



Musculoskeletal Abnormality Detection in Humerus Radiographs Using Deep Learning

Namit Chawla*, Nitika Kapoor

Computer Science and Engineering Department, Chandigarh University, Mohali, Punjab 140413, India

Corresponding Author Email: 18MCS1033@cuchd.in

<https://doi.org/10.18280/ria.340212>

Received: 26 December 2019

Accepted: 18 February 2020

Keywords:

deep learning, computer based diagnosis, image based diagnosis, ensemble learning, abnormality detection

ABSTRACT

Musculoskeletal radiographs bring a considerable amount of meticulous expertise in treating Bone diseases (BDs) or injuries. Usually, less experienced doctors are the first ones for assessment of radiographs and it is not surprising for humerus disorders being misdiagnosed. To take care of such misdiagnosis, Deep Learning and Machine Learning could play a major role in diagnosis of the musculoskeletal abnormalities. The presented paper intends to develop a better performing Computer Based Diagnosis (CBDs) model. First, some preprocessing techniques are performed on the chosen dataset of humerus radiographs, eliminating image size variability from the radiographs. Next, two architectures namely- DenseNet201 and Inception V3 were used to classify the given dataset as abnormal or normal. Later, ensemble techniques are applied to improve model's performance. The proposed technique is tested for the publicly available Musculoskeletal Radiographs (MURA) dataset and the qualifier results are compared with present results from the reference paper. For humerus radiographs, the accuracy achieved is 88.54%. Implementation results show the proposed method is a deserving strategy to classify bone disorders.

1. INTRODUCTION

Musculoskeletal disorders could be caused from situations or conditions resulting from accidents which could be sports injuries, accidents and many more. They could be present in one's musculoskeletal system due to genetics i.e. by birth and are common among children. These disorders are due to deformation or malformation of bones. Deformation could be from reshaping due to excess or unusual pressure. Malformation can be thought as of an error in an organ or tissue development. The abnormalities may affect one or a combination of bone and muscle development in limbs, skull etc. Also, proper detection and diagnosis are essential for further treatment.

These abnormalities invade the muscles, bones, ligaments, tendons, discs etc. About 20% - 30% of people around the world live with taunting musculoskeletal abnormalities. Various diagnosis tests are there for abnormality detection for different parts of a musculoskeletal system. For bones, radiographs or bone scans can be used. For muscular disorders, Electromyography (EMG) or biopsy can be used. And for joint disorders, radiography, Magnetic Resonance Imaging (MRI), arthroscopy and many more are available. As per context of the research area, diagnosis of bone disorders is somewhat contingent on radiographs of the bones. And these radiographs are then assessed by radiologists for the latter phase of treatment. A significantly large number of patients has made it tough and the process is time consuming for proper diagnosis and treatment. Thus, computer based abnormality detection can be of good use and more time efficient. The motivation is to proposed and build a computer based decision support system for more accurate diagnosis of such problems. Discrepancy rates of 26% major inter- observer and 32% intra-

observer were found in a study conducted in 2010, from Massachusetts General Hospital [1].

The aim is to develop a model or a system that classifies the radiographs as normal or abnormal. To do so, Image Based Diagnosis (IBDs) has been used rather than feature based diagnosis because image based diagnosis is known to be more accurate [2]. At first, a dataset named Musculoskeletal Radiographs (MURA) has been collected which consists of radiographic images of different parts of musculoskeletal system namely- elbow, finger forearm, hand, humerus, shoulder and wrist radiographs. This dataset has been used for both training and validation of the proposed CBDs model.

Previously, various machine learning algorithms has played a major contribution in medical image classification. Both Support Vector Machine (SVM) and Decision Forest have shown impressive outcome in image classification. Many clustering algorithms are very popular for image classification. In the same manner, Convolutional Neural Networks (CNN) has also been used abundantly in image classification problems. But, a large dataset (in ten thousand) is required for a CNN model so as to attain a decent level of accuracy.

