

# Knee Osteoarthritis Severity Prediction using an Attentive Multi-Scale Deep Convolutional Neural Network

Rohit Kumar Jain, Prasen Kumar Sharma, Sibaji Gaj, Arijit Sur and Palash Ghosh

**Abstract**—Knee Osteoarthritis (OA) is a destructive joint disease identified by joint stiffness, pain, and functional disability concerning millions of lives across the globe. It is generally assessed by evaluating physical symptoms, medical history, and other joint screening tests like radiographs, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scans. Unfortunately, the conventional methods are very subjective, which forms a barrier in detecting the disease progression at an early stage. This paper presents a deep learning-based framework, namely OsteoHRNet, that automatically assesses the Knee OA severity in terms of Kellgren and Lawrence (KL) grade classification from X-rays. As a primary novelty, the proposed approach is built upon one of the most recent deep models, called the High-Resolution Network (HRNet), to capture the multi-scale features of knee X-rays. In addition, we have also incorporated an attention mechanism to filter out the counterproductive features and boost the performance further. Our proposed model has achieved the best multi-class accuracy of 71.74% and MAE of 0.311 on the baseline cohort of the OAI dataset, which is a remarkable gain over the existing best-published works. We have also employed the Gradient-based Class Activation Maps (Grad-CAMs) visualization to justify the proposed network learning.

**Index Terms**—Classification, deep learning, kellgren lawrence grade, knee osteoarthritis, knee x-ray.

## I. INTRODUCTION

KNEE osteoarthritis is a common joint disorder caused by the eroding of the articular cartilage between the joints, which leaves the bones of the knee touching and rubbing against each other. In general, it occurs in the synovial joints and results from a combination of genetic factors, injury, and overuse [1]. Obesity, specific occupation, stress, trauma, age, gender, and family history are some well-defined risk factors [2]. The pain and stiffness in the joints begin to worsen by the rigorous activity and stress compared to other inflammatory

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

Rohit kumar Jain, Prasen Kumar Sharma, and Arijit Sur are with Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, India.

Sibaji Gaj is with Cleveland Clinic, Ohio, USA.

Palash Ghosh is with Department of Mathematics, Indian Institute of Technology Guwahati, India, and Centre for Quantitative Medicine, Duke-NUS Medical School, National University of Singapore, Singapore.

Corresponding author: Rohit Kumar Jain

Email: jkrohit03@gmail.com

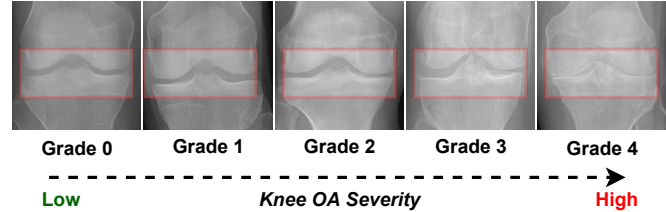


Fig. 1. Knee OA disease progression: A qualitative demonstration of sample X-rays and their corresponding KL grades.

arthritis where activity and exercising improve symptoms. It can also lead to instability, joint deformity, and reduction in joint functionality [2]. In addition, the distance between the knee joint begins to flatten out due to the loss of the cartilage, leading to the progression of knee OA [1]. The following key changes, described by the word LOSS, marks the progression of knee OA:

- L– “loss of joint space”, caused by the cartilage loss,
- O– “osteophytes formations”, projections that form along the margins of the joint,
- S– “subarticular sclerosis”, increase in bone density along the joint line, and
- S– “subchondral cysts”, caused due to holes in the bone filled with fluid along the joints [3].

Radiographic screening (X-Rays), MRI, and CT scans are a few of the common ways to detect the structural changes in the joint and diagnose knee OA’s biological condition. However, the traditional treatment for knee OA may not be effective enough to completely fix the disease in today’s time. Therefore, it is of utmost importance to detect the deformation of the joint at such a stage before which it becomes impossible to reverse the loss [4]. Generally, the knee OA severity is measured in terms of the World Health Organization (WHO) approved KL grading scale [5]. KL grading is a 5-point semi-quantitative progressive ordinal scale ranging from grade 0 (*low severity*) to 4 (*high severity*). Fig. 1 shows the disease progression along with its corresponding KL grade.

## A. Challenges

In general, a complete cure for this disease remains quite challenging to find, and OA management is mainly palliative [1], [6]. MRI screenings and CT scans are effective as they highlight the three-dimensional structure of the knee joints [7]

[8]. However, they have certain drawbacks, including limited availability, extreme device expenses, the time required in diagnosing, and the inclination to image ancient rarities [9], [10]. At the same time, X-Rays are the most effective and economically feasible way of diagnosing the disease, given the routine knee OA diagnosis. However, the currently adopted methods for assessing the disease progression from X-Ray images may not be much effective. They, in general, require a very skilled practitioner to analyze the radiographic scans accurately and are thus absolutely subjective. In most cases, the practitioners require multiple tests to quantify the condition accurately, which is generally time-consuming. The analysis may differ based on their expertise and sometimes may be inaccurate. Further, multiple tests may be costly for some of the patients.

A better and in-depth understanding of knee OA may result in timely prevention and treatment. It is believed that early treatments and preventive measures are the most effective way of managing knee OA. Unfortunately, there has been no significant and predominant way of identifying the disease at an early stage to date. Recently, the use of Machine Learning (ML) and Deep Convolutional Neural Networks (CNNs) for knee OA analysis have shown remarkable supremacy in detecting even the slightest differences in biological joint structural variations in the X-Rays [11].

Deep CNNs have been widely adopted in many medical imaging tasks, including classifications of COVID-19, pneumonia, tumor, bone fracture, polyps detection, etc. For *e.g.*, CheXNet [12], a 121-layers deep CNN, performed astonishingly better than the average performance of four specialists in assessing pneumonia using plain radiographs [13]. However, it is difficult to collect the medical images, as the collection and annotation of such data are challenged by the expert availability, and the data privacy concerns [13].

### B. The Osteoarthritis Initiative (OAI) Dataset

OAI is a distributed, observational study of patients, which is publicly available<sup>1</sup>. It facilitates the scientific and research community worldwide to work on knee OA progression and develop new treatments and techniques beneficial for its detection and treatment. In this work, we have utilized the data acquired from the OAI repository and made available by Chen *et al.* [14], [15]. The dataset comprises knee bilateral posterior-anterior fixed flexion radiographs of 4796 participants, including male and female subjects from the baseline cohort. Fig. 1 shows sample X-ray images pertaining to each KL grade.

