Abnormality Detection in Humerus Bone Radiographs Using DenseNet

Saksham Madan
School of Electronics Engineering
VIT University
Vellore, India 632014
sakshammadan@protonmail.com

Sudhansh Kesharwani
School of Electronics Engineering
VIT University
Vellore, India 632014
skeshu28@gmail.com

K Venkata Sai Akhil
School of Electronics Engineering
VIT University
Vellore, India 632014
akhilkatarirocks@gmail.com

Balaji S
School of Electrical Engineering
VIT University
Vellore, India 632014
sbalaji@vit.ac.in

Bharath K P
School of Electronics Engineering
VIT University
Vellore, India 632014
bharathkp25@gmail.com

Rajesh Kumar M
School of Electronics Engineering
VIT University
Vellore, India 632014
mrajeshkumar@vit.ac.in

2021 Innovations in Power and Advanced Computing Technologies (i-PACT) | 978-1-6654-2691-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/I-PACT52855.2021.9696904

***Abstract*— Treating of injuries and broken bones through reading musculoskeletal radiographs requires a great deal of expertise. It is common that less experienced doctors initially check the radiographs and have a high chance of getting it misdiagnosed. To avoid such misdiagnosis of abnormalities or injury in humerus bone, Deep Learning and Machine Learning algorithms can be applied. Although sophisticated deep learning models have surpassed human capacity under certain computer vision applications, rapid development in the field of medicine has been hampered by a lack of good model applicability and decent marked data, along with other things. This paper seeks to use the model comprehension and visualization methodology to analyze the deep convolution neural network feature removal procedure on the MURA dataset for the identification of anomalies. First, on the selected dataset of humerus radiographs, certain image pre-processing techniques are used to remove variations in size of the image from the radiographs. The following step was to identify the large data as abnormal or normal using the DenseNet-169 architecture. The suggested approach is a reliable technique for classifying bone disorders, according to the findings of the implementation.**

Keywords*— Deep Learning, Radiographs, Abnormality detection

I. INTRODUCTION

The involvement of technology in medical area has never been same before. With improvement of technology, lifespan of human being is increased significantly. One field that has particularly advanced the lot due to this is radiology department. X-ray image has now become the doctor's eye in determining the broken bones. But with the increasing population, the workload of radiologist has increased exponentially. With the improvement, the complexity and small details have also increased. In the United States of America in 2009-2010, close to 105 million emergency ambulatory visits to doctors and hospitals for occurrences of musculoskeletal as well as connective tissue disorders were carried out. Of those visits to healthcare facilities, 39 million were seen by main healthcare officials, 32.4 million were seen by surgeons, and 17 million were visited by medical experts in the field [1]. With the availability of good health service and less time to diagnose diseases, and due to limited training and abilities it might lead to inferior radiologist quality. As the amount of work for radiology department increasing, scientist and students have started taking interest in diagnostic assistance using AI and machine learning for better patient monitoring.

Years ago, the use of deep learning technologies would not have looked feasible due to the computational hardware constraints as well as limitation in the very techniques used to implement the methods of deep learning. But with the advent of open contribution by researchers across the globe to the field of computer vision and deep learning various architectures have been invented. One field of study which has gained the most is medical image analysis. Though increase in computational power may be a major factor in the improvement of results across the board, the invention of efficient techniques and solutions to previous problems have propelled this rise in application of AI in health care industry.

On comparison with earlier times, we have seen rise of computing power and availability of large datasets which has helped in making machine learning algorithms that can in exceedingly complicated operations, outclass human effectiveness in those assignments. All of this has propelled to create better models which can provide better accuracy and this will go a long way in helping the socio-economic problems of poor people unable to receive treatment as it was out of their economic means.

The aim is to create a model that can distinguish between normal and abnormal humerus radiographs. The first dataset chosen was Musculoskeletal Radiographs (MURA) [2], which comprises of x-rays of several areas of the musculoskeletal structure which include wrist, elbow, hands, shoulder, forearm and humerus x-rays. This dataset was used in the proposed model's training and validation. The MURA dataset includes 9,045 normal and 5,818 abnormal upper extremity musculoskeletal radiographic studies. MURA is one of the most comprehensive public radiographic image datasets available. The images of this datasets are passed through the designed model and interpreted. If the humerus bone was broken or contained any abnormality, the model indicated that this radiograph of humerus is different from the normal one.