Several pre-processing techniques such as resizing, filtering, rescaling have been applied on the given dataset. Two Deep Convolutional Neural Networks (Deep CNN) architectures (DenseNet201 and Inception-v3) [3, 4] have been used to construct a few models.

Contribution to this work can be summarized as follows:

- Some important preprocessing techniques are applied on radiographs.
- A set of benchmark architectures are applied to ensure finest performance on results.
- Several models are combined to form an ensemble to improve performance.

- Comparison and analysis are made between proposed technique and other famous classification techniques.

The rest part of the paper is structured as follows: the section 2 of the paper presents a brief description of various methods and the related work. The section 3 describes about the dataset being used for the experiment and the conceptual view of proposed framework. The section 4 of the paper, describes about the dataset and experimental setup. The section 5 of the paper, aims to disclose the results of the experiment and comparison with other present results. At last, the final section is conclusion of the paper.

2. RELATED WORK

The major part of the conduct has been centralized on discovering the abnormalities present in the musculoskeletal system using radiographs. Mondol et al. developed a Computer Aided Diagnosis (CADx) model with the help of Deep convolutional neural networks (Deep CNN) which helps in the diagnosis of the abnormalities. In this work, VGG-19 and ResNet architectures were employed to develop an underlying model for the research. The proposed architectures were an ensemble which was called as CADx model and was observed to perform comparatively better than both VGG-19 and ResNet architectures. The presented paper was concluded with the comparison of results from MURA and their proposed CADx model [5]. Saif et al. proposed a new architecture named capsule network. Capsule networks have the capability of being trained using very less number of training data and that is why, they can be employed for problems with relatively less number of images for a dataset [6]. Thian et al. proposed a model to understand how feasible and how well performing a CNN for fracture discernment and localization on radiographic wrist images could be. The dataset was split in 9:1 ratio for training and validation, for both front and back sides of radiographs. The research proposed a model based on Inception-ResNet and a Faster R-CNN architecture was accomplished as a final model. The proposed model was tested on a 524 radiographs of wrist [7]. Rajpurkar et al. trained a 169-layer Dense as a base model to identify and localize bone disorders(abnormalities). The model achieved a ROC score of 0.929, with 0.815 sensitivity and 0.887 specificity [8]. Chung et al. tried to find out the capability of artificial intelligence in the field of healthcare studies to identify and classify proximal humerus fractures using their own dataset. 1,891 images of normal shoulder’s radiographs and 4 proximal humerus fracture types were used in the experimental setup. These fractures were classified by specialists for standard reference and they were evaluated for the final results. The fracture types were classified after the exclusion of normal shoulder radiographs [9]. Spampinato et al. proposed several automated propositions for the assessment of skeletal bone age in an automated manner. This research proposed a CNN based model named as-BoNet for automatic bone age assessment. Several off the shelf CNN architectures were tested while the existing models were also fine-tuned. The results showed a mean variability between manual and automatic estimation of about 0.8 years which were very promising as per standards [10].

3. PROPOSED WORK

The Figure 1 presents the conceptual view of the model

being proposed. The MURA dataset consists of various normal and abnormal radiographs of bones. The data is fed to the Deep CNN framework and is split into two parts, training and validation. At first, Image preprocessing is applied to convert the Images to same size. In the following step, the preprocessed Images are fed to the CNN model of two architectures, DenseNet201 and Inception V3. After using these pre trained models, both the models are ensemble to form one of its own. After ensemble, model is altogether evaluated for classification conduct of the proposed model and if which one of the ensemble techniques is giving the finest results.