## II. RELATED DEVELOPMENTS

Several schemes have been developed for the Knee OA severity prediction in the past few years. Shamir *et al.* [16] utilized a weighted nearest neighbors algorithm that incorporated the hand-crafted features like Gabor filters, Chebyshev statistics, multi-scale histograms, etc. Antony *et al.* [17] proposed to utilize the transfer learning of the existing pre-trained deep CNNs. Later, Antony *et al.* [18] customized a deep CNN

from scratch and optimized the network using a weighted combination of the traditional cross-entropy and the mean squared error, which served as dual-objective learning. Tuilpin *et al.* [19] developed a method inspired from the deep Siamese network [20], for learning the similarity metric between the pair of radiographs. Gorriz *et al.* [21] developed an end-to-end attention-based network, bypassing the need to localize knee joint, to quantify the knee OA severity automatically. Chen *et al.* [15] proposed to utilize pre-trained VGG-19 [22] along with an adjustable ordinal loss for the proportionate penalty to the misclassification. Yong *et al.* [23] utilized the pretrained DenseNet-161 [24], along with an ordinal regression module (ORM), in order to treat the ordinality of the KL grading. They further optimized the network using the cumulative link (CL) loss function.

### A. Motivation

Deep CNNs are renowned for learning the highly correlated features in an image. In addition, it is a widely known fact that the first few layers of a deep CNN contribute to the learning of low-level features in an image. Whereas the last few layers contribute to the learning of the high-level features, enabling the final classification by adaptively learning spatial hierarchies of features [25]. While the low-level features are the minute details of an image, including points, lines, edges, etc., the high-level features comprise several low-level features, which make up the more prominent and robust structures for classification.

However, in general, the knee X-Rays do not comprise many edge or low-level structures. Due to a lack of such vital information, it may be difficult for a deep CNN to learn an efficient classification particularly, in the case of knee OA, where one KL grade is not very distinctive from the other unless carefully inspected (see Fig. 1). A few of the most recent state-of-the-art methods [15], [23] have directly utilized the existing popular image classification models in a plug-and-play fashion without supervising the network engineering relevant to the given problem. It should be mentioned that while a majority of those methods were built for a generic image classification problem, a few of them were explicitly designed using architectural search, *e.g.*, MobileNetV2 [26].

Moreover, for the knee OA severity classification, the presented best-performing deep CNNs were enormous in size, exceeding 500 MB [15], to be precise. As a result, such models may require substantially high computational resources, making it challenging to deploy in real-time environments. Therefore, it may be said that the direct usage of popular classification models may not be appropriate. Although some recent methods [19], [27], [21] [23], have started to design the models specific to knee OA given the amount of information present in the knee X-rays. However, they still lack in terms of accuracy and computational overhead. For *e.g.*, Zhang *et al.* [27] utilized the Convolutional Block Attention Module, namely CBAM [28], after every residual layer in their proposed architecture, which may not be computationally pleasant. The attention module has performed undoubtedly well in many high-level vision tasks. However, one must not

<sup>1</sup>Dataset source: <https://nda.nih.gov/oai/>

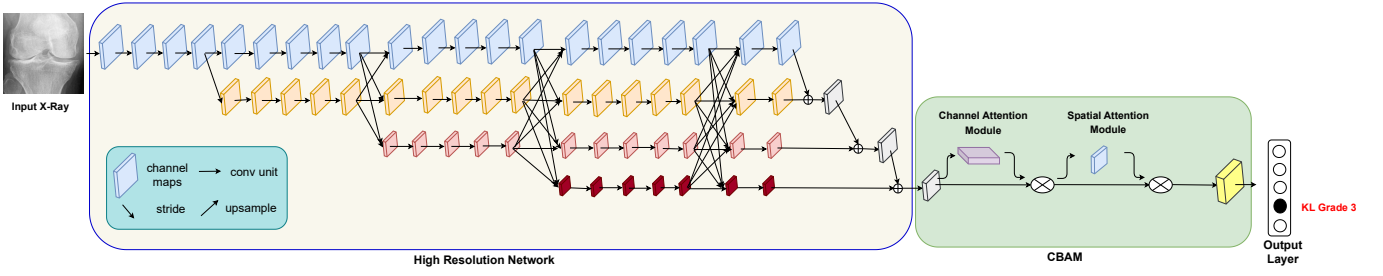


Fig. 2. An overview of the architecture of the proposed OsteoHRNet for the knee OA severity prediction. Blocks with different colors denote convolution features at different spatial scales. The proposed model takes knee X-Ray image as input and estimates the OA severity in terms of KL grade.

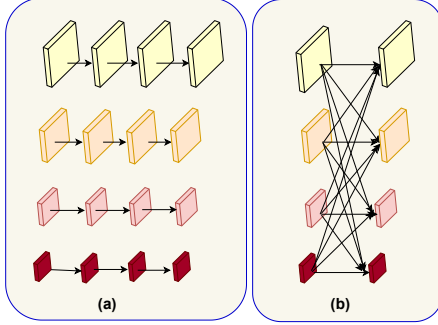


Fig. 3. **Connections in HRNet:** (a) Multi Resolution convolution in parallel, (b) Fusion of Multi-Resolution convolution. Different colors denote feature resolution at various scales.

overlook its computational overhead considering the presence of fully connected layers.

The applicability of deep CNNs in medical imaging heavily depends on the amount of data available for efficient learning. As an alternative, many deep learning-based methods have utilized the data augmentation techniques to further boost the performance, which has not been much considered in the existing works.

### B. Our Contributions

Based on the aforementioned drawbacks of the existing best-published works, our contributions are five-fold, as follows:

- 1) We propose an efficient deep CNN for the knee OA severity prediction in terms of KL grades using X-ray images. Unlike existing methods, our proposed scheme is not a blind plug-and-play of popular deep models. The proposed scheme has been built upon a high-resolution network; namely, HRNet [29], that takes the spatial scale of the X-Ray image into account for efficient classification.
- 2) We also propose to utilize the attention mechanism only once in the entire network to reduce the computational overhead and adaptive filtering of the counterproductive features just before classification.
- 3) Also, instead of relying on traditional entropy-based minimization, we have adopted the ordinal loss [15] to optimize the proposed scheme.
- 4) To further boost the performance of the proposed scheme, we have incorporated the data augmentation

techniques, which have not been much considered in any recent work so far.

- 5) Lastly, we present an extensive set of experiments and Grad-CAM [30] visualization to justify the importance of each module of the proposed framework.

The rest of the paper is organized as follows: Section III presents the proposed method and the adopted cost function. Section IV briefly describes the incorporated dataset, training details, competing methods, and evaluation metrics. Section V presents the quantitative and qualitative comparison against the best-published works. Section VI presents a brief discussion on the learning of proposed scheme in terms of Grad-CAM visualization of obtained results. Section VII demonstrates the ablation study against various components, and finally, the paper is concluded in Section VIII.

## III. PROPOSED METHOD

This section presents the details of the proposed model, followed by a brief description of the incorporated cost function. The proposed framework is built upon the HRNet and Convolution Block Attention Module (CBAM) in a serially cascaded manner. A descriptive representation of the proposed model is shown in Fig. 2.