The contribution of the paper is in effectiveness of the evaluation of the approach to identify the normal or abnormal radiograph in MURA using the proposed framework.

The following is how the remainder of our work in this paper is organised: The second section of this paper provides a concise introduction of different approaches as well as recent literature. Section 3 outlines the procedure as well as the suggested model. The data used is described in section 4. In the same section it attempts to explain the outcome of the research and compares them to other recent documentation. Finally, the paper's conclusion is presented in the section 5.

II. LITERATURE REVIEW

According to [3], the prevalence of Musculoskeletal Disorder's among the adult population in India was found to be in the range of 6.92% - 76.8%, their study showed that this case was mainly seen in both male and female gender, middle age, lower education status performing moderate work, and repetitive hand movements at work like employees in IT Industry. But from [4] there are around more than 1.7 billion individuals worldwide who got afflicted due to Musculoskeletal problems till now. As we know that there are many forms of Imaging for shoulder fracture problems like CT Scans (Computed Tomography), X-Ray Radiography, MRI Scans (Magnetic Resonance Imaging), Ultrasound Imaging, etc, which facilitates diagnosis based on suitable conditions. Imaging are two types which use radiation, and which do not where Radiography comes under the former one.

Many Image based Diagnosis Algorithms and models have been developed to find the probability of detecting Abnormality in Musculoskeletal Systems. Musculoskeletal Imaging is a diagnostic radiology which interprets medical images of bones and joints and respective soft tissues which is used in diagnosing many injuries and diseases. The diagnostic accuracy is the most important factor for a model to be in use to get an estimation with comparison to radiologist performance. A Radiologist finds the treatment plans on after evaluating the radiographs of the musculoskeletal systems as the diagnosis of these abnormalities are quite dependent on the radiographs. Reference [5] which was published in the 2014 on the radiological error categorization system found a clinically significant error category in a datasheet with around 1200-1300 errors due to the insufficiency of the decision factor in the model.

As we know, different machine learning techniques have already figured prominently in medical image classifications. Though the output efficiency after using Support Vector Machine [6] & Decision Forest Image Classification [7] methods in a model was considerate but they fail as they need more computation power and takes a lot of time to train a static model when they are fed with larger datasets or data with more noise when the targets classes get overlapped which leads to inaccuracy and instability. Reference [8] proposed a new deep learning architecture namely Capsule network (CapsNet) which not only involves a new intermediate building block between a neuron and a later which captures many features of the object but also introduces a dynamic routing between the multi-layered capsule networks. But it fails with respect to Convolutional Neural Networks (CNN) because of its crowding effect, higher computation power for implementation (i.e., Slow Learning) and mainly the inability in testing for larger image datasets (ImageNet) which makes it an expensive operation contrary to deep CNN. A model based on the densent-169 architecture was trained by Rajpurkar et al. to identify the abnormalities in Musculoskeletal system which achieved Cohen's Kappa Score 0.600 [2].

This motivated us to develop a more fast and efficient diagnosis using the computer-based system for a software Image diagnosis interpreter model with 169-layered Convolutional Neural Network with a set of benchmark architectures which not only has better computation power and speed in the detection of injuries and diseases but also can be employed for many other parts in musculoskeletal system like elbow, forearm, etc.

III. PROPOSED WORK

To predict the likelihood of abnormality in musculoskeletal radiographs, this paper proposes a 169-layer Convolutional Neural Network with DenseNet [9] architecture. The model was constructed using Keras under a Google Colaboratory environment. The model is pre-trained with weights from the ImageNet [10] dataset. This architecture was chosen due to its high recognition accuracy as stated in Bianco et al.'s [11] analysis. Researchers are interested in DenseNet because it solves the vanishing gradient problem, encourages function reuse, and reduces the number of parameters by a significant amount.