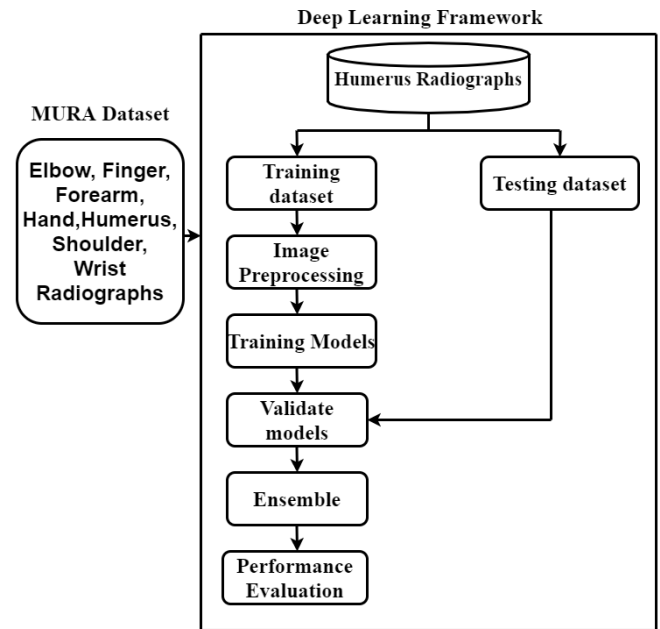


Figure 1. Conceptual view

3.1 Data collection

Dataset named MURA has been collected from the ML group of Stanford University [11], which is a large sized dataset and is widely accepted. The dataset consists of musculoskeletal radiographs which are composed of 14,863 studies of 12,173 number of patients. The dataset is split into two parts namely-training and validation. Both the training and validation datasets are having six categories, namely- Elbow, Forearm, Hand, Humerus, Wrist and Shoulder radiographs. Each category consists of two class labels, abnormal and normal.

3.2 Image preprocessing

The Image Preprocessing [12] step is carried out for every Image in the dataset. Image preprocessing in deep learning involves resizing of Images, conversions, filtering processes and Image rescaling. So, to be brief, it can be said that the Images are needed to be resized to a certain number and normalized because of the radiographs of variable sizes.

Images in the dataset are of variable sizes, which required them to be resized to a same size so as to abstain from Image processing difficulties. Conversion is up to the researchers, RGB color is maintained to detect color features from the Images. Then, Image filtering is carried out for the removal of noise (if any) present in the Images. Gaussian Filter is used

and upon observation the results are observed to improve.

To achieve better results and to avoid the problem of data overfitting, data augmentation [13] is used as a part of training process. It helps when there is less number training data available. Data augmentation adds some variability like rotation, flip, zoom, shift etc. to the images. It is observed that, with data augmentation the models performed better than those without data augmentation.

3.3 Model development

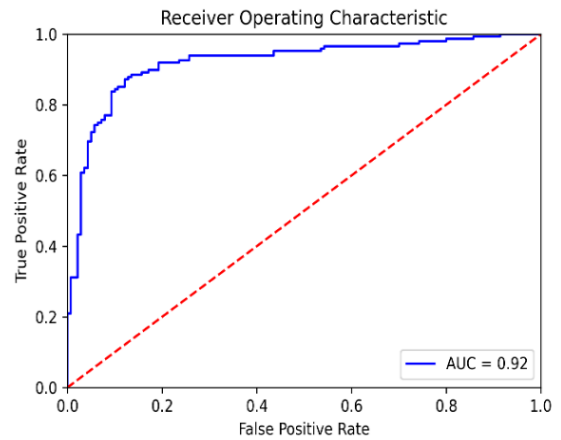
Before starting the training phase, a number of state of art of architectures in deep convolutional neural networks has been studied over the research period. DenseNet201 and Inception V3 architectures has been chosen to conduct the further process. These state of art architectures have already been trained on the ImageNet dataset which has greater than 22,000 object categories and about 15 million high resolution training images [14]. The dense CNN connects each layer to every other connecting layer in a feed forward manner. The architecture can be observed as an extension of ResNets. DenseNet evokes interest of researchers as they reduce the vanishing gradient problem, induces feature reuse and obviously reducing the number of parameters to a substantial amount. On the other hand, Inception-V3 model has about 25 million parameters and makes use of 5 billion operations to classify a single image.