### A. High Resolution Network

High-Resolution Network (HRNet) [29] is a novel and revolutionary multi-resolution deep CNN, which tends to maintain high-resolution feature representations throughout the network. It starts as a stream of 2D convolutions and subsequently adds up the high-to-low resolution streams to form the following stages. It then merges the multi-resolution streams in parallel for information exchange [29] as shown in Fig. 2 (marked as *High-Resolution Network*). HRNet tends to generate reliable multi-resolution representations with strong spatial sensitivity. It has been achieved by utilizing parallel connections instead of serial (see Fig. 3(a)) and recurrent fusion of the intermediate representations from multi-resolution streams (see Fig. 3(b)), as shown in Fig. 3. As a result, it enables the network to learn more highly correlated and semantically robust spatial features. This motivates us to incorporate HRNet for processing the knee X-Ray images, which lack such rich spatial features.

To formally define, let  $\mathcal{D}_{ij}$  denotes the sub-network in the  $i^{th}$  stage of  $j^{th}$  resolution index. The spatial resolution in this

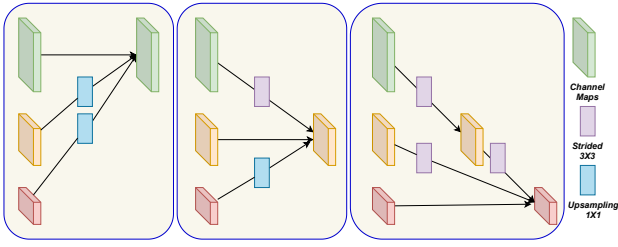


Fig. 4. Graphical demonstration of how HRNet fuses information from different resolutions.

branch is  $1/2^j - 1$  of that of the high-resolution (HR) branch. For *e.g.*, HRNet, which consists of four different resolution scales, can be illustrated as follows:

$$\begin{array}{ccccccc} \mathcal{D}_{11} & \rightarrow & \mathcal{D}_{21} & \rightarrow & \mathcal{D}_{31} & \rightarrow & \mathcal{D}_{41} \\ & \searrow & \mathcal{D}_{22} & \rightarrow & \mathcal{D}_{32} & \rightarrow & \mathcal{D}_{42} \\ & & & \searrow & \mathcal{D}_{33} & \rightarrow & \mathcal{D}_{43} \\ & & & & & \searrow & \mathcal{D}_{44}, \end{array} \quad (1)$$

Later, the obtained multi-resolution feature maps are fused to exchange the learned variscaled information, as shown in Fig. 4. For this, HRNet utilizes bilinear upsampling followed by the  $1 \times 1$  convolution to adjust the number of channels when transforming the lower resolution feature map to a higher resolution scale, or a strided  $3 \times 3$  convolution otherwise.

### B. Convolutional Block Attention Module

Convolutional Block Attention Module (CBAM) consists of two sequential sub-modules : (a) channel attention module, and (b) spatial attention module [28]. Given an input feature map,  $\mathbf{P} \in \mathbb{R}^{C \times H \times W}$ , CBAM sequentially infers a one-dimensional channel attention map  $Map_c \in \mathbb{R}^{C \times 1 \times 1}$  and a two-dimensional spatial attention map  $Map_s \in \mathbb{R}^{1 \times H \times W}$ . Thus we obtain a final refined attention map, here denoted as  $\mathbf{T}$ , and the comprehensive attention mechanism can be summarized as:

$$\begin{aligned} \mathbf{P}^c &= Map_c(\mathbf{P}) \otimes \mathbf{P}, \\ \mathbf{T} &= Map_s(\mathbf{P}^c) \otimes \mathbf{P}^c, \end{aligned} \quad (2)$$

where  $\otimes$  signifies element-wise multiplication.  $Map_c$  is first generated by making use of the cross-channel relationship of the features, as,

$$Map_c(\mathbf{P}) = g(MLP(\mathcal{A}(\mathbf{P}))) + MLP(\mathcal{M}(\mathbf{P})), \quad (3)$$

where  $g$ ,  $MLP$ ,  $\mathcal{A}$ , and  $\mathcal{M}$  denote sigmoid function, multi-layer perceptron, average pool and max pool, respectively.

Whereas, the  $Map_s$  is generated efficiently by performing  $\mathcal{M}$  and  $\mathcal{A}$  along the channel axis. Next, the pooled descriptors are concatenated together to generate a reliable and efficient feature descriptor by utilizing the inter-spatial correlation of the features. It can be written as,

$$Map_s(\mathbf{P}) = g(k^{7 \times 7}([\mathcal{A}(\mathbf{P}); \mathcal{M}(\mathbf{P})])), \quad (4)$$

where  $k^{7 \times 7}$  denotes the convolution operation with kernel of size  $7 \times 7$ .

### C. Network Architecture

We propose a deep CNN, called OsteoHRNet, that utilizes the HRNet as the backbone and is further empowered with an attention mechanism for the knee KL grade classification. CBAM is integrated at the end of the HRNet, followed by a fully connected (FC) output layer, as depicted in Fig. 2. It may be said that the integration of the CBAM module after HRNet has been beneficial in learning adaptive enriched features for an efficient KL grade classification. It can also be observed that the proposed one-time integration of CBAM is computationally pleasant, compared to the multiple additions in the existing work [27]. The resultant output from the CBAM is then fed into the final fully connected layer, which outputs the probabilities of the KL grade for the given input X-Ray image. HRNet has been considered for reliable feature extraction, whereas the capabilities of CBAM are leveraged to help the model better focus on relevant features.

### D. Cost Functions

A majority of the existing works on knee OA severity classification have considered the nominal nature of KL grades for classification. However, inspired by the idea of Chen *et al.* [15], we approach this task as an ordinal regression problem and therefore utilize the ordinal loss function instead of the traditional cross-entropy. The ordinal loss function used in this paper is a weighted ratio of the traditional cross-entropy. Given the ordinality in the KL grading, it must be acknowledged that extra information is provided by progressive grading. This approach penalizes the distant grade misclassification more than the nearby grade according to the penalty weights. For *e.g.*, a grade 1 classified as grade 3 is penalized more severely than it is classified as grade 2 and even more for being classified as grade 4. An ordinal matrix  $C_{n \times n}$  is considered as the penalty weights between the outcome and the true grade, i.e.,  $c_{uv}$  denotes the penalty weight for predicting a grade  $v$  as  $u$  with  $n = 5$ . In this study, with five KL grades to classify and  $c_{uu} = 1$ , the adopted ordinal loss can be written as

$$\mathcal{L}_o = \sum_{u=0}^{n-1} c_{uv} * q_u, \quad (5)$$

where  $u, v$  are the predicted and true KL grades of the input image, respectively,  $p_u$  is the output probability by the final output layer of the architecture with  $q_u = p_u$  if  $u \neq v$  and  $q_u = 1 - p_u$ , otherwise. We have utilized the following penalty matrix for our experimentation.

