multi-view radiograph images which are labeled manually as either normal or abnormal. The dataset, comprising of upper extremity radiograph images, is partitioned into training set of 36,808 images, validation set of 3,197 images and test set of 556 images.

3.2 Pre-processing and Augmentation

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

- Contrast Limited Adaptive Histogram Equalization (CLAHE) [8] to enhance the contrast of the radiograph images.
- Rescaling to resize the variable-sized images to $229 * 299$ pixel format.
- Normalization to convert the pixel values between 0 and 1 in order to reduce the computational complexity.

In order to prevent the model from overfitting problem, following augmentation techniques were applied, during the training stage, to introduce diversity in the dataset:

- Lateral Inversion of radiograph images with a probability of 0.5.
- Rotation of images randomly up to ± 30 degrees.
- Horizontal and vertical shift with range in the interval $[-0.2, +0.2]$.

3.3 Theoretical Background

3.3.1 Inception-ResNet-v2 Network

Inception-ResNet-v2 network [9] is a 164-layered deep convolutional neural network architecture pre-trained on ImageNet dataset. 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. The concept of an inception module in the Inception-ResNet network is based on 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.

3.3.2 Soft Attention Mechanism

The concept of attention mechanism is employed in neural network architectures 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 [10]. The concept is inspired from the implementation of skin lesion image classification [11] which showed improved performance in the results.

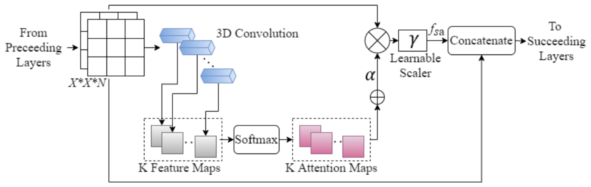

Figure 2: Soft Attention block unit

The feature tensor (t) that streams down the convolutional neural network is fed as input to the

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

soft-attention block unit. The soft attention map is calculated mathematically as:

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

Here, $t \in R^{hxwxd}$ represents a feature tensor that is used as input to a 3D convolutional layer, $W_k \in R^{hxwxdxK}$ represents the k^{th} 3D weight, 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.

3.3.3 Graph Convolutional Network

Graph Convolutional Network (GCN) [12] is one of the many variants of graph neural network family which operates on arbitrarily-structured graph data unlike traditional neural networks which can only be implemented on regular-structured data. The GCN learns the features by aggregating the features from the neighboring nodes. The weighted average of neighbor's feature vectors of each node is taken as represented in red color in the figure 3. 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.

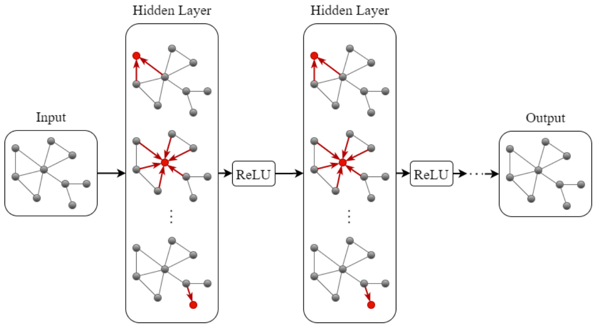

Figure 3: Schematic diagram of Graph Convolutional Network

The propagation rule for each GCN layer is summarized as:

$$H^{l+1} = \sigma(\hat{A}H^l W^l) \tag{2}$$

In equation 2, 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 with individual self-nodes taken 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.

3.4 Ensembled Network Model

The pre-processed radiograph images were fed to an Inception-ResNet-v2 network integrated with soft attention block for the extraction of feature vectors. The final classification layer of the pre-trained 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 were 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 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. After the fully connected layer, a final dense layer with softmax function was added relevant to the radiograph classification task.

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 cosine similarity metric. It characterizes the latent correlations of nodes and discovers the possible relationships among images. The cosine similarity was calculated between node i and j as:

$$cosine(X_i, X_j) = \frac{X_i.X_j}{|X_i| * |X_j|} \quad (3)$$

Here, $X_i \in R^{1xM}$ and $X_j \in R^{1xM}$ represent feature vectors of node i and j of extracted features $X \in R^{NxM}$. The adjacency matrix was constructed as:

$$A_{ij} = \begin{cases} 1, & \text{if } X_j \in knn(X_i) or X_i \in knn(X_j) \\ 0, & otherwise \end{cases} \quad (4)$$

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:

$$D_{ii} = \sum_{j=1}^{N} A_{ij} \quad (5)$$

Here, D_{ii} is an element of 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 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. After performing convolution on graph, the nodes were classified into different classes by using a dense layer with softmax function.