While model training, transfer learning [15] has been studied in which a model trained on a task is re-proposed for some other similar task. Deep CNN requires large sized dataset which is quite a problem when we have a smaller dataset. This problem is overcome with the help of transfer learning. Transferable weights are effective as they are tuned earlier with ImageNet dataset for the purpose of feature extraction where initialization with random weights might or might not be able to catch a weight that will be able to get required number of features to be extracted at a certain desired level. Thus the model training process needs to be started with transferable weights and not with any random weights. To incorporate transfer learning, pre-trained weights of ImageNet have been collected and initialized weights of it to start the training process.

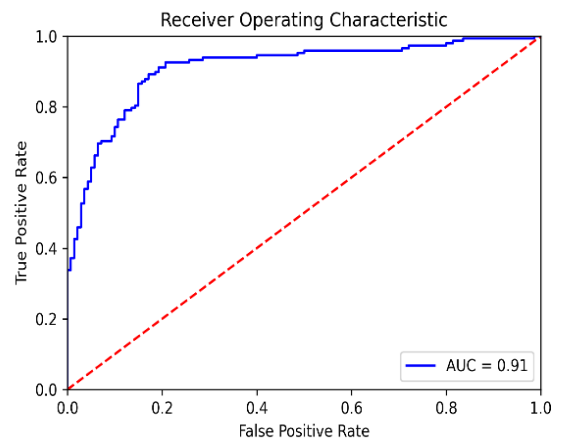
In both the architectures, models are trained with different learning rates [16] by using learning rate schedulers to improve model's performance and then the best common learning rate is chosen to build the final model for both the architectures. Stochastic gradient descent [17] with momentum is used as an optimizer for both models.

In Figure 2, ROC graphs of both DenseNet201 and Inception-V3 architectures are presented. Both Figures 2(a) and 2(b) for the architectures-DenseNet201 and Inception V3 respectively depicts the AUROC curves, which means to what degree the architectures are accomplishing the results in general. The ROC curve plots the true positive rates with respect to the false positive rates. Any architecture which has the finest positive rate, yields the best results amongst other architectures. Both the models performed as well as one another although DenseNet201 shows slightly better results but not by much of a margin.

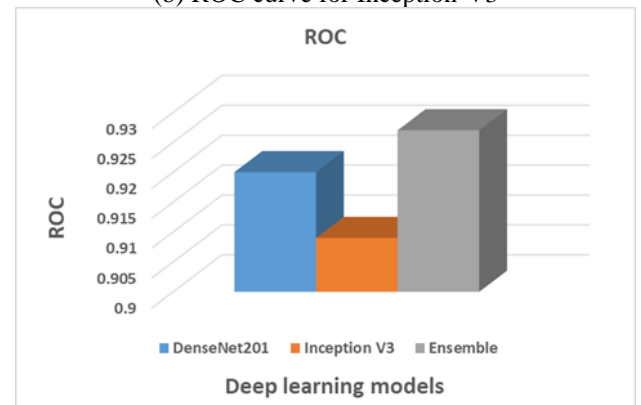
Figure 2(c) describes the performance comparison of both DenseNet201 and InceptionV3 architectures with the ensemble technique. The result shows a slight spike in the curve for the ensemble technique.



(a) ROC curve for DenseNet201



(b) ROC curve for Inception-V3



(c) ROC curve for comparison

Figure 2. ROC graph for DenseNet201, Inception V3 and Comparison of models respectively

3.4 Ensemble models

The best models from training are used as base models for the voting ensemble to achieve better performance metrics [18]. The prediction probability of abnormality is determined by the formula given below:

Prediction Probability,

$$P_p = \sum_{k=0}^m P_k * W_k \quad (1)$$

where,

P_k = probability assigned by the k^{th} classifier,

W_k = weight assigned by the k^{th} classifier.

If (Pp>0.5), the result is classified as 1 i.e. normal image else it will be classified as 0 i.e. abnormal image. After the completion of training phase of the models, testing phase is conducted to determine how well the model is performing on the testing dataset.