$$\begin{bmatrix} 1 & 3 & 6 & 7 & 9 \\ 4 & 1 & 4 & 5 & 7 \\ 6 & 4 & 1 & 3 & 5 \\ 9 & 7 & 4 & 1 & 4 \\ 11 & 9 & 7 & 5 & 1 \end{bmatrix}$$

## IV. EXPERIMENTAL DETAILS

### A. Dataset

We have utilized the X-ray radiographs acquired from the OAI repository which has been made available by Chen *et al.* [14]. The images obtained are of 4796 participants,



TABLE I  
DISTRIBUTION OF THE DATASET

Dataset	Grade0	Grade1	Grade2	Grade3	Grade4	Total
Training	2286	1046	1516	757	173	5778
Testing	639	296	447	223	51	1656
Validation	328	153	212	106	27	826
Total	3253	1495	2175	1086	251	8260

including men and women. Given that we focus primarily on the KL grades, radiographs with annotated KL grades from the baseline cohort are acquired to assess our method. The dataset of a total of 8260 radiographs, including the left and right knee, was split into train, test, and validation sets in the ratio of 7:2:1 with balanced distribution across all KL grades [14]. Table I shows the train, test, and validation distribution of the dataset.

### B. Training Details

The entire code is developed using Pytorch [31] framework, and all the experiments have been conducted on a 12GB Tesla K40c GPU. Furthermore, the training of all the experimental models was optimized using stochastic gradient descent (SGD) for 30 epochs with an initial learning rate of 5e-4. Additionally, owing to the GPU capacity, the batch size was set to 24.

### C. Competing Methods

In [15], the authors proposed to utilize the pre-trained VGG-19 [32] network with a novel ordinal loss function. Yong *et al.* [23] proposed to utilize the DenseNet-161 [24] with the ordinal regression module (ORM). We have compared the OsteoHRNet against the results obtained by the best-published studies mentioned above for a robust comparison.

### D. Evaluation Metrics

In this study, we have utilized the following three evaluation metrics to analyze and compare the performance of our proposed model : (a) Multi-class accuracy, (b) Quadratic Weighted Cohen's Kappa coefficient (QWK), and (c) Mean Absolute Error (MAE). Traditionally, multi-class accuracy is defined as the average number of outcomes matching the ground truth across all the classes. Accuracy for five classes with N instances is formulated as below

$$Accuracy = \frac{1}{N} \sum_{i=1}^5 \sum_{x:g(x)=i} F(g(x) = \hat{g}(x)), \quad (6)$$

where,  $F$  is a function which returns 1 if the prediction is correct and 0 otherwise.

MAE is the mean of the absolute error of the individual prediction over all the input instances. The error in the prediction value is determined by the difference between the predicted and the true value for that given instance. MAE for five classes with N instances can be expressed as below

$$MAE = \frac{\sum_{i=1}^N abs(y_i - \hat{y}_i)}{N}, \quad (7)$$

where,  $y_i$  &  $\hat{y}_i$  are the true and the predicted grade, respectively.

A weighted Cohen Kappa is a metric that accounts for the similarity between predictions and the actual values. The Kappa coefficient is a chance-adjusted index of agreement measuring the reliability of inter-annotator for qualitative prediction. The Quadratic Weighted Kappa (QWK) is evaluated using a predefined table of weights which measures the extent of non-alignment between the two raters. The greater the disagreement, the greater the weight.

$$\kappa = 1 - \frac{\sum_{p,\hat{p}} w_{p,\hat{p}} O_{p,\hat{p}}}{\sum_{p,\hat{p}} w_{p,\hat{p}} E_{p,\hat{p}}}, \quad (8)$$

$O$  is the contingency matrix for  $K$  classes such that  $O_{p,\hat{p}}$  denotes the count of  $\hat{p}$  grade images predicted as  $p$ . The weight,  $w$ , is defined as

$$w_{p\hat{p}} = \frac{(p - \hat{p})^2}{(1 - K)^2}. \quad (9)$$

Next,  $E$  is calculated as the normalized product between the predicted grade's and original grade's histogram vector. Of the three metrics, accuracy and QWK are positive in nature while MAE is negative in nature.

## V. RESULTS

### A. Comparison against State-of-the-Art Methods

It can be observed from Table II that the proposed method has outperformed the existing best-published works [15] [23] in terms of classification accuracy, MAE, and QWK. It should be mentioned that Yong *et al.* [23] reported the macro accuracy<sup>2</sup> and contingency matrix of their best model. For a fair comparison, equivalent to the above, we have reported their results in multi-class accuracy of 70.23%. Whereas Chen *et al.* [15] has reported the best multi-class accuracy of 69.69%. OsteoHRNet has reported a maximum multi-class accuracy of 71.74%, multi-class average accuracy of 70.52%, MAE of 0.311, and QWK of 0.869 which is a significant improvement over [23], [15]. Fig. 5 represents the confusion matrix obtained by using the proposed and existing methods [15], [23] which when fed with 1656 test images.

TABLE II  
QUANTITATIVE COMPARISON AGAINST THE EXISTING METHODS IN TERMS OF MULTI-CLASS ACCURACY, MAE, AND QWK.

Method	Accuracy	MAE	QWK
VGG 19 - Ordinal [15]	69.69 %	0.344	0.8460
DenseNet 161 - ORM [23]	70.23 %	0.330	0.8609
<b>OsteoHRNET</b>	<b>71.74 %</b>	<b>0.311</b>	<b>0.8690</b>

Furthermore, we have employed the Gradient-weighted Class Activation Maps (Grad CAM) [30] visualization technique to demonstrate the superiority of the proposed OsteoHRNet. It also helps in showcasing the most relevant regions the network has learned to focus on in the X-ray images. Figs. 6, 7, 8, 9, and 10 shows the qualitative comparison of