3.5 Evaluation Metrics

3.5.1 Qualitative Evaluation

The qualitative evaluation of the AGCNN model was done in two stages. First of all, soft attention map was extracted from the soft attention block of the network for localizing the key areas in radiograph images that the network was focusing on for making the prediction related to the classification task. Furthermore, rectangular bounding box for localizing the abnormality region was constructed from the contour of the generated attention map. Second of all, the node feature representations that were learned by each node in the GCN network were visualized using t-SNE visualization technique [13].

3.5.2 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.

4. Experimental Setup and Results

4.1 Experimental Setup

The pre-processed radiograph images were fed to the modified Inception-ResNet network for pre-training. The network model was trained with batch size of 32. Adam optimizer 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 IRv2 network, 128-dimensional feature vectors were extracted from the final fully connected 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 achieve the best result was explored by trying out different values. The extracted feature vectors and normalized adjacency matrix were fed as inputs to the two-layered GCN 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}.

4.2 Implementation on MURA

The performance of GCN was evaluated on the validation set of MURA dataset by varying the values of hyperparameter k.

Table 2: Performance results with varying *k*

k	Acc.	Sens.	Spec.	AUC
10	0.835	**0.789**	0.876	**0.897**
20	**0.839**	0.763	0.889	0.893
30	0.838	0.762	**0.908**	0.893

The maximum value of sensitivity and AUC score was achieved when the value of k was set to 10. Therefore, the graph structure, with k equal to 10, was used as input to the GCN network for further evaluation. The baseline model [6] was formed by ensembling the five best models which achieved the lowest validation loss. The baseline model was implemented on holdout test set of 556 images. 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 AGCNN model was calculated by averaging the results on those samples.

Table 3: Comparison of the network with the baseline

	Sens.	Spec.	AUC	Kappa
Baseline [6]	0.815	0.887	**0.929**	0.705
AGCNN	**0.82**	**0.89**	0.902	**0.711**

Table 3 shows that the ensembled network achieved better performance results on most of the metrics when compared with the baseline model.

4.3 Implementation on Xtremity

4.3.1 Localization of Abnormality Regions

The key areas of the radiograph images highlighting the regions of 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 box was constructed from the contour of normalized heatmap 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 shown in figure 4 with the soft attention mechanism showed the network's focusing ability of relevant features of the radiograph images.

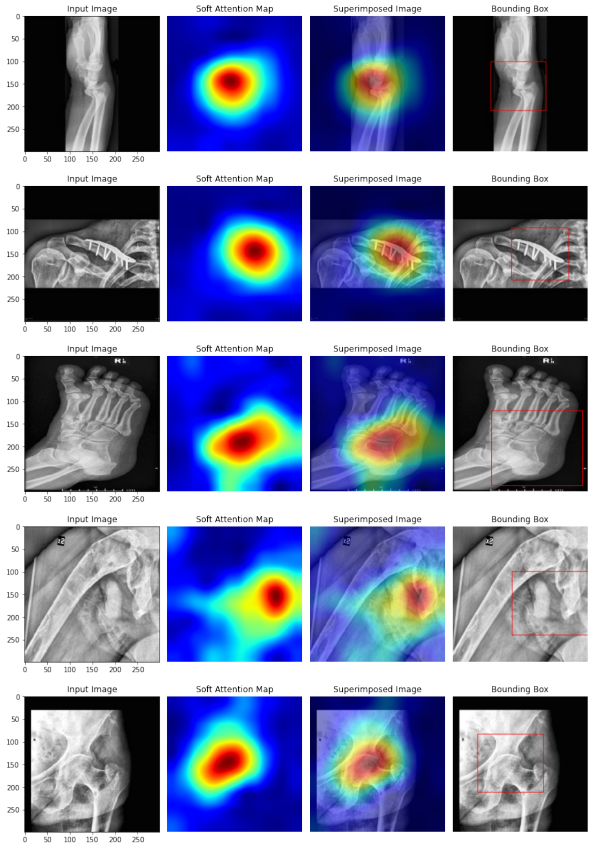

Figure 4: Localization Results

4.3.2 t-SNE visualization of Node Embeddings

The t-SNE visualization of node embeddings was done which illustrates the feature representations of nodes that were learned by GCN. The visualization was done to get a detailed picture of information that the network learnt about the nodes and their neighborhoods. The features of all nodes were extracted from the final graph convolution layer of GCN.