Figure 1 illustrates a conceptual view of the proposed framework. The data has been split into two parts: training and validation, and fed into the Dense Convolutional Neural Network system. Image pre-processing is used first to make the images of the same size. The pre-processed images are then fed into the DenseNet model in the next stage. After training and validation of the model, the performance evaluation of the model is done.

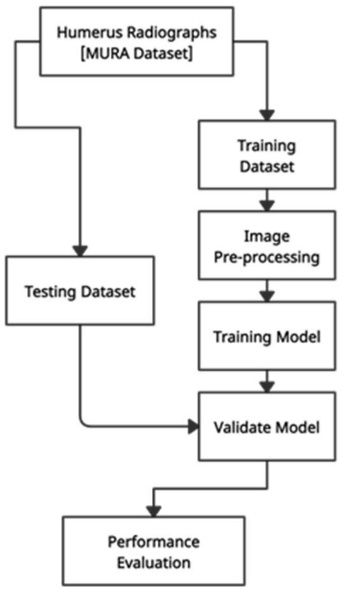

Fig. 1 Conceptual View of the Proposed Work.

A. Image Pre-processing

Since the radiographs are of varying sizes, the MURA dataset includes images that must be resized to a certain number and normalised. The images were standardized to 224x224. The RGB colour is maintained to detect colour features of the images. The resulting images are of shape (224, 244, 3) with 3 channels.

Data augmentation technique [12] is utilized as an important element of the training phase to improve performance and prevent the issue of data overfitting by generating synthetic data. We employed data augmentation techniques to increase the uniqueness of the data as well as the amount of data available, as this will aid in enhancing the model's accuracy by providing it with more information to work with, hence improving overall performance of the model. The dataset might contain images collected under a narrow range of variables, but it may be missing out on a number of variables which we have not considered. In this case, the synthetic data is useful in dealing with these kinds of situations.

Inversions and rotations of 30 degrees at random were applied to the images in the dataset during training, that is a geometric transformation is applied.

B. Network Architecture and Training

The network uses a Dense Convolutional Neural Network architecture, which connects each layer in a feed-forward system. Every layer inside a dense block gets feature maps from all succeeding layers as well as transfers their output to all successive layers, given to the network's feed-forward system. Concatenation is used to join feature maps from different layers. These interconnections generate a dense network of passageways that improve gradient movement. As previously said, as the network depth is greater, the issue of vanishing gradient is much more probable to emerge. The rationale being since the input data and gradient data are passed through numerous layers. This type of dense interconnection has now become comparable to each layer connected directly to input and loss function, reducing the vanishing gradient problem and eliminating the need for deep nets.

The model necessitates lower number of layers as a result of the dense interconnections, because there should be no necessity to grasp redundant feature maps, enabling the network's features to be used again. The dense neural network architecture uses narrow layers to give novel outcomes for feature maps with as few as 12 channels. Models with fewer as well as narrower layers have fewer parameters to master, enabling them simpler to train. A dense block is composed up of numerous dense layers that are connected by dense interconnections. The dense layers are linked together by dense logic, with each dense layer receiving feature maps from all previous layers and passing them on to all future layers.

We replaced the final completely connected layer with a single output layer, then used a sigmoid activation function on the estimates before calculating the loss. Deep learning networks' initial layers capture common features, while later layers concentrate on task-specific features. We exploit previous learning of common features captured in the initial layers and reduce the training data requirements by pre-training the model with weights from ImageNet [10].

The model was trained end-to-end for 12 epochs using a minibatch size of 32. Binary Focal Loss [13] was used as the loss function to be optimized, with the value of gamma(γ) as 2. Focal loss being:

$$FL\ (p_t) = -\alpha_t(1- p_t)^{\gamma} \log(p_t) \tag{1}$$

Where p_t being the probability of ground truth class and γ being the rate at which the easy to classify examples are down-weighted.

When the validation loss reaches a stage of little or no change after an epoch, the initial learning rate of 0.0001 is then decayed by a factor of ten. The model has used the Adam [14] optimizer with parameters of $\beta1 = 0.9$ and $\beta2 = 0.999$.