3.5 Metrics of performance

Evaluation is the key part of any work whether research oriented or non-research oriented. To estimate the ability of the model, multiple performance measures are evaluated. First, Confusion matrix is identified for actual and predicted values. The confusion matrix consists of True Negative (TN), True Positive (TP), False Negative (FN) and False Positive (FP) elements.

Metrics for evaluation are given as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{F1-measure} = \frac{2*Precision*Recall}{Precision+Recall} \quad (5)$$

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (6)$$

$$\text{FP Rate} = \frac{FP}{FP+TN} \quad (7)$$

$$\text{Cohen's Kappa} = \frac{\text{Accuracy}-\text{Expected}}{1-\text{Expected}} \quad (8)$$

$$\text{MCC} = \frac{TP*TN - FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (9)$$

For Matthew's correlation coefficient (MCC), value ranges from -1 to 1 with -1 being the worst value and +1 being the best value. The MCC score is said to be more informative than the F1 measure and accuracy score as it takes balance ratios of all the confusion matrix categories [19].

4. EXPERIMENTAL INVESTIGATION

This section of the paper gives a brief description about the dataset and experimental view of model.

4.1 Dataset description

MURA (musculoskeletal radiographs) is a large dataset of bone x-rays for bone disorder detection. MURA is one of the largest public radiographic image datasets available online. The dataset is made available by the Stanford ML group. Humerus radiographs are used to build the model and make predictions from that model. A total of 1288 radiographs (normal and abnormal combined) are used to train the model and 288 radiographs (normal and abnormal combined) are used to evaluate the performance measures of model.

4.2 Experimental setup

DenseNet201 and Inception-V3 architectures are used and

trained using python programming language. The main goal of the conduct is to determine how well the model is performing in terms of accuracy score, Cohen's kappa metric and make predictions of the classifier to compare with end results. The dataset is categorized into two parts training and validation data. Both the training and validation sets are further divided into abnormal and normal data. After preprocessing of training data, the training data is used to train the underlying model with a fully connected layer on top of it. The dataset is trained for abnormality being classified as 0 and normality being classified as 1. After that, the validation dataset is used for performance estimation of the model and to evaluate the results. Several performance metrics such as accuracy score, precision, F1-measure, sensitivity, specificity and Cohen's kappa score are used.

5. RESULTS AND DISCUSSION

This part of the paper gives description about the performance estimation, the results and the comparisons of the proposed architectures on the given dataset.

5.1 Performance assessment

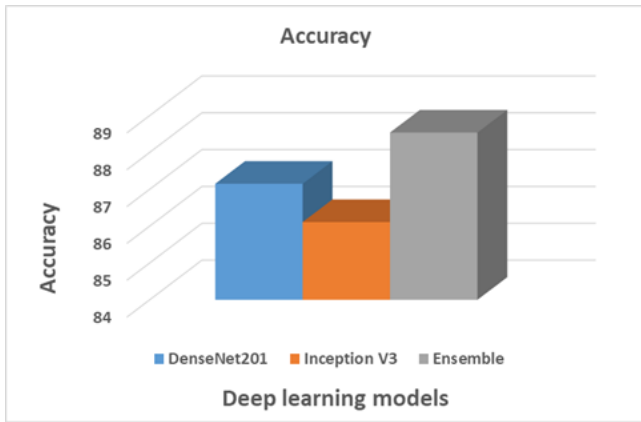
The performance of each of the architecture taken into consideration is computed for humerus radiographs. Confusion matrix for different Deep CNN models for humerus classification is illustrated below (Table 1):