<sup>2</sup>Macro accuracy: 88.09%

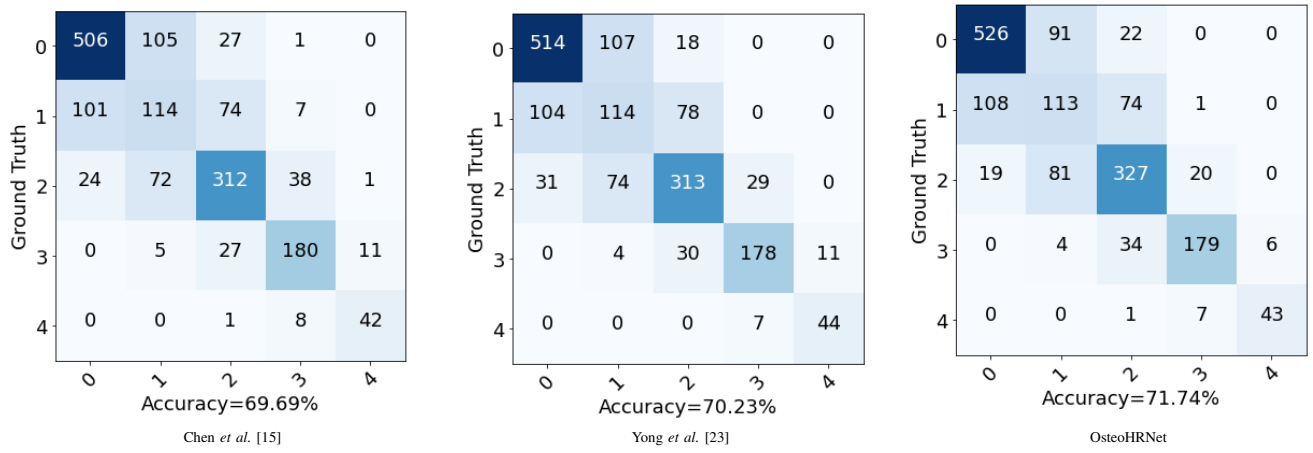


Fig. 5. Confusion matrices for KL grade prediction using different competing approaches [15], [23] and OsteoHRNet.

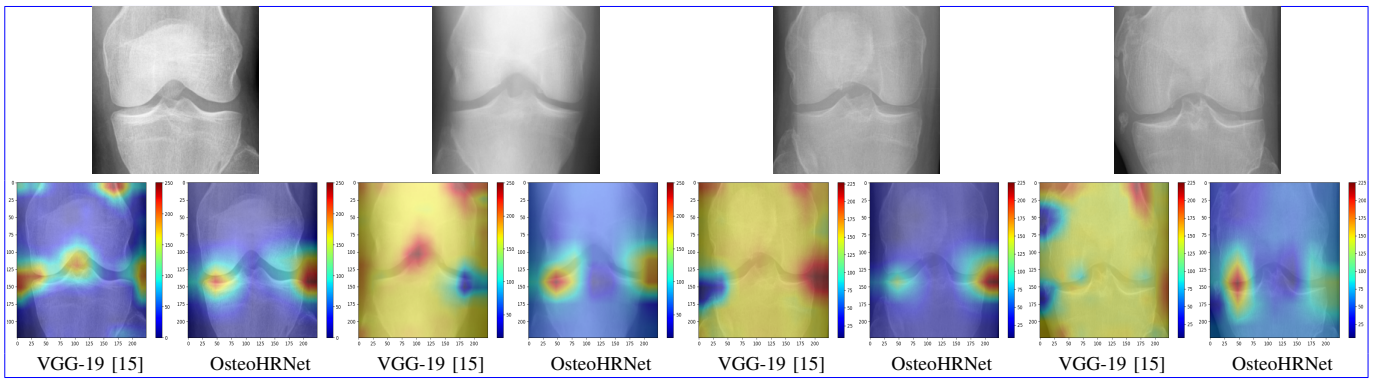


Fig. 6. Grad-CAM visualizations generated against KL grade 0 test images using Chen *et al.* [15] and OsteoHRNet.

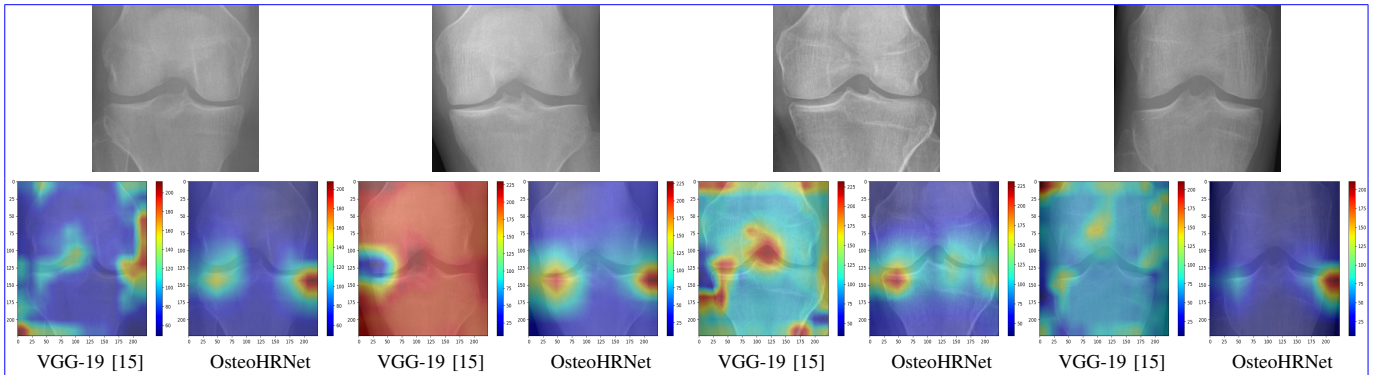


Fig. 7. Grad-CAM visualizations generated against KL grade 1 test images using Chen *et al.* [15] and OsteoHRNet.

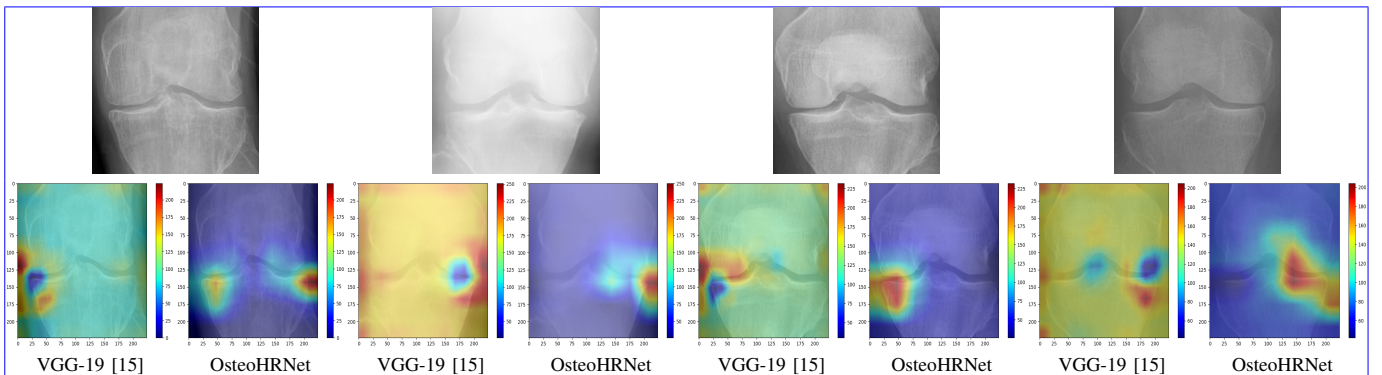


Fig. 8. Grad-CAM visualizations generated against KL grade 2 test images using Chen *et al.* [15] and OsteoHRNet.