C. Metrics of Performance Evaluation

Evaluation is an important aspect of any project, whether it is research-based or not. Multiple performance metrics have been assessed in order to estimate the model's capability. The confusion matrix of the proposed model is evaluated, from that True Negative, True Positive, False Negative, and False Positive values are noted. Using these the following metrics are calculated:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4}$$

$$\text{F1 Score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{5}$$

$$\text{Cohen's Kappa} = \frac{\text{Accuracy-Expected}}{\text{1-Expected}} \tag{6}$$

True Negative denoted as "TN" indicates the number of accurately identified negative cases. Similarly, the abbreviation "TP" stands for "True Positive," which denotes the number of correctly identified positive cases. The terms "FP" and "FN" stand for False Positive and False Negative values, respectively. "FP" stands for the number of actual negative cases categorised as positive, and "FN" stands for the number of actual positive cases categorised as negative.

IV. RESULT AND ANALYSIS

We have used the humerus bone radiographs from the MURA dataset to evaluate the model performance in this study made publicly available by Rajpurkar et al. [2]. gathered a large dataset of bone X-rays through an institutional review board-approved analysis and made it publicly available. MURA is de-identified, HIPAA-compliant musculoskeletal radiographs dataset from Stanford Hospital's Picture Archive and Communication System (PACS), consisting of seven types bone X-ray images.

Between 2001 and 2012, board-certified radiologists from Stanford Hospital manually labelled each radiographic study as normal or abnormal. The data was separated into two groups: training and validation. Figure 2 shows humerus bone radiographs which are present in the dataset. Table 1 illustrates the distribution of the humerus data from MURA. The Humerus data taken from the MURA dataset consist of 389 normal images and 338 abnormal images.

The labels used in Rajpurkar et al.'s [2] study is used in this study for training and testing. Table 2 displays the result of the proposed model and table 3 compares the Cohen's kappa against Rajpurkar et al.'s model. Figure 3 shows the confusion matrix obtained after performance evaluation of the proposed model.

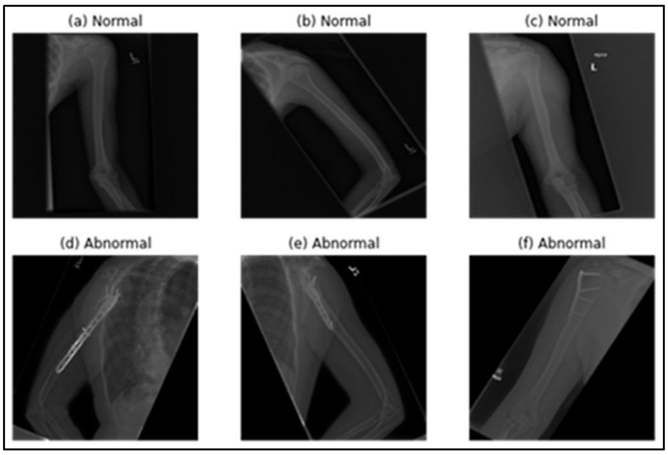

Fig. 2 Images of humerus bone radiograph: (a) to (c) are Normal whereas (d) to (f) are Abnormal.

TABLE I. HUMERUS DATASET

Radiograph	**Train**		**Validation**		**Total**
Humerus	***Normal***	***Abnormal***	***Normal***	***Abnormal***	727
	321	271	68	67	

TABLE II. PERFORMANCE OF PROPOSED MODEL

Accuracy	**Recall**	**Precision**	**F1 Score**
84.03%	0.821	0.846	0.833

TABLE III. COHEN'S KAPPA SCORE OBTAINED FOR HUMERUS BONE RADIOGRAPH

Model	**Cohen's Kappa**
Proposed Model	0.68
Rajpurkar et. al's Model	0.6
Saif et. al's Model	0.6

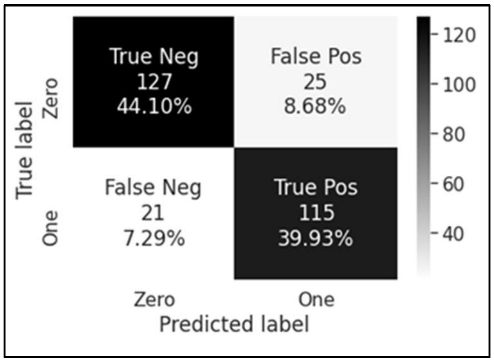

Fig. 3 Confusion Matrix of the Proposed Model

Since humerus radiographs are used for analysis, a Cohen Kappa comparison between the proposed model, [2] and [8] is done. Table 3 summarises the results of the comparison. Cohen's kappa statistic is considered to be more fit in providing more useful information in musculoskeletal research in healthcare [15,16].