Table 1. Confusion matrix parameters for humerus

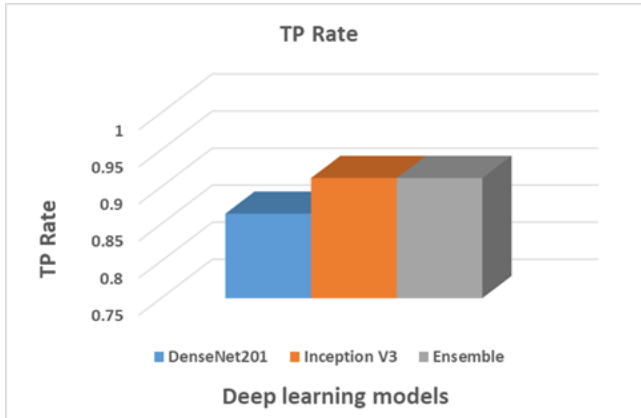
| | TP | TN | FP | FN |
|--------------|-----|-----|----|----|
| DenseNet201 | 123 | 128 | 17 | 20 |
| Inception V3 | 113 | 135 | 27 | 13 |
| Ensemble | 120 | 135 | 20 | 13 |

The performance results for baseline models (DenseNet201 and Inception V3) and ensemble model are presented in the Table 2. The metrics for evaluation namely- Accuracy, TP Rate, FP Rate, F1 measure, MCC score, AUC score, precision, specificity, and Cohen's Kappa score of the ensemble model is compared with the baseline models for the same dataset. The ensemble of both DenseNet201 and Inception V3 models obtained highest scores for metrics such as accuracy, TP Rate, F1 measure, MCC score, AUC score and Cohen's Kappa with values of 88.54%, 0.912, 0.892, 0.771, 0.927 and 0.770 respectively. Inception V3 performed better in case of FP rate with score of 0.192. And DenseNet201 performed better in case of precision and specificity with scores of 0.882 and 0.878.

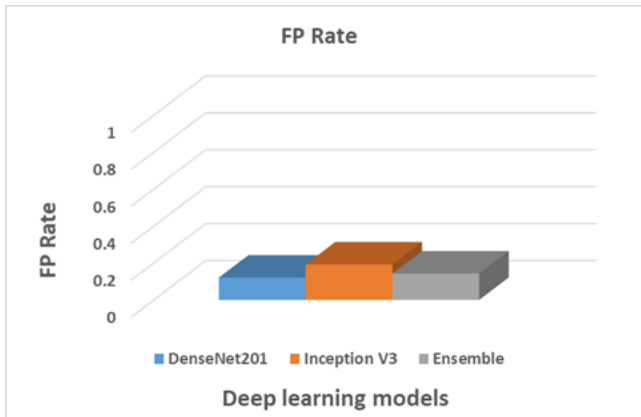
The rationale of Table 2, several graphs are devised to show the comparison of analysis of the models taken under consideration. In the subsequent Figure 3, a performance contrast of the architectures is given in a graphical manner. Figure 3(a) presents the accuracy measure of different models used for the experiment. In this figure, ensemble technique performed best amongst all, as it made use of base models for better classification. Figure 3(b) shows the TP rate for the models in comparison to each other, TP rate means the correctly classified instances. Figure 3(c) shows the FP rate, meaning the instances which are falsely classified. In Figure 4(d), F1-measure is compared for models and it can be observed, the ensemble technique performed better than both the individual models.



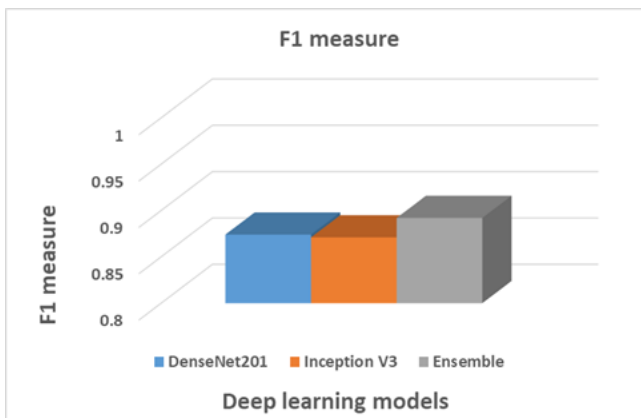
(a) Accuracy



(b) TP Rate



(c) FP Rate



(d) F1 measure

Figure 3. Comparison chart of models using Accuracy, TP rate, FP rate and F1 measure