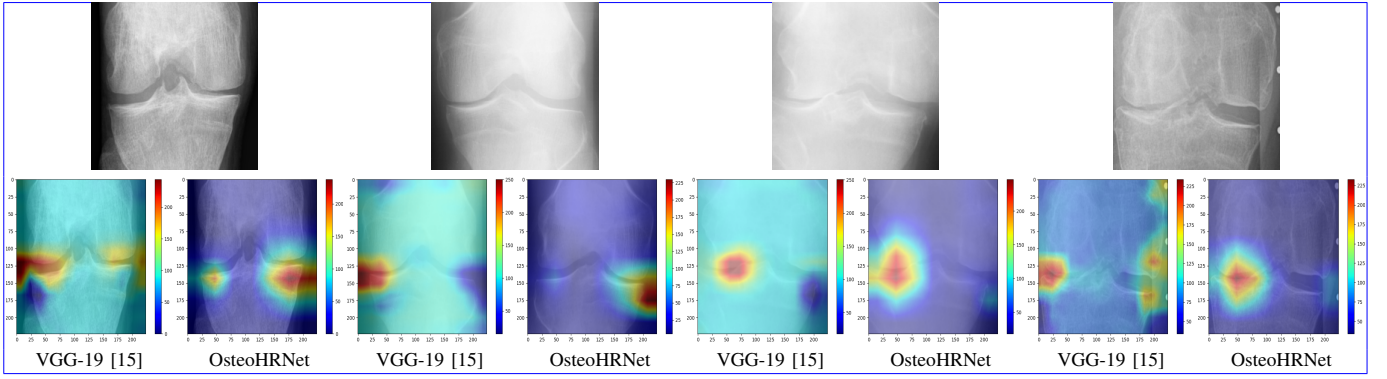


Fig. 9. Grad-CAM visualizations generated against KL grade 3 test images using Chen *et al.* [15] and OsteoHRNet.

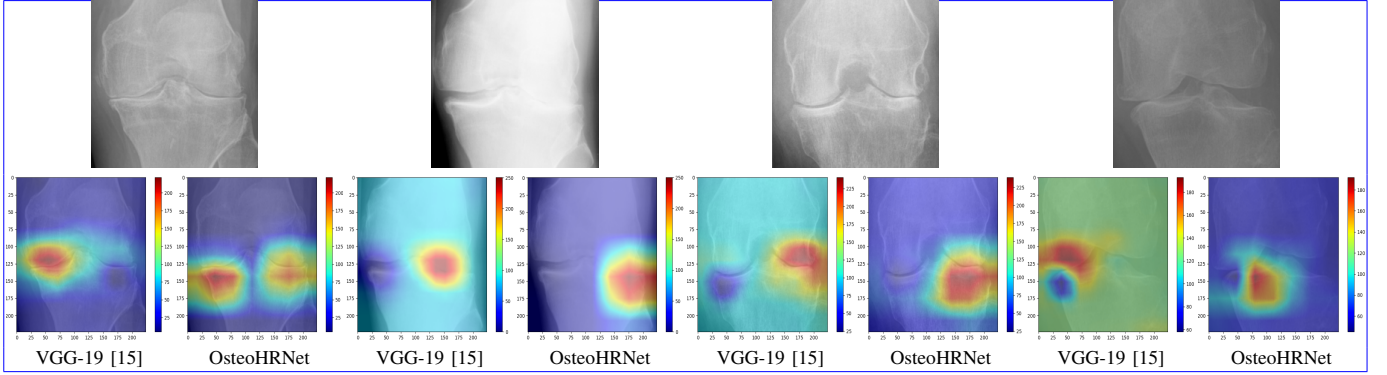


Fig. 10. Grad-CAM visualizations generated against KL grade 4 test images using Chen *et al.* [15] and OsteoHRNet.

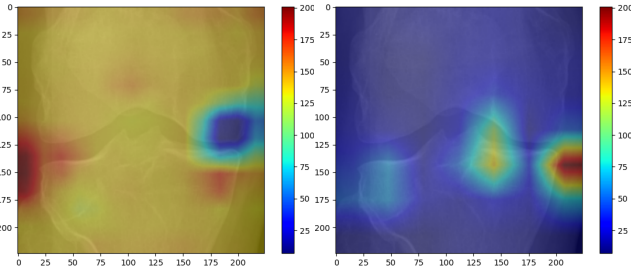


Fig. 11. Grad-CAM visualization for the incorrect classification by Chen *et al.* [15] (VGG-19; *left*) and proposed OsteoHRNet (*right*) for grade 2 radiograph.

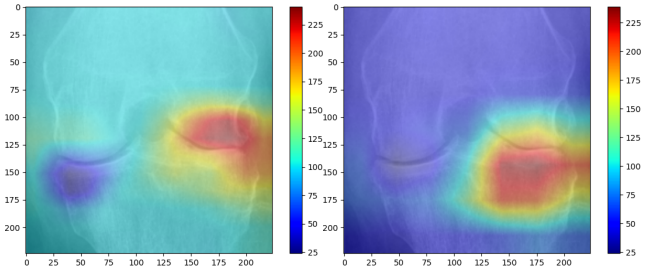


Fig. 12. Grad-CAM visualization for the incorrect classification by Chen *et al.* [15] (VGG-19; *left*) and proposed OsteoHRNet (*right*) for grade 4 radiograph.

the proposed model against the existing methods in terms of Grad-CAM visualization. It can be observed that the proposed OsteoHRNet considers both features and the area between the knee joints for an efficient severity classification (*denoted by the darker colors up the scales*). Moreover, it can be said that the decision-making of OsteoHRNet aligns in accordance with the actual real-world medical criterion of KL grade classification. The proposed model has efficiently learned the prominent features such as joint-space narrowing, osteophytes formations, and bone deformity, thus predicting the most relevant radiological KL grading. This validates the enriched and superior results obtained by the proposed OsteoHRNet model.

## VI. DISCUSSION

It is evident from Fig.5 that the OsteoHRNet has outperformed the previous works [15], [23], significantly. It should

be mentioned that the OsteoHRNet classifies the higher grade X-rays very accurately while reducing the misclassification between far away grades. In comparison to existing methods, there has been a significant increase in correct classifications for grade 2. Furthermore, the nearby misclassifications between higher grades (grade 2-grade 3, grade 3-grade 4) are minimum for the proposed method, which needs to be acknowledged. Also, by way of analysis using obtained Grad-CAM visualization of such incorrect classifications, it can be observed that OsteoHRNet is trying to locate joint space narrowing and osteophytes in accordance with the medical characteristics. At the same time, VGG-19 [15] is confused and focuses on the entire knee, giving importance to irrelevant features for KL grade classification, as seen in Figure 11, 12.