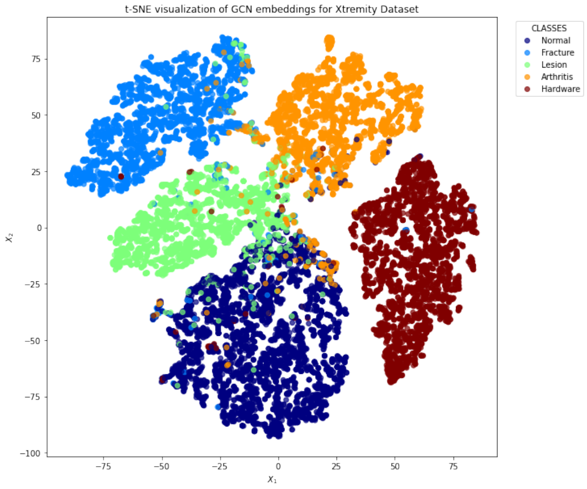

Figure 5: t-SNE visualization of the GCN node embeddings

Each node in the t-SNE visualization shown in figure 5 represents individual radiograph image. The visualization illustrates that the GCN features formed five distinguishable clusters representing five different classes. The fine-partitioned clusters represented that the network effectively classified the nodes, which in turn represent the radiograph images.

4.3.3 Quantitative Results

The performance of the GCN network was evaluated on the Xtremity dataset by varying the hyperparameter *k*.

Table 4: Performance results of the network with varying *k*

k	Accuracy	Precision	Recall	AUC
5	0.8763	0.8733	0.8652	0.9768
10	**0.8838**	**0.8797**	**0.8741**	0.9764
15	0.8797	0.8764	0.8681	**0.9769**
20	0.8783	0.8747	0.8677	0.9763

The maximum value was achieved when the value of *k* was set to 10. Therefore, the graph structure built with setting k equal to 10, was used as input to the GCN network for further evaluation.

4.3.4 Ablation Study

The ablation study of the AGCNN model was also carried out. Table 5 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 5: Ablation Study of the ensembled network model

Network	Acc.	Prec.	Recall	AUC
$IRv2$	0.853	0.848	0.843	0.971
$SA+IRv2$	0.872	0.868	0.862	0.975
$AGCNN$	**0.884**	**0.879**	**0.874**	**0.976**

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

4.3.5 Comparative Study

The comparative study was performed on five most popular 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 6 shows the performance results of different network architectures that were considered in the study.

Table 6: Comparison of the network with SOTA CNN architectures

Network	Acc.	Prec.	Rec.	AUC
$VGG16$[14]	0.759	0.748	0.739	0.93
$ResNet50v2$[15]	0.806	0.803	0.795	0.957
$Xception$[16]	0.823	0.82	0.810	0.964
$DenseNet121$[17]	0.82	0.824	0.811	0.964
$IRv2_{224x224}$	0.831	0.83	0.821	0.967
$AGCNN$	**0.884**	**0.879**	**0.874**	**0.976**

After observing the results of different pre-trained architectures, two findings were deduced. Firstly, the network models performed the classification task better with increasing depth of the network. In addition to the network depth, the width of network also contributed 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 ensembled AGCNN network showed better performance results related to the classification which proved that the ensembled network can outperform any single end-to-end pre-trained CNN architectures.

5. Conclusion

An ensembled AGCNN network model is successfully implemented for the multi-class classification of abnormalities in musculoskeletal radiograph images. The network achieved above par performance results despite the large variations in the extremity radiographs. The ensembled network model also outperformed various state-of-the-art end-to-end pre-trained CNN models. The localization task with soft attention map showed prominent areas representing the abnormality regions on the radiographic images. 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.

References

[1] Alarcos Cieza, Kate Causey, Kaloyan Kamenov, Sarah Wulf Hanson, Somnath Chatterji, and Theo 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] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[3] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.

[4] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

[5] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[6] Pranav Rajpurkar, Jeremy Irvin, Aarti Bagul, Daisy Ding, Tony Duan, Hershel Mehta, Brandon Yang, Kaylie Zhu, Dillon Laird, Robyn L Ball, et al. Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv preprint arXiv:1712.06957*, 2017.

[7] Maya Varma, Mandy Lu, Rachel Gardner, Jared Dunnmon, Nishith Khandwala, Pranav Rajpurkar, Jin Long, Christopher Beaulieu, Katie Shpanskaya, Li Fei-Fei, et al. Automated abnormality detection in lower extremity radiographs using deep learning. *Nature Machine Intelligence*, 1(12):578–583, 2019.

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

[9] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

[10] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR, 2015.

[11] Soumyya Kanti Datta, Mohammad Abuzar Shaikh, Sargur N Srihari, and Mingchen Gao. Soft-attention improves skin cancer classification performance. *arXiv preprint arXiv:2105.03358*, 2021.

[12] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[13] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.