Table 2. Performance comparison of ensemble with other models

| | DenseNet201 | InceptionV3 | Ensemble |
|---------------|--------------|--------------|--------------|
| Accuracy (%) | 87.15 | 86.11 | 88.54 |
| TP Rate | 0.864 | 0.912 | 0.912 |
| FP Rate | 0.121 | 0.192 | 0.143 |
| F1 measure | 0.874 | 0.871 | 0.892 |
| MCC | 0.713 | 0.725 | 0.771 |
| AUC | 0.920 | 0.909 | 0.927 |
| Precision | 0.882 | 0.833 | 0.871 |
| Specificity | 0.878 | 0.807 | 0.857 |
| Cohen's Kappa | 0.743 | 0.721 | 0.770 |

5.2 Discussion

Since, humerus radiographs are chosen for bone analysis, Cohen Kappa comparison is made amongst DenseNet201, Inception V3 and ensemble technique. The comparison results are presented in the Figure 4.

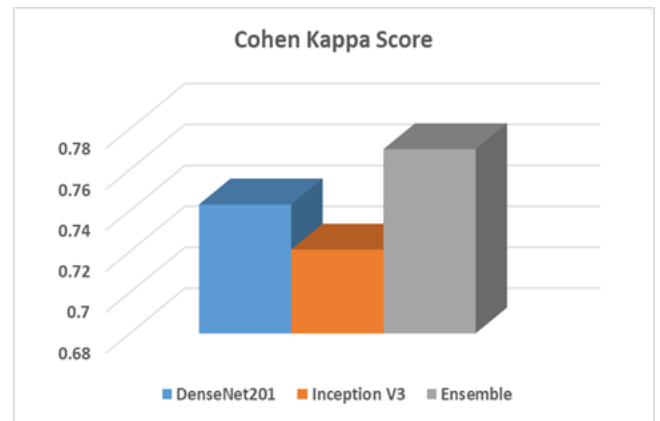


Figure 4. Cohen Kappa comparison graph

Comparison is made with the Computer Aided Diagnosis (CADx) model [5] and the proposed ensemble Model on Cohen's Kappa score, Accuracy score, F1-measure and MCC score. The results are presented in the Table 3. In case of musculoskeletal studies in healthcare, Cohen's kappa statistic is considered to be more robust and gives more valuable Information [20, 21].

Table 3. Comparison between CADx model and proposed Model on various performance metrics

| | CADx model | Proposed model |
|---------------------|------------|----------------|
| Cohen's Kappa Score | 0.743 | 0.77 |
| Accuracy (%) | 87.15 | 88.54 |
| F1-measure | 0.8737 | 0.892 |
| MCC | 0.7432 | 0.771 |

6. CONCLUSION

In today's revolution aligned environment, building computer based diagnosis systems and utilizing them to the whole extent plays a major role in the medical field for quick and efficient diagnosis. The most basic quality that can be thought of these systems is their efficiency, feasibility and high computation power in detecting diseases. While training the individual models, the best performing models are chosen amongst all and then, they are combined to predict results with

a voting classifier. On testing, the proposed ensemble model using the benchmark dataset, the accuracy, F1-measure, Sensitivity, Specificity, Cohen's Kappa and MCC score are found to be 88.54%, 0.892, 0.912, 0.857, 0.770 and 0.771 respectively. As compared to the CADx model, the proposed model performed better in all aspects of performance that are present. The proposed model can also be employed for other parts of the dataset such as elbow, forearm etc. The efficient framework based on ensemble learning techniques can be employed as a Computer based decision system (CBDs) for the decision making of abnormal and normal radiographs. In the future, these parameters will be enhanced by employing different Deep Learning models and ensembles for better results.