Owing to its superior network learning, our model is extremely relevant to the medical setting of KL grade clas-



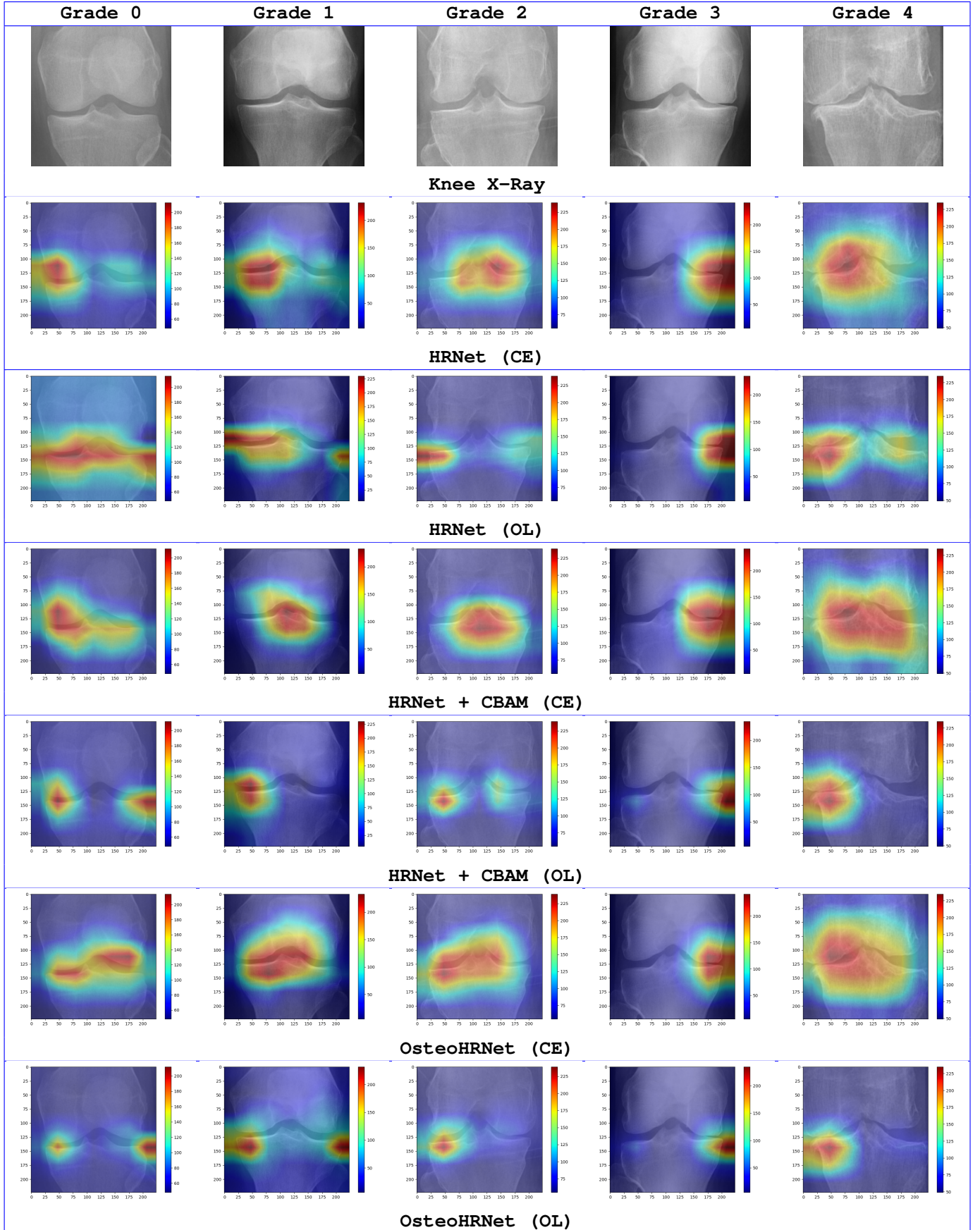


Fig. 13. Grad-CAM visualizations for the ablation study. CE stands for Cross-entropy and OL stands for Ordinal Loss



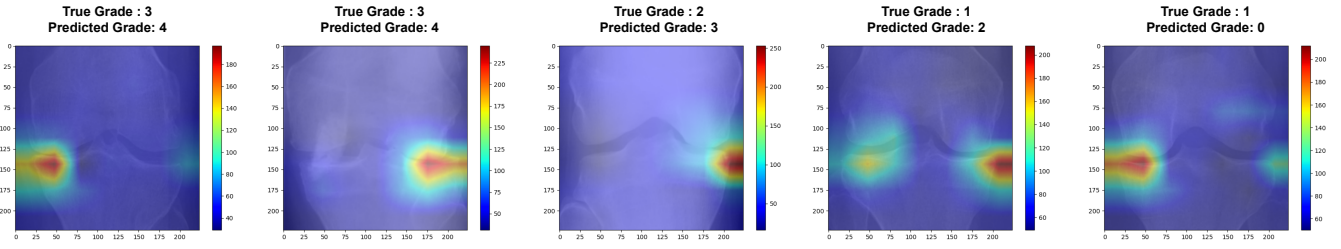


Fig. 14. Grad-CAM visualizations for the incorrectly classified radiographs obtained by using OsteoHRNet.

sification. Furthermore, the Grad-CAM visualization of our model can be extended for the use of the medical practitioner to provide confidence in the findings. However, our study has some limitations, and certain radiographs could not be correctly classified due to the lack of rich features in the radiographs. Fig. 14 shows nearby grade misclassifications, which to a great extent is unavoidable. But, there is high inter and intraobserver variability (correlation coefficient = 0.83) for manual knee KL grading [33]. Thus, our proposed fully automated KL grading method can be extended in clinical settings for getting reliable and reproducible OA grading.

## VII. ABLATION STUDY

TABLE III

EFFECTS OF DIFFERENT NETWORK MODULES & COST FUNCTION

Architecture	Cross Entropy		Ordinal Loss	
	Accuracy	MAE	Accuracy	MAE
<b>HRNet</b>	64.10 %	0.460	65.00 %	0.440
<b>HRNet + CBAM</b>	65.30 %	0.423	66.70 %	0.392
<b>OsteoHRNet</b>	69.90 %	0.373	<b>71.74 %</b>	<b>0.311</b>

This section presents an ablation study to demonstrate the contributions made by each sub-module of the proposed OsteoHRNet. For this, we have performed the following baselines:

- 1) **HRNet**: Original HRNet trained by utilizing the adopted dataset.
- 2) **HRNet + CBAM**: Original HRNet followed by the CBAM module trained using the adopted dataset.
- 3) **OsteoHRNet**: Original HRNet followed by the CBAM module trained using the adopted dataset. Further, during training, we have employed the data augmentation techniques to enhance the performance of the proposed model.