The evaluation of model performance on the humerus test data shows an accuracy of 84.028% with a weighted average value of Recall, Precision and F1 Score being 0.821, 0.846 and 0.833 respectively. The proposed model has a Cohen's kappa score of 0.680 compared to Cohen's kappa score of 0.600 obtained by Rajpurkar et al.'s model, performing significantly better. Reference [8] obtained a Cohen's kappa score of 0.600 for DenseNet [9] model trained on humerus bone radiograph obtained from the MURA dataset [2].

V. CONCLUSION

For fast and efficient diagnosis in the field of medicine, computer-based diagnosis systems will play a major and important role. Hence building efficient systems with feasibility, high computation power in detecting diseases and utilizing them to whole extent is important. On testing the proposed model using the test dataset, Cohen's Kappa score is found to be 0.680. In comparison to Rajpurkar et al.'s model, the proposed model outperforms it. Other sections of the dataset, such as the wrist, elbow, etc., can also be included for the proposed model.

The current research is confined to developing models to identify a normal and abnormal humerus radiograph in two classes. The purpose of this research is to demonstrate the potential of the methodologies utilized in developing models that may be used as preliminary screening tools to aid physicians. Even though the MURA is the largest open collection of musculoskeletal upper extremity radiographic images, since it is on the lesser part of the training curve's power law area thus performance may be enhanced with further samples to the training set. With larger datasets and new incoming relevant information produced every second, the suggested model can be retrained over time to increase the model's efficiency and performance for the specific radiological area.

REFERENCES

[1] L. Robert, "Musculoskeletal hits top 10 in primary care visits," *Patientcareonline.com*. [Online]. Available: https://www.patientcareonline.com/musculoskeletal-disorders/musculoskeletal-hits-top-10-primary-care-visits. [Accessed: 31-May-2021].

[2] P. Rajpurkar *et al.*, "MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs," *arXiv [physics.med-ph]*, 2018.

[3] J. Kishore and Director Professor & Head, Department of Community Medicine, VMMC & Safdarjung Hospital, New Delhi, "Prevalence of musculoskeletal disorders amongst adult population of India," *Epidemiol. int.*, vol. 04, no. 03, pp. 22–26, 2019.

[4] C. Paul, "The burden of musculoskeletal conditions at the start of the new millennium. Report of a WHO scientific group. Geneva: WHO technical report series, 919, 2003, pp. 218. ISBN: 92-4-120919-4," *Int. J. Epidemiol.*, vol. 34, no. 1, pp. 228–229, 2004.

[5] Y. W. Kim and L. T. Mansfield, "Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors," *AJR Am. J. Roentgenol.*, vol. 202, no. 3, pp. 465–470, 2014.

[6] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Trans. Med. Imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.

[7] A. Criminisi, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Found. Trends® Comput. Graph. Vis.*, vol. 7, no. 2–3, pp. 81–227, 2011.

[8] A. F. M. Saif, C. Shahnaz, W.-P. Zhu, and M. O. Ahmad, "Abnormality detection in musculoskeletal radiographs using capsule network," *IEEE Access*, vol. 7, pp. 81494–81503, 2019.

[9] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *arXiv [cs.CV]*, 2016.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[11] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *arXiv [cs.CV]*, 2018.

[12] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv [cs.CV]*, 2017.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv [cs.LG]*, 2014.

[15] J. Sim and C. C. Wright, "The kappa statistic in reliability studies: use, interpretation, and sample size requirements," *Phys. Ther.*, vol. 85, no. 3, pp. 257–268, 2005.

[16] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Fam. Med.*, vol. 37, no. 5, pp. 360–363, 2005.