REFERENCES

- [1] Abujudeh, H.H., Boland, G.W., Kaewlai, R., Rabiner, P., Halpern, E.F., Gazelle, G.S., Thrall, J.H. (2010). Abdominal and pelvic computed tomography (CT) interpretation: Discrepancy rates among experienced radiologists. *European Radiology*, 20(8): 1952-1957. <https://doi.org/10.1007/s00330-010-1763-1>
- [2] Kido, S., Hirano, Y., Hashimoto, N. (2018). Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, pp. 1-4. <https://doi.org/10.1109/IWAIT.2018.8369798>
- [3] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, pp. 2261-2269. <http://dx.doi.org/10.1109/CVPR.2017.243>
- [4] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 2818-2826. <http://dx.doi.org/10.1109/CVPR.2016.308>
- [5] Mondol, T.C., Iqbal, H., Hashem, M.M.A. (2019). Deep CNN-based ensemble CADx model for musculoskeletal abnormality detection from radiographs. 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, pp. 392-397. <http://dx.doi.org/10.1109/ICAEE48663.2019.8975455>
- [6] Saif, A.F.M., Shahnaz, C., Zhu, W.P., Ahmad, M.O. (2019). Abnormality detection in musculoskeletal radiographs using capsule network. *IEEE Access*, 7: 81494-81503. <http://dx.doi.org/10.1109/ACCESS.2019.2923008>
- [7] Thian, Y.L., Li, Y., Jagmohan, P., Sia, D., Chan, V.E.Y., Tan, R.T. (2019). Convolutional neural networks for automated fracture detection and localization on wrist radiographs. *Radiology: Artificial Intelligence*, 1(1): e180001. <http://dx.doi.org/10.1148/ryai.2019180001>
- [8] Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball, R.L., Langlotz, C., Shpanskaya, K., Lungren, M.P., Ng, A.Y. (2017). Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv preprint arXiv:1712.06957*.
- [9] Chung, S.W., Han, S.S., Lee, J.W., Oh, K.S., Kim, N.R., Yoon, J.P., Kim, J.P., Moon, S.H., Kwon, J., Lee, H.J., Kim, Y.J., Noh, Y.M. (2018). Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthopaedica*, 89(4): 468-473. <http://dx.doi.org/10.1080/17453674.2018.1453714>
- [10] Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., Leonardi, R. (2017). Deep learning for automated skeletal bone age assessment in X-ray images. *Medical Image Analysis*, 36: 41-51. <https://doi.org/10.1016/j.media.2016.10.010>
- [11] MURA dataset: Towards radiologist-level abnormality detection in musculoskeletal radiographs. (n.d.).Stanford Machine Learning Group. <https://stanfordmlgroup.github.io/competitions/mura>, accessed on 02 December 2019.
- [12] Image Preprocessing - Keras Documentation. (2020). from <https://keras.io/preprocessing/image/>, accessed on 3 December 2019.
- [13] Perez, L., Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- [14] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 1097-1105. <http://dx.doi.org/10.1145/3065386>
- [15] West, J., Ventura, D., Warnick, S. (2007). Spring research presentation: A theoretical foundation for inductive transfer. Brigham Young University, College of Physical and Mathematical Sciences, 1(8).
- [16] Zulkifli, H. (2018). Understanding learning rates and how it improves performance in deep learning. *Towards Data Science*, 21: 23.
- [17] Bottou, L. (1991). Stochastic gradient learning in neural networks. *Proceedings of Neuro-Nimes*, 91(8).
- [18] Dimitriadou, E., Weingessel, A., Hornik, K. (2001). Voting-merging: An ensemble method for clustering. In *International Conference on Artificial Neural Networks*, Springer, Berlin, Heidelberg, pp. 217-224. https://doi.org/10.1007/3-540-44668-0_31
- [19] Chicco, D., Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1): 6. <http://dx.doi.org/10.1186/s12864-019-6413-7>
- [20] Sim, J., Wright, C.C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3): 257-268. <http://dx.doi.org/10.1093/ptj/85.3.257>
- [21] Viera, A.J., Garrett, J.M. (2005). Understanding interobserver agreement: The kappa statistic. *Fam Med*, 37(5): 360-363.