It can be observed from Table III that the addition of the CBAM module and data augmentation techniques have immensely improved the performance compared to its curtailed baseline. The CBAM module might have adaptively learned the relevant features from the HRNet. Such features may have contributed more towards an efficient classification compared to the features learned by the original HRNet [29], VGG-19 [32], or DenseNet161 [24].

Fig. 13 demonstrates the Grad-CAM visualizations for our ablation study. It can be observed that the proposed OsteoHRNet has learned the robust features progressively on each

component of our proposed network. Thus, it is verified that each component of our network contributes to the final knee OA KL grade prediction.

## VIII. CONCLUSION

This paper proposes a novel OsteoHRNet by adopting the HRNet as the backbone and integrating the CBAM module for an improved knee OA severity prediction results from plain radiographs. The proposed network was able to perform exceptionally well and attain significant improvements over the previously proposed methods owing to the HRNet's capability to maintain high-resolution features throughout the network and its ability to capture reliable spatial features. The intermediate extracted features were significantly refined with the help of the attention mechanism; therefore, the radiographs with a similarity between classes and variations within classes could be distinguished better. Moreover, we have employed the Grad-CAM visualizations to validate that the model has learned the most relevant spatial features in the radiographs. In the future, we will work on the entire OAI multi-modal data and consider all the cohorts in our study.

## REFERENCES

- [1] H. Oka, S. Muraki, T. Akune, A. Mabuchi, T. Suzuki, H. Yoshida, S. Yamamoto, K. Nakamura, N. Yoshimura, and H. Kawaguchi, "Fully automatic quantification of knee osteoarthritis severity on plain radiographs," *Osteoarthritis and Cartilage*, vol. 16, no. 11, pp. 1300–1306, 2008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S106345840800085X>
- [2] P. N. S. H. M. E. M. G. F. G. A. . M. M. A. Lespasio, M. J., "Knee osteoarthritis: A primer," *The Permanente journal*, vol. 21, 2017.
- [3] A. R. H. R. . A. T. H. Audrey, H. X., "The truth behind subchondral cysts in osteoarthritis of the knee," *The open orthopaedics journal*, 8, 2014.
- [4] S.-B. K. U. . E. P. Michael, J. W., "The epidemiology, etiology, diagnosis, and treatment of osteoarthritis of the knee," *Deutsches Arzteblatt international* 107(9), 2010.
- [5] J. H. Kellgren and J. S. Lawrence, "Radiological assessment of osteoarthritis," *Annals of the Rheumatic Diseases*, vol. 16, no. 4, pp. 494–502, 1957. [Online]. Available: <https://ard.bmj.com/content/16/4/494>
- [6] L. Shamir, S. Ling, W. Scott, M. Hochberg, L. Ferrucci, and I. Goldberg, "Early detection of radiographic knee osteoarthritis using computer-aided analysis," *Osteoarthritis and Cartilage*, vol. 17, no. 10, pp. 1307–1312, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1063458409001101>
- [7] C. Peterfy, E. Schneider, and M. Nevitt, "The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee," *Osteoarthritis and Cartilage*, vol. 16, no. 12, pp. 1433–1441, 2008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1063458408002239>
- [8] S. Kashyap, H. Zhang, K. Rao, and M. Sonka, "Learning-based cost functions for 3-d and 4-d multi-surface multi-object segmentation of knee mri: Data from the osteoarthritis initiative," *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1103–1113, 2018.

- [9] H. Khalid, M. Hussain, M. A. Al Ghamdi, T. Khalid, K. Khalid, M. A. Khan, K. Fatima, K. Masood, S. H. Almotiri, M. S. Farooq, and A. Ahmed, "A comparative systematic literature review on knee bone reports from mri, x-rays and ct scans using deep learning and machine learning methodologies," *Diagnostics*, vol. 10, no. 8, 2020. [Online]. Available: <https://www.mdpi.com/2075-4418/10/8/518>
- [10] S.-N. M. B. J. L. B.-Z. S. A. P. R. B. P. K. B. W. N. R. V. J. L. P. R. M. . v. d. H. W. B. van Oudenaarde, K., "General practitioners referring adults to mr imaging for knee pain: A randomized controlled trial to assess cost-effectiveness." *Radiology* 288(1), 2018.
- [11] C. Kokkoti, S. Moustakidis, E. Papageorgiou, G. Giakas, and D. Tsaopoulos, "Machine learning in knee osteoarthritis: A review," *Osteoarthritis and Cartilage Open*, vol. 2, p. 100069, 05 2020.
- [12] 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," 2017.
- [13] J. S. Yadav, S.S., "Deep convolutional neural network based medical image classification for disease diagnosis." *J Big Data* 6, 113, 2019.
- [14] P. . Chen, ""knee osteoarthritis severity grading dataset".
- [15] P. Chen, L. Gao, X. Shi, K. Allen, and Y. Lin, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84–92, 2019.
- [16] L. Shamir, S. M. Ling, W. W. Scott, A. Bos, N. Orlov, T. J. Macura, D. M. Eckley, L. Ferrucci, and I. G. Goldberg, "Knee x-ray image analysis method for automated detection of osteoarthritis," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 407–415, 2009.
- [17] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," pp. 1195–1200, 2016.
- [18] N. E. O. J. Antony, K. McGuinness and K. Moran, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks." 2017.
- [19] T. J. R. E. L. P. S. S. Tiulpin, A., "Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach." *Scientific Reports.*, 2018.
- [20] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," vol. 1, pp. 539–546 vol. 1, 2005.
- [21] M. Górriz, J. Antony, K. McGuinness, X. Giró-i-Nieto, and N. O'Connor, "Assessing knee oa severity with cnn attention-based end-to-end architectures," vol. 102, pp. 197–214, 08–10 Jul 2019. [Online]. Available: <http://proceedings.mlr.press/v102/goriz19a.html>
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
- [23] C. Yong, K. Teo, B. Murphy, Y. Hum, Y. Tee, K. Xia, and k. w. lai, "Knee osteoarthritis severity classification with ordinal regression module," *Multimedia Tools and Applications*, 01 2021.
- [24] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2018.
- [25] N. M. D. R. e. a. Yamashita, R., "Convolutional neural networks: an overview and application in radiology." *Insights Imaging* 9, 611–629, 2018.
- [26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019.
- [27] B. Zhang, J. Tan, K. Cho, G. Chang, and C. M. Deniz, "Attention-based cnn for kl grade classification: Data from the osteoarthritis initiative," pp. 731–735, 2020.
- [28] S. Woo, J. Park, J.-Y. Lee, and I. Kweon, "Cbam: Convolutional block attention module," 07 2018.
- [29] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.
- [30] R. Rs, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that?" 11 2016.
- [31] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," pp. 8024–8035, 2019.
- [32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.
- [33] S. A. A. . F. N. D. Kohn, M. D., "Classifications in brief: Kellgren-lawrence classification of osteoarthritis." *Clinical orthopaedics and related research*, vol. 474(8), 2016.